UNIVERSITY OF THESSALY
SCHOOL OF ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Artificial Intelligence Tools for the Modern Electricity Markets

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Vasileios Laitsos

July 2025

UNIVERSITY OF THESSALY
SCHOOL OF ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Artificial Intelligence Tools for the Modern Electricity Markets

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Vasileios Laitsos

July 2025

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Εργαλεία Τεχνητής Νοημοσύνης για τις Σύγχρονες Αγορές Ηλεκτρικής Ενέργειας

Διατριβή η οποία υποβλήθηκε για τη μερική εκπλήρωση
των υποχρεώσεων απόκτησης του Διδακτορικού Διπλώματος

Βασίλειος Λαΐτσος

Ιούλιος 2025

UNIVERSITY OF THESSALY
SCHOOL OF ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Artificial Intelligence Tools for the Modern Electricity Markets

PhD Dissertation

Vasileios Laitsos

Advisory committee

Dimitrios Bargiotas, Professor, University of Thessaly (Supervisor)

Lefteris Tsoukalas, Professor, University of Thessaly

Aspassia Daskalopulu, Associate Professor, University of Thessaly

Examination committee

Dimitrios Bargiotas, Professor, University of Thessaly (Supervisor)

Lefteris Tsoukalas, Professor, University of Thessaly

Aspassia Daskalopulu, Associate Professor, University of Thessaly

Dimitrios Katsaros, Associate Professor, University of Thessaly

Antonios Alexandridis, Professor, University of Patras

Christos Manasis, Professor, University of Athens

Pandelis Biskas, Professor, Aristotle University of Thessaloniki

July 2025

DISCLAIMER ON ACADEMIC ETHICS
AND INTELLECTUAL PROPERTY RIGHTS

Being fully aware of the implications of copyright laws, I expressly state that this Ph.D. dissertation, as well as the electronic files and source codes developed or modified in the course of this dissertation, are solely the product of my personal work and do not infringe any rights of intellectual property, personality and personal data of third parties, do not contain work / contributions of third parties for which the permission of the authors / beneficiaries is required and are not a product of partial or complete plagiarism, while the sources used are limited to the bibliographic references only and meet the rules of scientific citing. The points where I have used ideas, text, files and / or sources of other authors are clearly mentioned in the text with the appropriate citation and the relevant complete reference is included in the bibliographic references section. I also declare that the results of the work have not been used to obtain another degree. I fully, individually and personally undertake all legal and administrative consequences that may arise in the event that it is proven, in the course of time, that this dissertation or part of it does not belong to me because it is a product of plagiarism.

The declarant

Vasileios Laitsos

Ευχαριστίες

I would first like to express my deepest gratitude to my Supervisor, Professor Dimitrios Bargiotas, for his guidance and oversight throughout the preparation of this PhD dissertation. Also, I would like to thank my colleague, Georgios Vontzos, for the exemplary cooperation that we have maintained over all these years. I would also like to especially thank the Advisory Committee and the Examination Committee for their valuable contribution.

I owe a heartfelt thank you to my parents for all they have done and continue to do for me.

Finally, I extend my thanks to Eurudiki for her constant support.

PhD Dissertation

Artificial Intelligence Tools for the Modern Electricity Markets

Vasileios Laitsos

Abstract

The electricity market is one of the most dynamically changing sectors on a global scale and it is undergoing remarkable reforms in many areas. In its early stages (e.g., solar and wind), new needs and challenges are emerging that must be addressed. Additionally, new technologies—such as the rapid development of artificial intelligence (AI)—are creating new possibilities that provide excellent tools for market operation. At the same time, new regulatory rules aimed at abolishing market monopolies and creating a unified, liberalized electricity market are generating new values and demands to ensure smooth operation.

In order to ensure the smooth operation of the wholesale electricity market, it is imperative to create and use advanced artificial intelligence tools capable of addressing these new challenges. The adoption of state-of-the-art AI mechanisms can create favorable conditions to predict both the electricity demand of end users and the fluctuations in electricity prices with optimal accuracy.

The main objective of this PhD dissertation is to develop advanced and novel algorithms for electricity load and price forecasting, specifically for the operation of modern wholesale market mechanisms. More precisely, the research focuses on developing algorithms and advanced Deep Learning models that can achieve very high short-term forecasting accuracy, serving as effective tools in innovative programs within the modern electricity market.

The research aims to cover various aspects of the modern electricity market by examining electricity mechanisms—specifically the Day Ahead Market, Intraday Market, and Balancing Market—and by exploring applications for active consumer participation such as Demand Side Management (which includes peak clipping, valley filling, load shifting, and load conservation) as well as Demand Response, in addition to forecasting the load of electric vehicles (EV Load) and investigating Transfer Learning with particular consideration given to the more challenging case of the day-ahead market.

This dissertation innovates in the following key components:

- Development and creation of advanced and novel Deep Learning models: These models can be used both for electricity load and price forecasting, as well as for other time series scientific tasks, such as problems in the economic and industrial sectors.

- Implementation of Novel Deep Learning Models for Electricity Demand Forecasting: These models can predict energy demand under various periods and conditions, assisting power producers adjust their generation and optimize operations.

- Implementation of novel Deep Learning Models for Wholesale Electricity Price Forecasting: This enables energy producers to make decisions regarding large-scale market operations and production, taking into account the forecasted prices.

- Analysis of the Price Formulation Framework in Wholesale Markets: By using data from different energy markets, trends, market players' strategies, and their interactions are analyzed to better understand how the final hourly prices are determined.

- Implementation of Transfer Learning Methods for the Day-Ahead Electricity Market: The developed models perform optimally in forecasting 24 hourly steps ahead, a challenging task in the current literature.

- Investigation of Demand Side Management and Demand Response Programs: These strategies are examined for their integration into market operations through active consumer participation. In this way, optimization functions to optimize the network operation and cost, covering both production/supply side are used.

- Development of New Methods for Forecasting the Load of Electric Vehicles: Considering five different datasets, the distinctive patterns of the EV-load curve are detected, isolated, and forecasted separately.

- Investigation of the new trends emerging in the operation of the modern electricity market through the study and presentation of various studies, which use different datasets and cover the full spectrum of forecasting algorithms.

keywords

Deep Learning; Machine Learning; Electricity Market; Exploratory Data Analysis; Price Forecasting; Load Forecasting; Smart Grids

Διδακτορική Διατριβή

Εργαλεία Τεχνητής Νοημοσύνης για τις Σύγχρονες Αγορές Ηλεκτρικής Ενέργειας

Βασίλειος Λαΐτσος

Περίληψη

Η αγορά ηλεκτρικής ενέργειας είναι ένας από τους τομείς που υφίστανται τις πιο δυναμικές αλλαγές σε παγκόσμιο επίπεδο και υπόκειται σε αξιοσημείωτες μεταρρυθμίσεις σε πολλούς τομείς. Στα αρχικά της στάδια (π.χ., ηλιακή και αιολική ενέργεια), ανακύπτουν νέες ανάγκες και προκλήσεις που πρέπει να αντιμετωπιστούν. Επιπλέον, νέες τεχνολογίες—όπως η ραγδαία εξέλιξη της τεχνητής νοημοσύνης—δημιουργούν νέες δυνατότητες που παρέχουν εξαιρετικά εργαλεία για τη λειτουργία της αγοράς. Ταυτόχρονα, νέοι ρυθμιστικοί κανόνες που στοχεύουν στην κατάργηση των μονοπωλίων της αγοράς και στη δημιουργία μίας ενιαίας, απελευθερωμένης αγοράς ηλεκτρικής ενέργειας παράγουν νέες αξίες και απαιτήσεις για τη διασφάλιση της ομαλής λειτουργίας της.

Προκειμένου να διασφαλιστεί η ομαλή λειτουργία της χονδρικής αγοράς ηλεκτρικής ενέργειας και να ανταποκριθεί αποτελεσματικά στις νέες εξελίξεις, είναι επιτακτική η δημιουργία και χρήση προηγμένων εργαλείων τεχνητής νοημοσύνης, ικανών να αντιμετωπίσουν αυτές τις νέες προκλήσεις. Η υιοθέτηση των πιο σύγχρονων μηχανισμών τεχνητής νοημοσύνης μπορεί να δημιουργήσει ευνοϊκές συνθήκες για την πρόβλεψη τόσο της ζήτησης ηλεκτρικής ενέργειας από τους τελικούς χρήστες όσο και των διακυμάνσεων των τιμών της ηλεκτρικής ενέργειας με βέλτιστη ακρίβεια.

Ο κύριος στόχος της διδακτορικής διατριβής είναι η ανάπτυξη προηγμένων αλγορίθμων Βαθιάς Μάθησης οι οποίοι θα μπορούν να χρησιμοποιηθούν στη λειτουργία των σύγχρονων μηχανισμών της χονδρικής αγοράς ενέργειας. Πιο συγκεκριμένα, η έρευνα επικεντρώνεται στην ανάπτυξη αλγορίθμων και καινοτόμων μοντέλων που μπορούν να επιτύχουν πολύ υψηλή ακρίβεια πρόβλεψης φορτίου και τιμής χονδρικής αγοράς ενέργειας σε βραχυπρόθεσμο ορίζοντα, λειτουργώντας ως αποτελεσματικά εργαλεία σε καινοτόμα προγράμματα της σύγχρονης αγοράς ηλεκτρικής ενέργειας.

Η έρευνα αναλύει σε βάθος όλο τα φάσμα που διέπει τη σύγχρονη αγορά ενέργειας. Πιο συγκεκριμένα, αναλύεται η Αγορά της Επόμενης Ημέρας (Day Ahead Market), η Ενδοημερή Αγορά (Intraday Market) και η Αγορά Εξισορρόπησης (Balancing Market). Επίσης, με σκοπό την μείωση του υπολογιστικού κόστους και χρόνου, μέθοδοι και στρατηγικές Βαθείας Μεταφερόμενης Μάθησης παρουσιάζονται εκτενώς με σκοπό την ενσωμάτωσή τους στα σύγχρονα υπολογιστικά συστήματα που διαχειρίζονται την λειτουργία των ενεργειακών συστημάτων.

Πρόσθετα, σημαντικό τμήμα της διατριβής επικεντρώνεται στις εφαρμογές ευελιξίας και ενεργής συμμετοχής των καταναλωτών στην λειτουργία της αγοράς, όπως τα προγράμματα Διαχείριση της Ζήτησης (που περιλαμβάνει το Peak Clipping, το Valley Filling, το Load Shifting και το Load Conservation) καθώς και η Απόκριση στη Ζήτηση (Demand Response). Τέλος, λόγω της συνεχώς αυξανόμενης ανάπτυξης του τομέα της ηλεκτροκίνησης, το τελευταίο τμήμα της διατριβής διερευνά την συμπεριφορά του φορτίου ηλεκτρικών οχημάτων (Electric Vehicle Load) και αναπτύσσει καινοτόμες λύσεις πρόβλεψης.

Η παρούσα διατριβή καινοτομεί στα ακόλουθους βασικούς πυλώνες:

- Ανάπτυξη και δημιουργία προηγμένων και καινοτόμων μοντέλων Βαθιάς Μάθησης: Αυτά τα μοντέλα μπορούν να χρησιμοποιηθούν τόσο για την πρόβλεψη του φορτίου και των τιμών του ηλεκτρισμού, όσο και για άλλα επιστημονικά προβλήματα χρονοσειρών, όπως είναι οι εφαρμογές στον οικονομικό και βιομηχανικό τομέα.

- Εφαρμογή καινοτόμων μοντέλων και εργαλείων βαθιάς μάθησης για την πρόβλεψη της ζήτησης ηλεκτρικής ενέργειας: Τα μοντέλα αυτά μπορούν να αξιοποιηθούν για την πρόβλεψη της ζήτησης ενέργειας σε διάφορες χρονικές περιόδους και υπό διαφορετικές συνθήκες, συμβάλλοντας στην υποστήριξη των παραγωγών ενέργειας ώστε να προσαρμόζουν την παραγωγή τους και να βελτιστοποιούν τις επιχειρησιακές τους λειτουργίες.

- Εφαρμογή καινοτόμων μοντέλων βαθιάς μάθησης για την πρόβλεψη των τιμών της χονδρικής αγοράς ηλεκτρικής ενέργειας: Με αυτόν τον τρόπο, οι παραγωγοί ενέργειας μπορούν να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με τις λειτουργίες μεγάλης κλίμακας στην αγορά και την παραγωγή, λαμβάνοντας υπόψη τις προβλεπόμενες τιμές.

- Ανάλυση του πλαισίου διαμόρφωσης των τιμών στις χονδρικές αγορές: Μέσω της ανάλυσης δεδομένων από διάφορες αγορές ενέργειας, εξετάζονται οι τάσεις, οι στρατηγικές των διαφόρων παραγόντων της αγοράς καθώς και οι αλληλεπιδράσεις τους, με στόχο την καλύτερη κατανόηση της διαδικασίας καθορισμού των τελικών ωριαίων τιμών.

- Εφαρμογή μεθόδων και στρατηγικών Μεταφερόμενης Μάθησης για την Αγορά Επόμενης Ημέρα: Τα αναπτυγμένα μοντέλα επιτυγχάνουν βέλτιστη απόδοση στην πρόβλεψη 24 ωριαίων βημάτων, μια πρόκληση που παραμένει ιδιαίτερα απαιτητική στην τρέχουσα βιβλιογραφία.

- Διερεύνηση των προγραμμάτων Διαχείρισης και Απόκρισης στη Ζήτηση: Εξετάζονται στρατηγικές για την ενσωμάτωση αυτών των προγραμμάτων στις λειτουργίες της αγοράς μέσω της ενεργούς συμμετοχής των καταναλωτών. Με αυτόν τον τρόπο, εφαρμόζονται λειτουργίες βελτιστοποίησης που στοχεύουν στη βελτίωση της λειτουργίας και στη μείωση του κόστους του δικτύου, καλύπτοντας τόσο την πλευρά παραγωγής όσο και την πλευρά προσφοράς.

- Ανάπτυξη νέων μεθόδων για την πρόβλεψη του φορτίου των ηλεκτρικών οχημάτων: Λαμβάνοντας υπόψη πέντε διαφορετικά σύνολα δεδομένων, εντοπίζονται, απομονώνονται και προβλέπονται ξεχωριστά τα διακριτά μοτίβα της καμπύλης φορτίου των ηλεκτρικών οχημάτων.

- Διερεύνηση των νέων τάσεων που εμφανίζονται στη λειτουργία της σύγχρονης αγοράς ηλεκτρικής ενέργειας μέσω της μελέτης και παρουσίασης διαφόρων ερευνών, οι οποίες χρησιμοποιούν διαφορετικά σύνολα δεδομένων και καλύπτουν το πλήρες φάσμα των αλγορίθμων πρόβλεψης.

keywords

Βαθιά Μάθηση, Μηχανική Μάθηση, Αγορά Ηλεκτρικής Ενέργειας, Διερευνητική Ανάλυση Δεδομένων, Πρόβλεψη Τιμών Αγοράς Ενέργειας, Πρόβλεψη Φορτίου, Ευφυή Δίκτυα

Table of contents

List of figures

List of tables

Abbreviations

AIC	Akaike Information Criteria
ANN	Artificial Neural Network
ADF	Augmented Dickey–Fuller
ACF	Autocorrelation Function
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ARX	Autoregressive with Exogenous Inputs
BPNN	Backpropagation Neural Network
BM	Balancing Market
BIC	Bayesian Information Criteria
BO	Bayesian Optimization
CNN	Convolutional Neural Network
CNN-GRU	Convolutional Neural Network with Gated Recurrent Unit
$C(\theta, x)$	Cost function of the consumer
$C_{\text{gen}}(t)$	Cost of Conventional Generator
$C_{\text{trans}}(t)$	Transfer Power Cost between Main Grid and Microgrid
DAM	Day-Ahead Market
DNN	Deep Neural Network
DTL	Deep Transfer Learning
DR	Demand Response
DSM	Demand Side Management
DES	Double Exponential Smoothing
ESN	Echo State Network
ELF	Electricity Load Forecasting

EPF	Electricity Price Forecasting
EMD	Empirical Mode Decomposition
ES	Exponential Smoothing
XGBoost	Extreme Gradient Boosting
ELM	Extreme Learning Machine
FFN	Feed Forward Network
GRU	Gated Recurrent Unit
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
IHEPC	Individual-Household-Electric-Power-Consumption
IDM	Intra-Day Market
IDA	Intra-Day Auction
KNN	K-Nearest Neighbors
LR	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MSE	Mean Squared Error
MLP	Multilayer Perceptron
MDTL	Multitask Deep Transfer Learning
NYISO	New York Independent System Operator
NARX	Nonlinear Autoregressive Exogenous Model
PJM	Pennsylvania–New Jersey–Maryland Interconnection
P_{PV}	Power of the Photovoltaic Array
P_{wind}	Power of the Wind Turbine
RF	Random Forest
RNN	Recurrent Neural Network
RMSE	Root Mean Squared Error
SHAP	SHapley Additive exPlanations
SARIMA	Seasonal Auto-Regressive Integrated Moving Average
SARIMAX	Seasonal Autoregressive Integrated Moving-Average with Exogenous Regressors
Seq2Seq	Sequence-to-Sequence
SG	Smart Grid
SR	Spinning Reserve

SVM	Support Vector Machine
SVR	Support Vector Regression
TCN	Temporal Convolutional Network
TOU	Time of Use
TL	Transfer Learning
TES	Triple Exponential Smoothing
V_1	Benefit Function of a Consumer
V_2	Utility Company Benefit Function

Chapter 1

Introduction

The modern power system is undergoing a rapid transformation driven by decarbonization, decentralization, and digitalization, posing, therefore, significant challenges. The integration of renewable energy sources (RES) such as solar and wind energy, while essential for reducing carbon emissions, introduces variability and uncertainty into the grid. These energy sources are weather dependent and cannot be fully controlled, making it difficult to maintain a balance between supply and demand. This variability requires the development of robust forecasting models and advanced grid management strategies to ensure balance between electricity supply and demand, stability and reliability.

Another pressing challenge is the decentralization of energy systems, exemplified by the proliferation of distributed energy resources (DER), such as rooftop solar panels, electric vehicles (EV) and energy storage systems. These assets shift the traditional unidirectional flow of electricity to a complex, bidirectional one. Managing this decentralized architecture requires real-time coordination and sophisticated control mechanisms, placing new demands on grid infrastructure and communication networks. Additionally, ensuring interoperability between diverse technologies is critical but remains a complex issue.

Economic and market challenges further complicate the modern power system. Traditional electricity markets were designed for centralized power generation and are not well-suited for handling the dynamics introduced by renewables and DERs. The rise of prosumers—consumers who also produce energy—necessitates novel pricing models, market mechanisms, and incentives. Ensuring fairness, efficiency, and transparency in these emerging markets is vital for encouraging participation while avoiding congestion and inefficiencies in the grid.

Also, cybersecurity and resilience are also paramount in today's power systems. Increased digitalization and reliance on smart technologies expose the grid to potential cyber threats. Attacks on critical infrastructure can disrupt operations, jeopardize reliability, and compromise sensitive data. Building a secure and resilient grid infrastructure involves addressing vulnerabilities, implementing advanced monitoring systems, and leveraging artificial intelligence to predict and mitigate potential threats.

Additionally, as electrification expands to include transportation and heating, the demand on electricity infrastructure increases dramatically. Scaling up capacity while maintaining affordability and accessibility, particularly for underserved populations, requires strategic investments and innovative approaches. Deep learning and AI-based solutions play a crucial role in tackling these multifaceted challenges, enabling predictive analytics, optimization, and decision-making capabilities that are essential for managing complexity in the modern electricity market.

Based on these consideration, this PhD dissertation aims to delve deeply into and thoroughly analyze the new dynamics that have emerged in modern electrical energy systems, as well as the challenges introduced by the new frameworks governing this market. In particular, the study seeks to explore how evolving technologies, regulatory changes, and shifting market conditions are reshaping the structure and operation of energy systems. The focus will be on identifying the critical factors driving these transformations, understanding their implications for stakeholders, and proposing strategies to address the complex challenges they present. By providing a comprehensive assessment, the research aspires to contribute valuable insights to the ongoing discourse on sustainable and efficient energy management in the contemporary era. To provide a deeper understanding of the purpose behind this dissertation, a detailed analysis follows regarding the new regulatory frameworks governing the energy-economic sector of the market, which are shaping new operating conditions for the systems. Additionally, the aspects of energy market liberalization are presented, and the new dynamics that have emerged and continue to develop on a global level are made clear.

So, the primary objective is to develop advanced and novel Deep Learning algorithms for forecasting electricity load and price, essential for the operation of modern wholesale electricity market mechanisms. Specifically, the research focuses on creating algorithms and advanced Deep Learning models that can achieve high forecasting accuracy over short-term horizons. These models are intended to serve as robust tools for use in innovative programs within the contemporary electricity market.

The goal is to comprehensively address the forecasting needs across all three main electricity markets: the Day Ahead Market, the Intraday Market, and the Balancing Market. These models are expected to be instrumental in applications that promote active consumer participation, such as Demand Side Management (DSM) and Demand Response (DR) programs. In addition, the developed tools focus on forecasting the load associated with electric vehicle (EV) charging, which is an increasingly important component of the overall electricity demand landscape.

Finally, the dissertation explores the domain of Transfer Learning, with particular attention to its application in the challenging context of the day-ahead market. This exploration aims to leverage Transfer Learning techniques to enhance forecasting capabilities in environments with limited data availability or where data characteristics vary significantly between regions or market segments.

1.1 Dissertation contribution

This dissertation presents several key scientific achievements in the realm of electricity market forecasting and demand management. More specifically, the main axes of contribution are the following:

- In the field of electricity load forecasting, advanced and novel Deep Learning models are developed. These models effectively predict energy demand across various time frames and under different conditions, assisting energy producers in optimizing production schedules and improving operational efficiency. Their accuracy and adaptability enhance the management of demand fluctuations.

- Innovations in the field of electricity price forecasting for the wholesale market through the development of innovative Deep Learning models and solutions. The exploitation of these models enables energy producers to make well-informed decisions for large-scale market transactions and production strategies. Furthermore, a comprehensive analysis of price formation mechanisms in wholesale markets is conducted. This includes studying trends, participant strategies, and market interactions, providing critical insights into the determinants of final hourly prices and fostering market transparency and efficiency.

- The implementation of Transfer Learning methods and the integration in the Deep Learning area for the day-ahead electricity market. This finding introduces models capable of forecasting 24 hourly steps ahead. This challenging task, rarely achieved in the existing literature, offers a more accurate and reliable forecasting tool crucial for day-ahead market operations.

- Investigation of Demand Side Management (DSM) and Demand Response (DR) strategies, emphasizing active consumer participation in market operations. Various DSM techniques, such as Peak Clipping, Valley Filling, Load Shifting, and Load Conservation, are explored. Additionally, an application is developed to simulate active consumer participation with load-shifting capabilities. This application employs a unified optimization function that optimizes network operations and costs, addressing production, transmission, and consumption aspects holistically.

- Electric Vehicle Load Management and Forecasting. New forecasting methods and tools for electric vehicle (EV) load are developed, leveraging data from five different datasets. The unique patterns of EV-load curves are identified, isolated, and forecasted separately. This work addresses the specific challenges of predicting EV-related demand patterns, which constitute an increasingly significant component of overall energy consumption.

- The state-of-the-art short-term forecasting approaches for electricity load and prices, underscoring the need for precise predictions of demand fluctuations and price volatility.

- An in-depth mathematical analysis of various forecasting models, including statistical, machine and deep learning, and ensemble approaches.

- The application of statistical, artificial intelligence, and hybrid models in the time-series analysis of diverse load and price data, crucial for enhancing forecast accuracy.

- An extended review of evolving trends in Electricity Load Forecasting (ELF) and Electricity Price Forecasting (EPF), incorporating economic and engineering perspectives that drive these trends.

These contributions collectively advance the state-of-the-art in electricity market forecasting and management. They provide practical tools and methodologies that can be directly applied in real-world scenarios, supporting the transition to more efficient and resilient energy systems.

1.2 Dissertation organization

The subsequent chapters of this dissertation will delve deeper into each of these aspects, presenting detailed methodologies, case studies, and empirical analyses. Specifically:

- Chapter 1: A general introduction to the dissertation topic and to the motivation and objectives of the work is provided, and all the scientific fields to which the work carried out contributes are analyzed.

- Chapter 2: An introduction to Modern Power Systems and Market Liberalization is provided. The new challenges arising from the transformations in the energy market are analyzed, and the all the mechanisms that constitute the modern power systems are presented in detail.

- Chapter 3: The new challenges and innovations introduced by Demand Side Management and Demand Response programs are presented in detail, along with two experimental applications.

- Chapter 4: It constitutes an extensive collection of techniques that are applied to time series forecasting problems. All the most widely used pre-processing methods are analyzed, and all the models—statistical, machine learning, deep learning, and ensemble—are presented, which are applied in contemporary research.

- Chapter 5: An introduction to the field of load forecasting is provided, and all statistical methods, Machine Learning, Deep Learning, and Ensemble methods are analyzed, ultimately outlining the reasons why the Deep Learning approach we study outperforms the Statistical and Machine Learning methods.

- Chapter 6: The field of electricity price forecasting is introduced, and emphasis is given to all the new models and innovations developed in this work.

- Chapter 7: A presentation of the field of Transfer Learning and its broader applications in market operations is provided.

- Chapter 8: An introduction to Electric Vehicles (EVs) is provided, and the need for accurate forecasting of EV load is analyzed for the operation of advanced programs in which consumers actively participate (Smart Charging Solutions for Demand Response).

- Chapter 9: The general conclusions of the research and its overall contribution to science are presented, while simultaneously emphasizing the proposals for future study that emerge.

Chapter 2

Introduction to Modern Power Systems and Market Liberilazitation

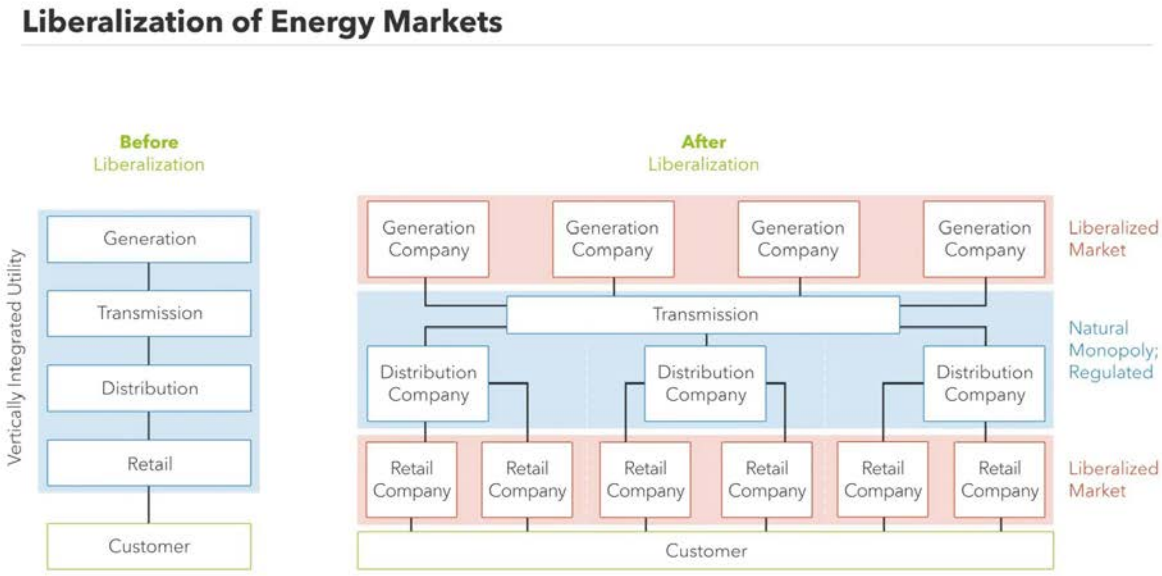
2.1 Key Drivers and Challenges of Market Liberalization

The restructuring of electricity markets aims to introduce competition among various electricity providers. Traditionally, these markets were dominated by state-owned or vertically integrated utilities that controlled generation, transmission, and distribution. Liberalization seeks to separate these functions and foster competition, particularly in the generation and retail segments. Economic efficiency is a primary driver, as competitive markets tend to be more efficient than monopolies, encouraging cost-cutting, efficiency, and innovation. Technological advancements, such as smart grids and renewable energy sources, have enabled the entry of more players into the market, challenging the old monopolistic models. Regulatory changes by governments and international bodies recognize the benefits of liberalization and have enacted measures to dismantle monopolies and promote competition. Additionally, consumer demand for more choices, better services, and lower prices is a significant factor in driving liberalization efforts.

Also, liberalization can lead to cost reductions as competition among electricity providers typically drives down prices, benefiting both consumers and businesses. It also spurs innovation, with new market entrants introducing advanced technologies and business models that support cleaner and more efficient energy solutions. The quality of service improves as multiple providers compete for customers, creating greater incentives to enhance service quality and customer satisfaction. Furthermore, a competitive market attracts private investment, which fosters the development of new infrastructure and technologies. Despite its advantages, liberalization poses several challenges and risks. Incumbent utilities may still exert significant market power, potentially stifling competition and manipulating prices. Effective regulatory oversight is crucial to prevent anti-competitive practices and ensure fair access to the market for new entrants. Liberalized markets can also be more volatile, with prices fluctuating due to supply and demand dynamics, potentially leading to instability. Additionally, the transition from a monopolistic to a competitive market can incur significant costs, including the need for new regulatory frameworks, infrastructure upgrades, and consumer education.

In addition, the deregulation and liberalization of the electricity market have been a cornerstone of the European Union's energy policy since the late 1980s. The closed nature of national energy markets was seen as incompatible with the EU's broader objective of creating a unified and competitive economic area. The main impetus for deregulation was to improve the efficiency of energy operations and the energy system as a whole, ultimately aiming to reduce electricity costs for consumers. This transformation also unlocked significant entrepreneurial opportunities within the energy sector, driving competition and fostering innovation to deliver more affordable retail electricity prices. Moreover, the liberalization of the electricity market had positive spillover effects on related sectors, such as the production of renewable electricity and the development of energy derivative markets, which added depth to the market ecosystem [1].

Market liberalization enabled participants to become more adaptable to rapid economic and environmental changes, including those posed by climate change. In several countries, the pace of deregulation outstripped national legislative timelines, allowing for faster progress than initially anticipated. Despite the initial goals of maintaining service standards and reducing prices in a transparent yet previously low-competitive environment, electricity prices have since risen. This indicates that the process of developing fully competitive electricity markets is still incomplete. Many EU citizens and businesses continue to face limited choices when it comes to electricity suppliers, as market fragmentation, high levels of vertical integration, and concentrated market power in the hands of a few major players hinder the full realization of a truly internal electricity market. With the opening and liberalization of the electricity market, electricity itself became a commodity that could be freely traded, further driving competition and efficiency. Figure 2.1 visualizes the wholesale energy market before and after liberalization.



Source: [2]

Figure 2.1: Liberalization of Energy Market

Parallel to these developments, climate change has increasingly become a global priority over the past few decades, as its impacts have intensified, causing serious economic and social disruptions across different regions. In response, countries worldwide have focused their energy policies on reducing carbon dioxide (CO2) emissions to promote sustainable economic growth and development. However, the short-term costs of adapting to and mitigating climate change are substantial, as nations set ambitious emission reduction targets that require significant investments in new infrastructure and cleaner technologies. Fossil fuels, such as natural gas, oil, and coal, remain the primary sources of CO2 emissions, making the transition to renewable energy a particularly difficult challenge for large economies [1].

Data from the Statistical Review of World Energy shows that CO2 emissions in the European Union (EU) fell from 4081 million tons in 2000 to 3466 million tons in 2018, while fossil fuel consumption decreased by approximately 15.32% over the same period. Fossil fuels, which have been the dominant energy source in driving economic growth, are largely responsible for the climate change impacts observed worldwide, including extreme weather events such as droughts, floods, and severe storms. To counter these effects, various international organizations and governments have proposed policies aimed at moving away from fossil fuels and toward more sustainable, renewable energy sources. The EU has been a leading force in these efforts, advocating for environmental policies that promote a greener and more sustainable future, characterized by lower emissions, efficient use of resources, and increased competitiveness that also benefits its citizens' well-being [3].

According to the International Energy Agency (IEA), the EU's carbon dioxide emissions declined from 4036 million tons in 2000 to 3475 million tons in 2017, representing 10.8% of the world's total emissions. Despite this progress, the energy sector still accounts for a significant portion (35.9%) of the EU's CO2 emissions due to its continued reliance on fossil fuels. However, the growing competition from renewable energy sources has begun to reshape the market. For the first time since 2008, energy prices in the EU fell in 2017, driven by increased competition from renewables. That year, renewable energy sources generated 30.7% of electricity and 19.7% of heating demand, with biomass, solar thermal, and geothermal energy playing key roles in heating [3].

The EU's relatively higher income levels have enabled member states to prioritize environmental protection, setting ambitious targets for CO2 emission reductions. One key policy has been the liberalization of the internal energy market, encompassing both natural gas and electricity sectors. The third energy package, approved in 2011, extended market liberalization to include both domestic and non-domestic markets. This policy aimed to create a more competitive electricity market, prioritizing consumers while encouraging the production of decarbonized electricity. The goal was to deliver a competitive, secure, sustainable, and accessible energy service for all.

In line with these efforts, the EU, in cooperation with the European Parliament, has set ambitious targets to improve energy efficiency, expand consumer choice in energy suppliers, and contribute to climate goals. These include a 40% reduction in greenhouse gas emissions by 2030 and the achievement of net-zero emissions by 2050. The transition to a low-carbon economy is expected to gain further momentum as new green technologies and products emerge, further driving sustainability efforts [4].

The liberalization of the internal energy market is crucial for increasing the consumption of renewable energy by removing trade barriers and other obstacles that have historically hindered the development of this market. Nevertheless, while this process offers numerous opportunities for innovation and growth, it also presents challenges. The market distortions that arise during the transition phase pose risks to the overall goal of reducing CO2 emissions. Careful management of these risks will be essential to ensure the long-term success of the

EU's decarbonization efforts.

2.2 European targets and policies on reducing carbon emissions

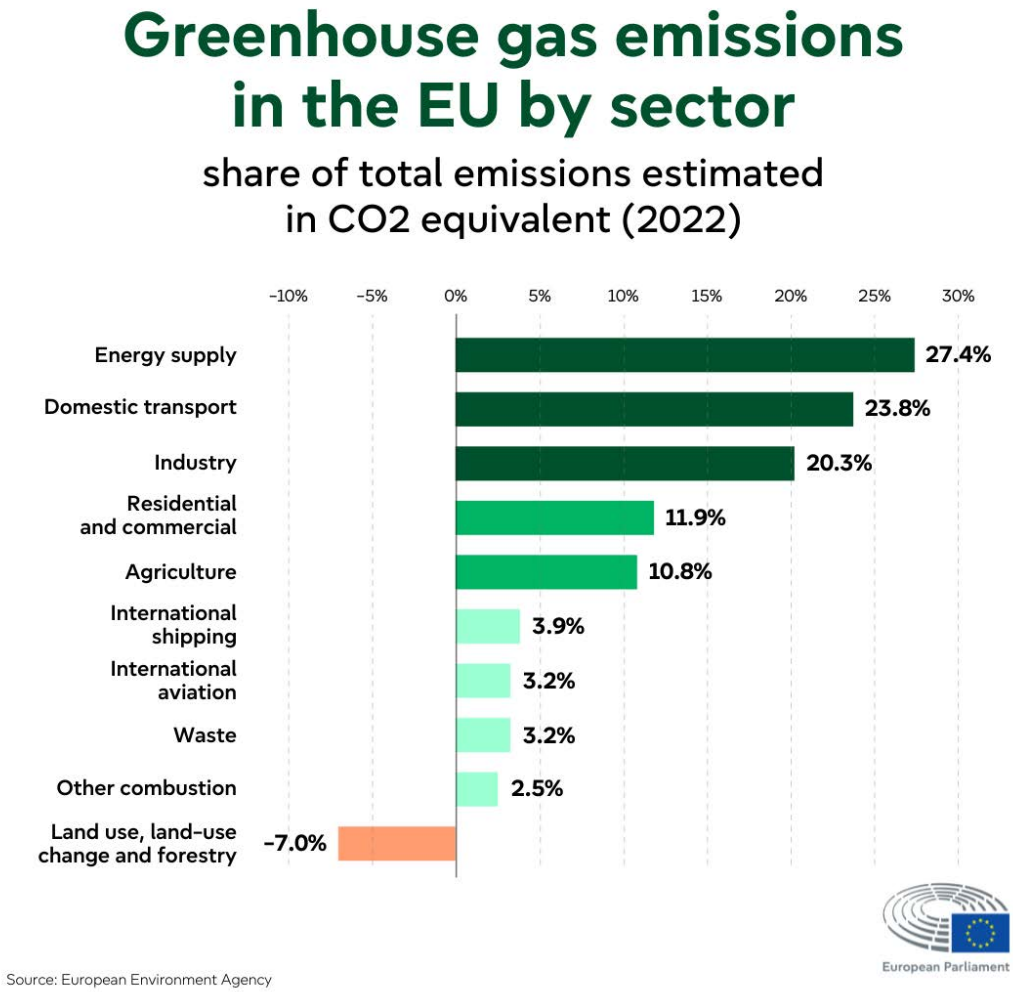
Climate change has a profound impact on Europe, with significant effects already evident across the continent. According to recent data, European temperatures have risen by approximately 1.2°C since pre-industrial times, a rate that is notably higher than the global average values. This warming trend has led to more frequent and severe weather events, including intense heatwaves, heavy rainfall, and prolonged droughts. For instance, Europe experienced its hottest summer on record in 2019, with temperatures reaching unprecedented levels. The continent is also seeing a marked increase in the frequency of extreme weather phenomena, such as storms and flooding, which have caused substantial economic and social damage. Also, sea levels around Europe are rising due to thermal expansion and melting ice, contributing to coastal erosion and increased flooding risks. These changes are not only altering natural ecosystems but also impacting agriculture, water resources, and public health. In response, European countries are investing heavily in climate adaptation and mitigation strategies, aiming to reduce greenhouse gas emissions and build resilience against future climate impacts [5].

In its 2023 Sixth Assessment Report, the Intergovernmental Panel on Climate Change (IPCC) highlighted that human activities have contributed approximately 1.1°C to global warming since the early 20th century. These activities primarily include burning fossil fuels like coal, oil, and gas, along with deforestation and agricultural practices.

An infographic detailing the European Union's greenhouse gas emissions for 2022 illustrates their distribution across major sectors and is presented in Figure 2.2. Energy supply contributed 27.4% of emissions, while domestic transport accounted for 23.8%. The industrial sector was responsible for 20.3%, with residential and commercial activities adding 11.9%, and agriculture contributing 10.8% [6].

On the other hand, land use, land-use change, and forestry played a mitigating role by absorbing carbon dioxide through vegetation. This sector offset emissions by an amount equivalent to 7% of the EU's total emissions [7].

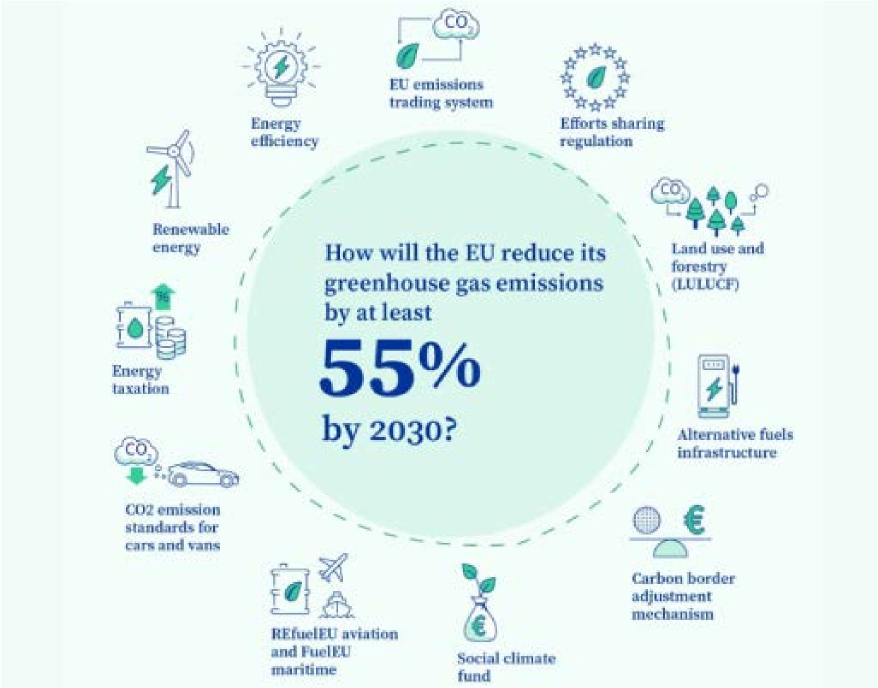
To combat climate change, the European Parliament enacted the European Climate Law,



Source: [6]

Figure 2.2: Greenhouse Gas European Emissions

which mandates a minimum 55% reduction in net greenhouse gas emissions by 2030, compared to the previous target of 40%. This law also establishes the goal of achieving climate neutrality by 2050 as legally binding. The European Climate Law is a cornerstone of the European Green Deal, the EU's comprehensive plan to transition to a sustainable and climate-neutral economy [8].



Source: [9]

Figure 2.3: Fit for 55 Package

2.3 Towards Climate Neutrality: The EU's Fit for 55 Legislative Package

In December 2019, European Union launched a broader new strategy in order to reduce the environmental pollution, named the European Green Deal. The Fit for 55 package, which is illustrated in Figure 2.3, is a set of proposals, introduced by the European Commission on July 14, 2021 and is part of the European Green Deal. This package is a legislative initiative under the European Green Deal, designed to achieve the updated climate goals. It consists of various interlinked laws and proposed measures focused on climate and energy. A key component of this package is the reform of the EU Emissions Trading System (ETS), the world's largest carbon market, which covers around 40% of EU greenhouse gas emissions. The ETS requires companies to purchase permits for each tonne of CO_2 they emit, incentivizing the reduction of emissions [8].

Also, it includes several key initiatives aimed at reducing emissions and aligning EU climate policies with its 2030 and 2050 targets. It proposes a comprehensive revision of the EU Emissions Trading System (EU ETS) to achieve a 61% emissions reduction in affected sectors by 2030, including phasing out free allowances in aviation and extending the system to maritime, road transport, and buildings. Member States' binding emissions reduction targets will increase, with an EU-wide goal of 40% reduction in greenhouse gases compared to 2005. The package sets a target for net removals of greenhouse gases through land use, forestry, and land use change at 310 million tonnes of CO2 equivalent by 2030 [8]. The renewable energy directive will be updated to require at least 40% of energy from renewable sources, while energy efficiency targets will rise to 36% for final energy consumption and 39% for primary energy. Infrastructure for alternative fuels and power supply will be expanded, and new CO2 emission standards will require a 100% reduction by 2035 for cars and vans, accelerating the shift to electric vehicles. A revision of the energy taxation directive will align energy product taxes with EU climate goals and rationalize tax exemptions. The carbon border adjustment mechanism (CBAM) will prevent emission leakage by imposing tariffs on carbon-intensive imports. The package also aims to reduce the aviation sector's environmental impact through sustainable fuels and cut ships' greenhouse gas intensity by up to 75% by 2050. Lastly, the Social Climate Fund, with €72.2 billion allocated for 2025-2032, will address the social impacts of emissions trading on Member States [10].

Additionally, to decrease energy consumption, which accounts for more than three-quarters of EU greenhouse gas emissions, new rules were adopted in July 2023. These rules require an 11.7% reduction in energy consumption at the EU level by 2030 compared to 2020 projections, along with an average annual savings rate of 1.5%. The heating and cooling of buildings, which constitute 40% of the EU's energy consumption, are also targeted for improvements, with aims to achieve zero-emission buildings by 2050 [11].

Finally, the EU is also enhancing the development of RES. By 2030, 42.5% of the EU's energy consumption should come from renewables, with a 45% aspirational target. The EU has fast-tracked permits for renewable energy projects and is phasing out funding for natural gas infrastructure, redirecting resources to hydrogen and offshore renewable energy projects [11].

2.4 Global Perspective

The liberalization of electricity markets worldwide has followed diverse paths, reflecting regional differences in infrastructure, regulatory frameworks, and economic development. In the European Union (EU), a series of directives, notably the Electricity Directive and the

Electricity Market Regulation, have progressively established a single electricity market. This has fostered greater integration, transparency, and competition across member states, driving down prices and encouraging innovation. Leading markets like the UK, Germany, and the Nordic countries have seen significant competition and consumer choice, particularly in renewable energy and smart grid technologies. The establishment of the European Internal Electricity Market (IEM) allows electricity to flow freely across borders, improving efficiency and enhancing energy security.

However, the pace of liberalization has been uneven across Europe. While countries like Germany and UK have embraced fully liberalized markets with a high degree of competition and private sector participation, others, particularly in Southern and Eastern Europe, have been slower to reform. In some cases, historical reliance on state-owned utilities, insufficient grid infrastructure, and regulatory hurdles have delayed the full implementation of competitive electricity markets. The European Commission has worked to address these disparities through initiatives like the Clean Energy Package and the Energy Union, aimed at harmonizing market rules, improving cross-border interconnectivity, and enhancing grid flexibility.

In countries such as France and Spain, market liberalization has occurred alongside strong state intervention. France's electricity market, for example, continues to be dominated by EDF, the state-owned utility, though competition has gradually increased through reforms. Spain has pursued liberalization through mechanisms such as capacity markets and subsidies for renewable energy, though regulatory challenges and high consumer prices remain contentious issues.

In addition to market competition, Europe's electricity liberalization has been shaped by its ambitious climate goals. The EU's Green Deal and the Fit for 55 package have placed decarbonization at the forefront of energy policy, pushing member states to accelerate the shift towards renewable energy sources. This has created new market opportunities for renewable energy developers and has also sparked innovation in energy storage, demand-side management, and grid modernization. The EU's carbon pricing mechanism, the Emissions Trading System (ETS), plays a crucial role in driving investments in low-carbon technologies and disincentivizing fossil fuel-based power generation.

In contrast, other regions such as parts of Asia and Africa are at earlier stages of liberalization, often hindered by infrastructure and regulatory challenges. In many developing countries, electricity markets remain heavily centralized and dominated by state-owned utilities, with limited private sector participation. These regions face significant barriers such as outdated infrastructure, financial constraints, and political instability, making it difficult to implement market reforms and attract investment. Efforts to introduce competition and private investment in these regions often require substantial reforms to regulatory frameworks, as well as investments in grid infrastructure and capacity building.

In the United States, the process of electricity market liberalization is not uniform across states. Texas has fully deregulated its electricity market, resulting in increased consumer choice, innovation, and a thriving renewable energy sector. Other states, such as California and New York, have partially deregulated markets, focusing on promoting clean energy and enhancing grid resilience. In contrast, many states in the Southeast and Midwest have retained more traditional, vertically integrated market structures, where state-regulated utilities control generation, transmission, and distribution. This patchwork of regulatory approaches reflects the complexities of transitioning towards more competitive electricity markets in a federal system, where energy policy is largely determined at the state level.

The variations in liberalization efforts, both within countries like the United States and across global regions, highlight the complexities and regional specificities in transitioning towards more competitive electricity markets.Local economic conditions, regulatory frameworks, and infrastructure capacities are key factors that determine how quickly and effectively these changes occur. For instance, Europe's efforts to integrate markets and reduce carbon emissions have introduced complex issues in grid management, market design, and energy equity. As a result, each country or region must customize its liberalization strategy to address challenges such as market structuring, consumer protection, infrastructure investment, and decarbonization. Moving forward, the continuously evolving nature of electricity markets will demand adaptive policies and regulatory measures that balance competition, sustainability, and energy security on a global scale.

2.5 Analysis of the Modern Electricity Market

In this section, we will delve into the modern functions of electricity markets and analyze all the sectors that comprise them.

2.5.1 Why electricity is trading?

Before proceeding with the analysis of the mechanisms governing the modern electricity market, let us first answer a fundamental question: Why is electricity traded? Liberalized markets for electricity emerged following deregulation efforts in the late 20th century, which replaced vertically integrated monopolies with competitive frameworks. Power traders, utilities, and often large industrial or commercial electricity consumers engage in day-ahead trading to manage their electricity production and consumption effectively. This process helps them balance their electricity portfolio, ensuring that the amount of electricity they plan to consume or produce matches their actual needs. This balancing act is managed within what is known as a balancing group.

Entities responsible for these balancing groups, known as balancing responsible parties (BRPs), must ensure that the electricity volume within their group—both sold or consumed and bought or generated—is consistently balanced. To manage their portfolios and mitigate risks over the long term, these parties typically utilize futures markets available on power exchanges. In these markets, they can secure contracts to buy or sell electricity months or even years in advance, helping to stabilize their balancing group ahead of the actual delivery date.

Based on the above and in what is analyzed in the dissertation, electricity trading encourages investments in power generation technologies and show the long-term trends in prices and demand. It helps keep the grid stable by setting prices for services such as frequency regulation and reserve capacity. Market systems—whether centralized auctions or decentralized peer-to-peer platforms—improve the flow of electricity across regions and support the integration of variable renewable energy sources.

2.5.2 Modern Energy Market Mechanisms Analysis

This section introduces the fundamental concepts and mechanisms of energy markets, focusing on how modern electricity markets operate. It will provide an in-depth overview of the market structures, the roles of various stakeholders, and the regulatory frameworks that govern these markets. The emphasis will be placed on the Target Model, which is the current regulatory framework adopted by the European Union for electricity markets and aims to create an integrated and efficient energy market across Europe, promoting competition, improving supply security and facilitating the integration of renewable energy sources.

General Overview of the Main Participants of Energy Market

Energy markets are platforms for buying and selling electricity, divided into key segments. Wholesale markets handle large-scale electricity transactions between producers, suppliers, and major consumers. Retail markets serve end users, including households and businesses, by providing electricity through various suppliers. Ancillary services markets support grid stability and reliability by facilitating the trade of essential services, such as frequency regulation and reserve capacity.

The primary participants in energy markets include:

- Producers: Entities that generate electricity using various energy sources (e.g., coal, natural gas, nuclear, renewables).
- Suppliers: Companies that purchase electricity in bulk and sell it to end consumers.
- Transmission System Operators (TSOs): Organizations responsible for transporting electricity over long distances through high-voltage networks.
- Distribution System Operators (DSOs): Companies that manage the distribution of electricity to consumers through low-voltage networks.
- Consumers: End users of electricity, including residential, commercial, and industrial users.
- Regulators: Government bodies and agencies that oversee market operations and enforce regulations to ensure fair competition and protect consumers.

2.5.3 Europe's energy legislation phases

Europe's integrated energy market took shape through four key legislative phases. The process began with the First Energy Package, which launched Directive 96/92/EC for electricity and Directive 98/30/EC for gas. These laws laid the groundwork for shifting national energy systems from closed, state-controlled models to competitive markets. The Second Energy Package deepened these changes with Directive 2003/54/EC and Regulation (EC) 1228/2003. The regulation, in particular, tackled cross-border electricity infrastructure and trade rules, pushing countries to work more closely together. Over time, these laws broke down monopolies and aligned standards across EU member states, fostering a more interconnected and fair energy network [12], [13].

During the third stage of European energy policy reform, a package comprising two directives and three regulations was launched to transform electricity markets. Central to these reforms was Directive 2009/72/EC, which established uniform operational standards for the internal electricity market. Complementing this, Regulation (EC) 714/2009 set out procedures for managing cross-border grid access, while Regulation (EC) 713/2009 created ACER—the Agency for the Cooperation of Energy Regulators—to streamline oversight among national authorities. Enacted in September 2009, these policies governed market operations until Directive 2009/72/EC was eventually replaced by revised electricity market regulations in January 2021. This phase notably advanced market integration by resolving regulatory fragmentation and bolstering supranational governance structures [14].

Regarding the fourth package, known as the Clean Energy for All Europeans (CEP), it proposed eight interconnected measures: four directives and four regulations. Central to electricity sector reforms are Regulation (EU) 2019/943, which defines core principles for the EU's integrated electricity market—emphasizing competitive wholesale markets and transparent grid management—and Directive (EU) 2019/944, outlining frameworks for electricity production, transmission, distribution, supply, and storage. The latter prioritizes consumer rights, enabling active participation through mechanisms like dynamic pricing and renewable energy communities. Adopted in mid-2019, these rules modernized the bloc's energy architecture, phasing out earlier frameworks to align with decarbonization goals and digital innovation in the sector [14].

The fifth energy package, that is named ”Fit for 55”, was analyzed in Chapter 1. It includes a set of proposals and Directives aimed at achieving higher targets for greenhouse gas emissions reduction and increasing the share of renewable energy sources (RES) in the EU energy mix by 2030. The ”Fit for 55” targets include a 55% reduction in greenhouse gas emissions and a 40% increase in RES by 2030, relative to 1990 levels. Key elements of this package include revisions to the Emissions Trading System to increase the emission reduction obligation, amendments to the Energy Efficiency Directive to improve overall energy efficiency, the introduction of a Carbon Border Adjustment Mechanism to protect EU industries, and incentives for investments in renewable hydrogen energy projects.

These efforts have been translated into network codes that establish principles for the efficient and reliable operation of the market. These codes describe the design of the European internal electricity market and include significant legislation such as Directive 2009/72/EC, Directive (EU) 2019/944, and Regulation (EU) 2019/943. Additionally, Regulation (EU) 714/2009 outlines conditions for network access for cross-border electricity exchanges. Specific guidelines, such as the Electricity Balancing Guideline (EB GL) and the System Operation Guideline (SO GL), provide detailed rules for integrating balancing energy markets and harmonizing rules for transmission system operators (TSOs), distribution system operators (DSOs), and significant grid users.

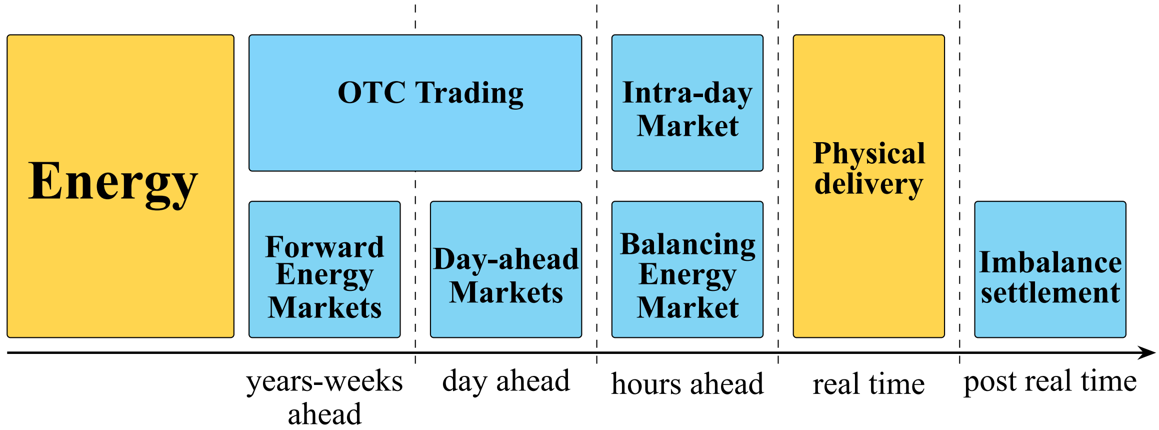
In Greece, these efforts towards an integrated energy market are embodied in the implementation of the Target Model, which has been in effect since November 1, 2020, based on Greek Ministry Laws 4425/2016 and 4512/2018 [15]. The reformation of the Greek electricity market began several years ago, with foundational legislation such as Law 2773/1999, which introduced the Regulatory Authority of Energy (RAE) to oversee and control the electricity market. Subsequent legislation, including Law 3426/2005, accelerated market liberalization, and Law 4001/2011 further reformed the sector by transitioning to the Independent Power Transmission Operator (IPTO) model and strengthening RAE's autonomy. Additional laws, such as 4336/2015, 4389/2016, and 4393/2016, introduced the NOME auctions to enhance market competition. The most recent regulatory reform is Law 1090/2018, which details the operation of the electricity market in Greece and establishes the Hellenic Energy Exchange (HEnEx) to oversee market transactions [14].

2.5.4 Target Model: Working Mechanism and Key Components

As analyzed above, the modern wholesale electricity market is a dynamic and complex system where electricity is bought and sold primarily through competitive bidding processes. It comprises various participants, including generators, retailers, and large consumers, who interact in day-ahead, intraday, and real-time markets to ensure an efficient and reliable supply of electricity. Market operators and power exchanges facilitate these transactions, ensuring transparency and fair pricing. The market incorporates advanced technologies and algorithms for demand forecasting, price setting, and grid management, while also integrating renewable energy sources to promote sustainability. Regulatory frameworks vary by region but aim to balance competition, reliability, and environmental goals.

Bilateral contracts affect electricity markets by providing stability and predictability in pricing for both buyers and sellers, reducing market volatility. They facilitate long-term planning and investment in energy infrastructure, and can help hedge against price fluctuations in the spot market. The general idea about Electricity Market working sequenceis presented in Figure 2.4.

The Target Model mechanism involves a structured approach utilizing four distinct markets: the Forward Market (FM), Day-Ahead Market (DAM), Intraday Market (IDM), and Balancing Market (BM). These markets collectively work to establish a unified electricity price across EU countries while managing the generation, transmission, delivery, and consumption of electricity at various stages.



Source: [14]

Figure 2.4: Energy Market Working Sequence

In terms of regulatory frameworks, they are essential for ensuring that energy markets function efficiently and fairly. The Target Model is designed to harmonize the rules and regulations governing electricity markets across member states. Key components of the Target Model include [16]:

- Price Coupling of Regions (PCR): Integrating national markets to facilitate cross-border electricity trading, enhancing market efficiency and price convergence.
- Capacity Allocation and Congestion Management (CACM): Rules for allocating transmission capacity and managing congestion to ensure optimal use of the grid.
- Balancing Markets: Mechanisms for maintaining the balance between electricity supply and demand in real-time, ensuring grid stability.

- Integration of Renewable Energy Sources: Policies and mechanisms to support the incorporation of renewable energy into the grid, promoting sustainability and reducing carbon emissions.

2.5.5 Price Coupling of Regions

Price Coupling of Regions (PCR), which was launched in June 2009, is a collaborative initiative by European Power Exchanges aimed at developing a single price coupling solution for calculating electricity prices across Europe, taking into account the constraints of transmission network capacity on a daily basis. This venture is crucial for achieving the overarching goal of the EU to create a harmonized European electricity market. A unified European electricity market is expected to increase liquidity, efficiency, and social welfare. The PCR initiative is open to other European power exchanges wishing to join and operates on three core principles: a common algorithm, robust operation, and the ability to handle decoupled energy exchanges. The common algorithm provides a fair and transparent determination of day-ahead electricity prices and a net position of a bidding zone across Europe. The algorithm was developed considering the specific characteristics of various energy markets throughout Europe and the constraints of the electricity network. It optimizes overall welfare and enhances transparency.

The project is operated by eight Power Exchanges: EPEX SPOT, GME, HEnEx, Nord Pool, OMIE, OPCOM, OTE, and TGE. The PCR is used to couple the following countries: Austria, Belgium, Czech Republic, Croatia, Denmark, Estonia, Finland, France, Germany, Hungary, Italy, Ireland, Latvia, Lithuania, Luxembourg, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom. The process is based on decentralized data sharing, ensuring robust and resilient operations. Finally, the PCR Matcher and Broker service enables the anonymous exchange of orders and transmission network constraints to compute bidding zone prices and other reference prices and net positions for all bidding zones [17].

2.5.6 Wholesale Energy Market Analysis

The purpose of this content is to elaborate in detail on the mechanisms and individual components that govern the operation of the modern electricity market.

2.5.7 Forward Market

The Forward Market (FM) functions by establishing electricity prices for future delivery, spanning from several years ahead up to the day before the actual delivery date. This market is designed to protect participants from the volatility of spot market prices by allowing them to lock in prices through long-term contracts. A crucial mechanism in the FM for mitigating risk from the Intraday Market (IDM) volatility is hedging. In this market framework, several financial instruments are used. Forward contracts are agreements to buy or sell a fixed amount of electricity at a specified price and future date. Unlike standardized exchanges, these contracts are traded over-the-counter (OTC), providing flexibility in terms of date and price customization. Futures contracts, on the other hand, are standardized agreements that obligate the buyer or seller to physically deliver electricity at a pre-agreed price and date, typically traded on futures exchanges such as the European Energy Exchange (EEX). These contracts standardize the quantity and quality of the electricity traded, offering greater certainty and ease of trading.

Another key instrument in the FM is options, which provide the right, but not the obligation, to buy or sell a specific quantity of electricity at a predetermined price during a future time frame. This flexibility allows the holder to decide whether to exercise the option based on market conditions, such as the availability of generating units and the pool price behavior. Unlike forward contracts, options come with an upfront cost that is non-refundable, known as the premium. The trading of futures and options on an exchange ensures that participants have clarity about prices, bids, offers, and access to clearing services, reducing the risk of trade defaults.

In the FM framework, electricity prices are set within bidding zones, which often span across national borders and may include multiple countries. Trading occurs both within and between these zones. Cross-zonal trading requires the allocation of transmission capacity, which is explicitly managed through auctions. Before engaging in cross-border electricity trades, market participants must first secure physical transmission rights (PTRs) in the forward capacity market. The Joint Allocation Office (JAO), comprising 25 TSOs from 22 EU countries, coordinates these allocations, serving as the Single Allocation Platform. From Greece, the Independent Power Transmission Operator (IPTO) is a member of JAO, holding a 5% stake in its share capital [16].

2.5.8 Day-Ahead Market

The Day-Ahead Market (DAM) is a crucial component of the Target Model for electricity trading. Transactions are scheduled for delivery on the day following the auction. At the close of each DAM session, the scheduled generation must match the forecasted demand, adjusted for net imports or exports. In this market, producers are required to participate, while other entities may choose to participate voluntarily. Each participant must submit a balanced portfolio, known as nominations, to the system operator, detailing the planned generation or consumption for each of their units. The price accepted in the DAM is set at the marginal price, which serves as the basis for the financial settlement of energy transactions [16].

Additionally, bilateral over-the-counter (OTC) contracts and contracts requiring physical delivery are included in the DAM, often originating from the Forward Market (FM) and explicitly allocated within the DAM framework. The integration of various EU countries' DAMs is facilitated through the implicit allocation of transmission capacity, meaning bids are accepted up to the point where congestion might occur.

The European Union is currently working towards the completion of the Single Day-Ahead Coupling (SDAC), a unified cross-zonal electricity market across Europe. This initiative is part of the Price Coupling of Regions (PCR) project and aligns with the Capacity Allocation and Congestion Management Guideline (CACM GL). The SDAC aims to optimize the allocation of limited cross-border transmission capacity by integrating the electricity markets of different regions using a common algorithm. This algorithm, known as the Pan-European Hybrid Electricity Market Integration Algorithm (EUPHEMIA), ensures compliance with CACM GL requirements. For effective market coupling, data from all involved Nominated Electricity Market Operators (NEMOs) and Transmission System Operators (TSOs) is processed, matching bids, offers, and network capacities with constraints. The resulting outputs, including matched trades, clearing prices, and scheduled exchanges, are then disseminated to NEMOs and TSOs. These operations are carried out within strict timelines to ensure optimal economic outcomes, high performance, and robustness. In 2023, a flow-based implicit allocation method will be implemented as part of the Core Flow-Based Market Coupling Project, enhancing the efficiency and reliability of the DAM [14], [18] .

In DAM, the bidding time period typically refers to the timeframe during which market participants submit their bids and offers for the next operating day. The general timelines and processes involved in the DAM, including the bid submission deadline and the delivery period are the followings [18]:

- Bid Submission Deadline: 12:00 PM (noon) on day D-1.
- Market Clearing and Results Publication: Usually by early evening on day D-1.
- Operating Day (Day D): The delivery period typically starts at 12:00 AM (midnight) and covers the full 24 hours of day D.

Figure 2.5 includes the DAM, as well as the IDA (Intra-Day Auction), which are detailed below.

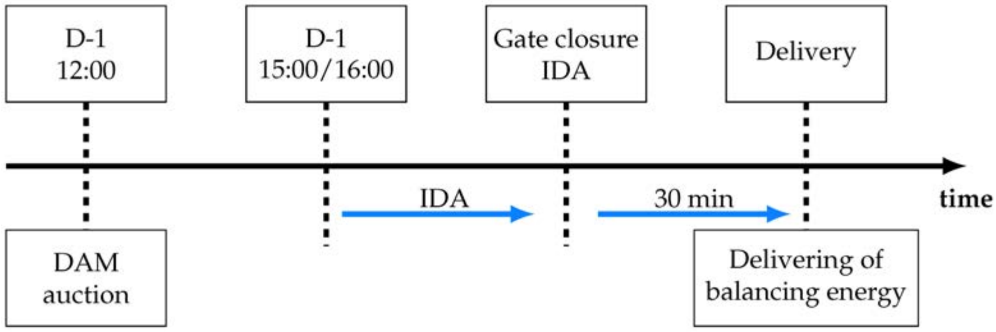


Figure 2.5: Day-Ahead Market (DAM) and Intra-Day Auctions (IDA)

This market aligns electricity supply and demand before production and delivery by gathering bids from producers and consumers to set hourly prices, ensuring market equilibrium. Market Pools or Power Exchanges publish these hourly prices for the following day. This information helps synchronize the anticipated electricity supply with expected customer usage. In the bidding process, participants indicate the volume of electricity they are willing to buy or sell each hour, along with the corresponding price levels in EUR/MWh (or USD/MWh).

The regulations for DAM trading can differ greatly between national markets. These variations include lead times, trading intervals, minimum tradable quantities, pricing mechanisms in the auction process, gate closure times, and other specific details. DAM results include the Market Clearing Price (MCP) for each market time unit of the delivery day, the net delivery position for each bidding zone, and the acceptance status and ratio of submitted orders, particularly for block orders. These results are subsequently integrated with the Intraday Market (IDM) through the exchange of essential information. This information includes (i) market schedules, which detail the energy schedule derived from the DAM, considering the accepted orders from generating units, Renewable Energy Sources (RES) portfolios, load portfolios, pumping units, RES feed-in-tariff (FiT) portfolios, rooftop photovoltaics, and the virtual balancing entity for the small connected system of Crete, and (ii) the scheduled imports and exports. The quantities from accepted sell and buy orders are recorded as physical delivery nominations and physical off-take nominations.

The clearing price for each market time unit is determined when the aggregated demand and supply curves intersect. The market operates continuously, covering 24 hours a day, 7 days a week. Nominated Electricity Market Operators (NEMOs) implement harmonized limits on the maximum and minimum clearing prices for the DAM as determined by the EUPHEMIA algorithm. According to ACER Decision 04/2017, the harmonized maximum MCP for the DAM is capped at +4000 EUR/MWh, while the minimum MCP is set at -500 EUR/MWh. These harmonized prices also consider the maximum value of lost load (VoLL) [19].

In Greece, clearing is conducted by EnExClear, based on the results from the DAM. EnExClear operates as a Clearing House for the Day-Ahead and Intraday Markets. It intervenes between the counterparties in transactions on the Energy Market (DAM and IDM) of the Hellenic Energy Exchange and assumes the role of buyer for every seller and seller for every buyer, handling the settlement of the financial aspects of these transactions.

Finally, the objective of the DAM auction, managed through the EUPHEMIA algorithm. This algorithm is trying to optimize the social welfare across the interconnected European DAMs. This entails maximizing the total surpluses from both sell and buy orders (DAM_S) within the integrated market, in addition to maximizing the congestion rent (DAM_CR) resulting from price disparities between neighboring bidding zones linked by interconnections.

2.5.9 Euphemia Algorithm

Euphemia is an algorithm designed to solve the problem of coupling day-ahead power markets in the Price Coupling of Regions (PCR) area. The development of Euphemia began in July 2011, based on the existing COSMOS algorithm, which had been in use in the Central Western Europe (CWE) region since November 2010. A stable version capable of covering the entire PCR scope was completed in July 2012. Since then, the algorithm has been continuously updated with corrective and evolutionary changes. Euphemia was first used in production on February 4, 2014, to couple the North Western Europe (NWE) and South-Western Europe markets in a synchronized mode. Subsequently, on February 25, 2015, the

Italian energy exchange GME was successfully integrated. On May 21, 2015, the Central Western Europe region was coupled for the first time using the Flow-based model. Additionally, on November 20, 2014, the 4M MC coupling was launched, integrating the markets of the Czech Republic, Hungary, Romania, and Slovakia [20].

Market participants begin by submitting their orders to their respective power exchanges. These orders are aggregated and processed by Euphemia, which determines which orders should be executed and which should be rejected, setting the corresponding prices. The primary goals are to maximize social welfare—which includes consumer surplus, producer surplus, and congestion rent across the regions—and to ensure that the resulting power flows do not exceed the capacity limits of the relevant network elements. This algorithm is capable of handling both standard and complex order types, adhering to all specified requirements. The algorithm quickly finds an initial viable solution and then iteratively improves it to enhance overall welfare. It is designed to be flexible, with no strict limits on the number of markets, orders, or network constraints, treating all orders of the same type equally [21].

2.5.10 Formulation of Marginal Price in DAM

In DAM, the marginal price, also known as the market clearing price (MCP), is determined through a market clearing process. This process involves the following steps [22]:

1. Bid Submission: Electricity generators submit bids indicating the quantity of electricity they are willing to supply at various price levels. Similarly, consumers or load-serving entities submit bids for the amount of electricity they are willing to purchase at different prices.

2. Bid Ranking: The market operator ranks the supply bids from the lowest to the highest price, creating a supply curve. This curve represents the order in which generation resources will be dispatched, starting with the least expensive options.

3. Demand Matching: The market operator matches the supply bids with the demand bids, starting from the lowest cost generation. The equilibrium point, where the total demand equals the total supply, determines the market balance.

4. Marginal Price Determination: The marginal price, or the system marginal price (SMP), is set at the highest price of the last unit of electricity required to meet the total demand.

All generators that are called upon to supply electricity receive this marginal price for the electricity they provide, irrespective of their bid price. Similarly, consumers pay this marginal price for all the electricity they consume.

5. Congestion and Locational Pricing: In some markets, transmission constraints can lead to different prices in different locations, known as Locational Marginal Prices (LMPs). LMPs reflect the cost of delivering electricity to specific locations, accounting for transmission losses and congestion. When transmission lines are congested, the cost of delivering electricity increases, resulting in higher LMPs in the congested areas.

2.5.11 Merit Order Principle

The merit order is a fundamental principle in electricity markets used to rank power plants based on their marginal costs, which is the cost of producing one additional unit of electricity. This ranking determines the sequence in which power plants are dispatched to supply electricity, starting with the cheapest available source. RES are positioned at the bottom of the merit order curve due to their very low marginal costs. Nuclear power follows, given its low running costs compared to fossil fuel plants. Fossil fuel plants like coal and natural gas are positioned higher up the curve, and the most expensive are the peak plants, such as those using diesel or gasoline, which are designed to cover short-term spikes in demand [23]. Figure 2.6 indicates the merit order principle formulation, as analyzed above.

Electricity producers submit bids indicating how much power they can provide at various price points, forming the supply curve. Concurrently, consumers submit demand bids, forming the demand curve. The market clearing price is set at the intersection of these two curves, determining the price at which all dispatched plants are compensated, irrespective of their individual costs. This market clearing price is critical as it reflects the most expensive marginal unit needed to meet demand. An increase in RES energy supply generally lowers the market clearing price, as these cheaper sources reduce the need for more expensive plants. This dynamic has led to decreased wholesale electricity prices in countries such as Germany and Belgium. However, these reductions in wholesale prices do not necessarily translate into lower prices for consumers due to other cost components, such as taxes, levies, and distribution fees [22], [24].

The difference between the market clearing price and the marginal cost for each plant is known as the inframarginal rent. These rents are essential for covering the fixed costs of power plants, ensuring the viability and profitability of power generation facilities. If the market clearing price consistently exceeds the sum of marginal and fixed costs, it signals an attractive environment for investment in new generation capacity, which is crucial for maintaining a secure and reliable electricity supply. To mitigate the investment risks associated with new generation capacity, grid operators may employ capacity mechanisms. These mechanisms provide additional payments to generators to ensure there is sufficient capacity available to meet peak demand, thereby enhancing the security of the electricity supply [25].

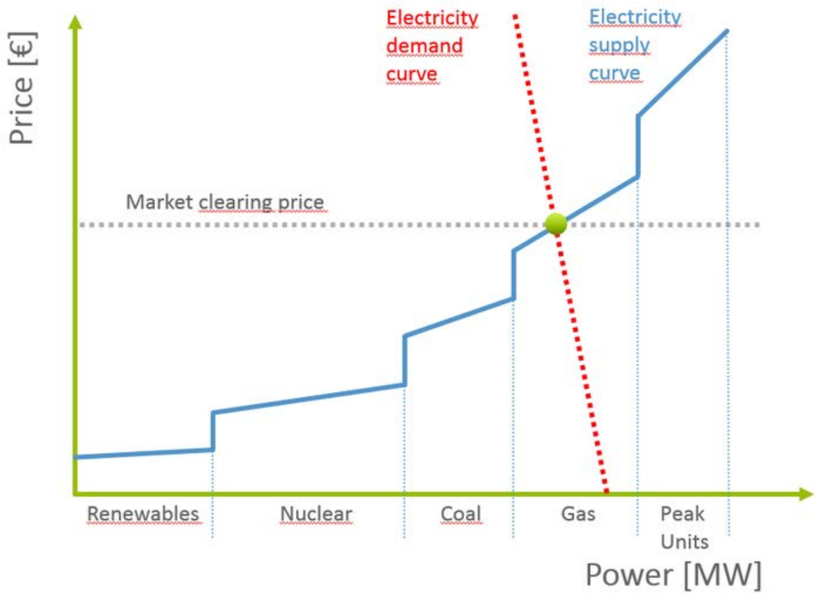


Figure 2.6: Merit Order Principle Formulation

2.5.12 Pricing Mechanisms in different DAMs

The specific rules and mechanisms for determining the marginal price can vary based on the market design and regulatory framework of each region. Some markets may include additional features such as capacity payments, ancillary services, and uplift payments to ensure reliability and encourage the availability of essential services beyond energy supply. The marginal price mechanism ensures the efficient use of resources by prioritizing the most cost-effective generation options. It also provides price signals to both generators and consumers, encouraging optimal production and consumption behaviors, such as increasing production during high-demand periods or reducing consumption when prices are high.

The main functions of day-ahead markets include balancing supply and demand by ensuring that anticipated electricity supply aligns with expected demand for the following day, maintaining market equilibrium through a fair and competitive auction process that matches bids and offers transparently, and optimizing resource allocation by signaling which generation plants should operate based on the forecasted prices.

In typical DAMs, from the bid submission deadline to the start of the delivery period, there are typically approximately 12 hours [26], [27], [28]. Tables 2.1, 2.2 and 2.3 present the typical DAM auctions of three different markets, the PJM Interconnection, the New York Independent System Operator and the California Independent System Operator, respectively.

Table 2.1: PJM Interconnection Schedule

PJM Interconnection	
Action	**Description**
Bid Submission Deadline	12:00 PM (noon) Eastern Time on day D-1
Market Results Published	Typically by early evening on day D-1
Delivery Period	12:00 AM (midnight) to 11:59 PM on day D

Table 2.2: NYISO Schedule

New York Independent System Operator	
Action	**Description**
Bid Submission Deadline	5:00 AM Eastern Time on day D-1
Market Results Published	By 10:00 AM Eastern Time on day D-1
Delivery Period	12:00 AM (midnight) to 11:59 PM on day D

Table 2.3: CAISO Schedule

California Independent System Operator	
Action	**Description**
Bid Submission Deadline	10:00 AM Pacific Time on day D-1
Market Results Published	By 1:00 PM Pacific Time on day D-1
Delivery Period	12:00 AM (midnight) to 11:59 PM on day D

2.5.13 Intraday Market

In Intraday Market (IDM) the electricity is trading on the same day as delivery. This market provides participants the opportunity to adjust their positions close to the time of physical delivery, particularly to correct deviations from the DAM and thus minimize their exposure to imbalance costs. About the main functions of IDM, these are [29]:

- Managing Imbalances: Allowing market participants to manage deviations from their scheduled energy deliveries, reducing penalties and costs associated with imbalances.
- Offer opportunities for market participants to optimize their positions and mitigate risks associated with imbalances.
- Enhancing Market Efficiency: Providing opportunities for market participants to adjust their positions closer to real-time conditions, improving overall market efficiency.

IDM nominations are integrated with day-ahead nominations, addressing upward and downward deviations. The financial settlement in the IDM is based on market equilibrium prices, which apply equally to both upward and downward adjustments. The IDM allows for adjustments to be made closer to the actual delivery time, accommodating changes in demand forecasts and generation availability. Here are the general timelines [30]:

- Continuous Trading: Intraday markets often operate on a continuous trading basis, allowing participants to trade up until a few hours before delivery.
- Gate Closure Times: These vary by market but are typically set a few hours before the delivery period starts. For example, gate closure might occur 1 to 2 hours before the delivery period.
- Delivery Period: The intraday market covers the hours remaining in the operating day (day D).

The methodology for pricing cross-zonal IDM capacity is outlined within the framework of the Cross-Border Intraday (XBID) trading and the Single Intraday Coupling (SIDC) solution. The XBID initiative is a collaborative effort to establish a common cross-border implicit continuous intraday trading solution across Europe, aiming to optimize the allocation of cross-border capacities. It is designed to enhance overall intraday trading efficiency and generate welfare benefits. This consortium includes several European Nominated Electricity Market Operators (NEMOs) like EPEXSPOT, GME, NordPool, OMIE, and various Transmission System Operators (TSOs). This solution supports both explicit allocation and implicit continuous trading across all bidding zone borders and aligns with the EU Target Model for an integrated IDM [31].

The adoption of the XBID project has led to the development of SIDC, which facilitates continuous cross-border trading throughout Europe, aiming to create a unified EU cross-zonal intraday electricity market. SIDC offers several benefits, including reducing the need for reserves and associated costs while maintaining system security, as it provides adequate time for system operation processes. Within SIDC, three pan-European auctions, known as Complementary Intraday Auctions (CRIDAs), are currently held and will eventually be replaced by regular intraday auctions. The IDM also utilizes an implicit continuous trade-matching algorithm, which operates on a first-come, first-serve basis, in line with the Capacity Allocation and Congestion Management Guideline (CACM GL). This algorithm integrates the IDMs across Europe [31].

The execution of CRIDAs for pricing cross-border capacity employs the EUPHEMIA algorithm, which calculates energy allocation, net positions, and electricity prices across Europe, thereby maximizing overall welfare and enhancing the transparency of price and power flow computations. Orders placed by market participants in one country can be matched with orders from participants in another, provided there is available cross-zonal capacity. SIDC supports both explicit and implicit continuous trading and adheres to the EU Target Model for an integrated cross-zonal IDM. The SIDC has been implemented in several phases, known as waves. The first go-live wave occurred in June 2018, including 15 countries, followed by a second wave in November 2019, which added seven more countries. A third wave, including Italy, occurred in September 2021, and Greece was expected to join the SIDC network by the end of 2022.

The primary objective of intraday trading is to maintain balance within a company's own balancing group through same-day trading. This practice helps fulfill balancing group commitments and minimizes potential imbalance costs. As power assets become more flexible, intraday trading allows for rapid power generation adjustments, optimizing profitability and maintaining system stability.

Mechanism and Operation of Single Intraday Coupling

The Single Intraday Coupling (SIDC) aims to establish a unified cross-border intraday electricity market across the European Union, enabling market participants to trade electricity continuously on the day it is needed. This market integration enhances efficiency by promoting competition, increasing liquidity, facilitating the sharing of generation resources, and allowing for adjustments in response to unexpected changes in consumption or generation outages. As renewable energy sources, particularly intermittent ones like solar, become more prevalent, the importance of intraday markets grows, as they allow for more precise balancing of supply and demand closer to real-time [32].

SIDC is a collaborative initiative involving Nominated Electricity Market Operators (NEMOs) and Transmission System Operators (TSOs). It builds upon the Cross Border Intraday Project (XBID), which introduced a platform for continuous trading in June 2018. Initially launched across 15 countries, SIDC has since expanded to include 25 countries by November 2022, following several phases of integration. The platform operates under the Capacity Allocation and Congestion Management (CACM) EU Target Model, which aims to standardize and optimize cross-border electricity trading across Europe. The SIDC platform utilizes three key components:

- Shared Order Book (SOB): Aggregates orders from market participants across different countries, facilitating the matching of buy and sell orders.

- Capacity Management Module (CMM): Manages the available cross-border transmission capacity, ensuring that trades do not exceed network limits.

- Shipping Module (SM): Handles the logistical aspects of electricity delivery, ensuring that physical exchanges match the financial transactions.

These components enable both explicit and implicit trading. In explicit trading, participants trade transmission capacity separately from the energy itself, while in implicit trading, capacity and energy are traded together. This system allows for seamless and efficient market operations, as orders can be matched across borders provided sufficient transmission capacity is available.

Market Dynamics and Matching Process

In the SIDC framework, the matching of orders is conducted on a first-come, first-served basis, prioritizing the highest buy prices and the lowest sell prices. When a market participant places an order, it is immediately compared with existing orders in the Shared Order Book (SOB). If a matching order is found, the trade is executed, and the involved capacities are updated in the Capacity Management Module (CMM). This real-time update mechanism ensures that the most current data is always used for decision-making, enhancing the market's responsiveness and reliability.

Impact of Renewable Energy and Market Inframarginal Rents

The increasing integration of renewable energy sources, which often have low or zero marginal costs, has significant implications for the SIDC and wholesale electricity prices. As these sources are incorporated into the market, they typically displace more expensive forms of generation, leading to lower overall market clearing prices. This phenomenon has been observed in markets like Germany and Belgium, where the growth of renewable energy has contributed to a downward trend in wholesale electricity prices. However, the reduction in wholesale prices does not always lead to lower consumer bills, as these also include taxes, levies, and distribution costs.

The concept of inframarginal rents—where power plants earn more than their operational costs— plays a crucial role in covering fixed costs and incentivizing investment in new generation capacity. The presence of inframarginal rents indicates a healthy market that can attract new investments, which are essential for maintaining and expanding the energy infrastructure. To further support investment, grid operators may implement capacity mechanisms, providing additional financial incentives to ensure sufficient capacity is available to meet peak demand and enhance system reliability.

Different Intraday Markets

Tables 2.4 and 2.5 present the timing auctions of European Power Exchange - EPEX SPOT and Nord Pool.

Table 2.4: EPEX SPOT (European Power Exchange) Trading Features

European Power Exchange - EPEX SPOT	
Action	**Description**
Continuous Trading	Available up to 5 minutes before delivery in some regions.
Gate Closure	Typically 30 minutes to 1 hour before delivery in most regions.
Delivery Period	Covers the remaining hours of the operating day.

Table 2.5: Nord Pool Trading Features

Nord Pool	
Action	**Description**
Continuous Trading	Available up to 1 hour before delivery.
Gate Closure	1 hour before the delivery period starts.
Delivery Period	Covers the remaining hours of the operating day.

2.5.14 Balancing Market

The primary function of the Balancing Market (BM) is to address imbalances between electricity supply and demand in real-time. In this context, the TSOs of each country procure both the necessary balancing capacity and energy to address real-time discrepancies between production and demand side. The BM can be categorized into three main submarkets [19]:

- Balancing Capacity Market: Ensures that sufficient reserves are available.
- Balancing Energy Market: Activates energy in real-time to maintain system balance while meeting energy and reserve demands.
- Imbalance Settlement Process: Allocates BM revenues and costs among participants.

In the balancing capacity market, BRPs (balancing responsible parties) are required to submit their forecasted energy production and consumption to the system operator one day before delivery. Contracts in this market can be secured as early as a year ahead of delivery or as late as one day prior. Once these agreements are in place, the same parties then offer balancing energy on the respective markets. In the balancing energy market, BSPs (balance service providers) present bids to the system operator, who procures the necessary energy to keep the system balanced, with the volume of energy activated being determined by real-time imbalances. Any differences between the BRPs' forecasts and the actual activated energy are reconciled during the imbalance settlement process, where the TSO is responsible for deploying reserves and setting tariffs for these imbalances.

The BM is used to ensure real-time balance between supply and demand. It involves the activation of reserves to handle unforeseen imbalances. Here are the general timelines:

- Real-Time Operation: The Balancing Market operates in real-time, with continuous adjustments being made to balance supply and demand.
- Activation of Reserves: Reserve power is activated as needed to address imbalances. This can occur at any time during the operating day.
- Settlement: Imbalances are settled after the operating day, based on the actual deviations from scheduled supply and demand.

Figure 2.7 visualizes the Balancing Market Auction, along with DAM and IDA.

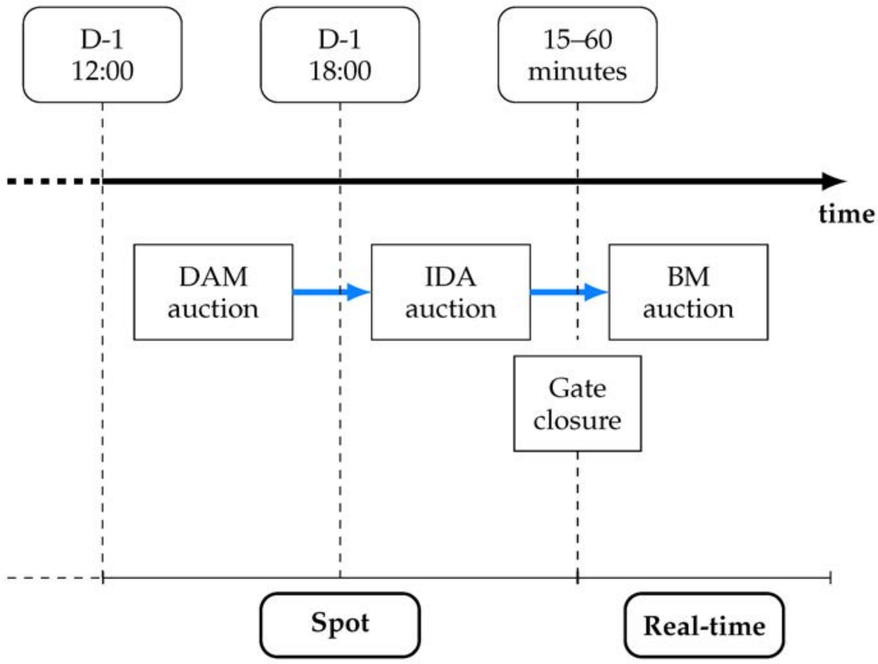


Figure 2.7: Balancing Market Auction

The BM utilizes various types of reserves, including frequency containment reserves (FCRs) for immediate frequency stabilization, manual frequency restoration reserves (mFRR) for restoring system balance within seconds to 15 minutes, automatic frequency restoration reserves (aFRR) for similar purposes but with automated control, and replacement reserves for addressing major imbalances when FRRs are insufficient. These reserves are activated as needed, with FCRs responding instantly, mFRR and aFRR handling mid-term imbalances, and replacement reserves addressing longer-term issues.

A key aspect of calculating imbalance prices is the imbalance settlement period, which determines the frequency of imbalance price signals sent to BRPs. These prices reflect real-time balancing costs, which are allocated to BRPs to encourage them to maintain balanced energy portfolios. BSPs providing upward regulation receive compensation based on either a marginal price or a bid price, depending on the pricing mechanism used. Conversely, BSPs providing downward regulation pay the respective downward regulation price or bid price.

The Electricity Balancing Guideline (EBGL) regulation and the recast of electricity regulation outline clear provisions for imbalance settlement, including establishing a harmonized time unit of 15 minutes for calculating BRP imbalances. A core component of the EB GL regulation is the implementation of platforms for the exchange of balancing energy, with projects like Trans-European Replacement Reserves Exchange (TERRE) for RR, Manually Activated Reserves Initiative (MARI) for mFRR, Platform for the International Coordination of Automated Frequency Restoration and Stable System Operation (PICASSO) for aFRR, and International Grid Control Cooperation (IGCC) for imbalance netting (IN).

Within the BM framework, redispatching of generation units also occurs. Redispatching involves system operators modifying the generation or load pattern to alter physical flows in the transmission system and alleviate congestion. This measure is necessary when market outcomes lead to potential violations of operational limits within a bidding zone. The Clean Energy Package (CEP) mandates that redispatching should be organized in a market-based manner. In most EU countries, generation units are legally required to participate in redispatching, and compensation is provided based on regulated prices or prior opportunity costs from the wholesale market.

Remuneration for activated redispatching, whether internal or cross-zonal, varies across EU countries. Different pricing mechanisms are used, including pay-as-bid pricing, regulated pricing based on market prices like day-ahead prices, or cost-based pricing that covers fuel and other related costs.

Concerning the remuneration of the aforementioned products, starting from the launch of the Balancing Market (BM) on November 1, 2020, until the initiation of continuous trading in the SIDC, the maximum and minimum price limits for balancing energy were set at +4,240 EUR/MWh and -4,240 EUR/MWh, respectively. Similarly, the price limits for balancing capacity were established at +3,000 EUR/MWh and 0 EUR/MWh. However, after the implementation of Continuous Intra-Day Trading (XBID), the price limits for balancing capacity changed to +9,999 EUR/MWh and 0 EUR/MWh. Additionally, following the introduction of Coordinated Redispatching and Countertrading Initiatives (CRIDAs) and before the Independent Power Transmission Operator (IPTO) integrates into the European platforms MARI or PICASSO, the price limits for balancing energy were adjusted to +9,999 EUR/MWh and -9,999 EUR/MWh [19].

In Greece, IPTO is working on reforming the existing BM through participation in the MARI and PICASSO European balancing platforms. These platforms aim to standardize local manual Frequency Restoration Reserve (mFRR) and automatic Frequency Restoration Reserve (aFRR) processes, in line with the standard products outlined in ACER Decision 01/2020. Both platforms are scheduled to become operational in Greece by 2022. Specifically for aFRR, the full activation time will be shortened from 7.5 minutes to 5 minutes by December 2024, with a minimum bid size of 1 MW, consistent with the French Transmission System Operator's requirements. Upon IPTO's integration into either MARI or PICASSO, the price limits for balancing energy will adjust to +15,000 EUR/MWh and -15,000 EUR/MWh, respectively, within a maximum of 48 months from the implementation deadline specified in Articles 20 and 21 of the Electricity Balancing Guideline (EBGL). Forty-eight months after this deadline, the price limits will further expand to +99,999 EUR/MWh and -99,999 EUR/MWh [19].

The final component of the Greek Balancing Market involves the imbalance settlement market, which operates post real-time to compensate or charge participants for any discrepancies between their actual energy usage and the final market schedule or dispatch orders. This process ensures a transparent calculation of any imbalances and their corresponding compensation. Bids not related to balancing purposes, such as those for redispatching or voltage control, are excluded from this imbalance price calculation. The imbalance settlement period, which has been set to 15 minutes since November 1, 2020, aligns with the requirements of ACER Decision 18/2020.

The pricing in the BM reflects the real-time cost of adjusting electricity supply or demand, often resulting in higher prices compared to the DAM due to the immediate nature of the required actions. The costs incurred in the BM are typically recovered through imbalance charges levied on market participants whose deviations from their scheduled positions necessitated the balancing actions. This mechanism incentivizes participants to adhere closely to their scheduled commitments and enhances the overall efficiency and reliability of the power system. Additionally, the BM plays a vital role in integrating renewable energy sources, which are often variable and less predictable. By providing a responsive and flexible platform for managing these variations, the BM supports the increased penetration of RES into the grid. Overall, the Balancing Market ensures the continuous, real-time alignment of electricity supply and demand, thereby maintaining grid stability and operational reliability while promoting efficient market behavior.

All the market mechanisms that were analyzed above constitute a cornerstone for the Demand Side Management and Demand Response programs, which are presented in the following chapter.

Chapter 3

Demand Side Management: Advanced Strategies and Innovations in the Energy Market

In this chapter, the field of Demand Side Management will be extensively presented. The importance of active consumer participation in the operation of the energy market will be analyzed, and then two experimental approaches will be presented. The first investigates an incentive-based demand response program, and the second involves an extensive simulation with direct load control on residential consumers.

3.1 Consumer Active Participation in the Operation of the Energy Market

Customer active participation in the energy market, often referred to as prosumerism, marks a paradigm shift in how consumers interact with the energy system. Traditionally, consumers were passive participants, merely consuming energy supplied by centralized utilities. However, technological advancements and evolving market structures are enabling consumers to become active participants, who not only consume but also produce, manage, and even sell energy. This transformation is driven by several key factors, including the proliferation of DERs, the rise of smart technologies, and the development of supportive regulatory frameworks.

First, the availability and affordability of DERs, such as rooftop solar panels, small-scale wind turbines, and home energy storage systems, are crucial enablers of customer active participation. These technologies allow consumers to generate their own electricity, thereby reducing their reliance on the grid and lowering their energy bills. In many cases, consumers can also sell excess energy back to the grid, participating in feed-in tariff schemes or similar programs, thereby generating additional revenue. This shift to decentralized generation disrupts traditional energy market dynamics. It transforms consumers into prosumers—both producers and consumers of energy—who can influence supply and demand. As prosumers, customers can help stabilize the grid by providing additional power during peak times or reducing consumption during periods of low generation. This decentralized approach can also enhance energy security and resilience, as it reduces the dependence on centralized power plants and large-scale infrastructure.

Also, Smart technologies Systems play a pivotal role in enabling and optimizing customer participation in the energy market. Smart meters, for example, provide real-time data on energy consumption and generation, enabling consumers to make informed decisions about their energy use. Home energy management systems (HEMS) can further optimize energy consumption by automating appliances based on real-time data and dynamic pricing signals, such as time-of-use rates or real-time market prices. The integration of Internet of Things (IoT) devices allows for more sophisticated energy management. For example, smart thermostats can adjust heating and cooling settings based on occupancy patterns, weather forecasts, and energy prices. Similarly, smart appliances can operate during off-peak hours when energy is cheaper and cleaner, further reducing costs and carbon footprints. These technologies not only empower consumers to reduce their energy bills but also contribute to grid stability and efficiency. By providing a more predictable and manageable load profile, consumers can help smooth demand peaks and valleys, reducing the need for expensive peaking power plants and enhancing the integration of renewable energy sources.

Additionally, active customer participation introduces new dynamics into the economic energy market and implications. It fosters competition and innovation, as traditional utilities are challenged to adapt to a more distributed and decentralized system. New business models are emerging, including energy aggregators who bundle the flexibility of numerous small prosumers to sell as a larger, more valuable resource to the grid. These aggregators can participate in demand response programs, providing grid services such as frequency regulation, voltage support, and spinning reserves.

Furthermore, prosumers can participate in peer-to-peer (P2P) energy trading platforms, where they can buy and sell energy directly with other consumers. This not only empowers consumers with more control over their energy sources and costs but also promotes local energy markets and can help reduce transmission losses and infrastructure costs [33].

3.2 Regulatory and Policy Framework

The growth of customer active participation necessitates the evolution of regulatory frameworks to accommodate and encourage this new dynamic. Policies that support the integration of DERs, such as net metering, feed-in tariffs, and incentives for energy storage, are critical. Regulators must also address the challenges associated with bi-directional flows of electricity and the need for enhanced grid infrastructure to manage distributed generation. Moreover, regulators face the challenge of ensuring fair access to these technologies and market opportunities. There is a risk that the benefits of active participation may be disproportionately accessible to wealthier consumers who can afford the upfront costs of DERs and smart technologies. Policymakers must therefore consider measures to ensure that low-income households can also participate, such as subsidies, financing options, or community energy programs.

Despite its potential benefits, active customer participation poses several challenges. From a technical perspective, the increased penetration of DERs and the variability of renewable energy sources can complicate grid management and require significant upgrades to existing infrastructure. There are also challenges related to data privacy and cybersecurity, as smart technologies and digital platforms collect and process vast amounts of sensitive data. From a market perspective, the integration of a large number of small, decentralized energy resources requires new market structures and rules. This includes developing mechanisms for valuing and compensating the services provided by prosumers, such as capacity, ancillary services, and environmental benefits. It also involves creating transparent and accessible platforms for P2P trading and aggregation.

Looking forward, the role of prosumers is likely to expand as technologies continue to advance and as market and regulatory frameworks evolve to support more active participation. This could lead to a more democratized and resilient energy system, where consumers are not just passive recipients of energy but active participants in a dynamic and interactive market ecosystem. The continued development of digital platforms, enhanced grid capabilities, and supportive policies will be crucial in realizing the full potential of customer active participation in the modern energy market.

3.3 Role of Flexibility in the Energy Market

The concept of flexibility and active customer participation in modern energy markets is integral to the transformation of the energy landscape. As the energy sector undergoes a significant transition towards sustainability, the roles of both flexibility and customer engagement have become critical in ensuring the stability, efficiency, and resilience of electricity systems. Flexibility refers to the ability of the electricity system to efficiently balance supply and demand under varying conditions. This capability is increasingly vital as the grid integrates a growing share of intermittent renewable energy sources, such as wind and solar power, which are characterized by their variability and unpredictability. Flexibility can be provided by various means, including demand-side response (DSR), energy storage systems, flexible generation, and interconnection with other grids [34].

The increasing share of renewable energy, such as wind and solar power, introduces variability into the electricity supply due to their dependence on weather conditions. This variability requires additional mechanisms to balance supply and demand. Flexibility in the energy market is achieved through various means, including [35]:

- Demand Response: This involves the adjustment of electricity consumption by end-users, either manually or automatically, in response to price signals or other incentives. This can include reducing or shifting demand during peak periods or increasing consumption during times of low demand and surplus generation. Energy storage systems, such as batteries, offer another avenue for flexibility by storing excess energy during periods of low demand and discharging it when demand is high. Additionally, flexible generation sources, like gas-fired power plants, can quickly ramp up or down their output to match demand fluctuations. Finally, interconnection with other grids allows for the sharing of resources across regions, enhancing overall system flexibility.

- Energy Storage: Technologies like batteries and pumped hydro storage can store excess energy when supply exceeds demand and release it when there is a deficit. Energy storage helps smooth out fluctuations in renewable energy generation and provides backup power.

- Flexible Generation: This includes power plants that can quickly ramp up or down their output to match the grid's needs. Examples include gas-fired power plants, which can respond rapidly to changes in demand.

- Interconnections and Grid Infrastructure: Enhanced grid infrastructure and interconnections between regions allow for the sharing of resources and balancing of supply and demand across wider areas. This reduces the reliance on local generation and increases system resilience.

Market mechanisms and regulatory frameworks play a vital role in promoting flexibility. For example, electricity markets can implement real-time pricing, which reflects the true cost of electricity at different times and encourages consumers to adjust their usage accordingly. Ancillary services markets can also be established to provide additional capacity and frequency regulation services. Despite the benefits, achieving flexibility poses several challenges. These include the need for significant investment in new technologies, regulatory adjustments, and the development of new business models. Furthermore, as the share of renewables continues to grow, the need for more sophisticated grid management and forecasting tools becomes increasingly important. The future of flexibility in the energy market will likely involve a greater emphasis on digitalization, smart grid technologies, and the integration of decentralized energy resources. These advancements will help create a more dynamic and responsive energy system, capable of meeting the demands of a decarbonized and electrified future.

3.3.1 Challenges and Problems of Flexibility

Despite the benefits, there are several challenges to achieving widespread flexibility and active customer participation. Technical challenges include the need for enhanced grid infrastructure and advanced control systems to manage the increased complexity and variability introduced by DERs and flexible resources. There are also issues related to data privacy and cybersecurity, as the increased use of digital technologies and data exchange in smart grids raises concerns about the protection of consumer information.

From a policy perspective, ensuring equitable access to the benefits of the modern energy market is crucial. There is a risk that the transition to a more flexible and participatory energy system could widen existing inequalities if only affluent customers can afford the necessary technologies, such as solar panels and home batteries. Thus, policy measures should aim to ensure that all consumers, regardless of income, can participate in and benefit from the evolving energy market.

Also, flexibility and active customer participation are pivotal to the successful transition to a modern, sustainable energy system. They offer significant benefits in terms of efficiency, cost savings, and enhanced grid stability. However, realizing these benefits requires overcoming technical, economic, and regulatory challenges. As the energy market continues to evolve, it will be essential to develop policies and technologies that support widespread participation and ensure that the benefits of the modern energy system are accessible to all.

3.4 Implementation and Challenges of Demand Side Management

Demand Side Management sector is a critical component of modern electricity markets, enabling active consumer participation and contributing to grid stability, economic efficiency, and the integration of renewable energy sources. These programs are expected to play a very significant role, as well as pose challenges, in modern energy systems. More specifically, regarding their implementation, we have the following:

- Smart Grids and Advanced Metering Infrastructure (AMI): The deployment of smart grids and AMI enables real-time monitoring and control of energy usage, facilitating the implementation of DSM strategies. Smart meters provide consumers with detailed information about their energy consumption patterns, allowing for more informed decisions.

- Automated Demand Response (ADR): ADR systems automatically adjust power consumption in response to signals from grid operators or utilities. This can involve pre-programmed responses in smart appliances, thermostats, and industrial equipment, enhancing the responsiveness and effectiveness of demand response programs.

- Consumer Education and Engagement: Educating consumers about the benefits and methods of DSM is crucial for its success. Programs that raise awareness and provide incentives for participation can lead to significant changes in consumption behavior.

- Regulatory Support and Market Incentives: Effective DSM requires supportive regulatory frameworks and market incentives that encourage both consumers and utilities to engage in demand-side activities. Policies that promote dynamic pricing, subsidies for energy-efficient technologies, and financial rewards for demand response participation are essential.

While DSM offers numerous benefits, there are also challenges to its widespread adoption. The adoption of advanced DSM programs faces several challenges. High technological and infrastructure costs can hinder implementation for both utilities and consumers. Encouraging widespread consumer participation is another hurdle, requiring effective communication, incentives, and educational efforts. Additionally, regulatory and market structures must adapt to support DSM, including the development of flexible pricing models and market rules that encourage active consumer engagement. Furthermore, the collection and use of detailed consumption data raise concerns about privacy and security, making it essential to establish robust data protection measures to ensure consumer trust and program success.

The following sections will explore the development and application of advanced DSM techniques and their impact on contemporary energy market operations.

3.4.1 Techniques and Categories of Demand Side Management

Six basic types have been established as techniques for DSM to control the consumer load curve, as shown in Figure 3.1.

1. Peak Clipping: This technique is designed to reduce peak load by either directly controlling consumer equipment or through pricing agreements that require customers to lower their consumption at specific times of the day [36].

2. Valley Filling: Programs using this method seek to boost energy usage during off-peak periods, thereby smoothing the overall load curve. As a result, power plant components—such as generators, transformers, and both transmission and distribution lines—operate

at approximately 80–90% of their rated capacity instead of only 15–20% during low-demand hours. This leads to increased efficiency and reduced operating costs thanks to an improved load factor [37].

3. Load Shifting: Load shifting involves moving consumption from peak to off-peak periods. While this reduces peak demand, the total energy consumed remains unchanged, making this one of the most crucial load management strategies [37], [36].

4. Load Reduction (Energy Conservation): Also known as energy conservation, this method focuses on decreasing electricity usage uniformly throughout most or all hours of the day. Although it manages overall consumption on a general basis—often including reductions in electricity prices and adjustments in usage patterns—it is not typically classified as a traditional load management technique [37], [36].

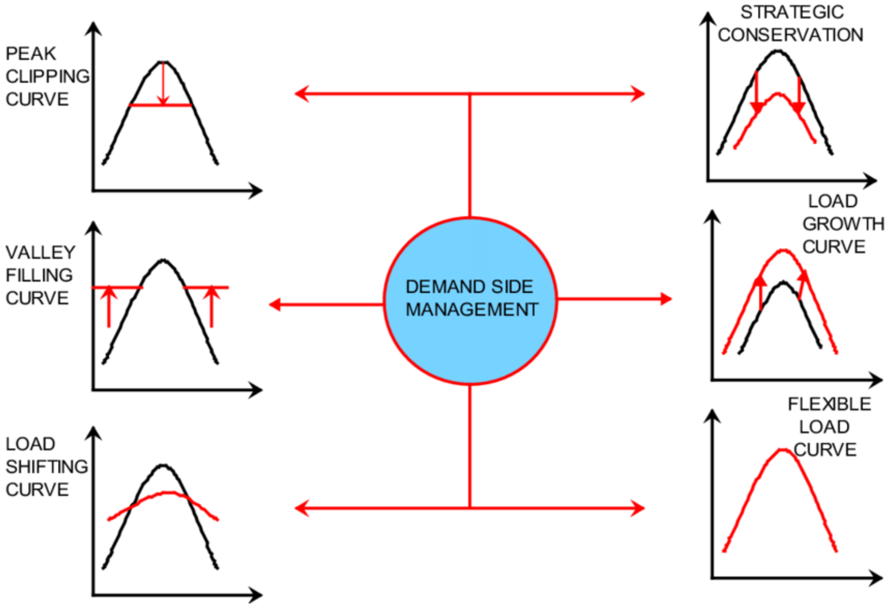
5. Load Growth (Load Building): This approach aims to increase market demand, thereby raising overall electricity sales. It does so by promoting new applications, such as investments in industrial automation systems and advanced electric vehicles. Figure 5 illustrates the resulting upward shift in the demand curve [36].

6. Flexible Load Shape: The core concept of this method is to establish contracts between utilities and consumers, whereby participants agree to modify their electricity usage when needed in exchange for financial incentives. These programs empower consumers to adjust their demand curves to either address immediate grid requirements or to help secure the system's energy reserves [36].

3.4.2 Categories of Demand Side Management

Based on the timing and impact of applied strategies on consumer energy usage, DSM can be categorized into four main types, the first three of which are presented in Figure 3.2 [39]:

1. Energy Efficiency (EE): Long-term measures aimed at reducing energy consumption without compromising performance. Energy Efficiency refers to measures taken by a consumer or the electricity transmission management company to mitigate energy waste. This includes permanent equipment upgrades, such as replacing an outdated, low-performance ventilation system with a modern, high-efficiency alternative, or improving a system's technical characteristics (e.g., adding additional insulation to a



Source: [38]

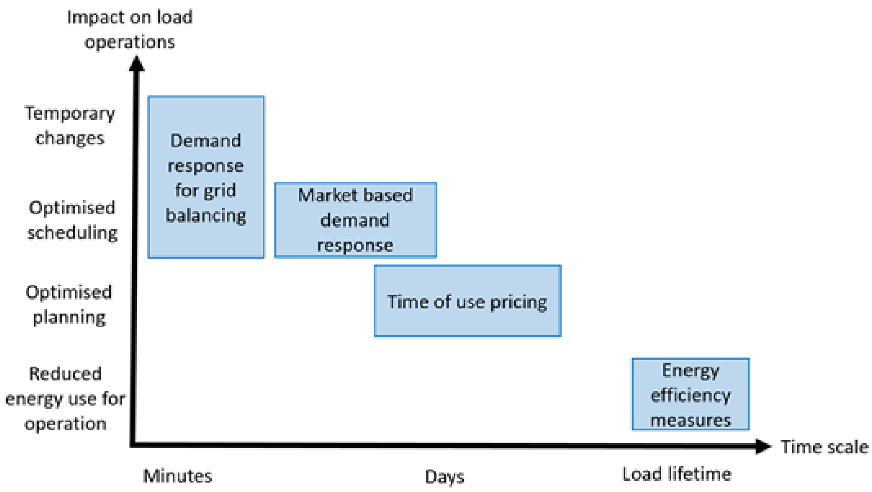
Figure 3.1: Demand Side Management Techniques

building) to reduce losses and enhance overall efficiency. These measures result in immediate and long-term energy savings, lower greenhouse gas emissions, and contribute to the efficient use of technical and financial resources.

2. Time of Use (TOU): The primary objective of TOU tariffs is to encourage consumers to shift their electricity usage away from peak periods, thereby optimizing energy production and improving equipment utilization. This shift extends the lifespan of energy infrastructure and reduces overall system costs. The variation in rates across different tariff zones reflects the expected benefits for all stakeholders involved in energy generation and distribution.

   Implementing variable electricity pricing based on demand fluctuations throughout the day. Time of Use (TOU) is a load management approach that involves measuring and billing electricity consumption based on the time it occurs. TOU tariffs are a common method of managing consumer demand, where electricity prices are higher during peak demand periods and lower during off-peak hours. These pricing schedules are predetermined and communicated to consumers in advance, enabling them to adjust their consumption habits accordingly.

   One key advantage of TOU programs is their ability to accommodate high electricity demand (e.g., heating) during low-load periods when market prices are below average.



Source: [41]

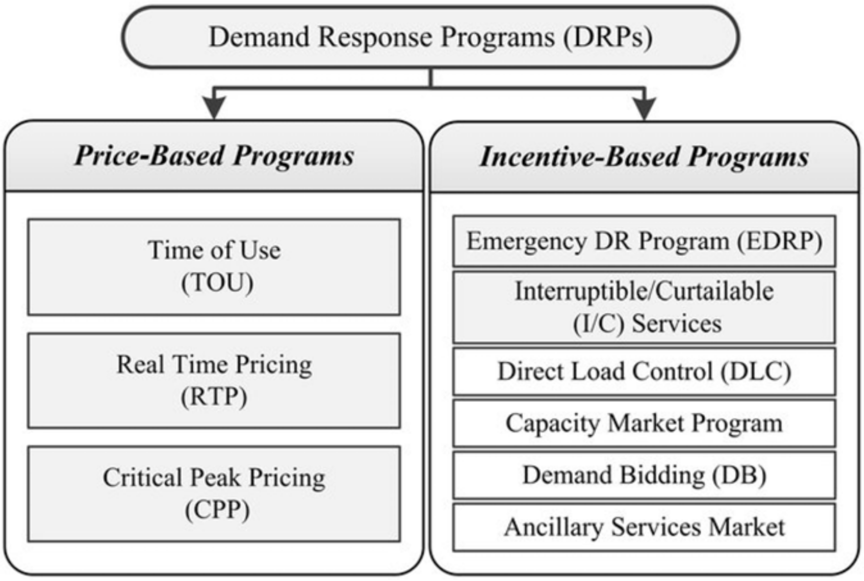
Figure 3.2: Categories of Demand Side Management

This enables electricity to remain competitively priced in comparison with alternative energy sources. Consumers benefit from significantly lower rates during off-peak hours, providing a financial incentive to adjust their consumption patterns. Empirical studies have demonstrated that TOU pricing effectively reduces peak energy demand. TOU can also be considered a basic form of Demand Response, though it lacks two-way communication and dynamic interaction between energy suppliers and consumers.

3. Demand Response (DR): Adjusting energy consumption patterns in response to grid conditions or price incentives. Is is categorized in Price or Market-based DR and Incentive-based DR programs, all of which are analyzed in detail below.

4. Spinning Reserve (SR): Utilizing reserve power sources to balance fluctuations in demand and supply. The term Spinning Reserve is widely used in the electricity systems community and refers to the reserve power output available to generate active power within a given timeframe in response to operational disturbances (e.g., generator failures, transmission line faults) or peak demand periods. The primary function of SR is to restore balance between generation and consumption, regulate frequency back to nominal levels, and maintain stable power exchange across interconnected systems. Reducing peak loads enhances reserve capacity, thereby increasing system security margins [40].

Additionally, based on their type and the manner in which participating consumers adjust their load profiles, DR programs are divided into two main groups: Price-based and Incentive-based programs. Figure 3.3 illustrates these two categories along with their respective sub-categories.



Source: [42]

Figure 3.3: Categorization of Demand Response programs

3.5 Comparison of Price-based and Incentive-based Programs

The key distinction between these two categories lies in their approach to demand reduction. In Incentive-based Demand Response programs, customers receive financial compensation for reducing a specific amount of load within a given time frame. Conversely, in Price-based programs, customers voluntarily adjust their consumption in response to economic signals that fluctuate based on market prices. Unlike Incentive-based programs, Price-based programs do not require participants to meet a predefined reduction target [39].

Price-based Programs

In Price-based programs, consumers modify their electricity usage based on energy prices provided by utilities. The primary types of Price-based programs include [43]:

- Time-Of-Use rates (TOU): A predetermined pricing structure where electricity rates vary depending on the time of consumption.

- Real-Time Pricing (RTP): Consumers are charged dynamic prices that fluctuate at short intervals, reflecting real-time market conditions.

- Critical Peak Pricing (CPP): Utilities set higher prices during anticipated peak demand periods or system emergencies to encourage reduced consumption.

Incentive-based Programs

Incentive-based programs offer financial rewards or direct control mechanisms to encourage customers to adjust their energy usage. The main types include [43]:

- Emergency Demand Response Programs (EDRP): Participants voluntarily reduce consumption in response to emergency signals.

- Interruptible/Curtailable rates (I/C): Customers agree to lower electricity prices in exchange for the obligation to reduce consumption during high-demand periods.

- Direct Load Control (DLC): Utilities or system operators remotely manage customer electricity usage by controlling appliances or equipment.

- Capacity Market Programs: Customers commit to providing load reductions when needed to help maintain grid stability.

- Demand Bidding Programs (DB): Consumers submit bids to limit their electricity usage at competitive prices.

- Ancillary Services Market: Customers contribute to power system reliability by providing services that help maintain grid stability and power quality.

3.6 Demand Response Programs

Demand Response emerged as an innovative strategy to maintain optimal grid capability and utility. Over the past two decades, these programs have attracted significant attention from the global research community because they offer novel ways to balance electricity supply and demand. Rather than requiring grid operators to activate additional industrial power sources during peak periods, Demand Response shifts the responsibility to consumers by curtailing demand. This approach enables utility operators to leverage the flexibility of industrial, commercial, and residential consumers, ensuring a smooth and efficient operation of the power system [44].

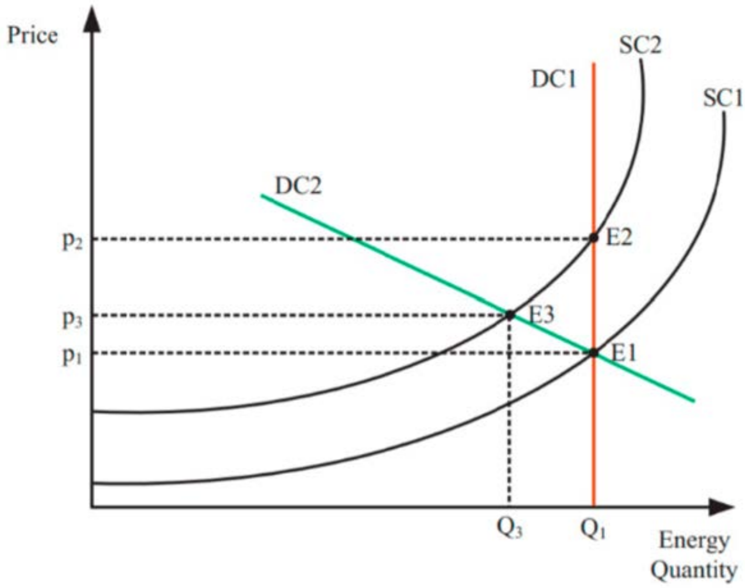
The term Demand Response Programs is derived from their function: they are initiated in response to signals—such as pricing adjustments, high wholesale prices, or compromised system reliability—that call for reduced electricity consumption. Moreover, Demand Response represents a modern solution whereby locally installed energy resources deliver mutual benefits to both the grid and consumers.

3.6.1 Participation of Demand Response Programs in the Wholesale Energy Market

Demand Response benefits both suppliers and consumers by limiting the ability of large market players to manipulate electricity prices. Participants gain increased market flexibility, allowing them to manage their consumption more efficiently and influence market conditions, especially through dynamic and market-based pricing schemes [45].

Figure 3.4 offers a simplified illustration of how a Demand Response program can affect wholesale electricity market prices. The market-clearing price is determined at the intersection of the cumulative supply and demand curves—the supply curve slopes upward while the demand curve slopes downward. Consumer-driven Demand Response has been shown to limit the pricing power of companies with a dominant market share. For example, during the California energy crisis of 2000–2001, a demand reduction of roughly 5% resulted in an electricity price drop of nearly 50%, demonstrating the exponential sensitivity of the system when operating near its maximum capacity. A modest decrease in demand can lead to a substantial reduction in production costs, ultimately lowering the price. The steepening of the supply curve as energy capacity increases reflects either the drive to maximize production profits or the high operating costs associated with generating large amounts of power.

The initial DC1 demand curve is vertical because it is initially assumed that DR programs are not implemented in the market. DR programs introduce a negative slope in the demand curve of participating customers and lead to a small reduction in demand and a fairly high reduction in the market price, as in the DC2 curve. The equilibrium value p_1 is the value at which the quantity offered and the quantity demanded are equal, and E_1 is the equilibrium point at which the curves DC1 and EC1 intersect.



Source: [46]

Figure 3.4: Impact of Demand Response on Electricity Market Prices

If consumer demand is represented by DC1, then the supply side will be able to manipulate and shape the price of electricity, thus shifting the supply curve to DC2 and increasing the market price to p_2, which corresponds to the new equilibrium point E_2. On the other hand, in case the demand side participates in DR market programs, it will be able to respond to prices, thus limiting price manipulation and achieving a different equilibrium point, E_3, at a lower value p_3.

In other words, the number of suppliers in the market is increasing through the improvement of the competition, and the prices are becoming more and more consumer-friendly.

3.6.2 Demand Response Benefits

Demand Response creates benefits mainly from saving resources that improve electricity supply significantly. It is essential to track the flow of these benefits through the market to determine who is earning and how much. It is therefore important to identify the key benefits that these programs offer, which are described below [45], [47]:

- Lower operating costs and savings on customer billing accounts.
- The lower prices in the wholesale market resulting from Demand Response (DR) create reduced supply costs for retailers, with the result that almost all retail customers usually

benefit from the savings of their accounts.

- Greater stability and robustness of the power system.

- Environmental benefits, including better land use, as a result of avoiding the installation of new electricity generation and distribution infrastructure [48].

- Real-time communication between the supply and the demand side.

- Sustainability: by shifting loads during peak hours and keeping the grid working steadily, DR programs help protect the system by managing real-time demand, achieving maximum efficiency, and ensuring back-up conditions [44].

3.7 Implementation of Incentive-Based Demand Response Program

3.7.1 Introduction

As global electricity demand continues to rise, traditional power systems struggle to meet production requirements reliably. Population growth drives higher energy consumption, while climate change necessitates a shift away from fossil fuels, further straining conventional energy infrastructures. These systems face several limitations, including slow response times due to mechanical switches, one-way communication between energy generation and consumption, and minimal data processing capabilities.

To address these challenges, advanced energy management strategies are being developed to enhance the flexibility of transmission and distribution networks, transforming conventional grids into smart grids. A Smart Grid (SG) integrates digital technologies with large-scale distribution networks to optimize energy use. Unlike traditional grids, which primarily function to transmit and distribute electricity from power plants to consumers, SGs incorporate sensors along transmission lines to monitor normal and faulty conditions in real time. Additionally, they leverage computational tools to improve operational efficiency and strategic planning.

3.7.2 Description of the Problem

The increasing demand for electricity contrasts with European-level objectives, which focus on sustainable development and environmental protection. In this context, optimizing electricity consumption has become a key area of research. An environmentally friendly approach to meeting rising demand involves integrating Renewable Energy Sources (RES) into the grid and encouraging active consumer participation in reducing peak demand, thereby smoothing the overall load profile.

This research explores the efficient and cost-effective use of electricity from a Demand Side Management (DSM) perspective and presents the implementation of a fully parameterized, explicitly constrained incentive-based demand response program. The program leverages the Particle Swarm Optimization algorithm, highlighting the benefits of RES integration while enabling two-way communication between energy production and consumption, as well as bidirectional power exchange between the main grid and RES.

The implementation described in the current research addresses mainly Load Shifting, Peak Clipping and Valley Filling. There are works in the area of Load Shifting and Peak Clipping, which demonstrate the effects of such techniques on system reliability. In constraint programming is used in order to construct an optimal schedule for home appliances. In [24] the MOGA algorithm is used for load scheduling. Our implementation regards the optimal power reduction of a system at a specific time period, where the system comprises a main grid, a microgrid of RES and a conventional generating unit, and consumers. Emphasis is placed on using as much as possible RES and the application of Demand Response when most of the load must be covered by conventional generation. The main value of our implementation is in the model construction via mathematical formulations and in the explicit use of constraints that guide the application of the optimization algorithm.

The Demand Side of a pilot incentive-based Demand Response (DR) program covers a one-day (24-hour) period. A first attempt was presented in [49]. The overall system consists of both Supply Side and Demand Side subsystems [48].

3.7.3 Supply Side

The Supply Side comprises the following two subsystems:

1. Microgrid: Microgrids have proven to be a critical technology for harnessing Renew-

able Energy Sources (RES), increasing network stability and reliability, and reducing the environmental carbon footprint. In this implementation, the microgrid includes a conventional diesel fuel generator, a wind turbine, and an array of photovoltaic cells.

2. Main Grid: The main grid is responsible for meeting the required demand when the power generated by the microgrid is insufficient to satisfy consumer needs. Additionally, if the microgrid produces excess power, it can sell the surplus back to the main grid, thereby generating financial profit.

3.7.4 Demand Side

The Demand Side consists of two low-consumption commercial consumers participating in an incentive-based Demand Response program. The pilot implementation engages these consumers actively throughout the 24-hour period, providing them with strong incentives to limit their energy demand. Figure 3.5 visualizes the high-level functionality of the system.

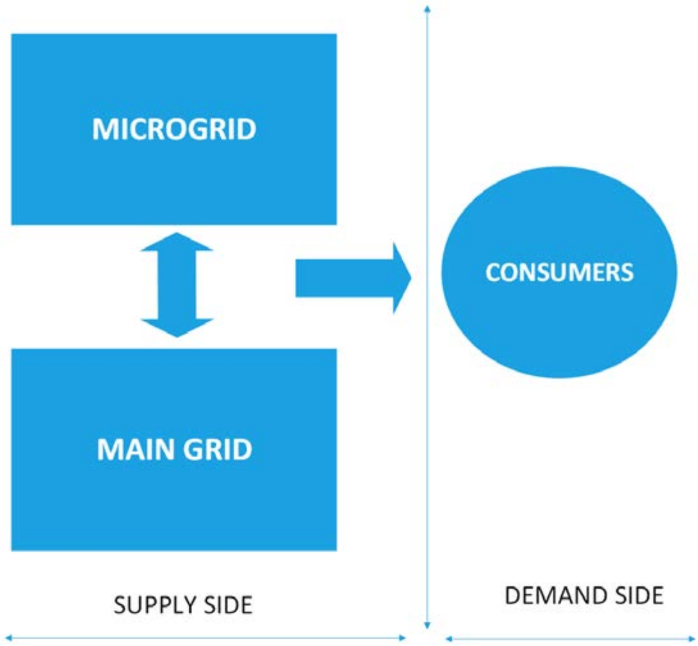


Figure 3.5: Overall System Functioning

In order to find the optimal and most cost-effective solution that will lead us to the ideal hourly energy mix of sources and consumption, the Particle Swarm Optimization (PSO) algorithm is utilized to model and solve the problem described in subsection 3.7.2. PSO is a computational method that finds an optimal solution by iteratively improving a candidate solution with respect to a given objective function [50], [51].

In our domain of investigation, we deploy PSO in order to determine the best strategy for the power system to satisfy the total load of the two commercial consumers over a 24-hour period, giving priority to the use of the Renewable Energy Sources (RES) of the microgrid (installed wind turbine and photovoltaic array). The aim is twofold:

1. Minimize both the cost of purchasing energy from the main grid and the production costs of the conventional diesel generator;

2. Maximize the financial benefit of the microgrid operator.

The mathematical formulation of the problem is presented below.

3.7.5 Mathematical Model Formulation Strategy of Supply and Demand Side

This subsection presents all mathematical models used in the PSO algorithm.

Mathematical Model of Photovoltaic Array

The hourly power generation of the photovoltaic array is given by the following equation [52, 53]:

$$P_{PV} = n_{PV} \cdot A_c \cdot H(pu) \tag{3.1}$$

where:

- n_{PV} is the efficiency of the photovoltaic array,
- A_c is the total area (in m^2) covered by the array,
- $H(pu)$ is the hourly solar radiation index (kW·h/m^2) on the PV array.

Mathematical Model of Wind Turbine

The output power of a wind turbine depends on both the magnitude and direction of the wind speed, i.e., the stochastic behavior of the wind, which is influenced by the turbine's location, air density, rotor blade geometry, and the efficiency of converting the wind's kinetic energy into electricity. The model to convert hourly wind speed into power is as follows [54]:

$$P_{wind} = 0.5 \cdot \eta_w \cdot \rho_{air} \cdot C_p \cdot A \cdot V^3 \tag{3.2}$$

where:

- P_{wind} is the power generated by the wind turbine,
- η_w is the efficiency of the wind turbine (as provided by the manufacturer),
- ρ_{air} is the air density,
- C_p is the power coefficient of the wind turbine (dependent on its geometry),
- A is the swept area of the wind turbine blades,
- V is the wind speed at a specific height.

Transfer Power Cost between Main Grid and Microgrid

The main grid compensates for any shortfall in electricity due to high consumer demand, covering the intermittent operation of RES. It is assumed that a trading program is in place, where power can be bought from the main grid in case of shortage or sold to it from the microgrid in case of surplus. When the microgrid supply cannot meet the total demand, extra energy is purchased from the main grid, and vice versa. Our objective is for RES to take precedence over other energy sources. The total cost at time t (with $t = 1, \ldots, 24$) for the transferred power between the main grid and the microgrid is expressed as:

$$C_{trans}(t) = \text{price} \cdot P_{trans}(t) \tag{3.3}$$

where:

- $C_{trans}(t)$ is the total cost of the transferred power at time t,
- price is the electricity charge per hour (€/kW),
- $P_{trans}(t)$ is the transferred power at time t.

This equation yields a positive value when power flows from the main grid to the microgrid ($P_{trans}(t) > 0$), a negative value when it flows from the microgrid to the main grid ($P_{trans}(t) < 0$), and zero when there is no transfer.

Hence, the first objective function, $fun1(t)$, is defined as:

$$fun1(t) = \sum_{t=1}^{24} C_{trans}(t) \tag{3.4}$$

Cost of Conventional Generator

The fuel cost function of the conventional generator, $C_{gen}(t)$, is modeled as a quadratic function:

$$C_{gen}(t) = k_i \cdot P_{gen}(t)^2 + m_i \cdot P_{gen}(t) \tag{3.5}$$

where:

- $P_{gen}(t)$ is the power generated by the conventional diesel generator,
- k_i and m_i are the fuel cost coefficients.

Thus, the second objective function, $fun2(t)$, is defined as:

$$fun2(t) = \sum_{t=1}^{24} C_{gen}(t) \tag{3.6}$$

Consumer Contract Design

The cost of reducing a consumer's power consumption depends on both the type of consumer and the magnitude of the reduction. The cost function for a customer, $C(\theta, x)$, is defined as [55]:

$$C(\theta, x) = K_1 \cdot x^2 + K_2 \cdot x - K_2 \cdot x \cdot \theta \tag{3.7}$$

where:

- K_1 and K_2 are cost coefficients,
- x is the reduction in consumption (in kW),
- θ is a continuous variable representing the type of customer, taking values in the closed interval [0, 1].

This variable allows us to model different customer types by assigning different cost values. For example, $\theta = 1$ is set for customers more willing to reduce their electricity consumption, whereas $\theta = 0$ is for those with the least inclination. As θ varies, the term $K_2 \cdot x \cdot \theta$ changes accordingly, resulting in different marginal costs. Consequently, customers with low θ incur lower marginal costs than those with higher θ, leading to greater marginal benefits from reducing their consumption. Note that θ can also take discrete values, a scenario that will be employed in subsequent modeling. For customers with zero reduction in consumption ($x = 0$), the cost $C(\theta, x = 0)$ is equal to 0.

3.8 Demand Response Program Overview

The parameter θ is not directly known by the utility company. Instead, based on its own estimation of the customer profiles it serves, the utility establishes an incentive function, $y(x)$, which specifies the monetary compensation offered to customers for reducing their power consumption by an amount x. In other words, this function indicates the total payment a customer will receive in exchange for a specific reduction in energy use.

Participants in the program independently decide on the quantity of power reduction, x, by referring to the provided incentive function. Naturally, no customer will opt to reduce consumption unless there is an economic reward for doing so. The consumer benefit function, denoted as V_1, is formulated as follows:

$$V_1 = y - c(\theta, x) = y - K_1 \cdot x^2 - K_2 \cdot x + K_2 \cdot x \cdot \theta \tag{3.8}$$

where:

- V_1 represents the consumer's net benefit,
- y is the incentive offered,
- K_1 and K_2 are cost coefficients,
- x is the reduction in power consumption (in kW),
- θ is a continuous parameter that characterizes the customer's type.

To encourage new customers to join, many Demand Response (DR) programs provide an initial fixed bonus as part of the total compensation for reducing consumption. In our implementation, participation is contingent on the following conditions, which are derived from financial contract theory [56]:

- A customer's decision to participate should yield a positive surplus; that is, $V_1 \geq 0$, ensuring that there is a clear monetary incentive to reduce consumption.
- The benefit function in (8) must be monotonic with respect to θ and non-decreasing with respect to x. Under the incentive compatibility requirement, each participant should be rewarded in proportion to the amount of demand reduction they achieve. This condition is mathematically expressed as:

$$V_{1,\theta} \geq V_{1,\theta^0} \quad (9) \tag{3.9}$$

where θ^0 represents the customer type if incorrectly declared.

Furthermore, the utility can assess the monetary impact of being unable to supply a certain amount of power to a consumer. This cost is characterized by the parameter λ, measured in €/kW, and can be computed using established optimal power flow methods. Essentially, λ quantifies the expense incurred by the utility for not delivering the required energy. With knowledge of λ, the utility defines its own benefit function, $V_2(\theta, \lambda)$, as follows:

$$V_2(\theta, \lambda) = \lambda \cdot x(\theta) - y(\theta) \tag{3.10}$$

Here:

- $V_2(\theta, \lambda)$ is the benefit (or net cost) to the utility,
- λ is the cost associated with energy shortfall,
- $x(\theta)$ is the reduction in consumption (in kW),
- $y(\theta)$ is the incentive payment corresponding to that reduction.

The utility's objective is to maximize its overall benefit from the DR program over a 24-hour period. Therefore, the third objective function is defined as:

$$fun3(t) = \sum_{t=1}^{24} \sum_{k=1}^{2} \Big\{ - V_2\big(\theta(k,t), \lambda(k,t)\big)\Big\} = \sum_{t=1}^{24} \sum_{k=1}^{2} \Big\{ y_{k,t} - \lambda_{k,t} \cdot x_{k,t} \Big\} \tag{3.11}$$

In summary, three objective functions are established:

1. To minimize the cost of energy purchased from the main grid.
2. To minimize the fuel expenses of the conventional generator.
3. To maximize the financial benefit to the microgrid operator.

Accordingly, the overall objective function is given by:

$$Obj_{total}(t) = z_1 \cdot \Big[fun1(t) + fun2(t) \Big] + z_2 \cdot fun3(t) \tag{3.12}$$

where z_1 and z_2 are weighting factors for the respective objectives.

3.8.1 Solving Strategy

In our study, we assign equal weights to the functions by setting $z_1 = z_2 = 0.5$. This choice reflects an equal priority between reducing costs for consumers (via z_1) and maximizing benefits for the utility company (via z_2). It is also possible to explore alternative scenarios by varying these weights (provided that $z_1 + z_2 = 1$), especially to analyze cases where consumer benefits are deemed more important than utility gains. Note that since consumer-related costs are also linked to environmental impacts, higher values of z_1 might represent scenarios where environmental protection is a primary concern.

The analysis covers a full day (24 hours) and considers the following decision variables for each hour:

- The active power output of the wind turbine, P_{WG}.
- The active power output of the photovoltaic array, P_{PV}.
- The power exchanged between the main grid and the microgrid, P_{trans}.
- The output power of the conventional generator, P_{gen}.
- The power reduction for each consumer in response to demand constraints, x_1 and x_2.
- The incentive payments for power reduction, y_1 and y_2.

3.9 Datasets

3.9.1 PV Array Generated Power Data

Solar data is obtained from the Photovoltaic Geographic Information System (PVGIS), which offers open access to solar radiation data for any location on Earth. This service supports both interconnected and autonomous grid systems across various technologies. In our study, the nominal power of the PV array is set to 30 kW. We use the PVGIS-SARAH database to retrieve hourly average solar radiation data. The installation site is chosen in the Thessaly region, with coordinates $39.371°$ latitude, $22.812°$ longitude, and an altitude of 216 m. System losses are assumed to be 14%. Table 3.1 presents data of the generated hourly PV power values for August 17, 2016.

Table 3.1: Hourly PV Power Generated Data

Time (Hours)	PV Power (kW)	Time (Hours)	PV Power (kW)
1	0	13	19.48
2	0	14	16.41
3	0	15	11.74
4	0	16	6.04
5	0.24	17	1.25
6	0.39	18	0
7	10.06	19	0
8	15.24	20	0
9	18.9	21	0
10	21.1	22	0
11	22.06	23	0
12	21.47	24	0

3.9.2 Conventional Generator

Table 3.2 presents the fuel cost coefficients of the conventional diesel generator, its minimum and maximum generated power, and the maximum rates of production increase and decrease [52].

Table 3.2: Conventional generator data

Parameter	Value
$P_{gen,min}$ (kW)	0
$P_{gen,max}$ (kW)	9
k_i	0.04
m_i	0.3
Max Increase Rate (kW)	8
Max Decrease Rate (kW)	8

3.9.3 Wind Turbine Power Prediction Data

The wind power generation data were adjusted to a nominal value of 22 kW [57] and are presented in Table 3.3.

Table 3.3: Wind Power Generation Data

Time (Hours)	Wind Power (kW)	Time (Hours)	Wind Power (kW)
1	17.56	13	21.02
2	16.50	14	20.05
3	16.25	15	20.67
4	17.48	16	20.98
5	18.48	17	19.37
6	19.42	18	19.61
7	19.82	19	19.70
8	19.35	20	18.72
9	20.08	21	17.21
10	19.01	22	16.75
11	20.04	23	16.03
12	21.68	24	16.90

3.9.4 Total Initial Consumer Demand

The total average hourly load demand of the two consumers is presented in Table 3.4.

Table 3.4: Total Average Initial Demand per Hour

Time (Hours)	Demand Power (kW)	Time (Hours)	Demand Power (kW)
1	31.83	13	39.67
2	31.40	14	41.70
3	31.17	15	42.10
4	31.00	16	41.67
5	31.17	17	40.70
6	32.10	18	40.07
7	32.97	19	38.63
8	34.10	20	36.40
9	37.53	21	34.10
10	38.33	22	32.80
11	40.03	23	32.50
12	41.17	24	32.00

3.9.5 Commercial Consumer

For the two commercial consumers, Table 3.5 lists the consumer type (represented by θ), daily power interrupt limits, and cost coefficients [52].

Table 3.5: Consumer Type, Daily Power Interrupt Limits, and Cost Coefficients

Consumer	θ	Max Limit/Day (kW)	K_1	K_2
Consumer 1	0.5	50	0.108	0.132
Consumer 2	0.6	60	0.184	0.164

3.9.6 Hourly Values of λ parameter

The parameter λ, defined as the "power outage" cost for the utility company, is determined via optimal power flow techniques [55]. Table 3.6 presents the hourly values of λ (€/kW) adjusted by [57], which are common to both consumers.

Table 3.6: Hourly Values of λ

Time (Hours)	λ **(€/kW)**	**Time (Hours)**	λ **(€/kW)**
1	0.157	13	0.73
2	0.140	14	0.78
3	0.220	15	0.85
4	0.376	16	0.71
5	0.450	17	0.68
6	0.470	18	0.63
7	0.504	19	0.58
8	0.535	20	0.42
9	0.670	21	0.38
10	0.616	22	0.301
11	0.638	23	0.253
12	0.682	24	0.142

3.10 Constraints

A central feature of our implementation is the explicit incorporation of constraints that guide the PSO algorithm in finding the optimal solution. All constraint equations refer to parameters that can be set for each run, thereby allowing experimentation with varying microgrid sizes, numbers of consumers, generation capacities, utility incentive budgets, policies for different consumer profiles, and more. The constraints are detailed as follows:

- Power Balance: For every hour t, the system must maintain power balance. The total power produced by the wind turbine, photovoltaic array, conventional generator, and main grid must equal the total required power of the two consumers. This is mathematically expressed as:

$$P_{gen}(t) + P_{PV}(t) + P_{WG}(t) + P_{trans}(t) = TD(t) - \sum_{k=1}^{2} P_{curtail,k}(t) \tag{3.13}$$

  where $TD(t)$ is the total power requested by the consumers and $P_{curtail,k}(t)$ is the curtailed power for consumer k (with $k = 1, 2$).

- Generation Limits: For each hour t, the output of both the photovoltaic array and the wind turbine must be within their operational limits:

$$0 \leq P_{PV}(t) \leq P_{PV,\max} \tag{3.14}$$

$$0 \leq P_{WG}(t) \leq P_{WG,\max} \tag{3.15}$$

- Main Grid Exchange Limit: The active power exchanged between the main grid and the microgrid must remain within specified limits:

$$|P_{trans}(t)| \leq P_{trans,\max} \tag{3.16}$$

- Conventional Generator Limits: The conventional generator's output must adhere to its operational bounds:

$$P_{gen,\min} \leq P_{gen}(t) \leq P_{gen,\max} \tag{3.17}$$

- Rate of Change of Generator Output: The change in generator output from time t to $t+1$ cannot be instantaneous but must be limited as follows:

$$-P_{limit,\min} \leq \Delta P_i \leq P_{limit,\max} \tag{3.18}$$

  where $\Delta P_i = P_{gen}(t+1) - P_{gen}(t)$, $P_{limit,\max}$ is the maximum allowable increase, and $P_{limit,\min}$ is the maximum allowable decrease.

- Daily Compensation Budget: The utility company, which knows the cost coefficients K_1 and K_2 for each customer, sets a daily total compensation budget (denoted as `Total ycost`) of €150. This constraint is expressed by:

$$\sum_{t=1}^{24} \sum_{k=1}^{2} y_{t,k} \leq \text{Total ycost} \tag{3.19}$$

- Daily Reduction Limit: The maximum daily power reduction capability for each consumer, denoted as Max Limit/Day$_k$, must be at least as great as the total power reduction achieved over the day:

$$\text{Max Limit/Day}_k \geq \sum_{t=1}^{24} x_{k,t}, \quad k = 1, 2 \tag{3.20}$$

- Exchange Power Pricing: For the pricing of power exchanged between the main grid and the microgrid, a rate of 0.12 €/kWh is applied [58].

3.11 Analysis of Results

The initial total hourly demand of the two commercial consumers of the program is presented in Table 3.4. It is noteworthy that the peak of demand appears between the hours 10 a.m.–7 p.m., which may be due to increased commercial needs the two consumers are required to cover during that time.

3.11.1 Conventional Generator Power

Table 3.7 presents the optimal hourly power produced by the conventional microgrid diesel generator for the period of 24 hours, with the generator presenting a continuous operation at all hours.

Table 3.7: Optimal Conventional Generated Power Over 24 Hours

Time (Hours)	**P_{gen} (kW)**	**Time (Hours)**	**P_{gen} (kW)**
1	6.68	13	1.18
2	6.90	14	6.17
3	6.92	15	1.02
4	5.52	16	8.95
5	8.69	17	7.29
6	8.44	18	7.21
7	5.26	19	6.93
8	0.69	20	5.68
9	6.21	21	8.89
10	4.42	22	8.05
11	2.89	23	8.47
12	1.43	24	7.10

3.11.2 Exchange Power between Main Grid and Microgrid

The optimal hourly exchange between the main grid and the microgrid, as a result of the simulation of our algorithm taking into consideration a specific scenario in terms of parameters and the imposed constraints, is presented in Table 3.8.

Table 3.8: Optimal Transferred Power Over 24 Hours

Time (Hours)	P_{trans} (kW)	Time (Hours)	P_{trans} (kW)
1	4	13	-4
2	4	14	-4
3	4	15	4
4	3.99	16	-4
5	4	17	4
6	4	18	4
7	4	19	4
8	4	20	4
9	-4	21	4
10	-4	22	4
11	-4	23	4
12	-4	24	4

3.12 Formulation of Supply and Demand Side Datasets with Demand Response Program

3.12.1 Wind Turbine and PV Generated Power

Figure 3.6 shows the optimal variation of the hourly produced power from the wind turbine. Also, Figure 3.7 shows the optimal variation of the hourly PV produced power.

3.12.2 Optimal Hourly Consumer Reduction Power

The optimal hourly values of the power reduction of the two consumers of the system, due to their participation in the Demand Response program, are presented in Table 3.9.

3.12.3 Hourly Incentive Payment

The hourly amounts that the two consumers receive as compensation for their participation in the Demand Response program are presented in Tables 3.10 and 3.11.

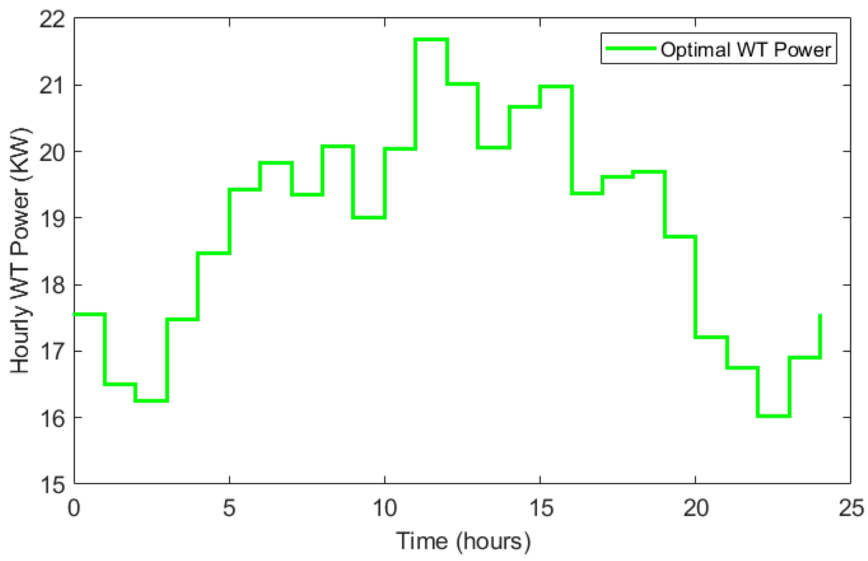


Figure 3.6: Optimal hourly Wind Turbine Generation

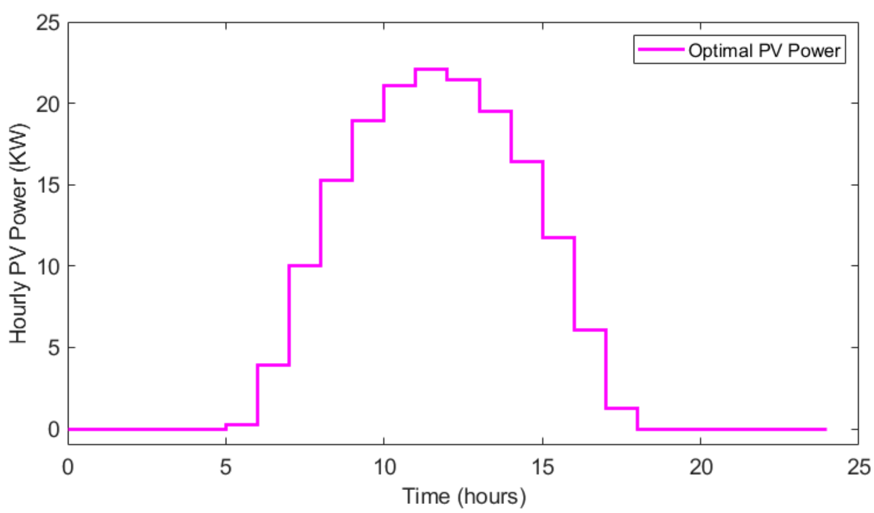


Figure 3.7: Optimal hourly PV Generation

3.13 Discussion

The results of the implementation present interesting data regarding the behavior of the microgrid and its cooperation with the main grid. According to Figure 3.8, the conventional diesel generator operates almost uninterruptedly throughout the 24-hour period, except at 8 o'clock where it presents almost zero output. However, it does not reach its nominal operating value at any point in time, because power demands are partly covered by the RES.

It is observed that during the hours when there is a lack of solar energy—and therefore the PV system shows zero output—the presence of the conventional generator is more intense, reaching an output power of up to 8.95 kW. The stronger presence of Renewable Energy

Table 3.9: Optimal Power Reduction of the Consumers Over 24 Hours

Time (Hours)	Consumer 1 (kW)	Consumer 2 (kW)	Time (Hours)	Consumer 1 (kW)	Consumer 2 (kW)
1	0	3.58	13	0	0
2	0	4	14	0	0
3	4	0	15	0	0
4	0	4	16	4	0
5	0	0	17	3.99	0
6	0	0	18	4	4
7	0	0	19	4	4
8	0	0	20	4	4
9	0	0	21	4	0
10	0	0	22	4	0
11	0	0	23	4	0
12	0	0	24	0	4

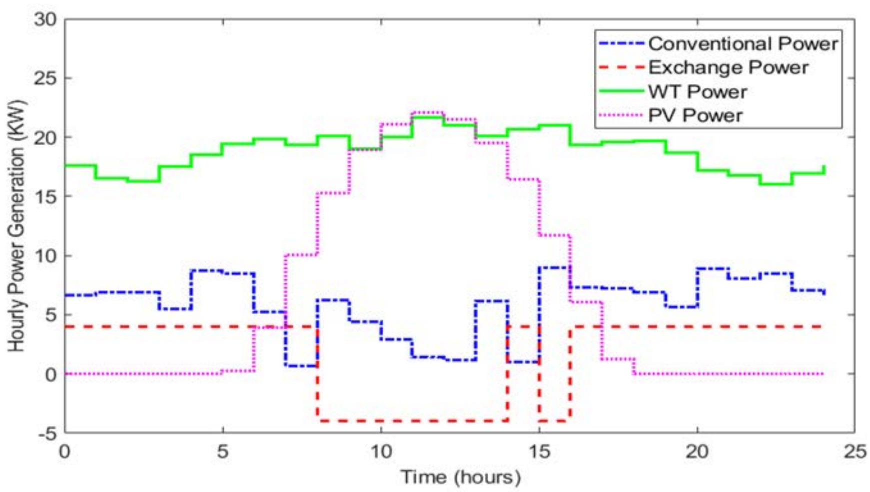


Figure 3.8: Total hourly power generation of all sources

Sources (RES) reduces the reliance on conventional forms of production, which is desirable.

However, during periods of low RES penetration, Demand Response (DR) plays a critical role and is most used in order to balance load demand and generation.

It is worth highlighting the curve of exchanged active power between the main grid and the microgrid. When it is positive, the microgrid purchases power from the main grid, and

Table 3.10: Hourly Incentive Payment for the Two Consumers (Hours 1-12)

Time (Hours)	Consumer 1 (€)	Consumer 2 (€)
1	0	2.6
2	0	3.22
3	1.99	0
4	0	3.22
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0

when it is negative, it sells excess power to it. According to Figure 3.8, during the first 8 hours the microgrid purchases energy to meet consumer demands. Then, for a period of 8 hours, the microgrid is able to sell energy back to the main grid, until the solar power output becomes zero and the need to buy energy for the rest of the day arises again. When the RES is operating at full capacity, there is surplus power available for sale to the main grid, especially when the photovoltaic array is fully operational.

As can be observed, wind and solar power are the primary sources of the microgrid, with the goal of operating them continuously over a 24-hour period, which is achieved in this case.

Regarding the participation of customers in the DR program, it is observed that both consumers are forced to reduce their demand when the PV array produces zero or relatively low output, since the combined power from the wind turbine and the grid input cannot meet the initial total hourly demand.

The total energy reduction in the total consumption of the two consumers and the corresponding monetary reward that each customer receives are presented in Table 3.12. The table also includes savings resulting from load shifting from on-peak to off-peak periods, due to participation in the demand response program, which offers the consumer a 30% return of the total compensation as an incentive.

Table 3.11: Hourly Incentive Payment for the Two Consumers (Hours 13-24)

Time (Hours)	Consumer 1 (€)	Consumer 2 (€)
13	0	0
14	0	0
15	0	0
16	1.99	0
17	1.99	0
18	1.99	6
19	6	3.21
20	1.99	3.22
21	6	0
22	6	0
23	1.99	0
24	0	3.21

Table 3.12: Total energy reduction and total compensation for each consumer

Consumer	Energy Reduction (kWh)	Total Compensation (€)	Incentive Saving (€)
Consumer 1	35.99	29.95	8.99
Consumer 2	27.58	24.70	7.41
Total	**63.57**	**54.65**	**16.40**

Both consumers are forced to reduce their consumption, without, however, having to reach the upper limits, as agreed with the program administrator. Thus, the system is able to respond to and handle unpredictable conditions of electricity shortage, which is a primary objective of large-scale and advanced Smart Grids.

This result is encouraging, as it enables the creation of backup conditions during emergencies, such as in the event of a temporary failure of the photovoltaic array during certain hours of the day. The parameters q are almost identical for the two consumers, which in turn contributes to the total final reduction in power and, correspondingly, to the monetary compensation for each one.

3.14 Summary and Future Work

This work discusses an implementation of an incentive-based Demand Response (DR) program, which employs the Particle Swarm Optimization (PSO) algorithm to encourage customers to reduce their energy consumption. Beyond the obvious benefits of replacing conventional power generation with Renewable Energy Sources (RES), as demonstrated through the implementation, the main contribution of this work to the broader area of Demand Side Management (DSM) lies in its modeling approach, particularly in the explicit incorporation of constraints.

All values specified in the constraint equations are parameterized, allowing experimentation both from the perspective of the customer and the utility company. From the customer's perspective, this is indirectly linked to the environmental impact of power generation. Such experimentation can be especially useful to policymakers—for instance, in determining appropriate incentive tariff values or evaluating how many RES units can be deployed to meet environmental targets such as net zero emissions by 2050.

Moreover, utility companies may offer tailored incentives to customers based on their past consumption patterns, the revenue they generate for the grid, and the time periods during which they typically consume energy (on-peak, mid-peak, or off-peak).

In the future, we intend to apply the model to scenarios with varying numbers of consumers, different consumer profiles, diverse incentive tariffs, and different levels of RES availability depending on geographical and climatic conditions. Through this, we hope to contribute to the development of a testbed for other incentive-based demand response programs.

The primary goal of DR is typically to flatten the demand curve, which often results in cost reductions for both suppliers and consumers, as well as lower emissions. Other studies utilize PSO similarly but with predefined target values. In contrast, our approach explores potential target values, with their achievability determined through further experimentation using our implementation.

3.15 Scalability and Impact of Direct Load Control: Enhancing Demand Side Management Efficiency

3.15.1 Summary

This experimental application employs the IEEE 13-bus test feeder and aims to simulate the operation of a system containing 629 residential consumers. In each household, there is an installed passive controller, which can be controlled via the system operator (Direct Load Control) and adjusts the consumer's demand. These consumers participate in a Direct Load Control Demand Response Program, according to which their household's demand is adjusted in real-time according to fluctuations in electricity prices. The goal of the simulation is to draw reliable conclusions regarding the program's impact on customer participation in reducing demand and, as a result, decreasing system overload.

3.15.2 Introduction and Key Features of IEEE 13-Bus Test Feeder

The IEEE 13-bus test feeder is a benchmark distribution system used for power flow analysis, fault studies, and smart grid research. It is part of the IEEE Distribution Test Feeder set and represents a real-world radial distribution system with certain meshed characteristics [59].

About the key features, this system is widely used for various applications, including power flow analysis for unbalanced distribution systems, short-circuit studies, and protection coordination. It also plays a critical role in voltage stability assessment and reactive power management, which are essential to maintain system reliability. In addition, the feeder is used to study the integration of distributed energy resources (DER), such as renewable energy sources, and to advance research in microgrids and smart grid technologies. More specifically, the key features are summarized below [60]:

- Network Topology: The IEEE 13-bus test feeder is composed of 13 buses arranged in a primarily radial structure. It includes both single-phase and three-phase components, allowing for unbalanced power flow studies. Additionally, it contains a small loop, making it useful for weakly meshed system analysis.

- Voltage Levels: The system operates at a nominal voltage level of 4.16 kV. Step-down transformers are included for voltage regulation, ensuring voltage stability across different sections of the feeder.

- Load Characteristics: The IEEE 13-bus test feeder includes a combination of constant power, constant impedance, and constant current loads. The presence of unbalanced loads provides realistic conditions for phase imbalance studies, making it a valuable benchmark for distribution system analysis.

- Line and Transformer Models: The feeder uses detailed line models with various conductor types and lengths to accurately represent distribution network characteristics. It also incorporates delta-wye and wye-wye transformers, which are essential for phase shifting analysis and evaluating different transformer configurations.

- Capacitor Bank and Voltage Regulation: A capacitor bank is included in the system for reactive power compensation, helping to manage power factor and voltage stability. Additionally, a voltage regulator is present to maintain proper voltage levels throughout the network, ensuring reliable system operation under different load conditions.

3.15.3 Software Environment

GridLAB-D is an open-source power distribution system simulation tool developed by the U.S. Department of Energy's Pacific Northwest National Laboratory (PNNL). It is designed to model large-scale electric power distribution networks with a focus on integrating renewable energy sources, demand response, and smart grid technologies. Unlike traditional power flow analysis tools, GridLAB-D incorporates advanced time-series simulation capabilities, allowing users to analyze not only steady-state conditions but also dynamic interactions between loads, distributed energy resources, and grid controls over time. The software supports modeling of residential and commercial buildings, electric vehicles, solar PV systems, and various grid automation components, making it a comprehensive tool for grid modernization studies.

One of the key strengths of GridLAB-D is its ability to perform high-fidelity simulations with a detailed representation of customer behavior, load variations, and power system controls. It includes built-in modules for load modeling, market operations, weather data integration, and voltage regulation, enabling researchers and utilities to evaluate the impact of emerging technologies on distribution networks. The software is designed to work with other grid analysis tools, such as OpenDSS and MATPOWER, through co-simulation techniques. Due to its extensible architecture, users can develop custom modules or integrate GridLAB-D with external applications to enhance its functionality. Its widespread adoption in academia and industry makes it a valuable resource for studying the transition to a more resilient and sustainable power grid [61].

3.15.4 Description of the Problem

In this task the IEEE 13-bus network is utilized. Figures 3.9 and 3.10 illustrate a segment of the network where nodes are represented by circles in a light grey hue, and the connections with overhead line or underground line between each node are depicted as light grey straight or dashed lines, respectively. The transformers within the network are highlighted using orange lines. Additionally, the graphic denotes distribution transformers and the triplex meters installed at each household using black dots and switch component with dark grey line. More specifically, there is a feeder transformer at node 650, the technical specifications of which are presented in Table 3.13:

Table 3.13: Feeder Transformer Specifications

Parameter	**Value**
Connection Type	WYE_WYE
Installation Type	PADMOUNT
Power Rating	5 MVA
Primary Voltage	33000 Volt
Secondary Voltage	2401.777 Volt
Impedance	0.00033+0.0022j [Ohm]

Also, there is a distribution transformer, the specifications of which are presented in Table 3.14 and two capacitor banks, the specifications of which are presented in Tables 3.15 and 3.16.

Moreover, as outlined in Table 3.17, a total of 629 residential units are employed within this test network. This table provides a detailed account of the number of houses attributed to each node and phase.

Each house object in the GridLAB-D model represents a well-insulated residential build-

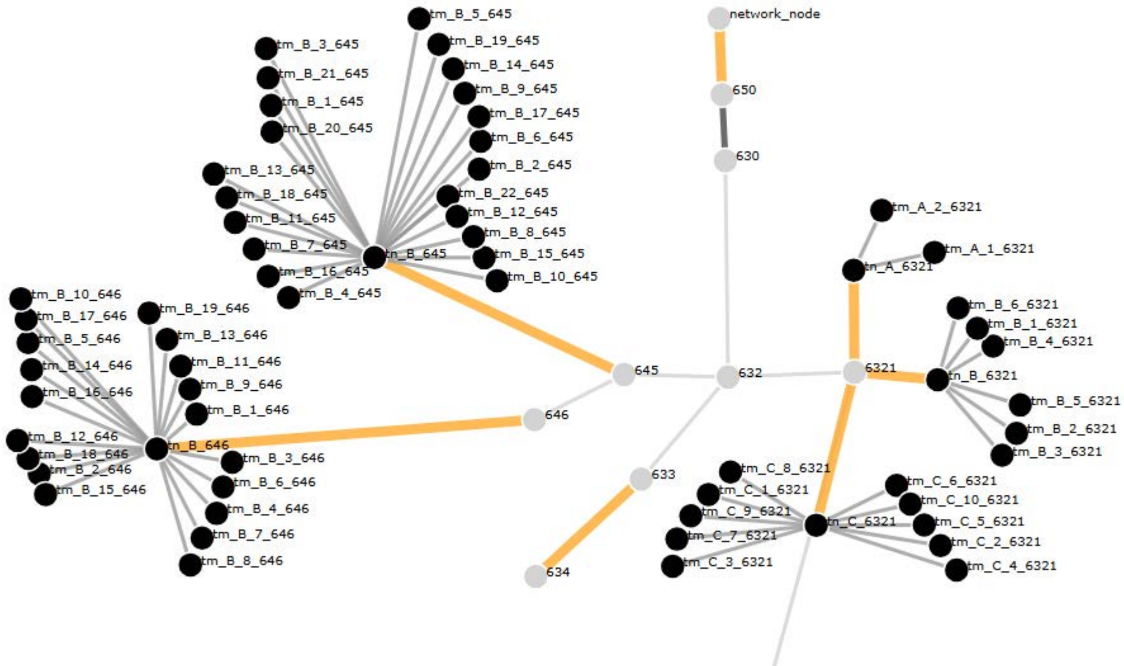


Figure 3.9: IEEE 13-bus test network

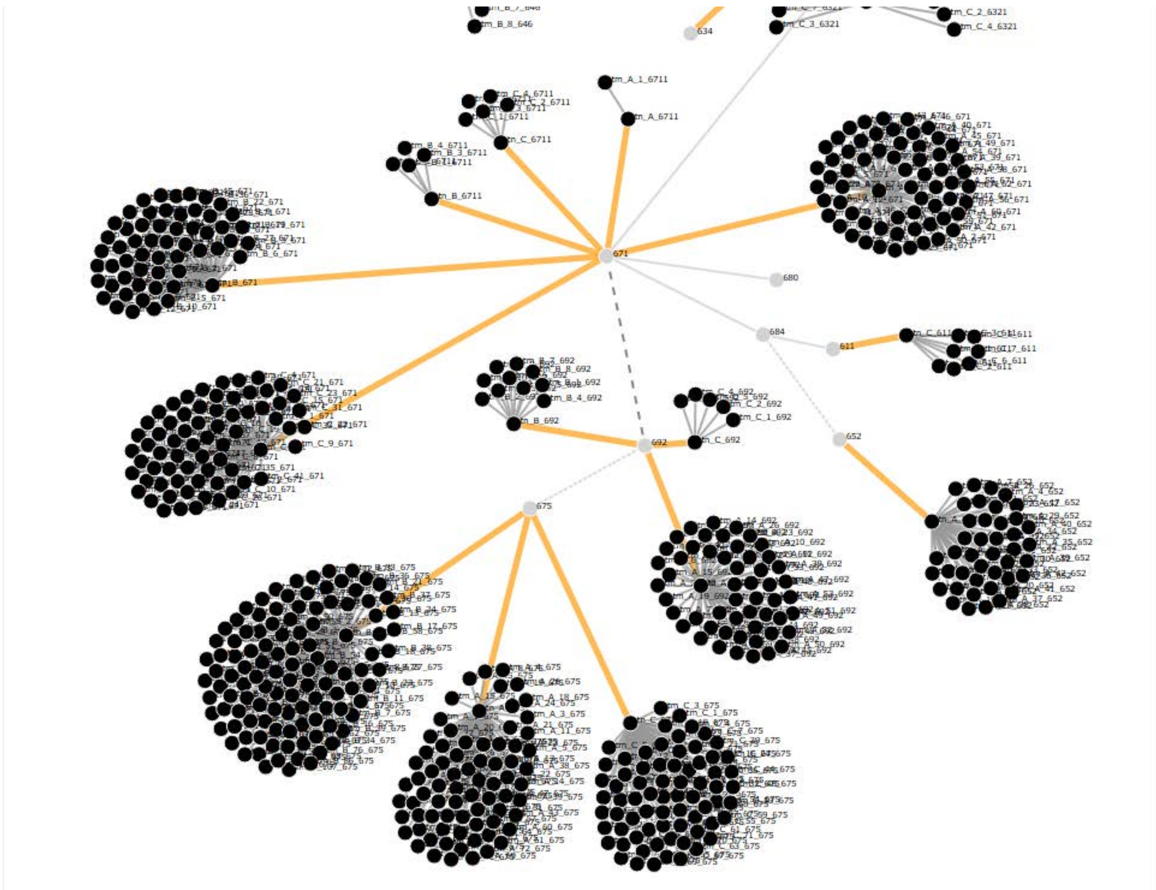


Figure 3.10: IEEE 13-bus test network

Table 3.14: Distribution Transformer Specifications

Parameter	**Value**
connect_type	SINGLE_PHASE_CENTER_TAPPED
install_type	POLETOP
shunt_impedance	10000+10000j [Ohm]
primary_voltage	2401.777 Volt
secondary_voltage	120 Volt
power_rating (kVA)	varies
impedance	0.00033+0.0022j [Ohm]

Table 3.15: Specifications of Capacitor 1

Attribute	**Value**
Phases	ABCN
Name	CAP1
PT Phase	ABCN
Parent	675
Phases Connected	ABCN
Control	VOLT
Voltage Set High	2700.0
Voltage Set Low	2250.0
Capacitor A	0.10 MVAr
Capacitor B	0.10 MVAr
Capacitor C	0.10 MVAr
Control Level	Individual
Time Delay	300.0
Dwell Time	0.0
Switch A	Closed
Switch B	Closed
Switch C	Closed
Nominal Voltage	2401.7771

Table 3.16: Specifications of Capacitor 2

Attribute	Value
Phases	ABCN
Name	CAP2
PT Phase	ABCN
Parent	675
Phases Connected	ABCN
Control	VOLT
Voltage Set High	2440.0
Voltage Set Low	2375.0
Capacitor A	0.05 MVAr
Capacitor B	0.05 MVAr
Capacitor C	0.05 MVAr
Control Level	Individual
Time Delay	300.0
Dwell Time	0.0
Switch A	Closed
Switch B	Closed
Switch C	Closed
Nominal Voltage	2401.7771

ing with an energy-efficient heating and cooling system. It utilizes a heat pump or gas for heating and an electric system (HVAC) for cooling, where electricity powered sources are controlled by passive controllers that adjust setpoints based on market conditions. The house is designed with a diverse scheduling mechanism to simulate realistic energy usage patterns. It also includes a ZIP load model for lighting and other electrical appliances, balancing different types of power consumption. Additionally, a water heater with a demand-based control strategy optimizes energy use while maintaining comfort levels. The thermal properties of the house ensure stable indoor temperatures, and the integration of market-based controls allows for dynamic responses to electricity pricing, improving overall efficiency and cost-effectiveness.

The simulation encompasses a total of 20 schedules dedicated to water heating and 8

Table 3.17: Total Houses per node and phase

node No	Phase A	Phase B	Phase C
645	0	22	0
646	0	19	0
652	42	0	0
671	63	60	72
675	74	113	71
632	2	6	10
611	0	0	7
692	53	10	5

schedules for heating and cooling. These are then allocated to the houses in a random manner.

The stubauction object in the GridLAB-D model represents a simplified market mechanism that updates every five minutes to simulate dynamic electricity pricing. It incorporates a player object, which reads price data from an external file and continuously loops through the dataset to provide real-time price signals. A recorder object is also included to track key market statistics, such as the 24-hour average price, standard deviation, and the most recent price, logging this data at one-minute intervals into a CSV file. This market model enables demand-responsive behavior in connected loads, allowing smart appliances, HVAC systems, and water heaters to optimize energy consumption based on price fluctuations. By integrating this pricing mechanism, the model supports transactive energy strategies, helping to balance grid demand and improve cost efficiency for consumers.

Passive Controller Description

The passive controller module functions as a demand-side control mechanism that passively adjusts a building's thermostat setpoint (like cooling temperature) in response to market signals. Instead of actively optimizing or forecasting, it uses statistical expectations and observed values to influence system behavior within predefined limits. In this case, the controller operates in RAMP mode, which means that it smoothly adjusts the cooling setpoint based on a market signal, most likely an electricity price value referred to as next.P. The degree to which the controller responds to the price signal is governed by the sensitivity parameter.

To make intelligent decisions, the controller needs two key inputs: One input provides a statistical expectation of the market signal, such as a 24-hour average value (avg24), helping the controller understand what a “normal” price looks like. The second input gives a real-time or forecasted value, such as the next anticipated market price (next.P), letting the controller compare current or near-future conditions to that normal baseline. The controller then calculates a control signal by determining how far the current or forecasted value deviates from the average, using parameters like ”range low” and ”range high” to define the meaningful range of variation. If the signal falls within a configurable band (”ramp low” to ”ramp high”), the thermostat setpoint is adjusted accordingly, for instance, increasing the setpoint slightly during high-price periods to reduce energy use, and lowering it during cheap electricity periods to take advantage of low costs. Figure 3.11 presents the above controller.

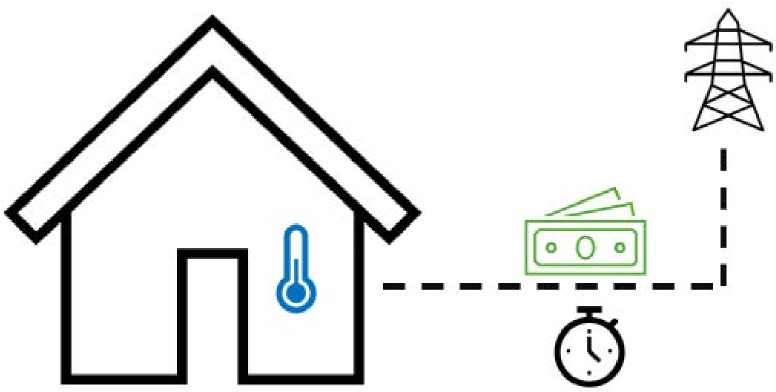


Figure 3.11: House Passive Controller

A separate flag lets the system know whether the controller is actively influencing the setpoint or not. Fundamentally, this controller facilitates a minimally invasive, market-sensitive approach to adjusting energy consumption patterns without necessitating continuous optimization, making it particularly suitable for environments where simplicity, predictability, and responsiveness are paramount.

3.15.5 Methodology and Results

We run the simulation for 24 hours on the 1st of August with two configuration, a) market price is steady and b) market price fluctuates based on the Table 3.18, which contains the hourly market prices of this day. The main goal of this simulation is to find if the demand response program helps to reduce the energy consumption.

Table 3.18: Hourly Market Prices

Hour	Market Price	Hour	Market Price
0	0.065139609	12	0.084471806
1	0.063279907	13	0.093280668
2	0.047926635	14	0.097517265
3	0.054980748	15	0.090978092
4	0.054144897	16	0.099465131
5	0.052692165	17	0.103445608
6	0.025675894	18	0.118779549
7	0.048756545	19	0.122554232
8	0.060731834	20	0.121734869
9	0.067444366	21	0.110559158
10	0.077139106	22	0.092833806
11	0.085129386	23	0.081005661

The simulation results indicate that, without direct control, the daily energy consumption of household HVAC systems reached 350 MW. However, by implementing the direct control program, consumption was reduced to 295 MW, representing a 15.7% decrease. More specifically, with regard to the active and reactive power of the transformer as well as the HVAC loads, the following observations can be made.

Active Power in each phase with and without Demand Response

Regarding the real power in each phase of the distribution transformer, it is observed that at most time points, the power reduction is noticeable. Figure 3.12 presents the total active power of the consumers without demand response in blue, and with demand response in orange. The reduction is particularly more pronounced in phases A and C, especially between timesteps 800–1000.

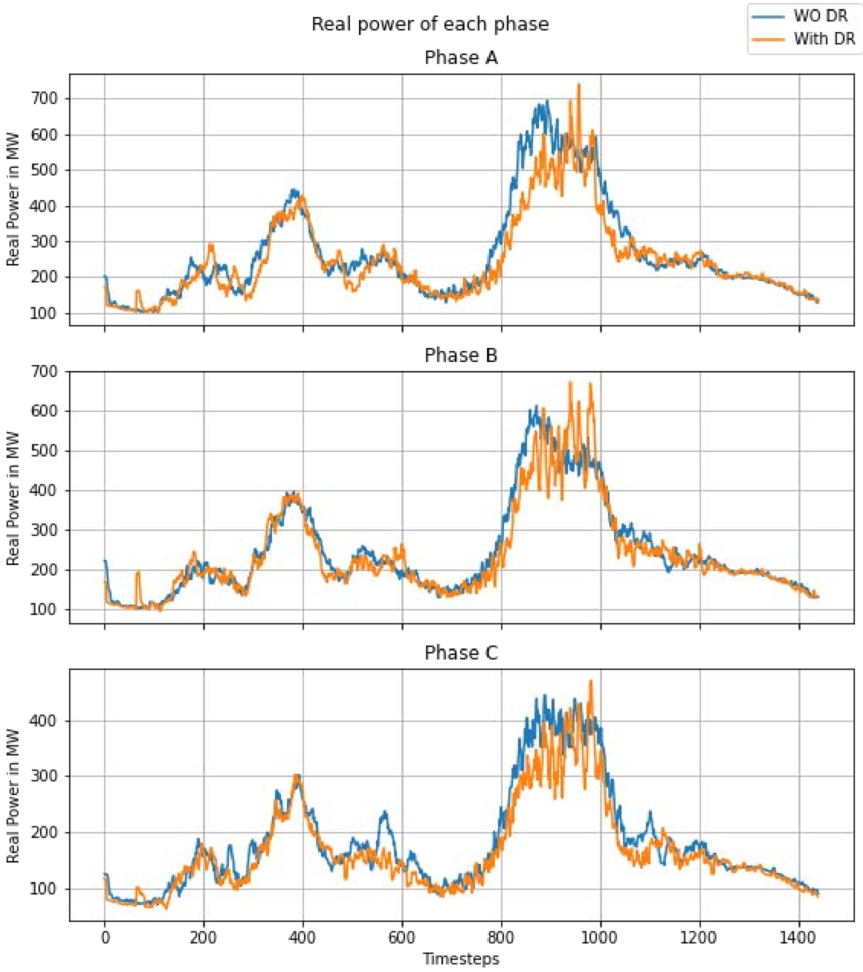


Figure 3.12: Total Active Power in each phase with and without Demand Response

Reactive Power in each phase with and without Demand Response

Regarding the total reactive power, which is presented in Figure 3.13 it is worth emphasizing that, in all three phases, the power factor shifts from inductive to capacitive under the influence of demand response. This change is attributed to the simultaneous reduction in HVAC consumption by residential users, combined with the operation of the two capacitors. This coordinated action significantly reduces the demand for reactive power, leading to a surplus, which in turn results in a capacitive power factor.

HVAC Power in each phase with and without Demand Response

In Figure 3.14, we observe the variation of total HVAC power in relation to the hourly change in temperature. In the first plot, it is evident that power consumption increases significantly in the case of demand response, reaching its peak just below 78°F. The second

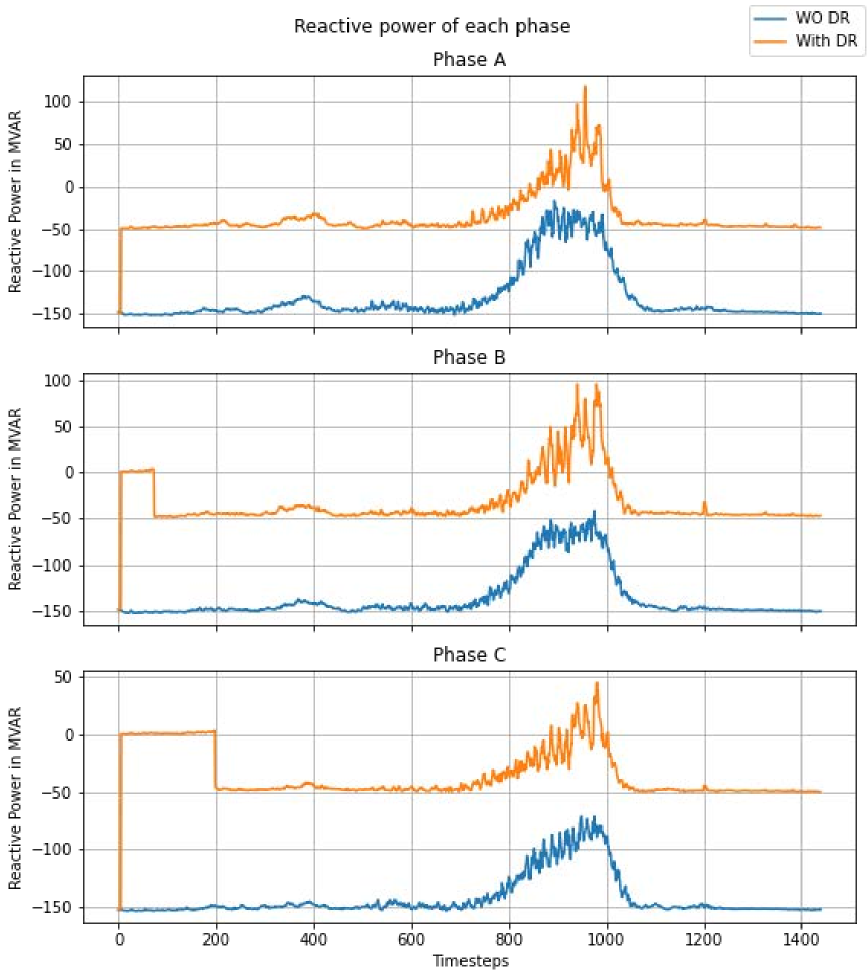


Figure 3.13: Total Reactive Power in each phase with and without Demand Response

plot, which perhaps provides the most reliable insight into the impact of demand response, shows that the total load is consistently lower under the DR scenario, with the only exception occurring between timesteps 900 and 1000.

3.15.6 Summary of the results and Conclusion

The study's simulations demonstrate that implementing a direct control strategy yields substantial energy savings in residential HVAC operation. Over a 24-hour period, uncontrolled HVAC systems consumed 350 MW, whereas the direct control program curtailed this to 295 MW—a reduction of 15.7%. Phase-wise analysis of the distribution transformer's real power confirms that demand response consistently lowers load in all phases, with the most pronounced decreases observed in Phases A and C between timesteps 800–1000. These results illustrate that targeted control not only shifts peak consumption but also smooths overall

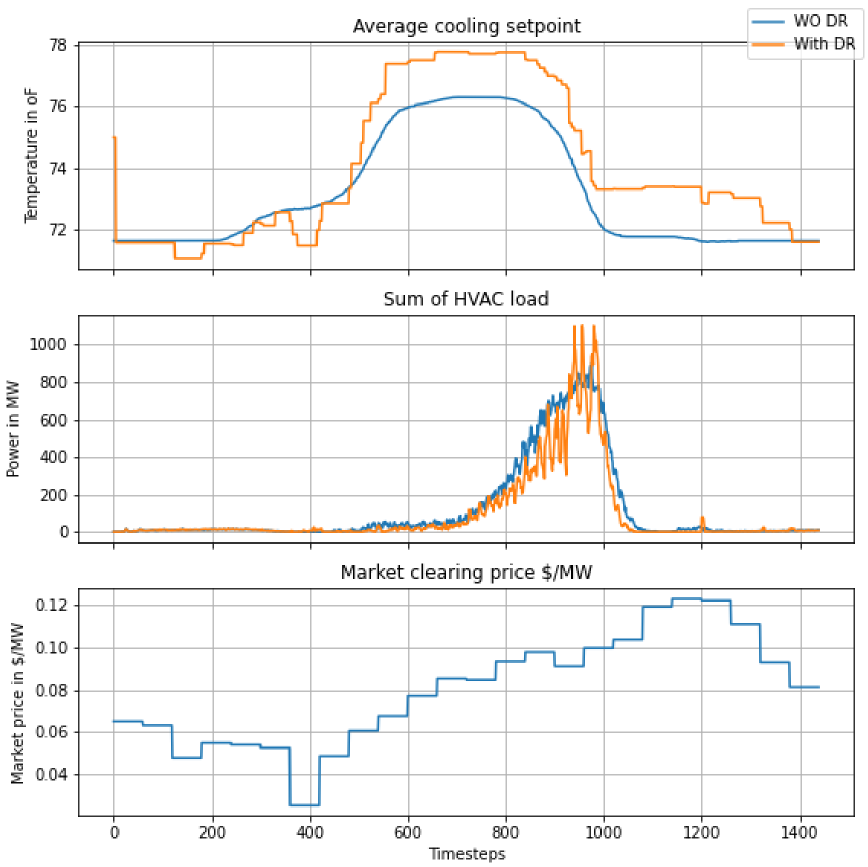


Figure 3.14: Total HVAC Power in each phase with and without Demand Response

demand, effectively flattening the daytime HVAC load profile.

Also, the coordinated operation of residential HVAC load shedding and the two capacitors fundamentally alters the network's reactive behavior. Under demand response, the transformer's power factor transitions from inductive to capacitive across all three phases, reflecting an excess of reactive capacity. This arises because reducing HVAC reactive draw, combined with capacitor injection, overcompensates for the network's prior inductive needs. Finally, when examining HVAC power against hourly temperature variations, demand response proves particularly effective: although momentary spikes occur—most notably just below 78 °F and between timesteps 900–1000—the overall HVAC load remains consistently lower under the DR scenario. These findings validate direct control as a dual-benefit mechanism, cutting both active and reactive power demands and enhancing grid stability.

Chapter 4

Time Series: Analysis and Forecasting Approaches

In this chapter, the field of time series will be introduced, their characteristics will be discussed, and a detailed analysis will be provided for all the models and algorithms used in forecasting problems, as well as the most common techniques applied for data preprocessing before they are used as inputs to the models.

4.1 Introduction to Big Data

Big Data is the sector that refers to the massive and complex datasets that exceed the capacity of traditional data processing systems to manage, analyze, and store efficiently. These datasets are defined by the ”three Vs”: volume, velocity, and variety. Volume pertains to the sheer scale of data generated daily, from sources such as social media, sensors, and digital transactions. Velocity is the rapid rate at which data is produced and needs to be processed, often in real-time. Variety captures the different formats of data, including structured, semi-structured, and unstructured forms such as text, images, videos, and logs. The growth of Big Data has been driven by advances in digital technology, ubiquitous computing, and the proliferation of data-generating devices.

The handling of these datasets introduces unique challenges that require advanced strategies for data collection, storage, and analysis. Traditional database systems are often inadequate for managing these large-scale datasets, leading to the development of new paradigms such as distributed computing frameworks, cloud storage solutions, and scalable data warehouses. Additionally, ensuring data quality and integrity becomes increasingly difficult as the volume and diversity of data grow. Addressing these challenges is crucial for leveraging these datasets to derive meaningful insights and support evidence-based decision-making across fields such as healthcare, environmental science, and business analytics. Also, this field requires not only technical innovation but also robust governance frameworks to ensure privacy, security, and ethical use of the data [62].

4.2 Time Series and challenges in Big Data

Time series data is a sequence of data points collected or recorded at successive points in time, typically at uniform intervals. Examples include stock prices, weather data, and sensor readings. Time series analysis is a critical component of Big Data analytics, offering insights into temporal patterns and trends. Time series data can be high-dimensional, especially when dealing with multiple sensors or financial instruments. Managing and analyzing such data requires sophisticated dimensionality reduction techniques. The volume of time series data can be enormous, necessitating efficient storage and retrieval mechanisms. Techniques such as data compression and indexing are often employed to handle large datasets. Time series data is often noisy and may contain missing values. Robust preprocessing techniques, including smoothing, interpolation, and outlier detection, are essential to prepare the data for analysis. Many applications require real-time analysis of time series data, such as monitoring industrial equipment or tracking financial markets. Stream processing frameworks like Apache Kafka and Apache Flink enable real-time data ingestion and analysis.

Also, these data presents several challenges, starting with its high dimensionality, especially when collected from multiple sensors or financial instruments. This complexity often requires advanced dimensionality reduction techniques to manage and analyze the data effectively. Additionally, the large volume of time series data necessitates efficient storage and retrieval methods. To handle extensive datasets efficiently, techniques such as data compression and indexing are crucial.

Another significant issue is the presence of noise and missing values, which can impact analysis. To address this, preprocessing methods like smoothing, interpolation, and outlier detection are essential for preparing the data. Furthermore, some applications, such as industrial monitoring or financial tracking, require real-time analysis. For these scenarios, stream processing frameworks like Apache Kafka and Apache Flink are employed to facilitate real-time data ingestion and analysis [63].

Finally, about the application, time series analysis finds extensive applications across various domains, leveraging its ability to uncover temporal patterns and trends. In financial markets, it is used for stock price prediction, algorithmic trading, and risk management, enabling competitive trading strategies. In healthcare, time series data from wearable devices and electronic health records support patient monitoring, anomaly detection, and disease outbreak prediction. Manufacturing benefits from time series analysis through predictive maintenance, quality control, and production optimization, reducing downtime and enhancing efficiency. The energy sector utilizes it for load forecasting, demand response, and grid management, ensuring efficient energy distribution. As part of Big Data analytics, time series analysis addresses the challenges of volume, velocity, and variety, requiring advanced methods for real-time processing and accurate predictions. While issues like data quality, integration, and privacy remain critical, advancements in statistical and machine learning techniques are expanding the potential of time series analysis in driving innovation across industries [64].

4.3 The Evolution and impact of Machine Learning in Big Data

The methods by which big data is processed have evolved over time, as the ability to collect and store an ever-increasing volume of data has grown more and more in recent years. Initially, the approaches to these kinds of problems included statistical methods, such as those analyzed above. In recent years, due to rapid technological progress, artificial intelligence models have almost completely taken over this particular field, gaining more and more applications. Machine learning, a subset of artificial intelligence, has undergone substantial evolution since its inception. In its early stages, machine learning was primarily focused on developing algorithms capable of identifying patterns and making predictions from data. The history of machine learning (ML) spans several decades, beginning with foundational theoretical work in the mid-20th century. In 1943, Warren McCulloch and Walter Pitts introduced the first mathematical model of neural networks, inspired by the workings of the human brain. The 1950s marked further progress with Arthur Samuel coining the term ”machine learning" in 1959 while developing programs that allowed computers to play checkers. Frank Rosenblatt's creation of the perceptron in 1957 represented a significant early step in enabling computers to classify and make decisions based on binary inputs. After a period of slower progress in the 1960s and 1970s, advances in computational power and the development of backpropagation algorithms in the 1980s reignited interest in neural networks. The 1990s emphasized data-driven approaches, introducing techniques like support vector machines and the use of recurrent neural networks. The 2000s saw the rise of large datasets and improved computational resources, enabling advancements in fields like image recognition. Deep learning techniques matured in the 2010s, transforming areas such as natural language processing, while recent years have seen widespread applications of ML across industries, marking its evolution from theory to impactful real-world technology [65], [66].

As the availability of large-scale data and computational power increased, the field of machine learning experienced a significant transformation. This shift was marked by the emergence of deep learning, a subset of machine learning characterized by neural networks with multiple layers. Deep learning models, inspired by the structure and function of the human brain, have demonstrated remarkable capabilities in learning from large and complex datasets. The development of convolutional neural networks (CNNs) has been pivotal in advancing image and video processing. CNNs excel at capturing spatial hierarchies in data, making them highly effective for tasks such as image recognition and object detection. Meanwhile, recurrent neural networks (RNNs) and their variants, such as long short-term memory (LSTM) networks, have significantly improved the analysis of sequential data, which is crucial for applications like speech recognition and natural language processing. The success of deep learning can be attributed to its ability to automatically learn hierarchical representations of data, reducing the need for manual feature engineering.

4.4 Applications of Machine Learning and Deep Learning in Power Systems

Machine learning (ML) and deep learning (DL) have found extensive applications in power systems, addressing various challenges and enhancing efficiency and reliability. One of the key applications is load forecasting, where ML models predict future electricity demand based on historical consumption data and other relevant factors. Accurate forecasting is essential for efficient grid management and planning. DL techniques, particularly RNN-based, are well-suited for handling time series data in power systems. These models can capture complex temporal dependencies, making them ideal for tasks such as predicting short-term and long-term energy demand, detecting anomalies in power consumption, and identifying patterns in renewable energy generation.

In the realm of predictive maintenance, ML algorithms can analyze data from sensors and other monitoring devices to predict equipment failures before they occur. This proactive approach helps in minimizing downtime and reducing maintenance costs. DL models, with their ability to process large volumes of data and detect subtle patterns, enhance the accuracy of predictive maintenance systems, also. Another significant application is in optimizing the integration of RES, such as solar and wind farms, into the grid. These models can predict the variability in renewable energy generation and optimize energy storage and distribution accordingly. This ensures a stable and reliable power supply while maximizing the utilization of renewable resources.

Moreover, these models employed in enhancing the security and stability of power systems. These technologies can detect cyber-attacks and other security threats by analyzing patterns of network traffic and system behavior. By identifying anomalies and potential vulnerabilities, machine learning models contribute to the resilience of power systems against cyber threats. Overall, the integration of machine learning and deep learning in power systems offers substantial benefits, including improved forecasting accuracy, enhanced predictive maintenance, optimized RES integration, and strengthened security. As these technologies continue to evolve, their applications in power systems are expected to expand, driving further advancements in the efficiency and reliability of energy management [67].

4.4.1 Time Series Preprocessing Techniques

Preprocessing time series data is a crucial step in preparing it for meaningful analysis and modeling. Unlike conventional datasets, time series data is inherently sequential, meaning that the order of observations carries significant temporal relationships. This characteristic necessitates specific preprocessing techniques to address issues such as missing data, outliers, non-stationarity, noise, and trends.

1. Handling Missing Data: Missing data is a common problem in time series, arising from sensor failures, human errors, or transmission issues. Various imputation techniques can be applied depending on the context and the nature of the data [68]:

- **Linear interpolation** and **polynomial interpolation** are effective for smoothly varying datasets.
- **Forward fill** and **backward fill** methods are commonly used in real-time applications where the last observed value holds relevance.
- For more complex cases, **model-based imputation methods**, such as those relying on **k-nearest neighbors** or **regression**, can provide more accurate estimates.

2. Smoothing Techniques: Smoothing is a key method to reduce noise in time series, which is often characterized by random fluctuations or irregularities. The key factors are the following [68]:

- **Moving Averages**: Involves averaging each data point with its neighbors. This technique effectively smooths out short-term fluctuations and highlights longer-term trends by creating a smoothed version of the original time series.
- **Exponential Smoothing**: Assigns exponentially decreasing weights to past observations, allowing more recent data to have a greater influence on the smoothed values. Variants include:
  - Simple Exponential Smoothing: Suitable for data without trend or seasonality.
  - Holt's Linear Trend Model: Accounts for linear trends.
  - Holt-Winters Seasonal Model: Incorporates both trends and seasonal patterns.
- **Kernel Smoothing**: Utilizes a kernel function to weigh data points, offering a flexible approach that adjusts to the structure of the data. This method helps in smoothing by considering the local data distribution around each point.

3. Normalization and Scaling: Normalization and scaling are essential preprocessing techniques for time series data, ensuring that the data is in a consistent format suitable for analysis and modeling. These methods are particularly important when working with algorithms sensitive to the scale of input features, such as distance-based methods and neural networks.

Normalization involves adjusting the values of time series data to a common scale without distorting differences in the range of values. Common normalization and scaling methods include [69]:

- **Min-Max Scaling:** This method scales the data to a specified range, often between 0 and 1, using the formula below:

  $$X' = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

  It is effective when the data does not contain significant outliers, as it preserves relationships between data points.

- **Standard Scaling (Z-Score Normalization):** This method transforms the data to have a mean of zero and a standard deviation of one. The formula is:

  $$X' = \frac{X - \mu}{\sigma}$$

  where μ is the mean and σ is the standard deviation of the time series. It is particularly useful for datasets with varying distributions and for algorithms that assume normally distributed data.

Other scaling techniques, such as robust scaling, address the presence of outliers by using the median and interquartile range (IQR) instead of mean and standard deviation. These preprocessing methods enhance model convergence, improve interpretability, and ensure consistency in time series forecasting tasks.

4. Data Cleaning: Data cleaning is a crucial step in the preprocessing pipeline for time series data, ensuring that the dataset is accurate, complete, and suitable for analysis or modeling. Time series data often contains noise, missing values, outliers, and inconsistencies that can degrade the performance of forecasting and analytical models. Effective data cleaning mitigates these issues, resulting in more reliable and interpretable results [69].

Key tasks in time series data cleaning include:

- **Outlier Detection and Removal:** Outliers, which deviate significantly from other data points, can distort model performance. Techniques to detect and handle outliers include z-score analysis, interquartile range (IQR) methods, and domain-specific thresholds.

- **Noise Reduction:** Time series data often includes random fluctuations or noise that can obscure meaningful patterns. Techniques like moving averages, smoothing filters, and wavelet transforms are commonly applied to reduce noise.

- **Handling Duplicates and Inconsistencies:** Duplicate entries and inconsistent formatting in timestamps or values can lead to inaccuracies. These issues are addressed by removing duplicate entries and standardizing formats to ensure uniformity.

5. Stationarity and Trend Removal: Time series models often assume stationarity, where the statistical properties of the series (mean, variance, etc.) remain constant over time. However, real-world time series frequently exhibit trends or seasonal patterns, violating this assumption. Differencing, which involves subtracting consecutive observations, is a common technique for removing trends and achieving stationarity. Seasonal decomposition methods, such as STL (Seasonal and Trend decomposition using Loess), further isolate the trend, seasonality, and residual components, enabling more focused analysis [70].

6. Feature Engineering: Feature engineering in time series analysis involves creating features that enhance model performance by capturing relevant patterns and dependencies [71].

- Temporal Features: Extracting components such as trend, seasonality, and cyclical patterns helps in capturing long-term movements and periodic fluctuations. Decomposing the time series into these components provides valuable insights into its underlying structure.

- Lag Features: Creating features based on previous time steps, or lag values, helps the model to learn temporal dependencies and patterns from past observations. This is crucial for capturing how past values influence future ones.

- Rolling Statistics: Features such as moving averages and rolling standard deviations smooth out short-term fluctuations and reveal longer-term trends. These statistics help to understand the data's underlying patterns more clearly.

- Differencing and Transformations: Techniques like differencing (calculating the difference between consecutive observations) help to stabilize the mean of the time series, while transformations (e.g., logarithmic or power transformations) can stabilize variance and make the data more suitable for modeling.

- External Features: Incorporating additional variables such as calendar features (day of the week, holidays), weather data, or economic indicators can provide additional context that influences the primary time series.

7. Resampling and Synchronization: Resampling is essential when time series data is recorded at different frequencies. Upsampling involves interpolating intermediate values to increase the frequency, while downsampling aggregates values (e.g., by calculating means or sums) to reduce the frequency. Synchronizing time series with inconsistent time stamps, such as aligning data collected from multiple sources, ensures uniformity, which is critical for comparative analysis or multi-variable modeling.

By addressing these preprocessing challenges, time series data can be transformed into a more manageable and informative form, paving the way for effective modeling and insightful analysis. Proper preprocessing ensures that temporal relationships are preserved while eliminating noise and inconsistencies, thereby enhancing the reliability of downstream applications [69].

4.4.2 Forecasting Methods

The methodologies used for timeseries forecasting tasks have been enhanced and progressed in high level the last five to ten years. Despite the fact that in the early years we saw statistical models being used, with advances in technology, there is an increasing trend toward the use of Machine Learning (ML), Deep Learning (DL), and Ensemble Methods, which appear to dominate over the former due to the flexibility and high performance they offer. Starting from traditional methods, there will be an extensive mathematical presentation of the most fundamental models, both statistical and those involving machine, deep, and ensemble learning methods. This section provides a detailed overview of the most common models used for these purposes. Initially, the statistical models are presented, followed by the machine and deep learning, and finally, the ensemble methods are analyzed. For a better understanding of the categories and the variety of the models, Figure 4.1 graphically presents the aforementioned.

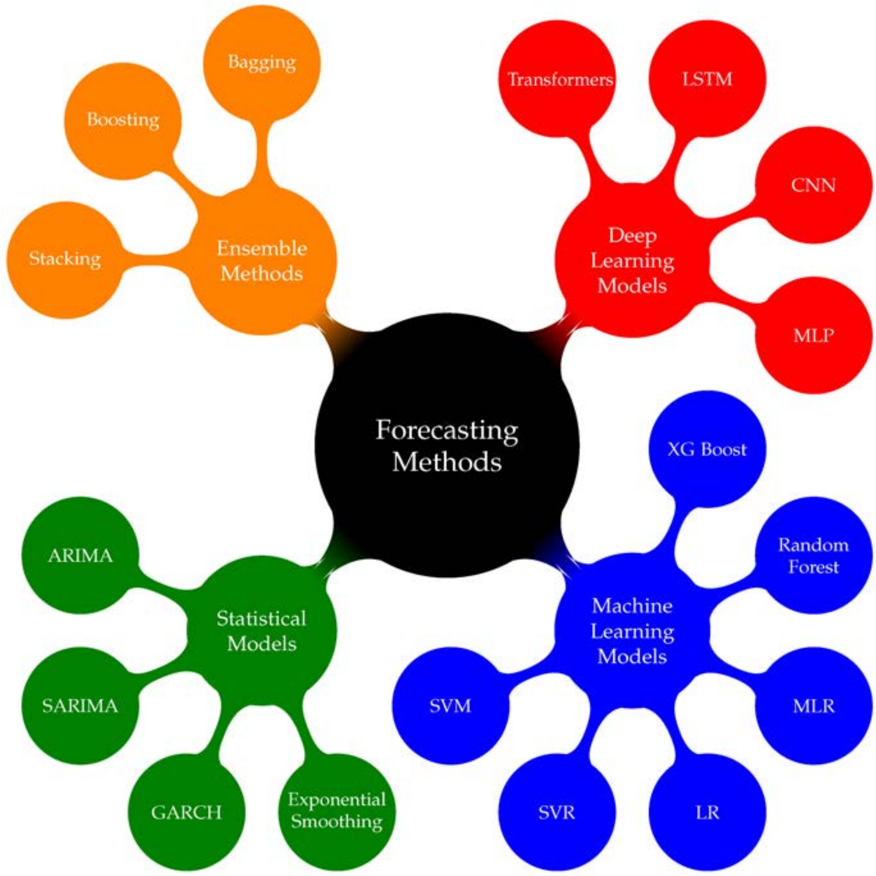


Figure 4.1: Timeseries Forecasting Methods

4.4.3 Statistical models

Statistical models are essential in ELF and EPF, providing robust frameworks to understand and predict trends in time series data. Among the most prevalent models are AutoRegressive Integrated Moving Average (ARIMA) models, which combine autoregression, differencing, and moving averages to model temporal dependencies and achieve stationarity. Seasonal ARIMA (SARIMA) models manage seasonal changes effectively by using seasonal differencing and seasonal autoregressive and moving average parts. Exponential Smoothing methods, like Holt-Winters, are popular for being simple and effective in dealing with seasonality and trends in time series. Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models are used to capture and predict volatility patterns, mainly in financial data. State Space Models and Kalman Filters offer a flexible way to model dynamic systems, especially for time-varying processes. These methods are important in the energy sector and are presented in detail below.

AutoRegressive Integrated Moving Average

AutoRegressive Integrated Moving Average (ARIMA) analyzes time series data and predicts future trends based on historical values. This model is characterized by three main parameters: p, d, and q. The first parameter p is the number of lag observations of the model, which constitutes the autoregressive part (AR). The second parameter d is the number of times the raw observations are differenced to achieve stationarity, which is the integrated part (I). The third parameter q refers to the Moving Average part of the model (MA) [72].

The basic equations of this model can be expressed as follows:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \tag{4.1}$$

where, y_t represents the value at time point t, c a is a constant, ϕ_i are the coefficients for the AR part, θ_i are the coefficients for the MA part, and ϵ_t is the error term at time t. The autoregressive (AR) part of the model means that the output is directly influenced by its past values and can be written as:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t \tag{4.2}$$

The integrated (I) part of the model is used for differencing the raw observations, in order to make the data more stationary. It can be expressed as:

$$y'_t = y_t - y_{t-1} \tag{4.3}$$

Finally, the MA part represents the dependency between an observation and a residual error. It can be written as:

$$y_t = c + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{4.4}$$

Seasonal AutoRegressive Integrated Moving Average

Seasonal AutoRegressive Integrated Moving Average (SARIMA), is an extension of the ARIMA model, because its capability to handle seasonality in time series. The model integrates both non-seasonal and seasonal components. Three parameters p, d, q are represent the seasonal part and four parameters P, D, Q, s are for the non-seasonal part. p is the number of non-seasonal AR terms, d is the number of non-seasonal differences, and q is the number of non-seasonal MA terms. For the seasonal components, P indicates the number of seasonal autoregressive terms, D represents the number of seasonal differences, and Q represents the number of seasonal moving average terms. Finally, the parameter s is the length of the seasonal cycle [73].

The main formulation of SARIMA can be expressed with the following equation [73]:

$$\begin{aligned} y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \\ + \sum_{k=1}^{P} \Phi_k Y_{t-ks} + \sum_{l=1}^{Q} \Theta_l \epsilon_{t-ls} + \epsilon_t \end{aligned} \tag{4.5}$$

where, ϕ_i are the non-seasonal AR factors, θ_i are the factors of MA part, Φ_k are the factors for the seasonal AR part, Θ_i are the MA seasonal factors, s is the seasonal period, and ϵ_t is the error term at time t.

Non-seasonal and Seasonal Parts: The non-seasonal part is similar to the ARIMA model and can be written as the equation (1). The seasonal part operates for the seasonal fluctuations and can be written as:

$$Y_t = \sum_{k=1}^{P} \Phi_k Y_{t-ks} + \sum_{l=1}^{Q} \Theta_l \epsilon_{t-ls} \tag{4.6}$$

where, Y_t represents the seasonal component of y_t.

Differencing: The seasonal and non-seasonal differencing part can be expressed as follows: The non-seasonal differencing can be expressed as the equation (3), while the seasonal differencing can be written as:

$$Y_t' = Y_t - Y_{t-s} \tag{4.7}$$

Exponential Smoothing

Exponential Smoothing (ES) method is suitable and effective in dealing with univariate datasets. It gives more importance to recent data values by applying weights that reduced in exponential way from the recent to the older data, so recent observations influence the forecast more than older ones.

There are three main exponential smoothing methods:

- Single Exponential Smoothing (SES): This method is suitable for time series without trend or seasonality.

- Double Exponential Smoothing (DES): Proper for data with a trend.

- Triple Exponential Smoothing (TES): Or Holt-Winters, which is suitable for data with both trend and seasonality patterns.

Single Exponential Smoothing (SES): The main equation for SES is:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t \tag{4.8}$$

where, $\hat{y}_{t+1}$ is the prediction for the time $t+1$, α is the smoothing parameter with $0 < \alpha < 1$, y_t is the actual and $\hat{y}_t$ is the predicted value at time t, respectively.

Double Exponential Smoothing (DES): DES introduces a second equation to account for trends:

$$\hat{y}_{t+1} = l_t + b_t \tag{4.9}$$

$$l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{4.10}$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \tag{4.11}$$

where, l_t is the level component at time t, b_t is the trend component at time t, and α and β are smoothing parameters, with $0 < \alpha, \beta < 1$.

Triple Exponential Smoothing (TES) - Holt-Winters: TES or Holt-Winters method, adds a seasonal component, as follows:

$$\hat{y}_{t+m} = l_t + b_t m + s_{t+m-s} \tag{4.12}$$

$$l_t = \alpha \frac{y_t}{s_{t-s}} + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{4.13}$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \tag{4.14}$$

$$s_t = \gamma \frac{y_t}{l_t} + (1-\gamma)s_{t-s} \tag{4.15}$$

where, l_t is the level component at time t, b_t represents the trend component, s_t is the seasonal component, s is the length of the seasonal cycle, α, β, and γ represent the smoothing variables, with values between 0 and 1, and m is the number of periods ahead to be forecasted [74].

Generalized Autoregressive Conditional Heteroskedasticity

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is a statistical model used to forecast future fluctuations in time series. It is mainly useful in financial applications, where GARCH predicts the volatility of financial markets. This model utilizes lagged terms of both the error and its variance and can be expressed through the mean and variance equations, as follows [75]:

The mean equation:

$$y_t = \mu + \epsilon_t \tag{4.16}$$

where, y_t represents the value at time t, μ is the mean, and ϵ_t is the error term.

The variance equation:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{4.17}$$

where, σ_t^2 is the conditional variance at time t, α_0 is a constant, α_i are the coefficients for the lagged squared residuals (ϵ_{t-i}^2), β_j are the coefficients for the lagged variances (σ_{t-j}^2), p is the order of the GARCH terms, and q is the order of the ARCH terms. Finally, the error term ϵ_t is typically modeled as:

$$\epsilon_t = \sigma_t z_t \tag{4.18}$$

where, z_t is a white noise process with zero mean and unit variance.

4.4.4 Machine learning models

Machine Learning (ML) models are effective in ELF and EPF tasks, and are capable of dealing with complex and nonlinear datasets that contain fluctuations and volatility. They are used for both regression and classification tasks and the most commonly models are presented below.

Support Vector Machine

Support Vector Machine (SVM) is a supervised learning model, of which the basic working mechanism is to find the hyperplane that best separates the data into two different groups, maximizing the margin among them. Given a set of training data $\{(x_i, y_i)\}_{i=1}^{n}$, where x_i is the feature vector and y_i is the label, the alrorithm aims to find the optimal hyperplane for the best data separation. There are two categories, the linear and non-linear approach [76].

Linear SVM: For linear SVM, the working function is given by:

$$f(x) = w \cdot x + b \tag{4.19}$$

where, w is the weight vector, and b is the bias term.

The optimization problem of finding the optimal hyperplane can be formulated as:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \tag{4.20}$$

subject to the constraints:

$$y_i(w \cdot x_i + b) \geq 1 \quad \text{for all } i = 1, 2, \ldots, n \tag{4.21}$$

Non-linear SVM: For non-linearly separable data, SVM can be extended using the kernel trick. The function becomes as follows:

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b \tag{4.22}$$

where, α_i are the Lagrange multipliers, and $K(x_i, x)$ is the kernel function, which computes the inner product between two feature vectors in a high-dimensional space.

Support Vector Regression

Support Vector Regression (SVR) is an extension of SVM model used for regression tasks. It aims to find a function that approximates the relationship between input features and continuous output values with minimal error, while maintaining model simplicity.

The objective of the model is to find a function $f(x) = w \cdot x + b$ that approximates the target values y_i within a margin of ϵ, while keeping the model complexity as low as possible. The SVR optimization problem is expressed as:

$$\min_{w,b,\xi_i,\xi_i^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{4.23}$$

Subject to the constraints:

$$\begin{aligned} & y_i - (w \cdot x_i + b) \leq \epsilon + \xi_i \\ & (w \cdot x_i + b) - y_i \leq \epsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \end{aligned} \tag{4.24}$$

where, w is the weight vector, b is the bias term, $\xi_i$and ξ_i^* are slack variables for points outside the margin, C is the regularization parameter, and n is the number of training samples [76].

Linear Regression

Linear Regression (LR) is a machine learning method, which assumes a linear relationship between an input variable x and the target variable y. The model can be represented as [77]:

$$y = \beta_0 + \beta_1 x + \epsilon \tag{4.25}$$

where, y is the target variable, β_0 is the intercept, β_1 is the slope (coefficient) of the input variable x, and ϵ is the error term.

The objective is to find the values of β_0 and β_1 that minimize the sum of squared residuals (the difference between the observed and predicted values):

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2 \tag{4.26}$$

where, n is the number of observations.

Multilinear Regression

Multilinear Regression (MLR) extends Linear Regression by modeling the relationship between a continuous target variable and multiple input features. It assumes a linear relationship between the target variable y and multiple input variables $\mathbf{x} = (x_1, x_2, \ldots, x_p)$. Similarly with Linear Regression, this model can be represented as [77]:

$$y = \beta_0 + \ldots + \beta_p x_p + \epsilon \tag{4.27}$$

where, y is the target variable, β_0 is the intercept, β_i are the coefficients of the input variables x_i, ϵ is the error term.

The objective is to find the values of $\beta_0, \beta_1, \ldots, \beta_p$ that minimize the sum of squared residuals:

$$\min_{\beta_0,\beta_1,\ldots,\beta_p} \sum_{i=1}^{n} (y_i - (\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}))^2 \tag{4.28}$$

where, n is the number of observations, and p is the number of input variables.

Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a very effective and scalable algorithm used for supervised learning tasks. It belongs to the ensemble learning family and is based on the gradient boosting framework. XGBoost combines the outputs of several models, typically decision trees. It trains each tree sequentially, with each tree correcting the errors made by the previous trees.

The objective function in XGBoost includes two parts: the loss function, which measures the difference between predicted and true values, and a regularization term, which manages model complexity to reduce overfitting.

$$\text{Obj}(\Theta) = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{4.29}$$

where, Θ represents the model parameters, $L(y_i, \hat{y}_i)$ is the loss function measuring the difference between the true label y_i and the predicted label $\hat{y}_i$, and $\Omega(f_k)$ is the regularization term for the k-th tree f_k.

About the Loss Function, XGBoost typically uses the mean squared error (MSE):

$$L(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \tag{4.30}$$

where, $\hat{y}_i$ is the predicted probability of class 1.

Finally, in order the models to avoid overfitting, the regularization term, consists of two parts: tree complexity term and leaf weight regularization term. It can be writtern as:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|w\|_2^2 \tag{4.31}$$

where, T is the number of leaves in the tree f_k, and w are the weights associated with the leaf nodes [78].

Random Forest

Random Forest (RF) is an ensemble learning algorithm, which is a combination of multiple decision trees. This model generates a set of decision trees during training and produces the final output by taking the mode of the average predictions from each individual tree. Each decision tree is trained on a random subset of the training data, and at each split in a tree, a random subset of features is considered. This randomness helps in making the model robust and reducing variance.

Mathematically, the prediction $\hat{y}$ of a Random Forest model is given by [79]:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t(x) \tag{4.32}$$

where, T is the number of trees in the forest, $f_t(x)$ is the prediction of the t-th tree for the input x. For classification, the mode of the class labels predicted by each tree is taken, while for regression, the average of all tree predictions is used.

4.4.5 Deep learning Models

Similar to ML algorithms, the most commonly used DL models are the presented below.

Multilayer Perceptron

The Multilayer Perceptron (MLP), which is presented in Figure 4.2, is a simple feedforward neural network (FNN) model that consists of one input layer of nodes, hidden layers, and one output layer, which creates the final prediction. Each node in each layer is connected to every node in the previous and next layer, through neurons, in which correspond synaptic weights, which are used for the transmission of information across the model. An activation function, such as the rectified linear unit (ReLU) is used in order to calculate the output of each node.

Given an input vector x, the output of a node j in the hidden layer h can be calculated as follows:

$$z_j^h = \sum_{i=1}^{n_{\text{input}}} w_{ij}^h x_i + b_j^h \tag{4.33}$$

$$a_j^h = f(z_j^h) \tag{4.34}$$

where, n_{input} is the number of nodes in the input layer, w_{ij}^h is the weight of the connection between the i-th node in the input layer and the j-th node in the hidden layer, b_j^h is the bias of the j-th node in the hidden layer, and f is the activation function.

Similarly, the output of a node k in the output layer can be computed as:

$$z_k^{\text{output}} = \sum_{j=1}^{n_{\text{hidden}}} w_{jk}^{\text{output}} a_j^h + b_k^{\text{output}} \tag{4.35}$$

$$a_k^{\text{output}} = f(z_k^{\text{output}}) \tag{4.36}$$

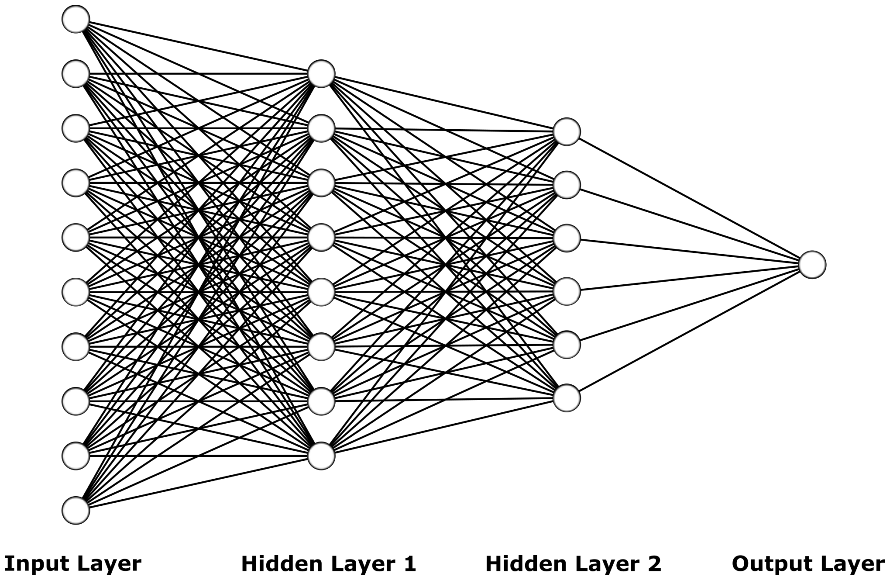


Figure 4.2: Multilayer Perceptron

The training of MLP model is carried out through backpropagation method, in which the error between actual and forecasting values is minimized by adjusting the weights and biases using gradient descent. The loss function used is based on each different task. Usually, mean squared error is used for time series forecasting and cross-entropy loss for classification tasks [80].

Long Short-Term Memory

Long Short-Term Memory (LSTM) is a type of recurrent based architecture, more suitable to manage sequential data, like energy consumption, and capture long-term dependencies between them from RNNs. LSTM networks, unlike regular RNNs, are capable of maintaining information for extended periods, which helps solve the vanishing gradient problem that often occurs in simple RNNs.

LSTM networks are composed of LSTM cells, which main structure are presented in Figure 4.3, and has three main gates:

- Forget Gate: Determines what information to discard from the cell state.
- Input Gate: Selects which new information to store in the cell state.
- Output Gate: Chooses what part of the cell state is proceed to the output.

The equations for the LSTM cell are as follows:

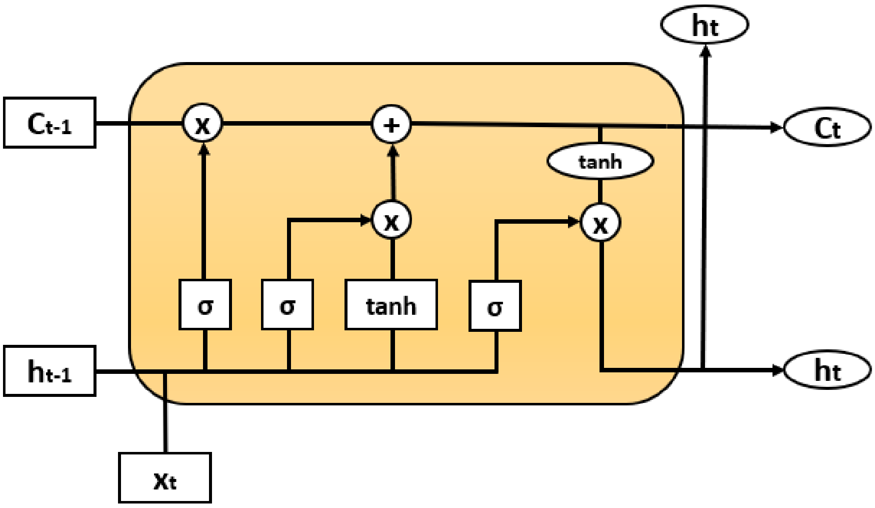


Figure 4.3: Long Short Term Memory Structure

Forget Gate

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{4.37}$$

Input Gate

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4.38}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{4.39}$$

Cell State Update

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{4.40}$$

Output Gate

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{4.41}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{4.42}$$

where, x_t is the input at time step t, h_{t-1} represents the hidden state from the previous time step, C_{t-1} is the cell state from the previous time step, f_t is the forget gate activation, i_t is the input gate activation, $\tilde{C}_t$ is the candidate cell state, C_t is the updated cell state, o_t is the output gate activation, h_t is the hidden state at time step t, W_f, W_i, W_C, W_o are the weight matrices, b_f, b_i, b_C, b_o are the bias vectors, σ is the sigmoid activation function, and $\tanh$ is the hyperbolic tangent activation function [81].

Convolutional Neural Network

A Convolutional Neural Network (CNN) includes convolutional layers, pooling layers, and fully connected layers. More specifically, its structure is presented in Figure 4.4 and includes the following:

Convolutional Layer: In this layer, convolution operations are applied to the input data through a series of learnable filters (also known as kernels). Each filter slides across the input, generating a feature map. The convolution process is defined as:

$$(f * x)(i, j) = \sum_{m} \sum_{n} x(i + m, j + n) \cdot f(m, n) \tag{4.43}$$

where, x is the input, f is the filter, and $*$ denotes convolution.

Pooling Layer: This layer reduces the spatial dimensions (height and width) of the feature maps, helping to decrease the number of parameters and computational load in the network. One common pooling operation is Max Pooling, which is defined as:

$$y(i, j) = \max_{m,n} x(i + m, j + n) \tag{4.44}$$

where, x is the input and y is the pooled output.

Final Fully Connected Layer: The fully connected layer creates the final prediction. Each neuron of this layer is connected to every neuron in the previous layer. The output is computed as:

$$y = W \cdot x + b \tag{4.45}$$

where, W is the weight matrix, x is the input vector, and b is the bias vector [81].

Hybrid CNN-LSTM Model

The CNN-LSTM model is a hybrid deep learning algorithm that uses the CNN layers for feature extraction of the input dataset and the LSTM model for sequence prediction, like multistep ahead forecasting of time series. Although the process of this model is divided into two parts, the operation of each of the two components-models separately is the same as

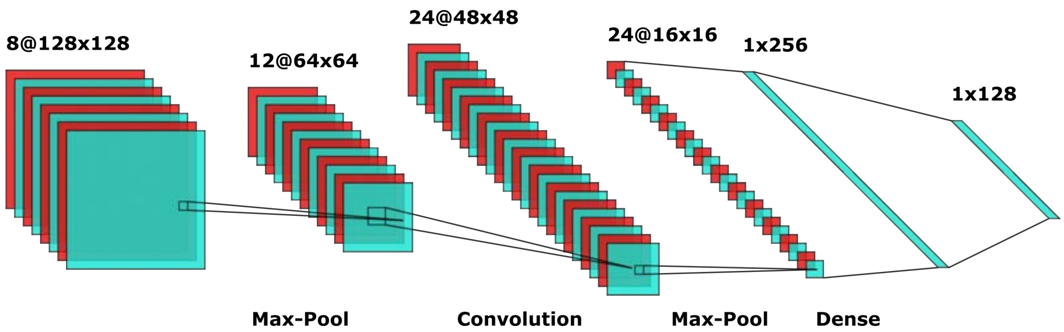


Figure 4.4: Convolutional Neural Network Architecture

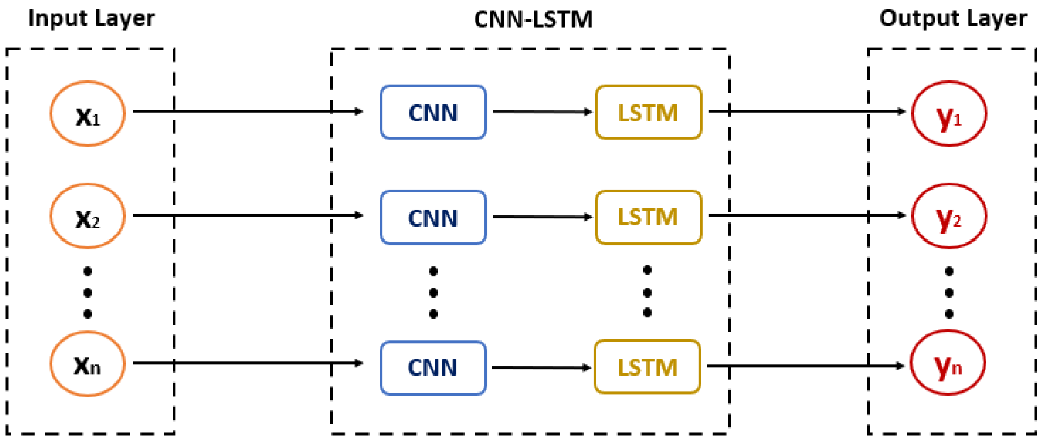


Figure 4.5: CNN-LSTM Model Architecture

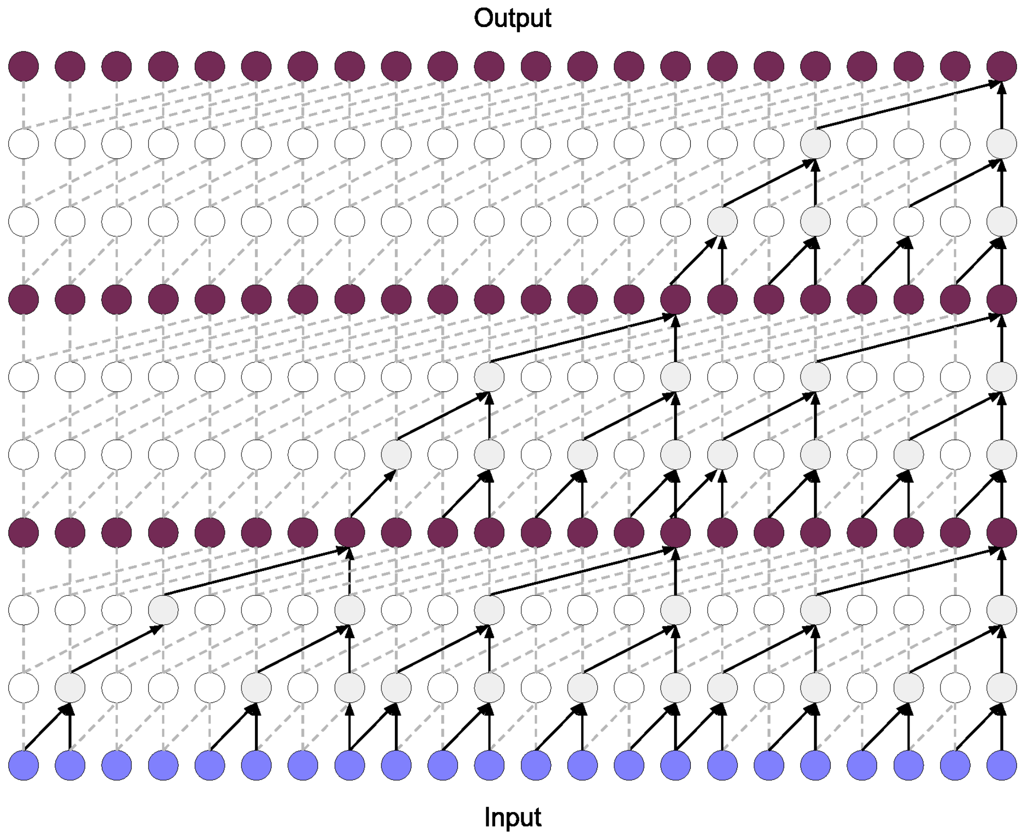
described above for LSTM and CNN algorithms, respectively. Figure 4.5 shows the CNN – LSTM's architecture.

Temporal Convolutional Networks

Temporal Convolutional Network (TCN) is a relatively new type of deep learning model that has a dilation of CNN- 1D layers with the same input and output length. The working principle of TCNs is the following [82]:

- At first, the model computes the low-level features using the CNN-1D modules encoding the spatial - temporal information,
- Secondly, the model feeds this information and classifies them using Recurrent Neural Networks.

Recent studies have proven that TCNs have great performance in predicting time series with multiple seasonalities and trends. Figure 4.6 presents the basic architecture of a TCN.



Source: [82]

Figure 4.6: Temporal Convolutional Network

Transformer Models

Transformer models have revolutionized time series forecasting with their ability to handle complex dependencies through self-attention mechanisms. It is a deep learning architecture designed for sequence-to-sequence tasks such as machine translation and text generation. Unlike earlier models that relied on recurrent or convolutional structures, the Transformer employs only attention mechanisms to process input sequences in parallel, significantly improving efficiency and scalability.

The Transformer consists of two main components: the encoder and the decoder.

- Encoder: Processes input sequences.

- Decoder: Generates output sequences from encoded representations.

The Encoder: The encoder is responsible for processing input sequences and generating context-aware representations. The process begins with input embeddings and positional encoding. Since the model lacks recurrence, it requires positional encoding to maintain order information. The input tokens are first converted into dense vector representations (embeddings) and then combined with positional encodings.

Within each encoder layer, a multi-head self-attention mechanism is employed. Each token attends to every other token in the input sequence, allowing the model to capture dependencies regardless of distance. This is achieved using scaled dot-product attention, which computes weighted sums of token representations based on their relevance. The output of the attention mechanism is then passed through an addition and normalization layer, ensuring stable gradient flow and improved learning. Following this, a fully connected feed-forward layer is applied independently to each token's representation, introducing non-linearity to the model. The encoder consists of multiple such layers stacked on top of each other, progressively refining token representations [83].

The Decoder: The decoder generates the output sequence one token at a time, utilizing the encoder's context-aware representations. Similar to the encoder, input embeddings are combined with positional encodings before being processed through multiple identical layers.

Each decoder layer begins with a masked multi-head self-attention mechanism. Unlike the encoder's self-attention, the decoder applies a mask to prevent positions from attending to future tokens, ensuring proper autoregressive generation. The output is then processed through an encoder-decoder attention mechanism, which allows the decoder to attend to the encoder's output and incorporate context from the input sequence. A feed-forward network and normalization layers further refine the token representations. The final processed representations are passed through a linear transformation and softmax activation to produce output probabilities over the vocabulary [83].

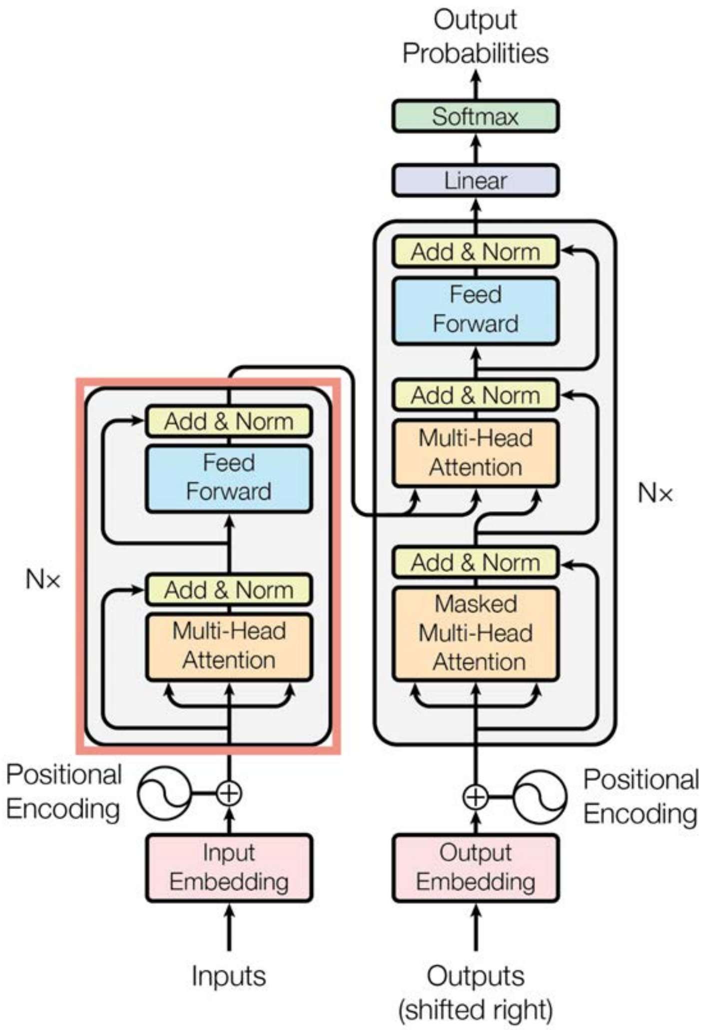
Each of these comprises multiple identical layers, containing key operations such as multi-head attention, feed-forward networks, and normalization layers [81]:

The core component is the Self-Attention Mechanism, which computes the attention scores using the following equations [83]:

Scaled Dot-Product Attention:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where, Q is the query matrix, K is the key matrix, V is the value matrix, and d_k is the dimension of the key vectors.



Source: [83]

Figure 4.7: Transformer Model Architecture

Multi-Head Attention:

$$\text{MultiHead}(Q, K, V) = \text{Concat}\left(\text{head}_1, \text{head}_2, \ldots, \text{head}_h\right) W^O \tag{4.46}$$

where, each head is computed as:

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{4.47}$$

and W_i^Q, W_i^K, W_i^V are parameter matrices for the i-th attention head, and W^O is the output matrix.

Figure 4.7 visualizes the typical structure of Transformer model.

Key Advantages of the Transformer Model: The Transformer model provides several advantages over traditional sequence-to-sequence architectures. Due to its parallelization capability, it processes entire sequences simultaneously, significantly reducing training time. The self-attention mechanism enables direct interactions between distant tokens, improving performance on long texts. Moreover, its scalability has led to widespread adoption, serving as the foundation for modern large-scale models such as BERT and GPT.

Ensemble Methods

Ensemble methods combine predictions from multiple models to improve accuracy. Techniques such as Random Forests and Gradient Boosting are commonly used:

$$\hat{y} = \sum_{m=1}^{M} w_m f_m(x) \tag{4.48}$$

where M is the number of models, w_m are weights, and $f_m(x)$ are the model predictions.

The choice of model for time series forecasting depends on various factors, including the nature of the data, the required forecast horizon, and the computational resources available. Traditional models offer simplicity and interpretability, while advanced ML and DL models provide the capacity to capture complex patterns and dependencies. The ongoing development in these areas continues to enhance the capabilities of forecasting systems, making them crucial tools in a wide range of applications.

The primary concept underlying ensemble methods is to capitalize on the advantages of different algorithms and combine their ultimate predictions, resulting in enhanced accuracy and robustness. These models can be applied to a variety of tasks, including statistical, as well as ML and DL models for classification, regression, and clustering tasks.

A key benefit of these algorithms is their capacity to decrease overfitting, which leads to bad forecasting performance and variance, which represents the level of fluctuation in predictions across different models within the ensemble, respectively. By combining the forecasts of several models, ensemble methods help minimize variance and produce more reliable predictions.

Regarding the disadvantages, they are typically more complicated to implement and interpret compared to individual models, and in most cases they require significantly more computational power. Their performance is highly dependent on the combination of models used, which makes identifying the best set of models a difficult task. Additionally, the ensemble's method quality is pivotal due to the fact that poorly performing models can negatively affect the ensemble final prediction, leading to issues such as overfitting or underfitting.

The most common techniques used to create ensemble models are bagging, boosting, and stacking, which are analyzed below [84].

Bagging

Bagging consists of training each model individually and then combining their predictions using a weighted average. This approach can be used with any type of model and is effective at reducing variance and preventing overfitting. Initially, several iterations of the dataset are executed, each containing distinct subsets of the initial data. After, these subsets are used to train models separately, and finally their predictions are integrated to produce the final output. By merging predictions from different models in a weighted average process, bagging improves both the accuracy and reliability of the final predictions [81, 84].

Boosting

Boosting is a technique where each model is trained to correct the errors in predictive data analysis. Models are trained using labeled data to make predictions about unlabeled data. This method improves the accuracy of the final predictions but is more vulnerable to overfitting in comparison to bagging method. Boosting begins with a simple model and gradually adds more complex models, each focusing on the data points that were wrongly labeled by previous models. Each of the different forecasts is merged using a weighted combination based on their final accuracy [81, 84].

Stacking

Stacking, also called stacked generalization, involves combining the predictions from various base models—sometimes known as first-level models or base learners—to produce a conclusive prediction. This approach entails training multiple base models on an identical dataset and utilizing their output as input for a higher-level model, termed a meta-model or second-level model, to produce the final outcome. The fundamental concept is to improve predictive accuracy by combining the outputs of various base models, rather than depending on a single model [81, 84].

4.4.6 Evaluation Metrics

In terms of related work, the evaluation metrics that are most frequently employed are included in Table 4.1 [85, 86, 87]:

Table 4.1: Common Evaluation Metrics for Time Series Forecasting

Metric	Equation	Description
Mean Absolute Error (MAE)	$MAE = \frac{1}{n}\sum_{t=1}^{n} \lvert y_t - \hat{y}_t \rvert$	Measures the average absolute difference between actual and predicted values. Lower values indicate better performance.
Mean Squared Error (MSE)	$MSE = \frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{y}_t)^2$	Penalizes larger errors more than MAE by squaring the differences, making it sensitive to outliers.
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{MSE}$	Provides the error in the same units as the original data, making interpretation easier.
Mean Absolute Percentage Error (MAPE)	$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left\lvert \frac{y_t - \hat{y}_t}{y_t} \right\rvert \times 100$	Expresses error as a percentage, useful for comparisons across datasets. Can be problematic if y_t is close to zero.
Symmetric Mean Absolute Percentage Error (sMAPE)	$sMAPE = \frac{100}{n}\sum_{t=1}^{n} \frac{\lvert y_t - \hat{y}_t \rvert}{(\lvert y_t \rvert + \lvert \hat{y}_t \rvert)/2}$	A symmetric version of MAPE that prevents issues with division by small values.
Mean Absolute Scaled Error (MASE)	$MASE = \frac{1}{n}\sum_{t=1}^{n} \frac{\lvert y_t - \hat{y}_t \rvert}{\frac{1}{n-1}\sum_{t=2}^{n} \lvert y_t - y_{t-1} \rvert}$	Scales errors relative to a naïve baseline model, more robust for intermittent demand.
R-squared (R^2)	$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$	Measures the proportion of variance explained by the model. Values close to 1 indicate a better fit.

where x_i, y_i, $\hat{x}_i$, and $\bar{y}$ are the forecasting values, the actual values, the mean of the forecasting values, and the mean of the actual values, respectively.

Apart from the above metrics, it is observed that some publications additionally use the following: Mean Arctangent Absolute Percentage Error (MAAPE), Mean Percentage Error (MPE), Error Sum of Squares (SSE), Sum of Squares (SSE), Mean of Mean Squared Error (FMSE), Symmetric Mean Absolute Percentage Error (sMAPE), Median Absolute Error (MedAE), Matthews Correlation Coefficient (MCC), Relative Root Mean Square (rRMSE), Normalized Root Mean Square Error (NRMSE), Coefficient of Variation (CV), and Accuracy, which is measured by the proportion of correct predictions out of the total number of predictions.

Chapter 5

Electricity Load Forecasting for the market mechanisms

In this chapter, we will delve into the field of Electricity Load Forecasting. Initially, the reasons and importance of accurate load forecasting will be highlighted. Next, we will proceed with an extensive literature review, in which notable works exploring this specific field will be analyzed, covering all types of models, from statistical, machine learning, deep learning, to ensemble methods. Following this, we will present two advanced implementations and emphasize valuable conclusions that arise.

5.1 The Importance of Electricity Load Forecasting in Modern Power Systems

The ability to accurately predict electricity is the cornerstone of modern power system planning and operation. Accurate predictions of future electricity demand are essential to ensure the reliability, efficiency, and economic viability of power systems. This section delves into the multifaceted importance of load forecasting, exploring its critical role in infrastructure development, operational management, integration of renewable energy sources, and the facilitation of competitive electricity markets [88].

The importance of ELF spans infrastructure development, operational efficiency, integration of renewable energy, market stability, demand-side management, reliability, and environmental sustainability. As the energy landscape continues to transform, the role of accurate and sophisticated load forecasting will become increasingly vital in ensuring resilient and efficient power systems. More specifically, the importance of accurate forecasting is analyzed on the following pillars.

5.1.1 Enhancing Operational Efficiency

In the short term, load forecasting is vital for the efficient operation of power systems. Day-to-day management relies on accurate demand predictions to schedule generation units effectively, optimize fuel usage, and minimize operational costs. By anticipating demand fluctuations, system operators can adjust generation schedules to maintain a balance between supply and demand, thereby reducing the need for expensive peaking power plants and minimizing operational inefficiencies.

5.1.2 Facilitating Renewable Energy Integration

The global shift towards renewable energy sources, such as wind and solar energy, introduces additional variability and uncertainty into power systems. The forecast of load, in conjunction with the forecasts of generation for renewable sources, is crucial to maintaining the stability of the grid. Accurate load predictions help schedule flexible resources, such as energy storage systems and demand response programs, to accommodate the intermittent nature of renewables. This integration ensures a reliable supply of electricity while promoting the use of cleaner energy sources.

5.1.3 Supporting Competitive Electricity Markets

In deregulated electricity markets, load forecasting plays a pivotal role in market operations. Market participants, including generators, retailers, and traders, rely on demand forecasts to make informed bidding and trading decisions. Accurate load forecasts contribute to efficient market clearing, price discovery, and risk management. For example, overestimation of demand can lead to oversupply and depressed market prices, while underestimation can cause supply shortages and price spikes. Thus, precise load forecasting is essential for the stability and efficiency of competitive electricity markets.

5.1.4 Enhancing Demand-Side Management

Load forecasting is integral to demand-side management (DSM) strategies, which aim to influence consumer demand to achieve desired objectives, such as peak load reduction and energy conservation. By predicting periods of high demand, utilities can implement DSM programs, such as time-of-use pricing and demand response initiatives, to encourage consumers to shift or reduce their energy usage. These programs not only alleviate stress on the power system during peak periods but also lead to cost savings for both utilities and consumers.

5.1.5 Foundation for Infrastructure Planning

Long-term load forecasting provides the necessary data for the development of strategic infrastructure. Power plants, transmission lines, and distribution networks require significant lead times for planning, approval, and construction. Accurate demand projections enable utilities to make informed decisions about when and where to expand capacity, ensuring that supply meets future demand without unnecessary overinvestment. Conversely, underestimating future load can lead to capacity shortages, jeopardizing system reliability, and leading to potential load curtailments. Therefore, precise load forecasting is indispensable for the timely and cost-effective expansion of the power system infrastructure.

5.1.6 Improving Reliability and Risk Management

Reliability is a paramount concern in power system operations. Load forecasting enables utilities to anticipate demand patterns and prepare for potential contingencies. By understanding expected load profiles, system operators can develop reserve strategies, plan maintenance activities, and implement emergency response plans effectively. Moreover, accurate load forecasts aid in risk assessment and management, allowing utilities to identify and mitigate potential threats to system stability and reliability.

5.1.7 Economic and Environmental Benefits

Accurate load forecasting contributes to the economic efficiency of power systems by optimizing resource allocation and reducing operational costs. By aligning generation with demand, utilities can minimize fuel consumption and reduce reliance on expensive peaking plants. Additionally, efficient load management leads to lower greenhouse gas emissions, as it enables greater integration of renewable energy sources and reduces the need for fossil fuel-based generation. Consequently, load forecasting not only supports economic objectives but also promotes environmental sustainability.

5.1.8 Technological Advancements

The field of load forecasting has evolved significantly with advancements in data analytics, machine learning, and artificial intelligence. Modern forecasting techniques leverage vast amounts of data, including historical load patterns, weather information, and socio-economic indicators, to improve prediction accuracy. The integration of smart grid technologies and the proliferation of Internet of Things (IoT) devices provide real-time data, enabling more dynamic and responsive forecasting models. As power systems continue to evolve, ongoing research and development in load forecasting methodologies will be essential to address emerging challenges and harness new opportunities.

5.2 Related Work on Electricity Load Forecasting

In this section, important related studies that have significantly contributed to the field of load forecasting are extensively analyzed. Initially, statistical models are presented, followed by publications focusing on ML and DL, and finally, emphasis is given to Ensemble Methods. Because most studies typically present models from both the ML and DL categories, it was chosen to present them in a common subsection, for a more substantial presentation of the related works.

5.2.1 Statistical Models

In the field of ELF, several methodologies and techniques have been investigated. These can be separated into two basic groups, the statistical and the modern models. Statistical models include models like AR, ARMA, ARIMA,model s with exogenous features, like ARMAX and ARIMAX, grey (GM) and ES. In tis way, Sethi et al. in [89] compare ARIMA method with other ML models for load prediction. Evaluation is done using RMSE and Mean Bias Error (MBE) and as case study is used a commercial building (amusement park) in San Diego. Chou et al. in [90] conduct a comparative analysis, using SARIMA model. The study use real-time consumption data collected from a smart grid infrastructure that is connected to a three-floor building in Xindian District, New Taipei City, Taiwan. Several metrics including R-squared, Maximum Absolute Error (MaxAE), RMSE, MAE and MAPE are used for the evaluation process. Hadri et al. in [91] conduct a comprehensive research employing ARIMA, SARIMA, XGBoost, Random Forest (RF), and LSTM. For the evaluation period, Mean Percentage Error (MPE), MAPE, MAE, and RMSE are utilized. Simultaneously, Vrablecová et al. in [92] apply ARIMA models in order to create short-term forecasts on public Irish CER dataset.

Some publications conduct comprehensive analysis of various statistical models used for ELF in different European Countries. Following the above, Pelka et al. in [93] investigate the performance of different statistical models, including ARIMA, ETS, and Prophet model. The target is to forecast monthly power demand, which estimates the correlation between past and future demand trends. As case study, electricity time series data for 35 European countries in a monthly resolution are used. In addition, Elamin et al. in [94] introduce a forecasting approach, utilizing a SARIMA model, incorporating both exogenous and interaction variables, to predict hourly load electricity demand. Finally, Mado et al. in [95] smooth the training and testing datasets of Java-Madura-Bali interconnection, Indonesia and propose a SARIMA model for short-term ELF, which considers double seasonal patterns in electricity consumption. The achieve MAPE equal to 1.56%.

5.2.2 Machine and Deep Learning Models

The ELF sector has undergone high transition towards the use of ML and DL models, which offer superior accuracy and adaptability compared to traditional statistical models. ML models are capable of handling with complex non-linear relationships in electricity consumption data. Meanwhile, DL models, including RNNs, LSTMs, and CNNs, excel in handling large volumes of data and uncovering intricate temporal patterns. These advancements play a critical role for improving the performance and sustainability of modern electricity systems, enabling better demand management, grid stability, and integration of RES. According to the above, Abumohsen et al. in [96] utilize several DL models, such as RNN, GRU and LSTM, in order to predict electricity consumption based on current measurements data. Forecasting models are crucial for reducing costs, optimizing resources, and improving load distribution for electric companies. Among the models tested, GRU exhibited superior performance, with

MSE equal to 0.00215, MAE of 0.03266 and R-squared equal to 90.228%. Li et al. in [97] introduce a model named TS2ARCformer. This model utilizes multi-dimensional, demonstrating significant forecasting performance, in comparison with to benchmark models, like LSTM and GRU The results show reductions of 43.2% and 37.8% in the aspect of MAPE.

Additionally, the model proposed in [98] utilizes neural network techniques enhanced with PSO to assess the Iranian power system. This approach, based on neural networks, showed lower prediction errors by effectively adapting to the underlying patterns of the consumption load. The evaluation was done using the RMSE metric, which was equal to 0.028, 0.045, 0.035, and 0.040, respectively, for multiple lengths of training input data. In studies [99, 100], empirical modal decomposition (EMD) was used in order to break down the original electricity consumption time series data into various intrinsic mode functions (IMFs) with different frequencies and amplitudes.

Also, in another study a hybrid prediction model is proposed. This model combines a temporal convolutional network (TCN) with variational mode decomposition (VMD). For better accuracy an error correction process is followed. The weighted permutation entropy process is followed, in order to determine the optimal decomposition number and penalty factor for VMD. The accuracy and performance of the proposed model are demonstrated using the Global Energy Competition 2014 dataset [101].

In the same way, Javed et al. in [102] research the short-term load forecasting problem, comparing eight state-of-the-art linear and non-linear models, which are ARX, KNN, ARMAX, Bagged Trees, SVM, OE, NN–PSO, and a single hidden dense layer ANN–LM and a two hidden dense layer ANN–LM, on real-time demand time series. Simultaneously, He et al. in [103], compare individual and decomposition-based forecasting methods. About the the individual models, these include LSTM, SVR, MLP, LR, and Random Forest Regressor. The decomposition-based methods include Empirical Mode Decomposition-LSTM (EMD-LSTM) and ensemble EEMD-LSTM models.

Additionally to the above, He et al. in [104] introduce a novel DNN architecture for short-term load forecasting. The model integrates various types parameters as input by employing appropriate NN elements designed to process each feature effectively. CNN components are used to extract features and patterns from the historical load time series, while RNN components are used to model the implicit dynamics. Additionally, Dense layers are employed to transform other types of features, improving the generalization of the models in understanding complex trends and patterns. Shi et al. in [105] suggest a model to tackle the unpredictability in electrical load patterns by applying DL and ML techniques to forecast household energy use in Ireland. This research introduces an innovative pooling-based deep RNN (PDRNN), which combines various load curves of consumers into a comprehensive set of inputs. The performance of the PDRNN is compared with ARIMA, SVR, and RNN models. The results demonstrate that the PDRNN method outperforms the other models, with RMSE values (in kWh) of 0.4505 for PDRNN, compared to 0.5593 for ARIMA, 0.518 for SVR, and 0.528 for RNN. Amarasinghe et al. in [106] suggested a framework to forecast short-term electricity consumption in residences by utilizing smart meter data from various properties. Following, Wand et al. in [107] propose an optimized SVM model using genetic algorithm (GA-SVM) in order to predict the hourly electricity demand of a commercial center located in Xin town, China.

Residential forecasting is an area that gathers a lot of scientific interest. Bessani et al. in [108] study the short-term residential forecasting task across multiple households employing Bayesian multivariate networks, which provide a structured, probabilistic approach to model complex dependencies between multiple variables, enabling efficient inference and decision-making even with incomplete data. This algorithm predicts the forthcoming household demand by analyzing historical consumption patterns, temperature, humidity, socioeconomic variables, and energy use. The assessment utilized actual data sourced from the Irish Intelligent Meter Project. According to the results, the proposed technique provides a robust single forecast model, which is capable of accommodating hundreds of households with various consumption patterns. The reported metrics include an MAE of 1.0085 kWh and a Mean Arctangent Absolute Percentage Error (MAAPE) of 0.5035, highlighting its effectiveness in accurate load forecasting.

Regarding RNN-based approaches, Caicedo-Vivas et al. in [109] analyze the issue of short-term load prediction by employing dataset from a Colombian grid operator. The proposed DL model utilizes a LSTM to predict load across a seven-day time horizon (168 hours ahead). It combines historical data from a specific region in Colombia with calendar attributes like holidays and the month's index relevant to the forecasting week. Unlike traditional LSTM approaches, this model enables LSTM cells to process multiple load measurements concurrently, with supplementary information (holidays and current month) appended to the LSTM output. This methodology improves the model's ability to capture and utilize temporal and contextual factors for more accurate load forecasts. Also, Kong et al. in [110] conduct a comparative study between the LSTM model and the extreme learning machine (ELM), back-propagation NN (BPNN), and KNN. They demonstrated significant reduction in prediction errors by employing the LSTM structure. The aggregated average MAPE for their forecasts was 8.18%, while the average MAPE for individual forecasts was 44.39%. This study underscores therich performance of LSTM models in short-term residential forecasting.

Following the above, Gao et al. in [99] investigate the impact of decomposition methods, which simplify complex data by breaking it down into underlying components, making it easier to analyze patterns, reduce dimensionality, and improve model interpretability. For this reason, the authors propose an EMD Gated Recurrent Unit with feature selection (EMD-GRU-FS) for short-term electricity forecasting, using Pearson correlation to identify the relationship between subseries and the original series. Experimental results indicated that this method achieved average prediction accuracies of 96.9%, 95.31%, 95.72%, and 97.17% across four datasets. Additionally, Yuan et al. in [100] present an enhanced model combining integrated EMD and LSTM for short-term load forecasting, where the recurrent model is used for feature extraction and creating load predictions. For the final consumers, short-term electricity consumption predictions were obtained by aggregating multiple target prediction results. The results show that the proposed model performs achieving MAPE 2.6249% in winter and 2.3047% in summer, respectively.

Some studies integrates Internet of Things (IoT) solutions in order to enhance ELF and improve localized forecasting, by gathering detailed, real-time data from devices like smart meters and local sensors. IoT-driven insights can optimize energy distribution, minimize supply mismatches, and support the integration of renewable sources, leading to a more reliable and efficient power grid. Besed on these, Imani et al. in [111] introduce a clustering forecasting approach in which divides the consumers' consumption based in two different features, collecting their data from an IoT connection from smart meters. As case study, an Irish dataset is used and the MLP model is proposed. Additionally, Hadri et al. in [91] developed load forecasting techniques that can be integrated with occupancy predictions and context-aware control of building appliances. The researchers investigated the accuracy of multiple methods for predicting energy consumption. They deployed an IoT and Big Data-based platform to collect real-time electricity values. The acquired data was used to build predictive models using XGBoost, RF, ARIMA, SARIMA and LSTM. The models' evaluation was done using metrics such as RMSE, MAE, MAPE, R and maximum absolute error (MaxAE). Finally, Raju et al. in [112] introduce an advanced and novel IoT method utilizing Machine Learning algorithms. LR, SVM, Gaussian Processes (GP) and Ensemble Methods are used and MAE, MSE and RMSE metrics are measured, in order to draw usefull conslusions for a smart home in which an Arduino Uno system has been installed.

Other works focuses on forecasting the energy usage of residential consumers. Accurate forecasting of residential consumers' load is critical because it enables efficient energy management, reduces operational costs, and improves grid stability. Based on these, in another study the authors conduct a comparative study between ARIMA and SARIMA, as well as RF, XGBoost, and LSTM models using MAPE, MAPE, Mean Percentage Error (MPE), and RMSE [91]. Li et al. in [113] developed a novel short-term forecasting model based on a modified LSTM model and an autoregressive, which is used for feature selection. The aim is to create forecasts for household electricity consumption.

Other studies explore how advanced techniques can be built into Deep Learning models to make their conclusions more accurate and reliable. Online learning is also essential in this context because it enables models to update continually as new data arrives. This constant updating allows the system to adjust to real-time changes in the market, adapting its responses to reflect current conditions. This method, known as online or continual learning, supports accurate, timely insights that are crucial for businesses operating in fast-changing, data-driven settings. Following these, Fekri et al. in [114] introduced an online adaptive RNN, which is capable of continuous learning from recently acquired data and adaptation to various patterns.

About CNN-based models, which have proven effective for timeseries forecasting by capturing local patterns and dependencies within sequential data, Andriopoulos et al. in [115] introduce an advanced CNN model that is concentrating on mathematical analysis of statistical data and advanced preprocessing of time series data for short-term load forecasting. This approach converts the data into a 2-D image format to leverage CNN-based models, which excel in handling stationarity and temporal volatility. The proposed model is compared with LSTM, considered the state-of-the-art solution for load forecasting. Following CNN models, Alanazi et al. in [116] present a novel approach, combining a residual CNN model and a layered Echo State Network (ESN), in order to handle efficiently both spatial and temporal patterns. Residential datasets from the Individual-Household-Electric-Power-Consumption (IHEPC) from Sceaux, Paris, and the Pennsylvania–New Jersey–Maryland (PJM) region are used. The developed model achieves MSE, MAE, and RMSE equal to 0.0713, 0.1978, and 0.2670 MW, respectively. Also, Jalali et al. in [117] study and propose a novel approach for short-term ELF using a deep neuroevolution algorithm, in order to develop a CNN model in automatic way, using an enhanced grey wolf optimizer (EGWO) as a modified advanced algorithm. Finally, Chitalia et al.in [118] present a strong forecasting framework in order to accommodate fluctuations in building operations across different types and locations of commercial buildings.

Another promising method used in load prediction studies are graph neural networks (GNNs), which capture dependencies within graph-structured data [119]. For instance, Vontzos et al. [120] used a GNN-RNN algorithm to forecast the consumption of a multistory educational building that exhibits in a graph the structure of the building zones with multiply methodologies. Additionally, Hu et al. [121] study incorparates urban building energy model that synthesizes the solar-based building interdependency and spatio-temporal graph convolutional network (ST-GCN) algorithm to predict hourly energy consumtpion of a university campus. Especially, we took a university campus located in the downtown area of Atlanta as an example to predict the hourly energy consumption. In terms of a region load forecasting Wei et al. [122] propose a method which incorporates limited dynamic time warping (LDTW) and Graph Convolutional Networks.

Finally, in many researches, weather data like temperature and humidity play a crucial role in ELF. In this way, Liao et al. in [123] propose the utilization of XGBoost model, using weekly historical data from a power plant in hourly resolution, incorporating weather factors such as temperature and humidity. The work additionally addresses the computational challenges, due to the complexity of the XGBoost hyperparameter optimization process.

5.2.3 Ensemble Methods

Ensemble models have gained interest because they improve predictive performance by combining multiple algorithms and minimize individual models' disadvantages. These models improve forecasting accuracy and generalization, making them effective for complex datasets, reducing overfitting and variance. Advances in computational power and accessible tools have made implementing ensemble methods easier, boosting their popularity.

Advanced combinations of different models is followed in Ensemble Learning methods. In this way, Guo et al. in [124] propose a novel approach to forecasting wind speed using a modified EMD-based Feed-Forward Neural Network (FNN). The advantage of the proposed model is based on the fact that is capable of addressing the nonlinear and non-stationary nature of wind speed data by decomposing the time series into multiple components and a residual series, using the EMD techniques. The data used are from Zhangye, China. Alobaidi et al. in [125] propose the implementation of an ensemble learning framework to tackle the issue of forecasting the daily mean demand on an individual household level for the day-ahead market. The proposed model incorporates a regression network, making the final model performing with very high accuracy on a dataset from France. Du et al. in [126] introduce an advanced hybrid algorithm that effectively retrieves accurate wind speed series using singular spectrum analysis. This algorithm demonstrates superior adaptive search and optimization capabilities compared to other methods, offering faster performance, fewer parameters, and lower computational cost. As case study, this work applies the model to wind speed data in 10 minutes resolution, from three wind farms located in Shandong Province, eastern China.

Simultaneously and following fuzzy logic, Ma et al. in [127] combine a denoising technique with a dynamic fuzzy neural network to meet the challenges in wind speed forecasting tasks. The raw data is smoothed using singular spectrum analysis, optimized by brain storm optimization, to produce a cleaner dataset. A generalized dynamic fuzzy neural network is then used for forecasting. The model's simpler and more compact neural network structure allows for faster learning and accurate predictions. Experimental results on wind speed data at 10-minute, 30-minute, and 60-minute intervals show that the model effectively approximates actual values and can serve as a reliable tool for smart grid planning. Also, Sadaei et al. in [128] combine fuzzy time series (FTS) and CNN models for short-term ELF. Multivariate consumption data, temperature and fuzzified version of load time series in hourly resolution, was converted into multi-channel images, in order to be used as input in the proposed model.

At the same way, Rafi et al. in [129] combine CNN and LSTM networks. As case study, the proposed method is applied to data from Bangladesh electricity network, in order to provide short-term forecasting. Finally, Wang et al. in [130] incorporate Q-learning with reinforcement learning. The aim for the final ensemble model is the creation of optimal ensemble weights, achieving predictions with the highest accuracy. As case study, various datasets of 25 households in the Austin area of the United States are used. The evaluation metrics show that the ensemble model surpass three single models in the testing period.

In summary, we observe that when combining predictions from strong models in ensemble algorithms, the final performance improves even further. Additionally, a significant advantage of Ensemble Methods over single models is that the potential poor performance of any individual model is minimized in the final outcome, leading to more reliable and stable results.

5.2.4 Aggregated Studies for Electricity Load Forecasting

In order to present the most of the studies analyzed above in a consolidated manner, Table 5.1 is provided.

Table 5.1: Aggregated table with resent publications for load forecasting.

Ref.	Year	Models	Metrics	Dataset Origin
[89]	2020	ARIMA, MLR, RPART, CTREE, BAGGING, RF	MAPE, R2	Electricity load data of a residential area
[90]	2018	ANN, SVM, RL, CR Tree, Voting, Bagging, SARIMA-MetaFA-LSSVR, SARIMA-PDO-LSSVR	R2, RMSE, MAPE, MaxAE	Real-time data collected from a smart grid installed in an experimental building
[91]	2019	ARIMA, SARIMA, XG-Boost, Random Forest (RF), Long Short-Term Memory (LSTM)	sMAPE, MAPE, MAE, RMSE	N/A
[92]	2018	Online SVR, MLP, ARIMA	MAPE	Public Irish CER dataset
[93]	2023	ARIMA, ETS, Prophet	MAPE	ENTSOE

Continued on next page...

Table 5.1 – continued from previous page.

Ref.	Year	Models	Metrics	Dataset Origin
[94]	2018	SARIMAX	RMSE, MAE, MAPE	Tokyo Electric Power Company (TEPCO)
[131]	2020	ARIMA, ARMA, SARIMAX	MSE, MAPE, MAE	Various datasets
[132]	2016	ANN. LR, MLR, Fuzzy Regression, ARIMA, ARMA, GARCH	Various Metrics	US utility
[95]	2022	DSARIMA	MAPE	Java-Madura-Bali interconnection, Indonesia
[96]	2023	LSTM, GRU, RNN	R-squared, MSE, MAE	Tubas Electricity Company, Palestine
[133]	2017	DNN, PSO	MAE, MSE	North Cyprus based energy company, Cyprus Turkish Electricity Authority (Kib-Tek)
[99]	2019	GRU, SVR, RF, EMD-GRU, EMD-SVR, EMD-RF	MAPE, RMSE	U.S. Public Utilities Data Set, Region in Northwest China
[100]	2022	RNN, EMD, LSTM, EMD-LSTM	MAPE, RMSE	Region in Guangdong Province, China
[103]	2019	LSTM, SVR, MLPR, LR, RFR, EMD-LSTM, EEMD-LSTM	RMSE, MAE, MAPE (%), R2	Hubei Province, China

Continued on next page...

Table 5.1 – continued from previous page.

Ref.	Year	Models	Metrics	Dataset Origin
[101]	2022	TCN, VMD, LSTM, GRU, VMD-LSTM, VMD-GRU, VMD-TCN	RMSE, MAPE, R2	Global Energy Competition 2014 dataset
[104]	2017	LR, SVR, DNN, CNN	MAPE, MAE	Hourly loads of a North China city
[105]	2018	ARIMA, SVR, RNN, DRNN, PDRNN	RMSE, MAE, NRMSE	920 smart metered customers from Ireland
[106]	2017	CNN, LSTM, SVM, ANN	RMSE	Benchmark dataset of electricity consumption for a single residential customer
[108]	2020	Bayesian networks	NRMSE, MAE, MedAE, MAAPE	Real data from the Irish smart meter project
[110]	2019	LSTM, BPNN, KNN, ELM	MAPE	Real residential smart meter data
[109]	2023	LSTM, XG-BOOST	MAPE	Historical load values from a region in Colombia
[102]	2021	OE, ARX, ARMAX, SVM, ANN–PSO, KNN, ANN–LM	RMSE, MAPE, MAE	Real-time electrical load and meteorological data of the city of Lahore in Pakistan
[115]	2021	CNN, LSTM, MLP, ANN	MAE, RMSE	Three open-source, publicly available datasets

Continued on next page...

Table 5.1 – continued from previous page.

Ref.	Year	Models	Metrics	Dataset Origin
[123]	2019	XGBOOST, LSTM	MSE, RMSE, MAPE	State Grid Corporation
[113]	2021	Persistence, SW-ARIMA, SW-ETS, SW-SVR, ConvLSTM, CLSAF	Accuracy	Manhattan, New York, residential building
[114]	2021	MLP, KNN, RNN	MSE, MAE	Five residential consumers provided by London Hydro
[117]	2021	CNN, EGWO, LSTM, SVR, MLP	RMSE, MAE, MAPE	Australian Energy Market Operator
[118]	2020	LSTM, BiLSTM	RMSE, MAPE, CV	Bangkok, Thailand; Hyderabad, India; Virginia, USA; New York, USA; and Massachusetts, USA
[97]	2023	LSTM, GRU, TS2ARCformer, TCN	MAPE, MSE, MAE	National College Student Mathematical Contest in Electrical Engineering
[124]	2012	EMD-FNN	MSE, MAE, MAPE	Zhangye, China
[125]	2018	IANN, Bagging Ensemble Model	RMSE, MAPE, rRMSE	Individual household energy consumption
[126]	2016	XNN, GRNN, BPNN, PSO-BP, APSOSA-BP	AE, MAE, MAPE, RMSE	Wind farms in Shandong Province, eastern China

Continued on next page...

Table 5.1 – continued from previous page.

Ref.	Year	Models	Metrics	Dataset Origin
[127]	2017	FNN, SSA-GDFNN, BSO-SSA-GDFNN	FE, MAPE, MAE, RMSE	Historical short-term wind speed data collected from Shandong Peninsula of China
[128]	2019	A hybrid model of CNN and FTS	MAPE	N/A
[129]	2021	CNN, LSTM, CNN-LSTM, XGBOOST, RBFN	MAE, RMSE, MAPE	Bangladesh power system
[130]	2023	HI-TCN, HI-ELM, HI-GRU, HI-BPNN	MAE, MAPE, RMSE	Pecan Street datasets contain the load power data of 25 households in the Austin area of the United States
[116]	2023	CNN, CNN-LSTM, CNN-GRU, ESN	MSE, MAE, RMSE, MAPE, NRMSE	Individual-Household-Electric-Power-Consumption (IHEPC) dataset from residential houses in Sceaux, Paris, and Pennsylvania–New Jersey–Maryland (PJM)
[107]	2023	SVM, GA-SVM	MAPE, CV-RMSE, R2	Data of a large public building in Xi town, China
[111]	2018	MLP	MAPE	Irish social science data archive (ISSDA)

Continued on next page...

Table 5.1 – continued from previous page.

Ref.	Year	Models	Metrics	Dataset Origin
[112]	2020	LR, SVM, Fine Tree (FT) regression, Ensemble Bagged Regressor, Gaussian Process Regressor, Ensemble Boosted Regressor	RMSE, MSE, MAE	Domestic data collected from IoT system
[120]	2024	GCN-LSTM, MLP, CNN, LSTM	MAE, MSE, CV-RMSE, R^2	Academic building in Bangkok, Thailand
[121]	2022	Average, LR, MLP, XGBoost, GRU, ST-GCN	RMSE, MAPE	University campus in Atlanta, Georgia
[122]	2024	SVR, LSTM, LSTM-Att, TCN, Informer, GCN-LSTM	MAPE, RMSE, R^2	Substations data in area of Shanghai

5.3 Consumer Categories

Energy consumption patterns vary significantly across different sectors, each reflecting distinct operational characteristics and usage behaviors. In this section, we present a detailed discussion of the load profiles observed in the commercial, industrial, and residential sectors. We also emphasize the particular challenges encountered in forecasting residential loads due to their inherent instability and unpredictability [134], [135].

5.3.1 Commercial

Commercial consumers stem from establishments such as offices, retail centers, hotels, and other service-related facilities. The energy demand in this sector is largely dictated by business hours and customer activity. Typically, commercial buildings exhibit a pronounced increase in consumption during the daytime when the facilities are in full operation. For instance, office buildings and shopping malls experience well-defined peaks corresponding to work and shopping hours, followed by gradual declines as business operations wind down. Although external factors like seasonal variations and special events may introduce minor fluctuations, the overall demand profile remains relatively consistent. This regularity facilitates the development of effective forecasting models and allows energy providers to optimize grid management and operational planning [134].

5.3.2 Industrial

Industrial energy consumption is generally characterized by its scale and stability. Factories, manufacturing plants, and other industrial facilities often operate on fixed production schedules or continuous shifts, which result in a steady and predictable load profile. While some variability can occur due to production cycles, maintenance periods, or seasonal changes, the baseline demand remains largely constant [135]. The predictability of industrial loads enables the implementation of rigorous energy management strategies and supports the integration of efficient forecasting techniques. In many cases, these facilities also employ advanced energy management systems that further contribute to maintaining a stable consumption pattern, thus reducing the complexity involved in grid operation [134].

5.3.3 Residential

Residential consumers arise from individual households and encompass a wide range of activities, including heating, cooling, appliance usage, and lighting. Unlike the structured operations seen in commercial and industrial sectors, residential consumption is influenced by a variety of factors such as individual behavioral patterns, weather conditions, and lifestyle differences [135]. Typically, residential load profiles exhibit multiple daily peaks—often during the morning and evening hours—corresponding to routine activities like cooking, leisure, and personal care. However, these peaks are neither uniform nor entirely predictable, as they can vary significantly from day to day or across different neighborhoods [134].

5.4 Challenges in Residential Load Forecasting

The forecasting of residential loads is particularly challenging due to the high degree of variability inherent in individual consumption patterns. The decentralized nature of household energy use means that aggregated residential demand can be highly erratic, with sudden spikes or drops that are difficult to anticipate. Additionally, the increasing penetration of distributed energy resources, such as rooftop solar panels and home energy storage systems, further complicates the load profile by introducing additional, often intermittent, sources of power. These factors combine to create an environment where conventional forecasting methods may fall short, necessitating the development of advanced models that can effectively capture the stochastic behavior of residential loads.

So, while commercial and industrial loads benefit from structured operational patterns that lead to relatively stable and predictable demand profiles, residential loads present a unique challenge. Their inherent variability and the influence of numerous external factors make accurate forecasting a complex task. Addressing these challenges is critical for enhancing grid reliability and ensuring that energy management strategies are robust enough to handle the dynamic nature of modern residential consumption [136].

5.5 Enhanced Automated Deep Learning Application for Short Term Load Forecasting

5.5.1 Description of the problem

In recent times, the power sector has become a focal point of extensive scientific interest, driven by a convergence of factors such as mounting global concerns surrounding climate change, the persistent increase in electricity prices within the wholesale energy market, and the surge in investments catalyzed by technological advancements across diverse sectors. These evolving challenges have necessitated the emergence of new imperatives aimed at effectively managing energy resources, ensuring grid stability, bolstering reliability, and making informed decisions. One area that has garnered particular attention is the accurate prediction of end-user electricity load, which has emerged as a critical facet in the pursuit of efficient energy management. To tackle this challenge, machine and deep learning models have emerged as popular and promising approaches, owing to their remarkable effectiveness in handling complex time series data. In this research work, the development of an algorithmic model, which leverages an automated process to provide highly accurate predictions of electricity load specifically tailored for the island of Thira in Greece, is introduced. Through the implementation of an automated application, an array of deep learning forecasting models has been meticulously crafted, encompassing the Multilayer Perceptron, Long Short-Term Memory (LSTM), One Dimensional Convolutional Neural Network (CNN-1D), hybrid CNN-LSTM, Temporal Convolutional Network (TCN), and an innovative hybrid model called the Convolutional LSTM Encoder-Decoder. Through evaluation of prediction accuracy, it is observed satisfactory performance across all the models considered, with the proposed hybrid model showcasing the highest level of accuracy. These findings underscore the profound significance of employing deep learning techniques for precisely forecasting electricity demand, thereby offering valuable insights to tackle the multifaceted challenges encountered within the power sector. By adopting advanced forecasting methodologies, the electricity sector is moving towards greater efficiency, resilience and sustainability.

5.5.2 Main Findings

The Deep Learning topology we propose in this work for STLF is a hybrid model named Convolutional LSTM Encoder-Decoder which was used by other researchers to forecast global total electron content [137] and to predict the El Niño-related Oceanic Niño Index (ONI) and El Niño events [138]. Generally, our research contributions can be summarized as follows:

1. Proposal of a hybrid model called Convolutional LSTM Encoder-Decoder for power production time series.

2. Presentation of an automated STLF algorithm which incorporates data pre-processing techniques, training, and optimization several AI models, testing and prediction of the results.

The remaining part of the work is organized as follows: Section 5.6 is the Case Study of the work, and analyzes the dataset, the essential procedures related to the pre-processing of the data used for forecasting, the architecture of the artificial intelligence models, and the automated STLF algorithm. In Section 5.7, the experiments of the case study are illustrated and the results of experiments are discussed. Finally, Section 5.8 contains the conslusions, the summary of this work and suggests some directions for future work.

5.5.3 Proposed Approach: Convolutional LSTM Encoder-Decoder Model

For the proposed model, at the beginning, a description of Convolutional LSTM Network (ConvLSTM) was made and then the basic operation of Encoder-Decoder Mechanism is analysed.

ConvLSTM model uses Recurrent Layers, such as simple RNN and LSTM Networks but, in contrast with these types of models, the internal matrix multiplications are exchanged with convolution operations, as this model has convolutional structures in both the input-to-state and state-to-state transitions. As a result, the data that flows through the ConvLSTM cells keeps the input dimension (3D in our case) instead of being just a 1D vector with features [139]. The basic architecture of ConvLSTM model is presented in Figure 5.1.

Additionally, the proposed model is a hybrid method that uses an Encoder-Decoder mechanism for the hourly prediction of electricity demand. The Encoder-Decoder mechanism provides efficient techniques for Sequence-to-Sequence (Seq2Seq) forecasting, such as Natural Language Processing (NLP) and time series forecasting, as well as Image Recognition and Sentiment Analysis. An Encoder-Decoder model (Figure 5.2) consists of three main parts [140]:

- *The Encoder Component*: It is the first component of the Network and its main function the feature extraction of the input dataset and this is the reason that is called ‘encoder’. It receives an input the sequence and it passes the information values into the internal state vectors or Encoder Vectors, creating in this way a hidden state.

- *The Encoder Vector*: The encoder vector is the last hidden state of the Network which converts the 2D output of the RNN (ConvLSTM in our case) model to a high length 3D vector, in order to help the Decoder Component to make better predictions.

- *The Decoder Component*: The Decoder Component consists of one or a stack of several recurrent units, each one of them predicting an output y at a timestep t.

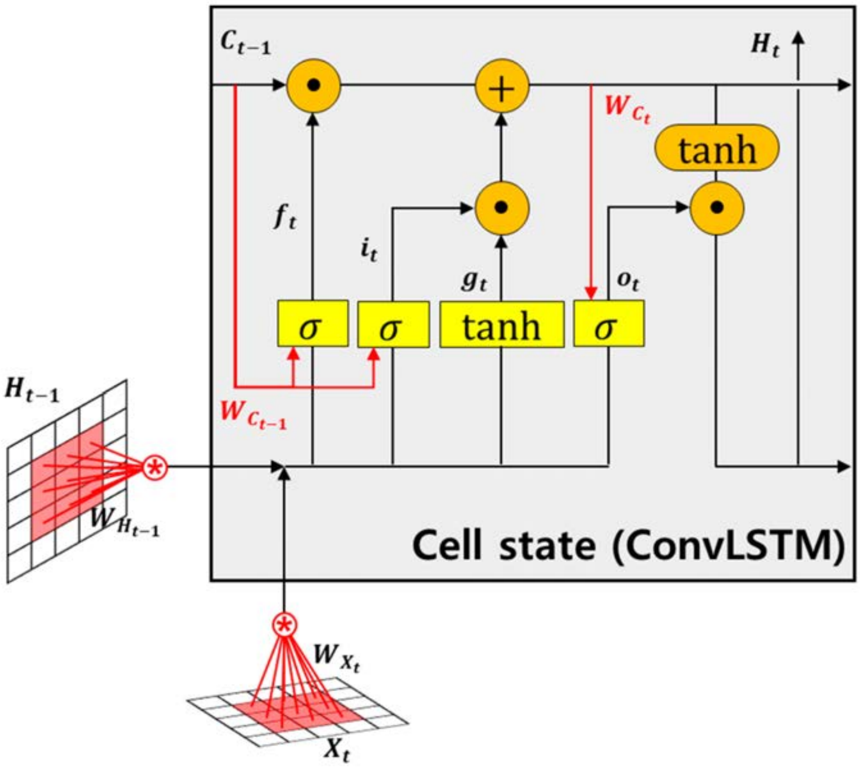


Figure 5.1: Basic architecture of ConvLSTM model

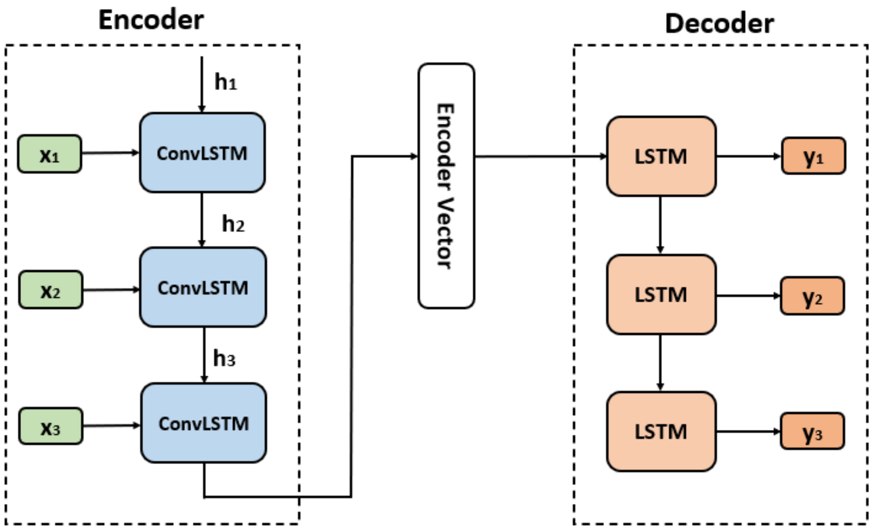


Figure 5.2: Encoder Decoder Mechanism

One of the main advantages that makes this model stands out from others is that the input and output vector length may be different, allowing this model to have effective performance in big Seq2Seq problems and video captioning.

5.5.4 Automated STLF Algorithm

Using Deep Learning techniques, a DL-based forecasting algorithm was implemented, in order to easily and efficiently find the best forecasting model for every kind of time series datasets. The function of the algorithm includes the following steps:

1. The algorithm loads the dataset.

2. The algorithm proceeds to Preprocessing Step, in which the dataset is normalized using Min-Max Scaling algorithm and the Cyclical Time Features are created using One-Hot Encoding.

3. The dataset is splitted into Training, Validation and Testing sets.

4. Every Deep Learning Model is trained and evaluated using Bayesian Optimization Algorithm.

5. Selection of the best model in terms of higher prediction accuracy to be used from the algorithm for future predictions.

The Automated STLF Algorithm is described also in Figure 5.3.

5.6 Case Study

In this section, we present the case study used to test the proposed approach. The final dataset used for our experiments incorporates weather conditions and power demand from the Greek Cycladic Island, in Aegean Sea, Thira also known as Santorini. Thira island is a non-interconnected island to the Greek electrical bulk network and the power production is solely based on fossil fuel generators. The dataset period is three years from January 2017 to December 2019. The specific time period was chosen because the power requirements followed a normal distribution and there were no special events (such as COVID-19) that would not lead to safe conclusions. Working with this dataset, no such limitations were observed. Regarding the challenges encountered, these came mainly from the parameterization and adaptation of the algorithms created and their ability to assimilate and train on a dataset with multiple seasonality, such as the one studied.

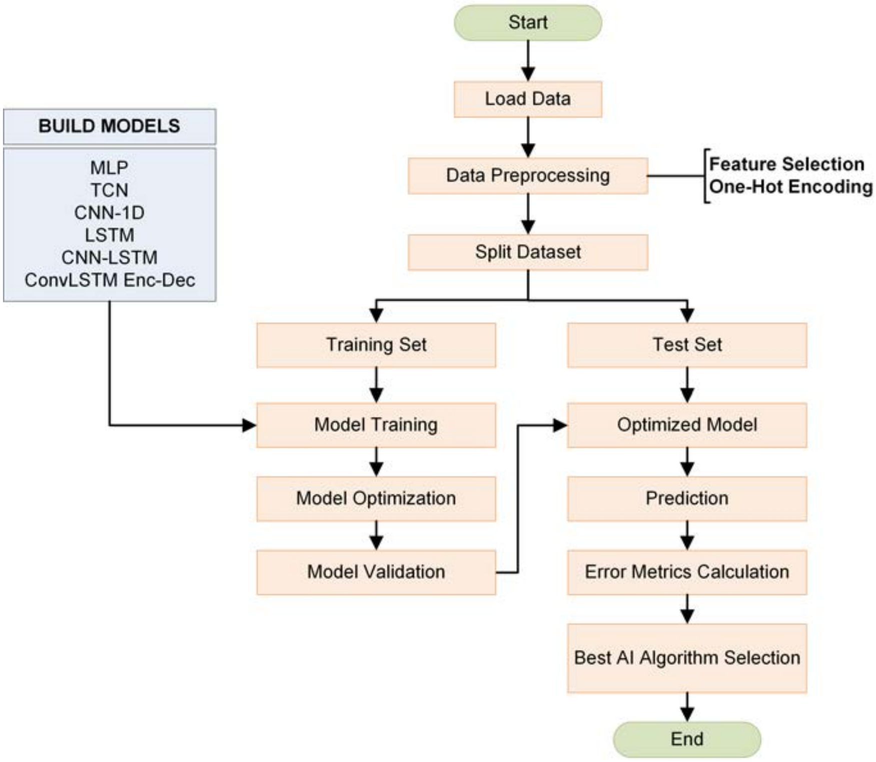


Figure 5.3: Automated STLF Algorithm

5.6.1 Climate Dataset

The climate data was collected from Santorini's airport Meteorological Aerodrome Reports (METAR) [141, 142]. The airport's International Civil Aviation Organization (ICAO) code is LGSR. A typical report usually is published from airport's weather station every hour or half-hour and aggregates actual values and current conditions for the temperature, dew point, wind direction and speed, precipitation, cloud cover and heights, visibility, and barometric pressure. For the current work we retrieve, from METARs, hourly values for temperature in ^{o}C, dewpoint in ^{o}C, wind speed in m/s and degrees of wind direction. Interpolation was applied to infill climate data missing values.

5.6.2 Electrical Power Dataset and Exploratory Analysis

Power production dataset contains the actual hourly generated energy from island's unique thermal station and is referred to entire island power demand [143]. Thira exhibits multiple seasonality throughout year, due to change of numbers of residents and tourists on the island which affects dramatically the energy demand between different months. Additionally current energy dataset presents nil number of missing values, fact that leads to realistic experiments and safe conclusions.

Table 5.2 presents a descriptive analysis of the power consumption data set. Based on the above table, it is observed that the dataset contains a total of 25944 values. Also, its mean value and standard deviation are 23.31 MW and 9.95 MW, respectively. The minimum value is 8.50 MW and the maximum 51.52 MW. Finally, regarding the intermediate values, it turns out that 25% of the dataset values are smaller than 14.70 MW, 50% smaller than 21.15 MW and 75% smaller than 31 MW, respectively.

Table 5.2: Power consumption descriptive analysis

Metric	**Value**
Number of samples	25944
Mean value	23.31 MW
Standard deviation	9.95 MW
Minimum value	8.50 MW
Percentile 25%	14.70 MW
Percentile 50%	21.15 MW
Percentile 75%	31.00 MW
Maximum value	51.52 MW

In order to have a better observation of the daily distribution of power consumption, a boxplot was created. Figure 5.4 presents the variation of average hourly power per day. It is worth noting that the peak demand is at 01:00 pm and 08:00 pm every day. The event is more intense on summer days, due to the increased number of tourists on the island.

Correlation Heatmap

Figure 5.5 presents the Correlation Heatmap which visualizes the strength of relationships between all the numerical variables that have been created. For instance, it is clear that "temperature" variable has a very strong relationship with "power" variable, which is equal to 0.81 with maximum possible value of 1.

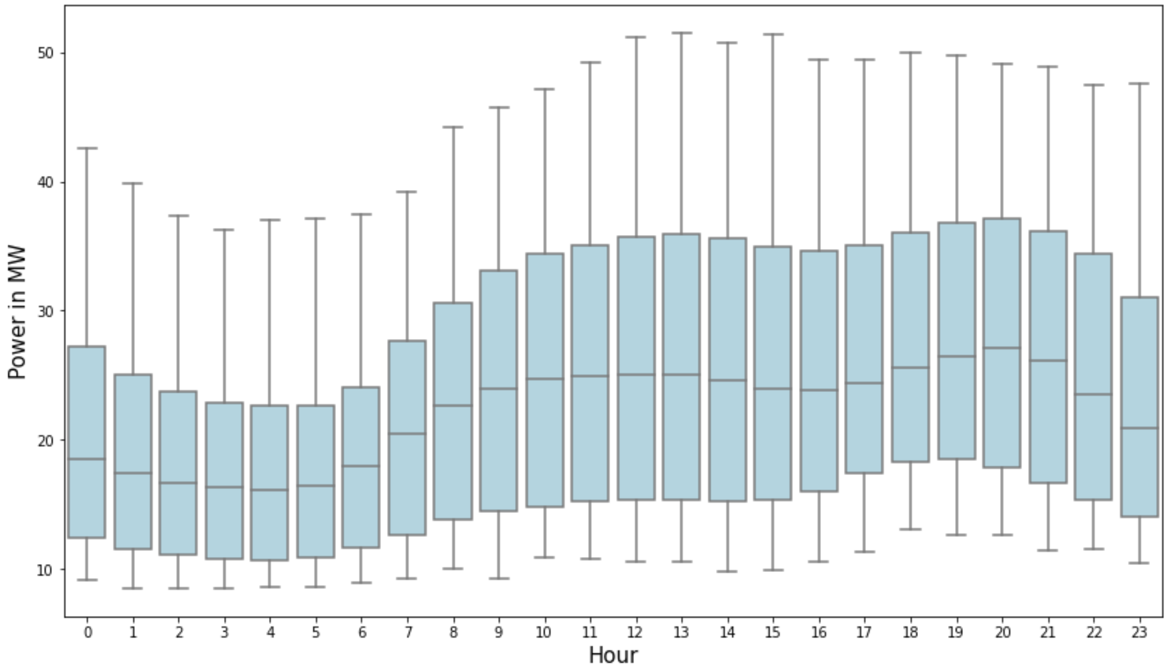


Figure 5.4: Boxplot of the variation of daily average hourly consumption for the reviewed time period

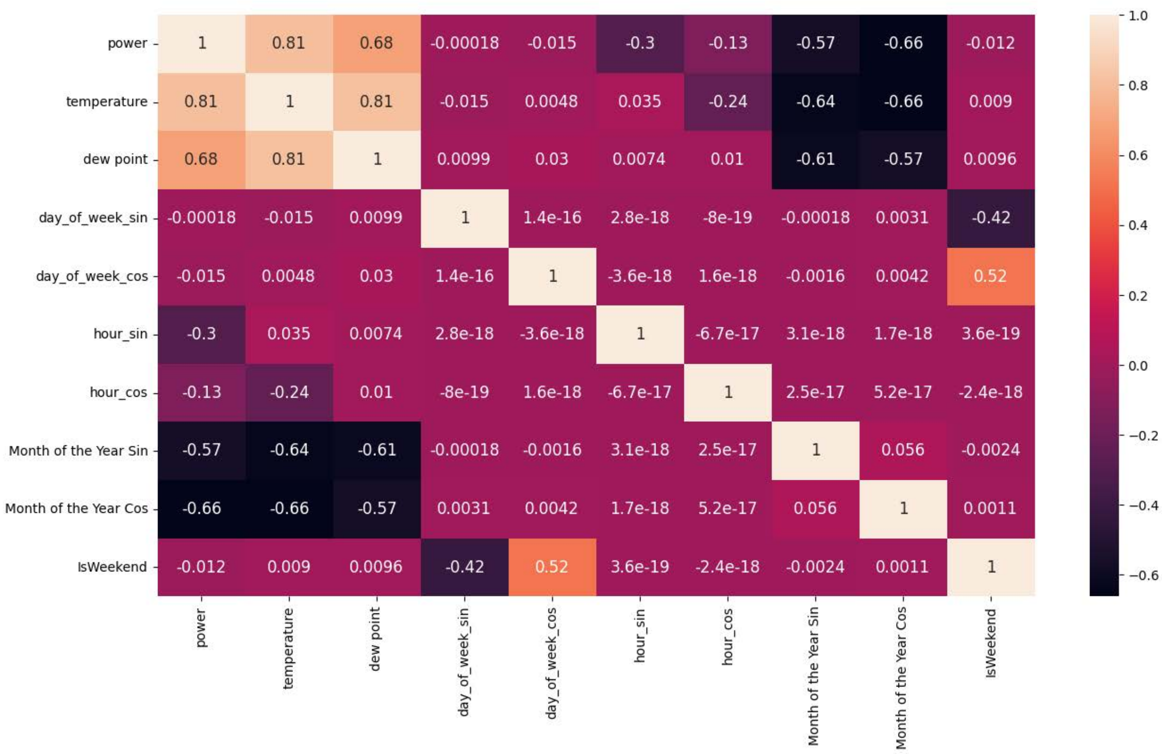


Figure 5.5: Feature Correlation Heatmap

5.6.3 Feature Selection

Several input features were studied and evaluated for this study to understand the most significant for predicting electricity demand. A total of ten input features were studied for predicting electricity demand one hour ahead. The features used as input variables are the same for all the Deep Learning Models and are shown in Table 5.3:

Table 5.3: Deep learning models input variables

Input Variable	Description
Power	The sequence of 24 hours of load values for 1 day.
Temperature	The sequence of 24 hours of temperature values for 1 day.
Dew Point	The sequence of 24 hours of dew point values for 1 day.
Cos of Day of Week / Sin of Day of Week	The Day of the Week converted by One-Hot Encoding to sine and cosine type.
Cos of Hour of Day / Sin of Hour of Day	The Hour of Day (1-24), converted by One-Hot Encoding to sine and cosine type.
Cos of Month of Year / Sin of Month of Year	The Month of the Year converted by One-Hot Encoding to sine and cosine type.
IsWeekend	A dummy variable “Is Weekend”, which takes 0 for working days and 1 for weekends.

5.6.4 Data Prepossessing

Min-Max Scaling

The prepossessing technique for all the dataset that used in this work is Min-Max Scaling, which scales all the data points between 0 - 1. For this reason, two different scalers were used, one for input and one for output datasets. The mathematical formulation is given by the follow equation:

$$X_{scaled} = \frac{X_{actual} - X_{min}}{X_{max} - X_{min}} \tag{5.1}$$

Where X_{actual} is the real value of a data, and X_{min} and X_{max} is the minimum and maximum value of the dataset, accordingly. The main reason Min-Max Scaling is used is that it helps the deep learning models to be trained more efficiently at the training period and converge faster to the optimal solution of the loss function.

One-Hot Encoding

One-Hot encoding is a mathematical methodology that converts categorical to numerical vectors and transforms the numerical data to cyclical, using trigonometric transformation. Using this technique, day of the week, hour of the day and also month of the year were converted to Sin and Cosine type.

5.6.5 Data Partitioning

The data was split into a training and test set while maintaining the temporal order of the data. The data from 1 January 2017 to 30 September 2019, were used as the training set and the last 10% of this were used for the validation of the models. The data from 1 October 2019 to 31 December 2019, were used as the test set. The specific interval is not random but was chosen because throughout its range it has both an interval with a trend and an interval which is stationary. Thus, the deep learning models are tested in the most difficult case that could be extracted from our dataset. Figure 5.6 presents the visualization plot of the power consumption for all the datasets and the individual partitioning sections.

5.6.6 Model Architectures

This subsection details the architectures of each model used in this work, in order to have a clear understanding of the optimal parameters that were extracted. The hyperparameters of all models were chosen after applying Bayesian Hyperparameter Optimization Algorithm with maximum 20 iterations in every search. Initially, several trials were performed with different values of batch sizes. Due to faster converge and higher accuracy of the DL algorithms, batch size equal to 128 was chosen for every model created in this work. More specifically, the main architecture components and the search space of every hyperparameter that were optimized for each model are presented in Table 5.4. The optimal extracted parameters are presented below:

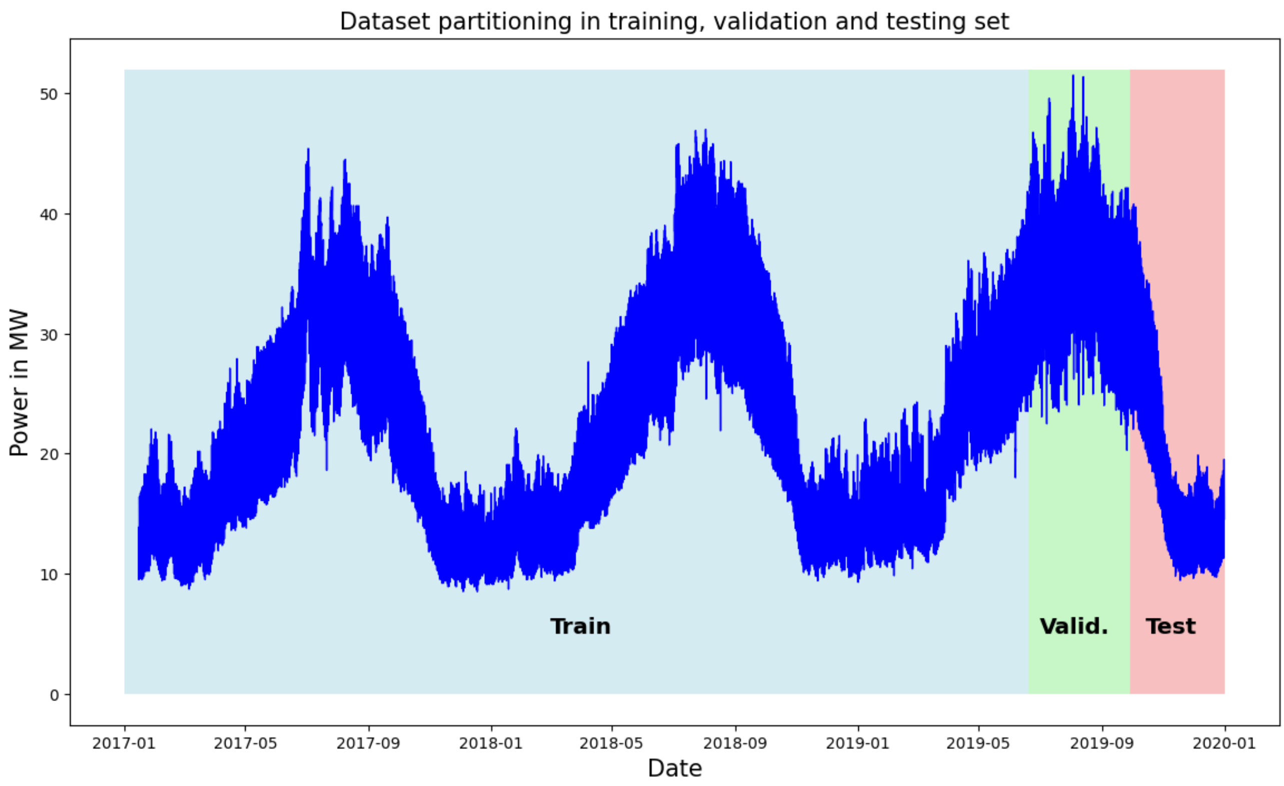


Figure 5.6: Dataset partitioning in training, validation and testing set

LSTM Model

The LSTM model that performs the best accuracy in the optimization process has the following hyperparameters:

- Units of lstm network = 256
- Batch size = 128
- ReLU activation function for the LSTM Module, as well as the output Dense Layer
- Optimizer = Adam
- Learning rate = 0.0010
- Epochs = 100

CNN-LSTM Model

The hyperparameters of the hybrid CNN-LSTM model have been resulted are presented below:

- filters = 64
- kernel size = 2
- ReLU activation function for the both the CNN and LSTM Module
- 48 units for the LSTM Module
- MaxPooling1D with pool size=2
- Batch size = 128
- Optimizer = Adam
- Learning rate = 0.0010
- Epochs = 100

Multilayer Perceptron Model

After converging to the optimal forecasting accuracy, the hyperparameters for the MLP model are formulated as follow:

- 480 Neurons for the Input Layer
- Batch size = 128
- ReLU activation function for the Input Layer, as well as the output Dense Layer
- Optimizer = Adam
- Learning rate = 0.0011
- Epochs = 100

Temporal Convolution Network

- Filters = 256
- Dilations = [1, 2, 4, 8, 16, 32]
- Batch size = 128

- Optimizer = Adam
- Learning rate = 0.0010
- Epochs = 100

CNN–1D Model

- Filters = 64
- Kernel size = 3
- MaxPooling1D (pool size=2)
- Neurons of Dense Layer = 16
- ReLU activation function for the CNN Module, as well as the output Dense Layer
- Optimizer = Adam
- Learning rate = 0.0010
- Epochs = 100

ConvLSTM Encoder- Decoder Model

This is the hybrid proposed model of this research that outperforms the other deep learning models. The hyperparameters that compose the architecture of this model are the followings:

- Filters = 64
- Kernel size = (1,4)
- ReLU activation function for the ConvLSTM2D Module, for the LSTM Layer, as well as the output Dense Layer
- Encoder Vector with size = 1
- 64 Units for the LSTM module
- Optimizer = Adam
- Learning rate = 0.0068

- Epochs = 100

A visualization of each model is presented in Figure 5.7.

5.6.7 Software Environment

The experiments presented in this study were implemented in Python Language, using the Open-Source software library Tensorflow 2 and the high-level API, Keras. Pandas and Numpy libraries were used for data analysis and visualization of the problem. The project was executed on Google Colab Pro platform, using a GPU with the following characteristics: NVIDIA- SMI 460.32.03, Driver Version: 460.32.03, CUDA Version: 11.2, RAM: 25.45 GB and Disk: 166.77 GB.

Regarding the time required to train each model, one of the primary factors influencing this process is the architecture and the level of complexity of each model, and more specifically the number of trainable parameters which are required to train each algorithm. Among the models examined, the MLP (Multi-Layer Perceptron) stands out as the simplest and fastest to train, followed by CNN-1D, CNN-LSTM, LSTM, ConvLSTM Encoder Decoder, and TCN. Despite their architectural differences, the training times do not exhibit significant variations. Once the optimization procedure is completed and the optimal hyperparameters for each algorithm are determined, the training of each model takes no more than 20 minutes. Finally, regarding the time required for each model to make a complete prediction, i.e., the inference time, it does not exceed 1 minute. So, this remarkably short training time enables real-time predictions using these models, extending their applicability beyond the academic realm pursued in this work.

5.6.8 Performance Metrics (Evaluation Metrics)

In this section, we present the performance metrics used in the evaluation of the hourly power prediction models. Mean Absolute Error (MAE) [144] is a commonly used metric with regression models, comparing actual and predicted values, which gives a measure of average error. MAE is calculated by the following formula where y_i is the predicted value and x_i is the actual value in a set of n samples:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{5.2}$$

Moreover, Mean Absolute Percentage Error (MAPE) [145] is used as a measure of quality for regression and time series models because it is explains intuitively the relative error. MAPE is computed by the following formula which uses the same parameters as the MAE calculation:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}|\frac{y_i - x_i}{x_i}| \quad (5.3)$$

Table 5.4: Main optimization parameters of each model

Model Type	**Model Hyperparameters**
LSTM	• LSTM module with search space of the units: min_value=16, max_value=256, step=16 • Optimizer = Adam • Learning rate with search space: min_value=0.0010, max_value=0.010, sampling=”log”
MLP	• 1 Dense layer with search space of Neurons: min_value=16, max_value=512, step=16 • Dropout layer with search space: min_value=0, max_value=0.5, step=0.05 • Optimizer = Adam • Learning rate with search space: min_value=0.0010, max_value=0.010, sampling=”log”

Model Type	Model Hyperparameters
CNN-1D	• Filters of CNN with range: min_value=64, max_value=128, step=16 • CNN kernel size with search space: min_value=4, max_value=8, step=2 • Neurons of Dense Layer with search space: min_value=24, max_value=120, step=12 • Optimizer = Adam • Learning rate with search space: min_value=0.0010, max_value=0.010, sampling="log"
CNN-LSTM	• Filters of CNN module with range: min_value=64, max_value=128, step=16 • Kernel size with search space: min_value=1, max_value=3, step=1 • LSTM units with search space: min_value=16, max_value=256, step=16 • Optimizer = Adam • Learning rate with search space: min_value=0.0010, max_value=0.010, sampling="log"

Model Type	Model Hyperparameters
TCN	• Filters of TCN with search space: min_value=128, max_value=256, step=16 • Dilations = [1, 2, 4, 8, 16, 32] • Optimizer = Adam • Learning rate with search space: min_value=0.0010, max_value=0.010, sampling="log"
Conv LSTM Encoder-Decoder	• Filters of ConvLSTM with range: min_value=64, max_value=128, step=16 • Kernel size of ConvLSTM with range: min_value=2, max_value=4, step=1 • Repeated Vector with range: min_value=1, max_value=3, step=1 • LSTM module with search space of the units: min_value=16, max_value=128, step=16 • Optimizer = Adam • Learning rate with search space: min_value=0.0010, max_value=0.010, sampling="log"

Furthermore, Mean Squared Error (MSE) [144] and Root Mean Squared Error (RMSE) [146] error metrics are included in the performance evaluation of this study. The MSE metric measures the average squared difference between the predicted and the true values. In

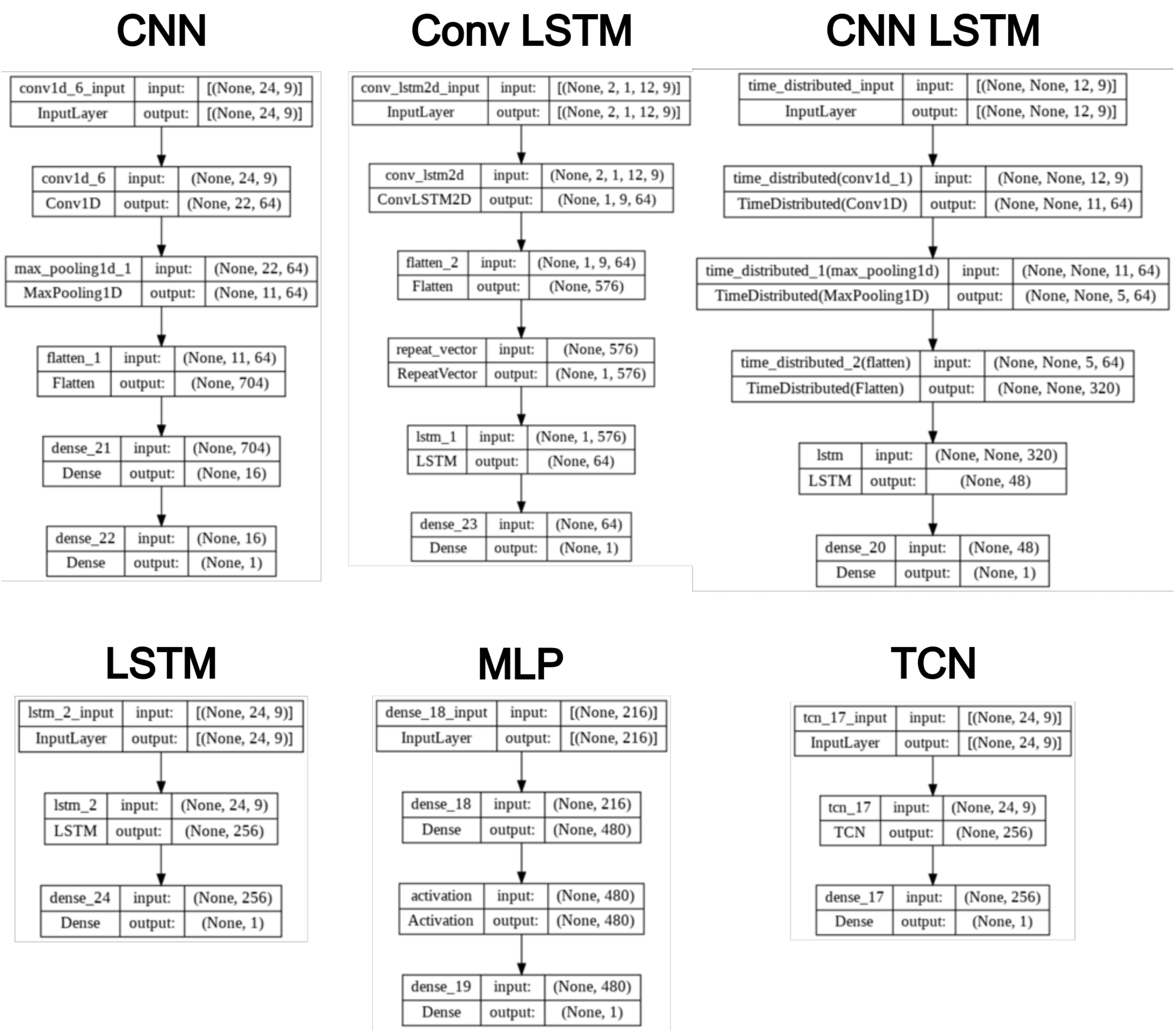


Figure 5.7: Models visualization

this study, we used MSE as the loss function for the training of the neural network prediction models due to the characteristic that gives a higher weight on extreme error values. Additionally, RMSE metric is the square root of the MSE, use the same parameters as the previously mention evaluation metrics and the computation formulas are the following:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 \tag{5.4}$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(y_i - x_i)^2}{n}} \tag{5.5}$$

Additionally, the coefficient of determination or R squared is used as a performance metric on this study, which represents the proportion of variance (of y) that has been explained by the independent variables in the model. It indicates a proper model fit and measures how unobserved samples are possible to be forecasted by the model, through the proportion of explained variance. R squared can be more informative than the metrics mentioned earlier in regression analysis evaluation [146]. R squared is calculated by the following formula where y_i is the predicted value, x_i is the actual value and $\bar{x}$ is the mean of the actual values in a set of n samples:

$$R^2 = 1 - \frac{\text{sum squared regression (SSR)}}{\text{total sum of squares (TSS)}}, = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(\bar{x} - x_i)^2} \tag{5.6}$$

5.7 Analysis and Discussion of the Results

In the presented work, an automated application was implemented in order to find the optimal forecasting model for hourly electricity data in MWs. For this purpose, six artificial deep learning models were created and optimized as shown in the tables above. Essentially, an application was implemented which will be able to run not only on electric power data but will be used in any kind of time series forecasting problems.

The results obtained from the algorithmic experiments are presented and an analysis of their results was made. In Figure 5.8 the daily hourly demand prediction in MWs for the three-month period, from 1/10/2019 to 31/12/2019 is presented.

More specifically, Table 5.5 presents the values of R^2, MAE, MSE, RMSE and MAPE for each model. For better understanding of the results, Table 5.5 has been sorted by MAPE and in descending order.

Based on the experiments carried out, it is easy to see that the worst models are the CNN-1D and TCN networks and it seems that the models contain Recurrent Networks, such as LSTM and the two hybrid models implemented, show the best accuracy. So, an important finding is that the models that contain only convolution operation present the worst accuracy, a fact that leads to the conclusion that this mechanism doesn't allow the models to extract meaningful features from the input time series data compared to the other models.

More specifically, CNN-1D and TCN have MAE 0.4936 and 0.4576 MW and RMSE 0.7282 and 0.7096 MW, accordingly, while MLP achieves MAE equal to 0.4477 and RMSE equal to 0.6784 MW. LSTM performs a MAE 0.4275 and RMSE 0.6747 MW and hybrid CNN-LSTM network has MAE 0.4412 and RMSE 0.6866 MW, accordingly.

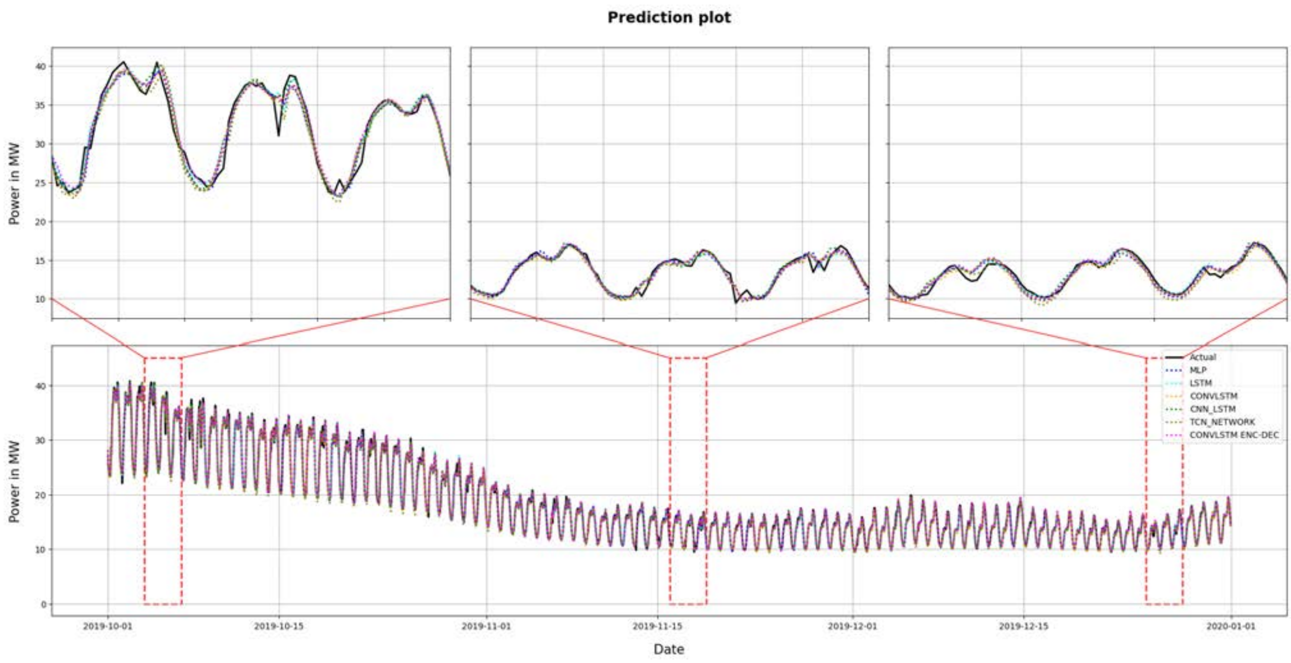


Figure 5.8: The prediction results of each model. Lower subfigure represents the whole testing dataset. Upper subfigures describe different three – day predictions plot from the same training dataset

Table 5.5: Evaluation metrics.

Method	R^2 (MW)	MAE (MW)	MSE (MW)	RMSE (MW)	MAPE (%)
CNN-1D	0.9906	0.4936	0.5302	0.7282	2.66%
TCN	0.9910	0.4576	0.5036	0.7096	2.40%
MLP	0.9919	0.4477	0.4602	0.6784	2.38%
CNN-LSTM	0.9917	0.4412	0.4714	0.6866	2.31%
LSTM	0.9920	0.4275	0.4552	0.6747	2.25%
ConvLSTM Encoder Decoder	**0.9921**	**0.4142**	**0.4454**	**0.6674**	**2.15%**

The proposed model, which has not been used extensively in electricity forecasting datasets is a hybrid model, called Convolutional LSTM Encoder-Decoder. The advantage of this model compared to the other implemented models and mainly to the hybrid CNN- LSTM model is the use of a ConvLSTM Network, which offers better feature extraction and accuracy. Thus, the profile of time series with multiple seasonality, such as the one studied in our case is recognized in a more efficient way, making this model capable of performing MAE and RMSE 0.4142 and 0.6674, respectively. Figure 5.9 presents the comparative values of R^2, MAE, MSE and RMSE for each of the six models proposed and illustrated.

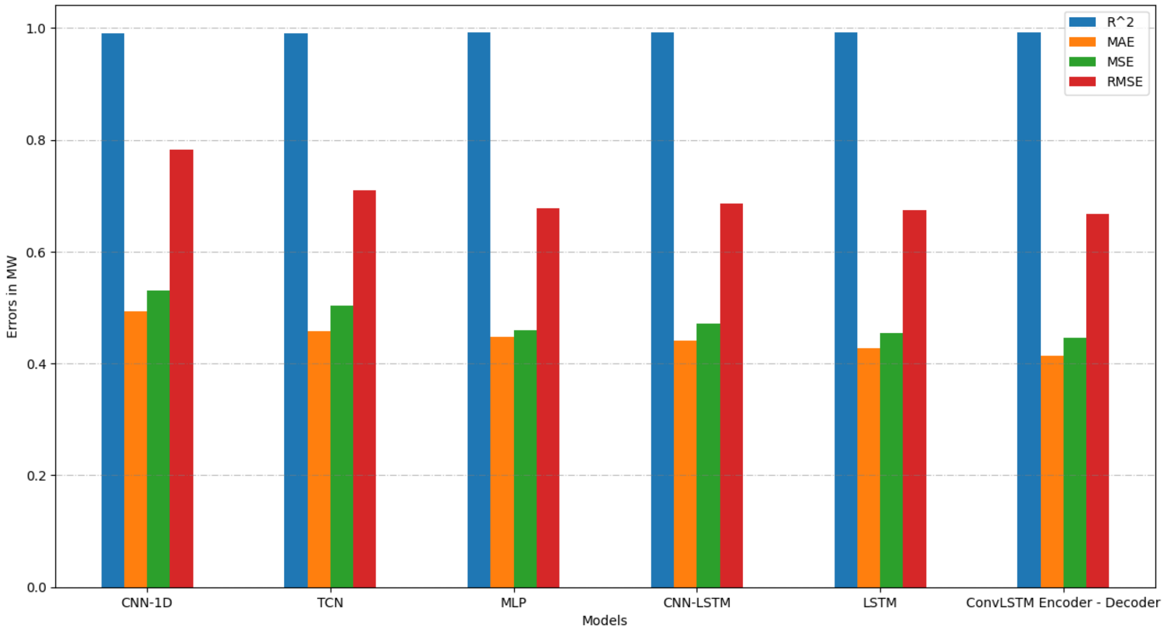


Figure 5.9: Comparison of the models in terms of R2, MAE, MSE and RMSE error metrics

5.8 Summary and Future Work

In recent years, the energy sector has experienced substantial transformations due to continuous surge in demand and to energy crises. These developments have given rise to new requirements aimed at addressing emerging situations and ensuring the effective and efficient operation of the power grid. Accurate demand forecasting is a key factor for the reliable operation and robustness of modern power systems. Especially, in short-term load forecasting, the efficiency is a cornerstone for market flexibility which ensures the balance between supply and consumption in real time. For those reasons, this study introduces an innovative and automated Deep Learning forecasting model for use in the energy sector. The load forecasting results of five enhanced models, which are the CNN-1D, LSTM, CNN-LSTM, TCN and MLP, are produced and compared with those generated by a proposed forecasting hybrid model, named ConvLSTM Encoder-Decoder.

Based on the results obtained and presented, the proposed ConvLSTM Encoder-Decoder, compared with the other methods, leads to better results, achieving a MAPE equal to 2.15%.

LSTM follows with 2.25%, hybrid CNN-LSTM with 2.31%, MLP with 2.38%, TCN with 2.40% and lastly CNN-1D, with a MAPE equal to 2.66%. The differences, in terms of the MAPE, are quite small among them most likely due to the extensive optimization process carried out in all models.

Regarding the contribution of the current research, it is emphasized the implementation of a Deep Learning automated algorithm, which is capable to identify the optimal model for predicting future time series data across various problems. While the present study focused on load forecasting, the proposed approach could also be applied to predict quantities in systems involving renewable energy sources, like wind and solar power. Moreover, the system has potential in predicting building energy consumption in general and specifically in commercial, industrial and military facilities, considering factors such as climate data, indoor environmental information and occupants' behavior, which applies to time series problems.

In addition, through the suggested load forecasting model, a new and innovative deep learning model is presented that achieves the higher forecasting accuracy among the models studied. This proposed hybrid model consists of a ConvLSTM component, which encodes the information contained in the input time series vectors. Then, the information passing through an Encoder Vector, is distributed in an LSTM model, whose target is to predict future data.

Considering the aforementioned contribution of the current research, it becomes apparent that the developed methodology distinguishes itself from the counterparts presented in the literature review. While the referenced publications primarily focus on comparing outcomes derived from various statistical and machine learning methods, this work takes a different approach. The selection of the best model is accomplished through an automated process, introducing flexibility and performance for potential users of the developed application. Furthermore, an advantage of the proposed approach lies in the introduction of a state of the art model (ConvLSTM Encoder Decoder) that has not been applied extensively before to load forecasting problems.

In terms of future research directions stemming from the current research, some noteworthy propositions emerge. The automated model could be enhanced by incorporating additional efficient models such as the Exponential Smoothing LSTM (ESLSTM) and advanced Multi-Head Transformer. A notable challenge lies in applying the models developed to medium and long-term multistep prediction tasks to evaluate their stability over extended sequences. Furthermore, testing the applicability of the proposed approach to other diverse types of time series data, can also be a significant challenge. Moreover, conducting a comparative analysis of prediction accuracy in time series tasks by combining the models used in this publications with Custom Activation Functions represents another potential avenue for future investigation. Lastly, exploring the utilization of these algorithms in Demand Side Management and Demand Response problems is a challenging prospect, given the high level of accuracy required for short-term forecasting in such problems.

5.9 Short-term residential Load forecasting for Smart Grids

5.9.1 Summary of the problem

In recent years, the electricity sector has attracted considerable scientific interest due to global regulatory initiatives promoting sustainability and decarbonization. Transmission and distribution system operators (DSOs) are facing one of the greatest challenges in their history: the energy transition of the power system. Sustainable development and technological advancements, particularly in artificial intelligence and machine learning, play a crucial role in transforming traditional power networks into smart grids. A key aspect of this transformation is the accurate prediction of residential electricity consumption.

Smart metering technologies in smart grids generate large volumes of consumption data, facilitating energy demand forecasting. This forecasting is essential for managing electricity demand and supporting utilities in load planning. The objective of this work is to analyze residential consumer behavior and assess the accuracy of demand predictions, which significantly enhances the operation of smart grids and their interaction with electricity markets. The assets and consumers connected to the distribution grid can also participate in electricity markets by providing additional flexibility services.

This study focuses on residential load forecasting for small-scale consumers on the Greek island of Skiathos. To achieve this, state-of-the-art algorithms are applied to perform short-term forecasting of electricity consumption. The dataset used consists of hourly-resolution data aggregated from 15 individual low-voltage consumers. Five different models are developed to enable a reliable comparative analysis.

The first model is a multilayer perceptron (MLP), consisting of fully connected layers. Next, a convolutional neural network (CNN) model is designed, followed by a single dense layer (CNN-Dense). To further explore the problem, a Long Short-Term Memory (LSTM) model and a Temporal Convolutional Network (TCN) are implemented, both specifically designed for time-series forecasting. Additionally, an ensemble model combining CNN and LSTM predictions is developed. Bayesian Optimization is applied to fine-tune the hyperparameters of the models for optimal performance. Finally, a Naïve model is implemented as a benchmark to compare the deep learning models against a statistical approach.

The work is structured as follows: First, a literature review provides an overview of related scientific studies. Then, the methodology and operation of the five forecasting models are outlined. An Exploratory Data Analysis (EDA) is conducted to examine key characteristics of the demand time series, including seasonal patterns and peak consumption periods. Following this, the forecasting errors are analyzed through comparative metrics to evaluate the accuracy of the models. The study concludes with a discussion of key findings and suggestions for future research.

5.9.2 Introduction

The energy sector has garnered significant attention from both researchers and investors in recent years, owing to its rapid growth and transformative changes. Emerging investments necessitate the deployment of new equipment, accompanied by the formulation of novel rules and legislation designed to safeguard the environment and address the greenhouse effect. These developments underscore the sector's heightened importance and establish essential conditions for its oversight [147]. Simultaneously, a widespread and profound energy crisis has unfolded on a worldwide level, primarily attributed to the ongoing conflict in Northern Europe. This geopolitical situation has led to the suspension of operations by large industrial units and businesses, driven by escalating operating costs. The surge in global electricity demand has further exacerbated uncertainties and energy-related risks [81]. Among all these advances, the short-term prediction of household electricity consumption emerges as a fundamental factor for the seamless operation of modern electrical networks [132]. This task involves forecasting the short-term trends in the electricity consumption of residential consumers, and it plays a crucial role in ensuring the efficient functioning of contemporary electrical grids. A precise understanding of short-term electricity consumption patterns enables proactive resource management, optimising grid performance, and enhancing overall reliability [148]. In terms of buildings power usage prediction, different practices were applied to overcome the difficulties due to its high energy consumption, nonlinearity of data, and dynamic occupant behavior [149]. On one hand, statistical methods are used, as Awerkin et al. [150] propose a hybrid model which consists of statistical parametric methods of Fourier analysis and stochastic differential equations. The findings of the current research conclude that this model is more suitable for short-term predictions in small time intervals rather than long-term forecasting horizon. On the other hand, in recent power prediction works, experts utilise machine learning (ML) and deep learning (DL) algorithms. Dinesh et al. in [151] present a new method for forecasting energy consumption in a house based on non-intrusive load monitoring (NILM), and affinity-aggregation spectral clustering is introduced, with the idea of extending it to forecast consumption in a greater number of houses, such as a micro-grid. Furthermore, Biswas et al. in [152] developed Artificial Neural Network models based on the Levenberg-Marquardt and OWO-Newton algorithms, which produce show promising prediction results using data from TxAIRE Research houses. Kim and Cho in [153] processed data an apartment with a resolution of one minute with a sliding window technique to develop a 60-minute time frame and set the next 60 minutes as a prediction. The multi-variable time series is converted to a two-dimensional array to feed a hybrid forecasting model (Convolutional Neural – Long-Short Term Memory Networks). Furthermore, Ullah et al. in [154] examine the use of an intelligent hybrid approach that incorporates a convolutional neural network (CNN) with a multilayer bidirectional long-short-term memory (M-BDLSTM) method to effectively learn the sequence pattern of the predicted data. The main contribution of this work lies in the direct application of Deep Learning models to the final demand curve, which is an important aspect in economic or market contexts. The specificity of this application is underscored using a limited sample of consumers, deliberately avoiding assumptions and data smoothing. Despite these constraints, the research asserts that not only is the proposed task feasible but can be applied and implemented effectively under real-world operational conditions, such as its integration into the systems developed by the DSO and TSO. From the DSO perspective, the provision of local flexibility services, such as congestion management and voltage control is a pivotal role challenge, and from the TSO point of view, these small consumers will be enabled through IoT devices to participate in a secure manner directly into the balancing market, for adjusting the frequency of the grid. This emphasis on practicality suggests a focus on the applicability and robustness of the DL models in addressing real-world scenarios, offering insights into the potential implications of the findings for broader contexts beyond the study's limited data. This work is organized as follows: Section 5.10

discusses the main methodology and the data collection process. Section 5.11 presents in detal the datasets used for this task. Section 5.12 analyzes the forecasting models. Section 5.13 shows the results of the short-term residential power prediction, and Section 5.14 concludes this work, discussing the main conclusions and future study proposals based on this work.

5.10 Methodology

5.10.1 Datasets

The data set of this study refers to the power consumption of 15 residential consumers. These properties are located on Skiathos Island – Greece. Skiathos Island is a famous tourist destination during the summer period, so the load curve of consumers presents increased usage. Due to the privacy of client data, we studied the aggregated value of the 15 residences, as presented in Figure 5.10. The investigation period is from 1-08-2020 to 30-06-2023. Data were provided to the researchers of the current study from the Hellenic Electricity Distribution Network Operator [155].

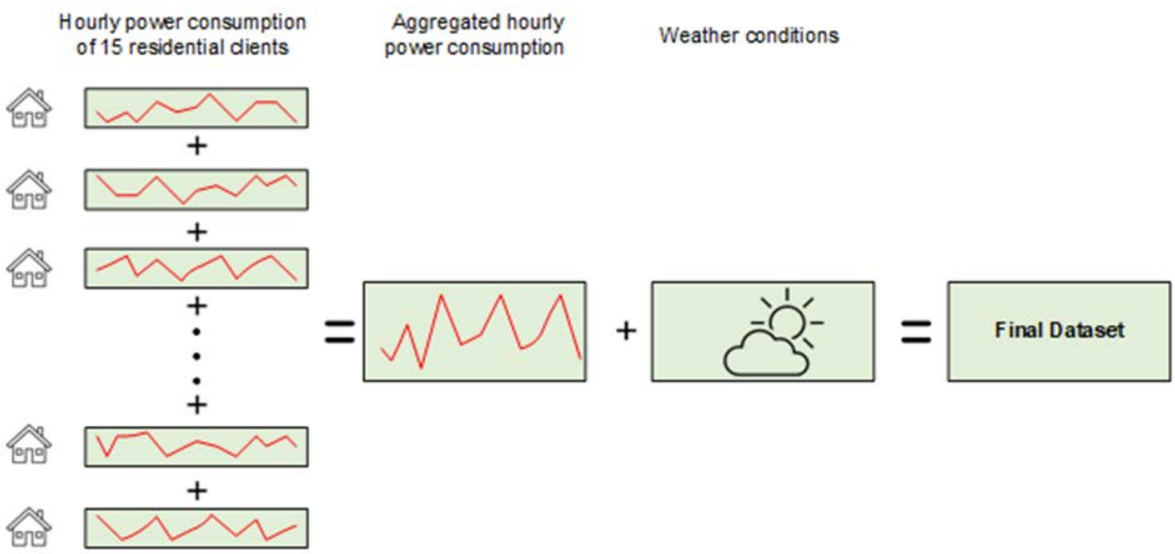


Figure 5.10: Creation of the dataset

In addition, the weather conditions of the island are recovered from the meteorological station of the Skiathos airport. The International Civil Aviation Organization (ICAO) code for this airport is LGSK. In general, airport meteorological stations publish reports twice per hour, named METAR, on the weather conditions that occur on the runway. These reports include elements such as temperature, humidity, wind speed, and direction, among others, which are also the elements that were used for this work.

5.10.2 Software Environment

This work's trials were conducted in Python language, version 3.10, leveraging the Open-Source software library Tensorflow 2.14.0 and the high-level API Keras 2.15.0, for training and testing the deep learning algorithms. Additionally, Pandas 2.1.0 and Numpy 1.26.0 libraries for data analysis. Furthermore, the Seaborn and Matplotlib libraries were used for visualisation purposes of the exploratory analysis and prediction results. The trials were carried out on Google Colab Pro platform, which utilizes GPU NVIDIA (Version 460.32.03), RAM: 25.45 GB, disk space: 166.77 GB, and CUDA (Version 11.2).

5.10.3 Prediction Evaluation

For this study, the following three most common prediction performance metrics in regression analysis are used:

Mean Absolute Error (MAE): This metric estimates the average of the absolute differences between the forecasted and true values. MAE is calculated as:

$$MAE = \frac{\sum |y - \hat{y}|}{n} \tag{5.7}$$

Root Mean Squared Error (RMSE): This metric calculates the square root of the average of the squared differences between the predicted and true values. RMSE is calculated as:

$$RMSE = \sqrt{\frac{\sum (y - \hat{y})^2}{n}} \tag{5.8}$$

Mean Absolute Percentage Error (MAPE): This metric computes the average of the absolute percentage differences between the predicted and actual values. MAPE is calculated as:

$$MAPE = \frac{\sum \left| \frac{y - \hat{y}}{y} \right|}{n} \times 100 \tag{5.9}$$

where y is the actual value, $\hat{y}$ is the predicted value, and n is the number of observations.

5.11 Data Analysis and Results

In this section, the analysis of the studied data is initially presented, followed by a focus on the fundamental architectures and hyperparameters of each model. Finally, a detailed presentation of the prediction results of each algorithm is provided.

5.11.1 Exploratory Data Analysis

Concerning the variation in consumption data, the boxplots in Figures 5.11, 5.12 and 5.13 illustrate the total average consumption per day of the month, per day of the week, and per month, respectively. It is observed that the months with the highest consumption are July and August, a phenomenon that is logical due to the increased temperatures during the summer period. Notably, during the day, the average hourly consumption follows a consistent trend, as does the average consumption during the midday hours per week. These variations are indicative of the non-uniform fluctuation exhibited by the data and the absence of specific patterns.

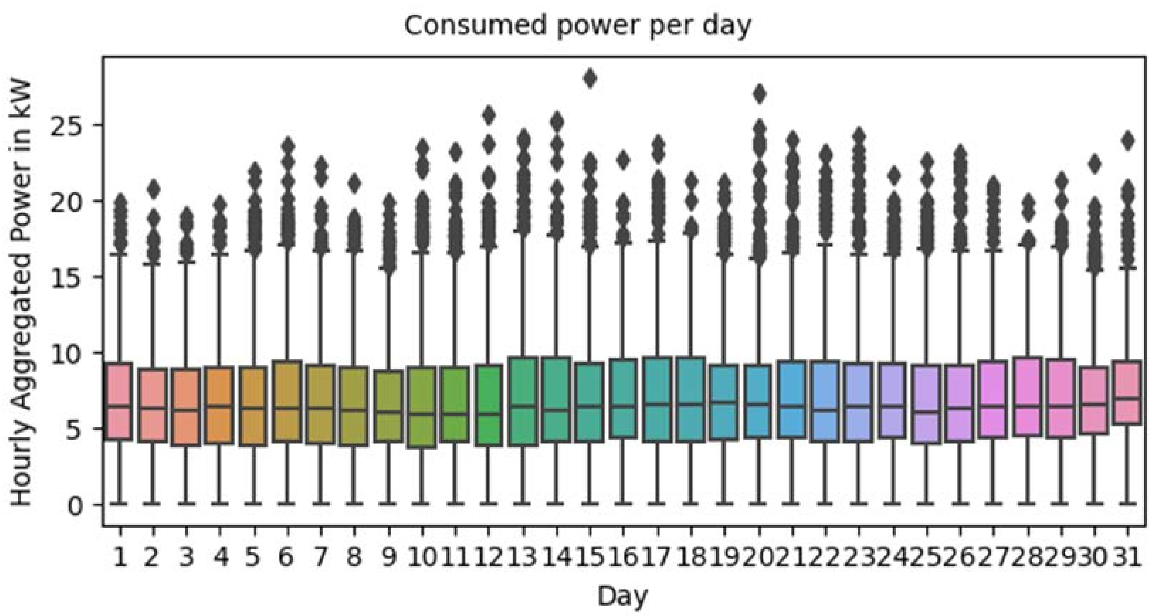


Figure 5.11: Daily total average consumption

Figure 5.14 presents the correlation between the variables of weather characteristics and electricity consumption. None of the variables exhibit a high correlation, a fact that led us to utilize univariate prediction methods.

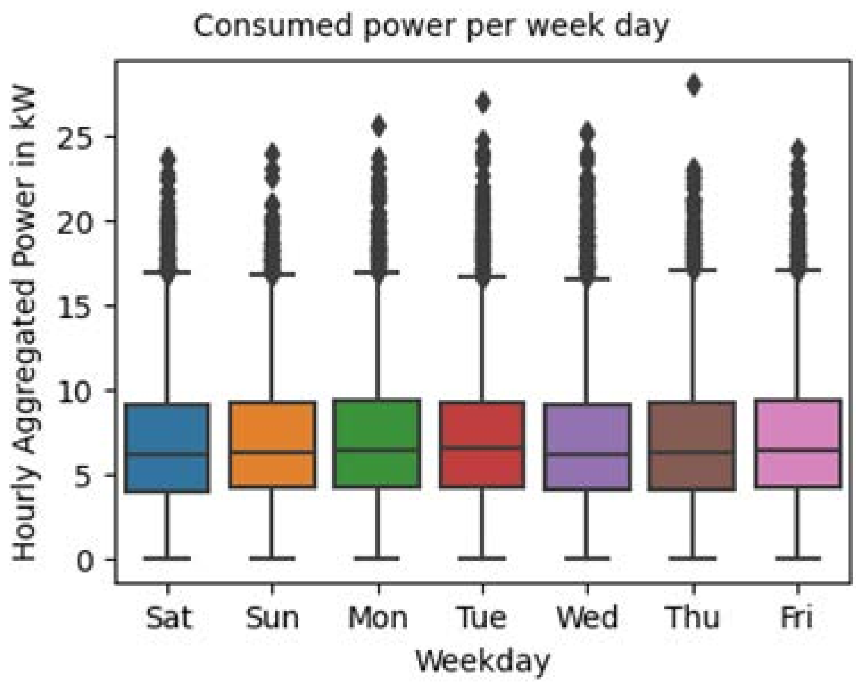


Figure 5.12: Weekly consumption

5.11.2 Feature Engineering

With the aim of creating appropriate features for the optimal training of models, the One-Hot Encoding technique is employed. This is a scientific approach that transforms categorical

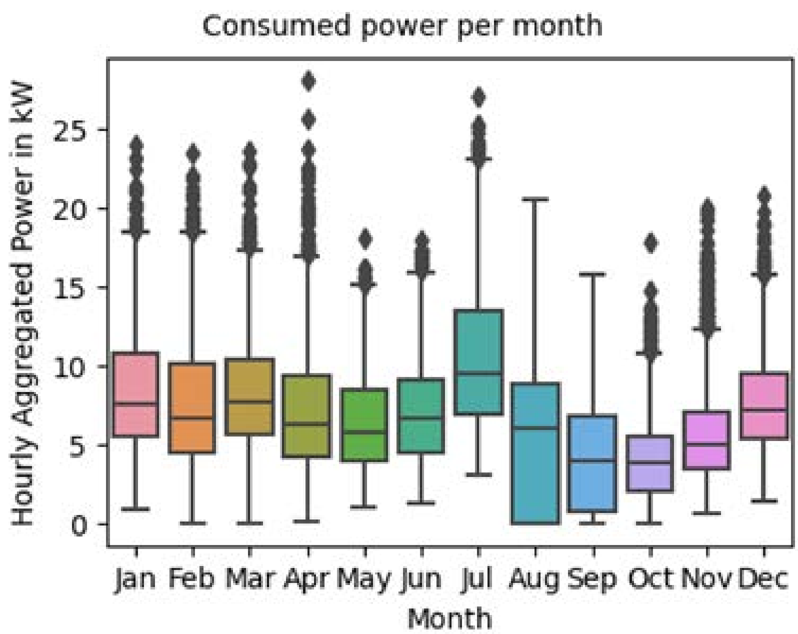


Figure 5.13: Monthly consumption

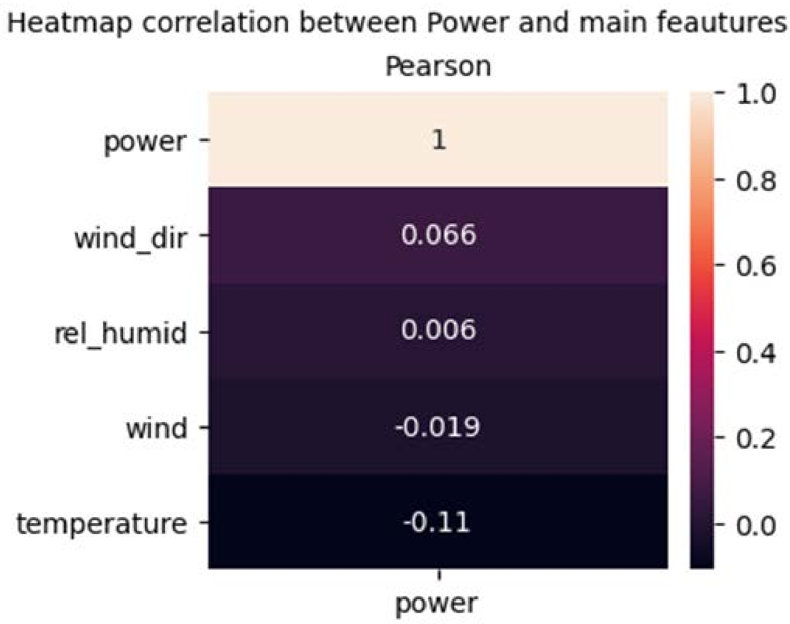


Figure 5.14: Correlation Heatmap

data into numerical vectors and converts numerical information into cyclical patterns through trigonometric transformation. This method converted weekdays, hours, and months into sine and cosine representations. Specifically, the features created are as follows:

- **Hourly Consumption**: The historical 24 hours of values for each predicted hour.
- **Sin and Cos of the day**: Sin and cosine representation for the day.
- **Sin and Cos of the hour**: Sin and cosine representation for the hour.
- **Sin and Cos of the month**: Sin and cosine representation for the month.
- **Weekend**: Dummy variable where 0 represents working days and 1 represents week-ends.

5.12 Forecasting Models

5.12.1 Deep Learning Architectures

The hyperparameters for all models were determined through the application of the Bayesian Hyperparameter Optimization Algorithm. For every algorithm, Adam Optimizer was chosen due to its adaptive learning rate, efficient handling of sparse gradients, fast convergence, and

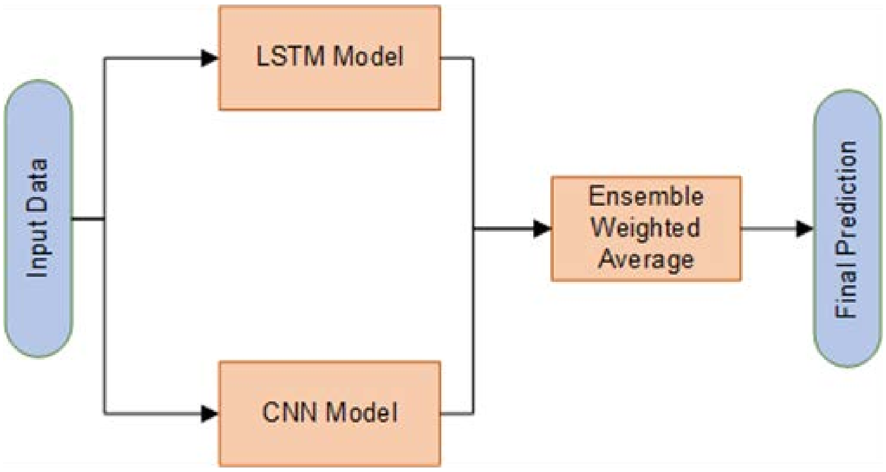


Figure 5.15: Weighted Average Ensemble Model

robustness to noisy data. The specific details of the final hyperparameters are outlined as follows:

- **MLP Model**: Batch size = 64, Learning rate = 0.0010128, Neurons for the Dense Layer.
- **CNN-1D Model**: Filters = 64, Learning rate = 0.0010, Kernel size = 3, MaxPooling1D (pool size = 2), Neurons of Dense Layer = 16, ReLU activation function for CNN and Dense Layer.
- **LSTM Model**: Batch size = 128, Learning rate = 0.0010, Units of LSTM network = 48, ReLU activation function.
- **TCN Model**: Batch size = 128, Learning rate = 0.0010, Filters = 256, Dilations = [1, 2, 4, 8, 16, 32].
- **Weighted Average Ensemble Model**: Batch size = 128, Learning rate = 0.0010, Filters = 64, Kernel size = 3, 48 units for the LSTM Module, MaxPooling1D with pool size = 2, Contribution of LSTM: $W^{lstm} = 0.5$, Contribution of CNN: $W^{cnn} = 0.5$. This model is visualized in Figure 5.15.

5.12.2 Naïve model

The Naïve model is a statistical approach that, for each forecast hour, considers the average value of the same hour one week ago and the same hour two weeks ago. The underlying assumption of this method is that the time series exhibits specific patterns, such as seasonality at the monthly, weekly, and daily levels. This approach serves as a benchmark model for comparison in time series forecasting tasks.

5.13 Results

Data used in this work were divided into training and testing sets. As the testing set, the last six months of the initial dataset were selected. The previously mentioned models were trained, and the values of the relevant metrics used to evaluate the performance of each algorithm are presented in Table 5.6.

It is observed that all Deep Learning models consistently outperform the benchmark statistical model used. According to MAE, the best-performing model is the Ensemble, with a value of 1.580 kW. Additionally, the three individual models, CNN, LSTM, and MLP, exhibit very close prediction accuracy values, with MAE scores of 1.606 kW, 1.586 kW, and 1.624 kW, respectively. The poorest prediction model is Naïve (Benchmark), achieving MAE, RMSE, and MAPE values of 2.608 kW, 3.344 kW, and 28.50%, respectively. This fact implies the superiority of DL and ML models over the statistical model, as the latter is unable to capture the non-linear variations in time series and the abrupt changes in instantaneous values.

For better representation of the results, Figure 5.16 illustrates the predicted values in comparison to the actual values for a one-week period for every evaluation month. It becomes evident that the fluctuation of actual hourly consumption does not follow a specific pattern, presents multiple seasonality between the duration of the day, and also exhibits volatility, a factor that complicates the operation and prediction of the models further.

5.14 Summary of the Results and Future Work

In this work, an extensive attempt was made to forecast the consumption of household consumers from Skiathos Island. Hourly data from 15 households are used. The problem is that having only a few households makes it hard to create a reliable prediction model. Also, the amount of electricity used by homes changes based on the time of the month, week, and day. This happens because people's electricity usage follows a stochastic pattern at times.

Table 5.6: Performance metrics of the models

Model	MAE (kW)	RMSE (kW)	MAPE (%)
Naïve (Benchmark)	2.608	3.344	28.50
CNN	1.606	2.114	16.25
LSTM	1.586	**2.084**	16.38
MLP	1.624	2.092	17.55
TCN	1.691	2.192	17.67
Ensemble	**1.580**	2.109	**15.62**

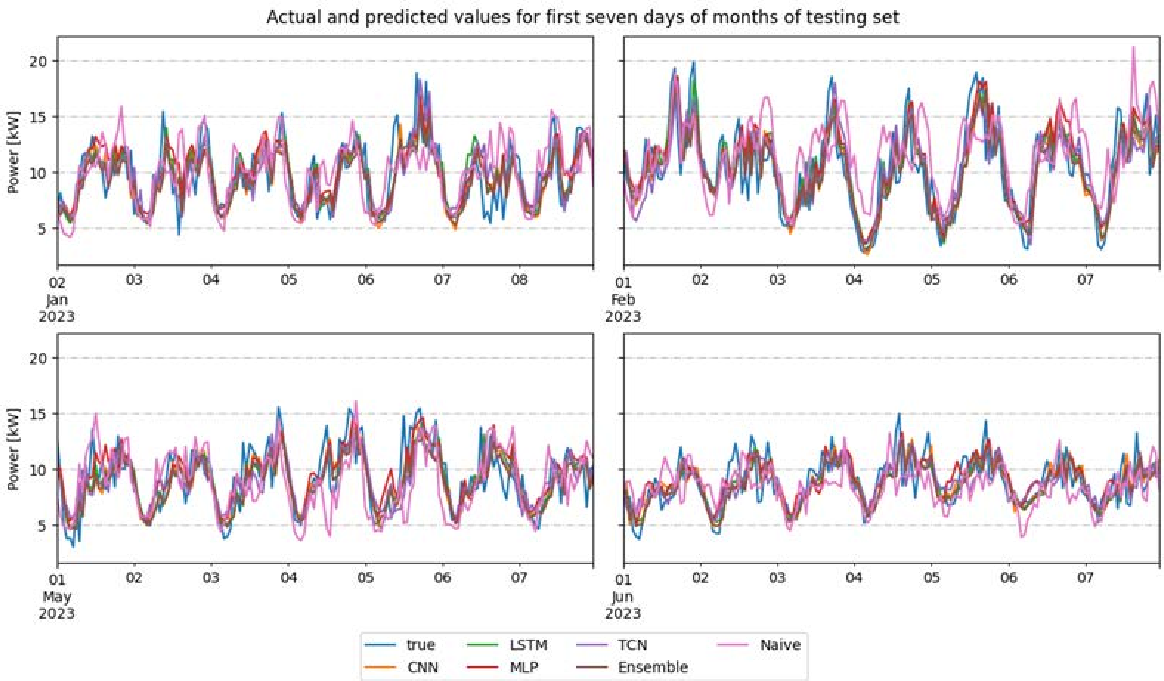


Figure 5.16: Actual and Forecasting Results for first 7 days of each month of the testing set

The research findings demonstrate that deep learning models surpass statistical models, for instance, the naive model employed, in accurately capturing the consumer behavior curve. Another significant conclusion is that, for a limited consumer sample, weather phenomena, including temperature, humidity, and wind speed, do not significantly impact the hourly variation in consumption. This assertion is supported by the low correlation observed among these specific variables in our research.

Looking ahead to future research ideas based on this study, a key focus is on creating Artificial Intelligence models tailored for predicting consumer behavior. The close connection between household activities and electricity consumption is an important area that needs in-depth exploration. The strong link observed between household activities and electricity consumption highlights the potential success of Machine and Deep Learning models in predicting and understanding consumer energy usage.

Moreover, there is a significant area gaining interest worldwide in research circles: Demand Side Management (DSM) and Demand Response (DR) programs. These efforts, aimed at improving energy usage and enhancing power grid reliability, offer another area for thorough exploration. Companies, such as DSOs in various European countries, are actively involved in these initiatives to achieve intelligent energy conservation.

Consequently, the amalgamation of the algorithms meticulously developed within this research work holds promise as a crucial determinant for the successful implementation and functioning of such advanced programs. The potential ramifications of incorporating these algorithms extend beyond theoretical considerations, suggesting a practical application in the operationalization of sophisticated energy management initiatives.

Chapter 6

Wholesale Electricity Price Forecasting

6.1 Challenges and Solutions in Electricity Price Forecasting

Similar to ELF methods, in this section, studies that have significantly contributed to the field of price forecasting are extensively analyzed. Initially, statistical models are presented, followed by publications focusing on ML and DL, and finally, emphasis is given to Ensemble Methods. Similar to ELF, for a more substantial presentation of the related works, the related studies utilize ML and DL models are presented together.

6.1.1 Statistical Models

Similar to ELF approaches, the statistical models used to predict electricity prices rely on similar algorithms. Initially, traditional time series forecasting methods, like simple regression techniques, moving averages and smoothing methods were employed. Then, more complex and advanced models like dynamic regression, models, Box-Jenkins, and hybrid statistical models [156] to detect trends and seasonality in electricity price time series data were investigated. Various works, including [157, 158, 159, 160], have explored ELF using the above methods. These foundational models paved the way for the creation of more powerful forecasting techniques. Skopal et al. in [157] propose a hybrid combination of ARIMA and a MLP model. The statistical model identified the linear trends, while the neural network handled the modeling of the nonlinear residuals. This approach leaded in lower forecast metrics, measured by the MAPE and mean absolute deviation (MAD), in comparison to the single models. Similarly, other stydies including [158, 159], propose multiple autoregressive models, like SARIMA and GARCH, as weel as more advanced hybrid algorithms to predict electricity prices, evaluating their performance through statistical metrics like R^2 and MAPE. Their results highlight the effectiveness of these hybrid models. Additionally, a hybrid model combining ARIMA and multivariate linear regression and Holt-Winters is developed in [160]. For the evaluation process, datasets from the Iberian electricity market are used and MAPE is used. The results show that the combined model achieves satisfactory performance compared to other benchmark models. Marcjazz et al. in [161] focuse on the selection of optimal calibration windows for day-ahead EPF. The research utilizes six years of data from three major power markets and involves fitting four autoregressive expert models to both raw and transformed price time series. Also, Sanchez et al. in [162] investigate an ARIMA model using multiple input data for the Pennsylvania-New Jersey-Maryland (PJM) and Spanish electricity market, in order to generate 24 hours ahead forecasts.

In terms of statistical methods, the authors in [163] use ARIMA, SARIMA, SARIMAX, and multiple linear regression (MLR) forecasting methods to forecast day-ahead electricity prices for Germany, based on a variation of input features. In this article [164], a collaborative framework is suggested between humans and machines to predict the day-ahead electricity price in wholesale markets. This framework combines numerous statistical and machine learning models, various exogenous features, a combination of several timeseries decomposition techniques, and a collection of timeseries characteristics derived from signal processing and timeseries analysis methods. The researchers in this work [165] apply traditional inferential statistical methods, among others, for the prediction of electricity prices and proposed an innovative solution to forecast electricity spot prices. In recent literature statistical methods are mostly used as benchmark models to compare with Machine, Deep, and Ensemble Learning models, which are presented below.

Other studies [166, 167, 168] have focused on comparing different time series analysis methods for EPF. Abunofal et al. in [166] conduct a comparative analysis between different autoregressive models, like ARIMA, SARIMA and SARIMAX, which utilizes exogenous features as input. The aim is to create forecasts for the German DAM. Between the models, the SARIMAX, achieved the highest forecast accuracy. Also, Banitalebi et al. in [167] compare DES and TES for EPF from variability, utilizing elastic net regularization. The forecasted results show that TES provides exceptional performance, and regularization helps minimize the MSE. These outcomes aid in making well-informed decisions for power generation planning within the market. Additionally, various autoregressive statistical models and their derivatives have produced accurate forecasting results, as discussed in [168]. The study provides detailed computational procedures, in detail results, and evaluation metrics, like MAPE, highlighting promising outcomes and areas for further consideration.

Khairuddin et al. in [169] explore the Imbalance Cost Pass-Through (ICPT) within Malaysia's incentive-based regulation (IBR), emphasizing its role in enabling power producers to adjust tariffs in response to changing fuel prices, thereby strengthening economic resilience in electricity generation. The Energy Commission's biannual reviews of generation costs lead to either rebates or surcharges for customers. The study proposes forecasting ICPT prices to aid stakeholders in Peninsular Malaysia, utilizing three models and evaluating performance via the MAPE. Findings indicate that ARIMA is the most accurate forecasting method in comparison to MA and LSSVM, with minimal price differences observed. The forecasted ICPT prices align closely with actual tariffs, suggesting a potential price drop in the upcoming announcement. This research could positively impact Malaysia's energy sector sustainability.

GARCH approaches are advantageous in ELF tasks because they effectively capture price volatility and temporal dependencies, allowing for more accurate predictions of sudden price spikes and fluctuations in highly volatile electricity markets. Based on these, Entezari et al. in [170] examine methods to quantify risk and forecast the bias of electricity prices to revert to their historical average in Portugal's energy market. The data were used are from January 2016 to December 2021. They apply both symmetric and asymmetric GARCH models to mitigate data fluctuation peaks, with the asymmetric model capturing the distinct effects of positive and negative shocks. The MSGARCH model, estimated with two distinct states, successfully captures the shifts between low and high volatility periods, as well as the prolonged impact of shocks during high volatility. The advantage of this method is that helps power producers predict price movements and adjust their operations with high accuracy accordingly, leading to improved market stability and efficiency. Finally, authors in [171] research the financial impact and sustainability of a RES. More specifically, they use a PV plant with nominal power equal to 5 MWs, which is located in Montenegro. This PV plant has direct market access instead of a long-term Power Purchase Agreement (PPA). Using ARIMA for price forecasting, it finds that this riskier strategy is more profitable and offers a shorter payback period. The study concludes that direct market access is a better option to enhance project returns and support energy transformation in developing countries.

6.1.2 Machine and Deep Learning Models

The integration of ML and DL models into wholesale EPF problems has been gaining increasing application in recent years. Gonzalez et al. in [172] examine the effectiveness of two advanced and hybrid models for forecasting next-day spot electricity prices on the APX-UK power exchange. The first model combines a fundamental supply stack model with an econometric model incorporating price driver data. The second model extends this approach by including logistic smooth transition regression (LSTR) to handle regime changes during structural shifts. The evaluation results in the test set from both hybrid models, particularly the hybrid-LSTR, perform better than those from non-hybrid SARMA, SARMAX, and LSTR models. The models' prediction intervals (PIs) are assessed by comparing nominal to actual coverage, though no formal tests are conducted. The LSTR model achieves the best results, followed by the hybrid-ARX and SARMAX models. However, the hybrid-LSTR model tends to produce too many exceeding prices due to overly narrow PIs.

Moreover, regarding Machine Learning (ML) applications, the authors in [173] present a three-stage model for short-term electricity price forecasting using ensemble empirical mode decomposition (EEMD) and extreme learning machine (ELM). In [174], a probabilistic forecasting method for day-ahead electricity prices is proposed using SHapley Additive exPlanations (SHAP) feature selection and Long- and Short-term Time-series Network (LSTNet) quantile regression. The study in [175] discusses the use of artificial neural networks to predict the price and demand of the day of electricity, compares different neural network architectures and finds that methods that combine fully connected and recurrent or temporal convolutional layers present better performance. Following the previous, a model for electricity price forecasting is proposed in [176], which improves precision by combining trend decomposition, temporal convolutional network, and neural basis expansion analysis (STL-TCN-NBEATS). The authors in [177] propose using Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) classifiers for short-term electricity price and load forecasting. The modified SVM achieves an accuracy of 88.2740% for price forecasting, while the modified KNN achieves an accuracy of 89.8605% for load forecasting. In [178], a novel hybrid machine learning algorithm is presented to predict the day-ahead EPF, integrating linear regression with Automatic Relevance Determination (ARD) and ensemble bagging Extra

Tree Regression (ETR) models. Recognizing that each model of EPF has its own strengths and weaknesses, the combination of several models presents more precise predictions and overcomes the constraints associated with any individual model. The researchers in [179] optimized an electricity price prediction model with inverted and discrete particle swarm optimization, which utilized variational mode decomposition and the K-means clustering algorithm. Additionally, the authors of [180] perform simulations for four primary regression techniques, including Support Vector Machine, Gaussian Processes Regression, Regression Trees, and Multi-Layer Perceptron. Based on performance, such as MAE, RMSE, R, and the total number of percentage error anomalies, Support Vector Machine (SVM) stands as the optimal selection for predicting the price of the electricity market and, finally, [181] proposes a combination of single and hybrid machine learning models for predicting the short-term price of electricity. The hybrid models outperformed the single models, with the Logistic Regression - CatBoost model achieving the lowest MAE and RMSE values.

At the same time, open-source platforms provide great flexibility for such problems. Medina et al. in [182] investigate the application of advanced GPT tools, such as OpenAI's ChatGPT, to improve electricity price prediction in Spain's complex energy market. By analyzing energy news and expert reports, they generate supplementary variables that enhance predictive models. Their research includes two methodologies: employing in-context example prompts and fine-tuning GPT models. The results show that insights derived from GPT models closely match post-publication price changes, indicating that these methods can significantly boost prediction accuracy and aid market stakeholders and companies in making informed decisions amidst electricity price volatility. Simultaneously, [133] investigates the problem of short-term ELF using PSO methodologies in order to optimize back propagation feedforward networks. The results comparison is based on MAE and MSE, achieving 8.69% and 1.62%, respectively.

Furthermore, on the implementation of deep learning (DL) techniques in electricity forecasting tasks, [183] discusses electricity price forecasting using deep learning algorithms such as long short-term memory networks (LSTM), deep neural networks (DNN), convolutional neural networks (CNN), CNN-LSTM, and CNN-LSTM-DNN. The accuracy of the models is evaluated using MAE and MAPE metrics. In addition, [184] presents a hybrid model that combines discrete wavelet transform (MODWT) denoising and empirical mode decomposition (EMD) with sequence-to-sequence LSTM neural networks for forecasting electricity prices. The proposed model outperforms other forecasting models such as LSTM and stacked LSTM. The authors in [185], present a modular day-ahead electricity price forecasting approach that incorporates convolutional and recurrent neural networks to predict electricity prices. An optimized heterogeneous structure LSTM model is presented in [186] to forecast the electricity price, which improves accuracy and stability compared to traditional models. Researchers in [187] propose a hybrid neural network model for short-term load forecasting, combining a temporal convolutional network (TCN) and a gated recurrent unit (GRU). The model is optimized using the AdaBelief optimizer. The authors in [188] investigate an efficient method for forecasting the daily electricity price using a LSTM neural network model. [189] is about electricity price forecasting based on an enhanced convolutional neural network, while [190] proposes a novel methodology for probabilistic energy price forecast using Bayesian deep learning techniques. It focuses on providing prediction intervals and densities to enable enhanced bidding and planning strategies considering uncertainty explicitly. Following, in [191] a hybrid model is proposed for electricity price forecasting, which decomposes the price series, selects significant subseries, predicts each subseries using the GRU network, calculates the best weights using the particle swarm optimization algorithm (PSO), and combines the results to obtain the final price prediction. Furthermore, in [192] a short-term electricity price forecasting method based on data mining is introduced, using a fuzzy classification miner and an LSTM model. In [193] a model is presented for the forecast of electricity prices based on Gated Recurrent Units (GRU). The model considers factors such as temperature, wind speed, and daily activities to increase the accuracy of the forecast. The authors in [194] proposed a feedforward deep neural network with the integration of an autoregression module to minimize error on day-ahead price forecasting task.

Some works carry out extensive comparisons between ML and DL models. Conte et al. in [195] explore forecasting techniques to enhance sustainable electricity management in Brazil by predicting critical variables such as the price of differences settlement (PLD) and wind speed. They employ two approaches: the first combines deep neural networks (LSTM and MLP) optimized with a canonical genetic algorithm (GA), and the second uses machine committees. One committee includes MLP, decision tree, linear regression, and SVM, while the other consists of MLP, LSTM, SVM, and ARIMA. Their findings indicate that the hybrid GA + LSTM algorithm performs exceptionally well in PLD prediction, achieving a mean square error (MSE) of 4.68, and in wind speed forecasting, with an MSE of 1.26. These methodologies aim to enhance decision-making in Brazil's electricity market.

Transformer models are gaining more and more interest in EPF sector, mainly due to their attention mechanisms, which allow them to focus selectively on relevant time steps, enables parallelization and capture intricate temporal dependencies in sequential data. Laitsos et al. in [196] utilize advanced DL algorithms, implementing four different novel models: a CNN incorporating an attention mechanism, a hybrid CNN-GRU with attention, a soft voting ensemble, and a stacking ensemble model were introduced. Additionally, they developed an optimized transformer model known as the Multi-Head Attention model and the Perceptron model is selected as benchmark. The results reveal substantial predictive accuracy, particularly with the hybrid CNN-GRU model, which achieved a MAPE of 6.333%. The soft voting ensemble and Multi-Head Attention models also demonstrated strong performance, with MAPEs of 6.125% and 6.889%, respectively. These findings underscore the potential of transformer models and attention mechanisms in EPF, suggesting promising directions for future research.

Additionally, the influence of conventional energy forms plays a significant role in shaping the final hourly price. Jin et al. in [197] aim to enhance electricity market price forecasting by accounting for the discrete influences of fuel costs, large-scale generator costs, and demand. They introduce two innovative techniques: one focuses on feature generation and forecasting transformed variables based on fuel cost per unit and the maximum of the Probability Density Function (PDF) derived from Kernel Density Estimation (KDE). The second technique involves demand decomposition followed by feature selection, verified by gain or SHapley Additive exPlanations (SHAP) values. In their case study in the Korean Electricity Market, they observed improvements across all indicators. Feature generation using fuel cost per unit improved forecasting accuracy by reducing monthly volatility and error, achieving a MAPE of 3.83%. The KDE-based method achieved a MAPE of 3.82%. Combining both techniques resulted in the highest accuracy with a MAPE of 3.49%, while using decomposed data alone had a MAPE of 4.41%.

Baskan et al. in [198] explore the use of ML and DL methods to predict day-ahead electricity prices over a 7-day horizon in the German spot market, focusing on the challenge of increasing price volatility. The study underscores the potential of ML models to deliver reliable predictions in volatile and unpredictable market conditions, presenting a compelling alternative to traditional models that rely heavily on human modeling and assumptions. Through both qualitative and quantitative analysis, the research evaluates model performance based on forecast horizon and robustness determined by hyperparameters. Three distinct test scenarios are manually selected for evaluation. Various models are trained, optimized, and compared using standard performance metrics. The results show that DL models outperform tree-based and statistical models, demonstrating superior effectiveness in managing volatile energy prices.

Other studies address the complexity of data that includes numerous block bids and offers per hour. Pinhão et al. in [199] present an innovative methodology for EPF by predicting the underlying dynamics of supply and demand curves derived from market auctions. The proposed models integrate several seasonal effects and key market variables such as wind generation and load. The evaluation metrics show the superior performance compared to benchmark methods, emphasizing the importance of integrating market bids in EPF.

Some studies examine the integration of explainable AI into Deep Learning models in order to deliver more precise and dependable conclusions. Heistrene et al. in [200] integrate Shapley additive explanations (SHAP) with LSTM and CNN models in order to introduce an Error Compensation EPF method. Data from Italian Power Market are used, and RMSE, MAE and MAPE metrics are used. Also, Heistrene et. al. in [201] introduce an explainable AI Deep Learning method using a CNN model, in order to enhance the predictions of electricity prices from Italy's and ERCOT's electricity market. The main findings of these studies are the important role of Feature Importance, which help identify the impact of each input feature on model predictions and the models' interpretability, which helps researchers understand the functionality pf each model.

Additionally, Duo et al. in [202] introduce a day-ahead forecasting method for China's Shanxi electricity market prices, using a hybrid extreme learning machine (ELM), which randomly initializes the weights of the hidden layer and only requires the calculation of the output weights, making it significantly faster compared to traditional neural networks. Initially, they preprocess trading data by eliminating weakly correlated features using the Spearman correlation coefficient. Next, they develop a day-screening model based on grey correlation theory to filter out irrelevant samples. In order to improve the accuracy, they optimize the parameters of the regularized ELM (RELM) using the Marine Predators Algorithm (MPA). The proposed model demonstrates superior performance compared to other existing models, helping to develop strong quotation strategies.

Finally, Saglam et al. in [203] investigate the influence of exogenous independent variables on peak load forecasting by employing multiple combinations of multilinear regression (MLR) models. Data from 1980 to 2020 are used, incorporating factors such as import/export figures, population, and gross domestic product (GDP) to predict instantaneous electricity peak load. The performance of these techniques is evaluated throurg MAE, MSE, MAPE, RMSE, and R-squared. The results show that the ANN and GRO models achieve the highest forecasting accuracy.

6.1.3 Ensemble Methods

Ensemble models various advantages, as they leverage the capabilities of both statistical algorithms and machine and deep learning. Kushagra Bhatia et al. in [204] investigate the influence the impact of RES on price forecasts, using them in in the training process of the models. The study proposes a bootstrap aggregated-stack generalized architecture for very short-term EPF, aimed at helping market participants develop real-time strategies. Guo et al. in [205] introduce an advanced ensemble model, which is called BND-FOA-LSSVM, by combining three district algorithms. These are Beveridge-Nelson decomposition (BND), fruit fly optimization algorithm (FOA), and least square support vector machine (LSSVM). The results show that the proposed model achieves a MAPE of 3.48%, a RMSE of 11.18 Yuan/MWh, and a MAE of 9.95 Yuan/MWh. These metrics are significantly smaller compared to those of the ARIMA, LSSVM, BND-LSSVM, FOA-LSSVM, and EMD-FOA-LSSVM models.

Ensemble Learning (EL) is gaining more and more interest in electricity price forecasting research. The authors of [206] propose a hybrid decomposition approach that incorporates ensemble empirical modal decomposition (EEMD) and wavelet packet decomposition (WPD) within a quadratic framework, coupled with a deep extreme learning machine (DELM) optimized using an improved hunter-prey optimization algorithm (TPHO). The combination of these techniques improves the precision of electricity price forecasting. Researchers in [207] propose a short-term electricity price prediction methodology that utilizes days with similar load consumption among the historical data, and the Ensemble Empirical Mode Decomposition (EEMD) technique. A Gated Recurrent Unit Neural Network (GRU-NN) is used as the machine learning model and is combined with a feature selection model. The proposed method outperforms existing techniques in terms of accuracy.

Yang et al. in [208] research and create a data-driven ensemble DL model (GHTnet) to capture the temporal variation of real-time electricity price values. For this reason, an advanced CNN module based on inception architecture is developed, which is called GoogLeNet. The advantage of this model is the capability of capturing the high-frequency features of data used, while the inclusion of time series summary statistics has been shown to enhance the forecasting of volatile price spikes. Datasets from 49 generators from the NYISO are used for the evaluation of the performance of the models. The final results show that the psoposed model achieves significant improvements over that of state-of-the-art benchmark methods, improving the MAPE by 17.34% approximately.

Also, Jahangir et al. in [209] propose a novel method for EPF by using Grey correlation analysis to select important parameters, denoising data with deep neural networks, implementing dimension reduction, and utilizing rough ANNs. Additionally, Khajeh et al. in [210] focuse on using ANNs with an improved iterative training algorithm for EPF in such markets. This work introduces a novel forecasting approach for short-term electricity price prediction, addressing the inherent complexity, nonlinearity, volatility, and time-dependent nature of electricity prices. The proposed strategy incorporates two key innovations: a two-stage feature selection algorithm and an iterative training algorithm, both designed to enhance the accuracy and reliability of the proposed model.

Following the above, Zhang et al. in [211] introduce a novel hybrid model based on wavelet transform (WT), ARIMA and least squares SVM (LSSVM). The hyperparameters of the model are optimized by PSO algorithm, making it capable of achieving high performance in EPF tasks. The proposed method is tested using data from New South Wales (NSW) in Australia's national electricity market. The result show that this method gives more accurate and effective results compared to other EPF methods.

Finally, Jaime et al. in [212] propose a hybrid EPF algorithm that utilizes a two-dimensional approach. In the vertical dimension, the final forecasts were generated by combining the outputs of XGBoost and LSTM models in a stacked array. In the horizontal dimension, the model leveraged available market data at the time of forecasting to create price predictions ranging from 1 hour to 96 hours ahead. Additionally, a new input feature, derived from the slope of thermal power generation output, significantly improved the accuracy of the model.

6.1.4 Aggregated Studies for Electricity Price Forecasting

In order to present the most of the studies analyzed above in a consolidated manner, Table 6.1 follows.

Table 6.1: Aggregated Table with resent publications for price forecasting.

Ref.	Year	Models	Metrics	Dataset Origin
[157]	2015	DNN, ARIMA	MAPE, MAD, MFE	Czech electricity spot prices
[158]	2013	ARMA, GARCH	RMSE, MAPE	Electricity prices in the electricity market of New England
[159]	2021	SARIMA, GARCH	MAPE, RMSE	Indian Energy Exchange (IEX)
[160]	2019	ARIMA, Holt-Winters, MLR	MAPE	Iberian electricity market
[166]	2021	ARIMA, SARIMA, SARIMAX	MAE, RMSE	EEX (European Energy Exchange)
[167]	2021	Exponential Smoothing	MSE	Ontario, Canada
[168]	2020	ARIMA, ARIMAX	MAPE	Belgorod region in European part of Russian coincides
[169]	2024	ARIMA, MA, LSSVM	MAPE	Malaysia, the Energy Commission
[172]	2012	NAIVE, SARMA, SARMAX, LSTR	MAPE	APX power exchange for Great Britain
[182]	2024	Generative pre-trained transformers (GPT)	MCC	Electricity Spanish market

Continued on next page...

Table 6.1 – continued from previous page.

Ref.	Year	Models	Metrics	Dataset Origin
[196]	2024	CNN Attention, Multi-Head Attention, CNN-GRU Attention, Stacking Ensemble, Soft Voting Ensemble, Perceptron	MSE, RMSE, MAPE	HEDNO (Hellenic Electricity Distribution Network Operator)
[170]	2024	GARCH, MS-GARCH	Coefficient, Standard Error, t-Value, p-Value	Portugal's energy market
[195]	2024	MLP, LSTM, SVM, ARIMA, GA, PLD	MSE	National electric system of Brazil
[197]	2023	XGBoost and Decomposition Techniques	MAPE, RMSE, MDE	Korean power system
[198]	2023	LSTM, CNN-LSTM, ARIMA, Random Forests, Gradient Boosted Trees, Naive, SVM, KNN, Decision Trees	MAE, RMSE	German spot market
[199]	2022	XGBOOST, Market Curves, XGBOOST-Market Curves	MAE, WMAE	OMIE (Electricity Market Operator, Spain)

Continued on next page...

Table 6.1 – continued from previous page.

Ref.	Year	Models	Metrics	Dataset Origin
[203]	2024	ANN, SVR, MLR, PSO, DO, GRO	MAE, MSE, MAPE, RMSE	Turkish Statistical Institute (TSI), World Bank Open Database
[202]	2022	CRITIC-GRAMPA-RELM, MPA-RELM, GA-ELM, GA-SVM, ELM, SVM	SSE, MSE, RMSE, MAE	Shanxi spot market
[171]	2022	NNAR, ARIMA	ME, RMSE, MAE, MPE, MAPE	Hungarian Power Exchang
[204]	2021	LASO, ANN, RANDOM FOREST, RVM, XGBOOST-RVM	MAAPE, RMSE, MAE	ENTSO-E Transparency Platform
[208]	2022	ARIMA, Lasso, CNN, ANN, GRU, LSTNet, HFnet, Uber, W-GRU, W-HFnet, GHTnet (5,9)	MAPE, MAE	New York Independent System Operator (NYISO)
[209]	2019	DR-R-SDAE, LSTM, R-DAE, SDAE, CNN, SVM	RMSE, MAE, MAPE	Ontario, Canada

Continued on next page...

Table 6.1 – continued from previous page.

Ref.	Year	Models	Metrics	Dataset Origin
[210]	2017	RNN, PCA with CNN, MLP with LM, CA with CNN, SVC-RFE with CNN	MAPE	Pennsylvania–New Jersey–Maryland (PJM) electricity markets
[211]	2012	ANN, ANN-ARIMA	MAPE, MAE, RMSE	Australian national electricity market
[212]	2023	DTR, LSTM, XGBOOST, Hybrid	MAE, RMSE	Several market players in the Alberta electricity market
[161]	2018	Several AR models	MAE	Nord Pool (NP; Northern Europe), PJM Interconnection (Northeastern United States), EPEX Germany and Austria (Central Europe)
[205]	2017	f LSSVM, BND-LSSVM, FOA-LSSVM	MAPE, MAE, RMSE	The daily electricity price in the northeast power grid of China
[162]	2016	several types of ARIMA	FMSE	Spanish electricity market
[200]	2024	XAI CNN, LSTM, XG-BOOST	RMSE, MAE, MAPE	Italian Electricity Market, ERCOT Electricity Market
[201]	2023	CNN-based models	MAE	Italian Electricity Market, ERCOT Electricity Market

6.2 Data-Driven Techniques for Short-term Electricity Price Forecasting Through Novel Deep Learning Approaches with Attention Mechanisms

The electricity market is constantly evolving, driven by factors such as market liberalization, increasing use of Renewable Energy Sources (RES), and various economic and political influences. These dynamics make it challenging to predict wholesale electricity prices. Accurate short-term forecasting is crucial to maintaining system balance and addressing anomalies such as negative prices and deviations from predictions. This research investigates short-term electricity price forecasting using historical timeseries data and employed advanced deep learning algorithms. First, four deep learning models are implemented and proposed, which are: A Convolutional Neural Network (CNN) with an integrated attention mechanism, a hybrid CNN followed by a Gated Recurrent Unit model (CNN-GRU) with attention mechanism, and two Ensemble Learning models, which are a Soft Voting Ensemble and a Stacking Ensemble model. Also, the optimized version of a Transformer model, the Multi-Head Attention model is introduced. Finally, the Perceptron model is used as a benchmark for comparison. Our results show excellent prediction accuracy, particularly in the hybrid CNN-GRU model with attention, achieving a mean absolute percentage error (MAPE) of 6.333%. The soft voting ensemble model and the Multi-Head Attention model also perform well, with MAPEs of 6.125% and 6.889%, respectively. These findings are significant as previous studies have not shown high performance with Transformer models and attention mechanisms. The presented results offer promising insights for future research in this field.

6.3 Introduction to the Electricity Market and the Contribution of this work

In the current age characterized by our increasing dependence on electricity for various aspects of daily life, the precision of electricity price forecasting (EPF) has emerged as a matter of paramount importance. With the ever-growing demand for electricity spurred by factors such as population growth, industrial expansion, and the widespread adoption of electric vehicles, the need for precise and reliable electricity price predictions is more pronounced than ever. This research work embarks on an exploration of the critical role that accurate electricity price forecasting plays, delving into the multifaceted factors and methodologies that impact this pivotal aspect of the energy sector. A thorough understanding of electricity price trends is not only crucial for stakeholders within the energy market, but also has significant relevance for policy makers, consumers and advocates of environmental sustainability. This understanding directly influences financial decision-making, cost management, sustainability initiatives, and the broader economic welfare of societies in today's globally interconnected world. This research seeks to illuminate the inherent challenges and the increasing significance of electricity price forecasting in the context of our contemporary energy landscape.

The significance and advancement of short-term electricity price prediction have garnered significant attention in recent years. The introduction of innovative market methodologies, including intra-day timeframes, has further accentuated the demand for precise forecasts. This, in turn, enables market stakeholders to fine-tune their strategies and enhance the efficiency of their energy trading activities within shorter time intervals. These forecasts are of essential importance to market participants, allowing them to competently oversee their assets, mitigate risks, and capitalize on opportunities arising from the dynamic nature of market conditions.

Also, accurate price predictions in the electricity sector are imperative for various critical facets. For utilities, these predictions are fundamental in resource planning, allowing them to efficiently allocate resources to meet anticipated demand. Additionally, in energy trading, precise forecasts play a crucial role in guiding decisions to buy and sell electricity among market participants. The stability of the grid is crucially influenced by accurate predictions, helping operators anticipate and manage sudden fluctuations in demand or supply [213]. In addition, for the integration of renewable energy sources, such predictions aid in balancing electricity supply and demand, ensuring grid stability in the face of intermittent energy generation [214] . Cost-saving measures for consumers, achieved by shifting usage to lower-priced periods, are facilitated by such forecasts.

Moreover, these forecasts are relevant for investment decisions, particularly for investors in energy projects who rely on price expectations to assess profitability. In particular, environmental impact is also influenced by accurate price predictions, prompting energy companies to opt for more eco-friendly energy generation methods during high-price periods. In conclusion, the precision of electricity price predictions profoundly shapes efficient and informed decision making across resource management, trading, grid stability, renewable energy integration, investment planning, and environmental considerations. Forecasts are integral in promoting effective electricity system management and influencing consumer behaviors. Several timeseries methods have been used to predict electricity prices, which will be presented below. In contrast to existing methods that aim to solve the electricity forecasting problem and are presented in the literature review, the methods proposed in this work employ complex attention mechanisms, making the models proposed capable of assimilating extremely difficult fluctuations in price timeseries, capturing both the positive and negative values of the data.

Regarding the contribution of this work, the following key points are highlighted:

1. Novel and advanced deep learning models are developed and proposed for electricity price forecasting approaches using timeseries data. The first is the Convolutional Neural Network (CNN) with an integrated attention mechanism. The second is the hybrid CNN followed by a Gated Recurrent Unit (CNN-GRU) with integrated attention mechanism, the third is the Soft Voting Ensemble model, which is an Ensemble Learning model that combines the predictions of the CNN-GRU with attention mechanism model and a Transformer model, the Multi-Head Attention model, creating a weighted output, and the fourth is a Stacking Ensemble model, which combines the outputs of the CNN-GRU with attention and the Transformer model in a Dense Layer, creating the final prediction. These architectures are robust in predicting data that does not follow specific patterns, such as those under study. More specifically, the proposed Soft Voting Ensemble model exhibits optimal predictions. These four models not previously utilized in electricity price prediction instill optimism for future studies based on their performance.

2. Also, an enhanced and optimized version of the Multi-Head Attention Model is introduced in such a way as to exhibit very satisfactory predictive accuracy, compared to other works, where the Transformers do not perform as well. This model demonstrates exceptionally high prediction accuracy within the electricity price framework.

3. Explainability Process: The research contributes to the field by employing an Explainability process, elucidating the impact of variables introduced into the models. This analysis provides insights into the factors that influence prediction outcomes, aids in

understanding model behavior, and improves the interpretability of results.

4. Resolution of Sign-Change Challenges: The research addresses a common challenge in existing algorithms, namely their inability to capture abrupt sign changes in prediction time series. Considering the correct parameters and conducting thorough preprocessing analysis, the proposed models demonstrate improved performance in handling such variations.

In order to present a comprehensive view of the work carried out, this research is structured as follows. Section 6.4 presents a general description of the mechanisms of the electricity market. Section 6.5 clarifies the fundamental architectures of the models being developed. In Section 6.6, an in-depth analysis of all the data and correlations under examination is carried out. In Section 6.7, the results of the predictions are presented and the respective metrics are compared with each other. Finally, in Section 6.8, the key conclusions arising from the study are outlined, along with notable suggestions for future research.

6.4 The electricity market

In order to make the electricity market more understandable, this section introduces both the mechanism and the three main markets that constitute the wholesale energy market: the day-ahead, intra-day, and the balance market.

The electricity market [215] is a complex system that brings together electricity generators, distributors, retailers and consumers to facilitate the production and consumption of electricity. The working mechanism of the electricity market can be broadly categorized into two parts: the wholesale market and the retail market.

In the wholesale market, electricity is traded between generators and retailers, who submit bids and offers for the supply of electricity. The market clearing price is then determined by matching the bids and offers, and the generators that offer to supply electricity at or below the clearing price are dispatched to supply the required amount of electricity. The wholesale market operates in advance, which means that the market clearing price is set one day in advance.

In the retail market, retailers purchase electricity from the wholesale market and sell it to consumers at retail prices. Consumers can purchase electricity from their retailer or generate their own electricity through distributed generation technologies, such as rooftop solar panels. Retailers also provide various value-added services to consumers, such as billing, customer service, and energy efficiency advice. The electricity market is regulated by various government agencies to ensure that it operates fairly and efficiently. Regulatory bodies oversee market design, market operations, and market participants to ensure that they comply with rules and regulations.

The electricity Day Ahead Market (DAM) is a crucial component of the electricity market structure, where electricity is bought and sold a day before actual delivery. It functions as a platform for electricity producers and consumers, enabling them to trade electricity for delivery on the following day. Participants, including power generators, distributors, and large consumers, submit bids specifying the quantity of electricity they are willing to buy or sell at various price levels for each hour of the following day. The market operator then matches the bids to determine the clearing price for each hour, considering the supply and demand conditions. This transparent and competitive market allows market participants to manage their electricity needs, balance supply and demand, and mitigate price volatility while ensuring efficient and reliable power delivery.

The Intra-Day Market (IDM) is like a special marketplace for electricity that operates during the day. It is where electricity is bought and sold for very short periods, like hours or even minutes, before it is actually used. This helps electricity companies and users adjust their plans and make sure that they have enough electricity to meet the needs of the moment. It is a bit like a last-minute electricity store, where you can quickly get what you need to keep the lights on and machines running smoothly during the day.

The Balancing Market involves the continuous matching of electricity supply with demand in real time to ensure grid stability. It is a crucial process that addresses inherent variability and unpredictability in power consumption and generation. Market operators use advanced technologies and forecasting tools to anticipate fluctuations in demand and supply, adjusting electricity generation or consumption accordingly. This involves dispatching additional power when demand exceeds forecasts or activating demand response mechanisms to curtail consumption during surplus periods. Balancing the electricity market is essential to maintain a reliable and resilient power system, to prevent overloads or blackouts, and to optimize the use of diverse energy sources in the most efficient manner.

6.5 Description of the models

In this section, the architecture of the deep learning models developed and explored is presented in detail. Initially, an introduction to the application of Transformers in timeseries forecasting was made. After, the four novel models proposed are analyzed, namely the CNN with attention mechanism, the Hybrid CNN-GRU with attention mechanism, and the two Ensemble Learning models, the Soft Voting Ensemble and the Stacking Ensemble model. Following that, the Optimized Version of the Multi-head Attention model is presented, and finally, the benchmark Perceptron model is analyzed.

6.5.1 Transformers for Time Series Forecasting

Time series forecasting aims to predict future values based on historical data. Classical models, such as Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory Networks (LSTMs), have dominated this field. However, these models struggle with long-term dependencies, irregular patterns, and computational inefficiencies.

The Transformer model, originally introduced by Vaswani et al. in [216], have demonstrated exceptional performance in natural language processing (NLP) tasks. Its ability to capture long-range dependencies using self-attention mechanisms has made it an attractive choice for time series forecasting. Below, we delve into the architecture of Transformers, their mathematical foundations, and their application in forecasting tasks.

6.5.2 Transformer Architecture

A standard Transformer consists of an encoder-decoder structure, where each component is composed of multi-head self-attention layers and feed-forward networks. In time series forecasting, variations such as Informer [217] and Temporal Fusion Transformers (TFT) [218] have been developed to enhance efficiency. More specifically, the main architecture of Transformer model consists of the following mechanisms:

Self-Attention Mechanism

The core mechanism of Transformers is self-attention, which allows the model to focus on different parts of the input sequence dynamically. Given an input sequence $X \in \mathbb{R}^{T \times d}$, where

T is the number of time steps and d is the feature dimension, the self-attention operation is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6.1}$$

where: - $Q = XW^Q$, $K = XW^K$, $V = XW^V$ are the query, key, and value matrices, - W^Q, W^K, W^V are learnable parameter matrices, - d_k is the dimension of the key vectors.

The scaled dot-product attention allows the model to weigh the importance of different time steps dynamically.

Multi-Head Attention

Instead of using a single attention mechanism, Transformers employ multiple attention heads to capture diverse patterns:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, ..., \text{head}_h)W^O \tag{6.2}$$

where:

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{6.3}$$

Positional Encoding

Unlike recurrent models, Transformers lack an inherent notion of sequence order. To incorporate temporal information, a positional encoding function is used:

$$PE(t, 2i) = \sin(t/10000^{2i/d}) \tag{6.4}$$

$$PE(t, 2i+1) = \cos(t/10000^{2i/d}) \tag{6.5}$$

where t represents the position and i is the dimension index.

6.5.3 CNN with attention mechanism model

A CNN with an attention mechanism [219] synergistically integrates convolutional neural networks (CNNs) and attention mechanisms to improve the model's proficiency in extracting crucial features from 2-D images. In addition, the CNN with Attention Mechanism model is designed to capture complex patterns in time series data, particularly focusing on electricity price prediction. The attention mechanism enhances the model's ability to highlight relevant features. The key components of the model are described below:

Attention Mechanism

The attention mechanism computes attention scores (α) for each element in the input sequence, providing a weighted sum for context representation. For the input sequence X and the attention weights α, the context vector C is calculated as [220]:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{T} \exp(e_j)} \tag{6.6}$$

$$C = \sum_{i=1}^{T} \alpha_i \cdot X_i \tag{6.7}$$

where the weights α_{ij} are used to obtain the final weighted value.

The model utilizes the context vector in conjunction with the output of a Convolutional Neural Network (CNN). The attended feature map ($\hat{Z}$) is obtained by convolving the attention-weighted context vector with the CNN output:

$$\hat{Z}_{i,j} = \sum_{m=1}^{H} \sum_{n=1}^{W} \alpha_{m,n} \cdot Z_{i+m,j+n} \tag{6.8}$$

6.5.4 Hybrid CNN-GRU with attention mechanism model

The Hybrid CNN-GRU with Attention Mechanism model represents an innovative approach for time series prediction, particularly in the context of electricity price forecasting. Combining CNN, GRU, and Attention mechanisms, this model exhibits a robust ability to capture intricate patterns in temporal data. Attention Mechanism The attention mechanism computes attention scores (α) to emphasize relevant elements in the input sequence. For input X and attention weights α, the context vector C is calculated as [220]:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{T} \exp(e_j)} \tag{6.9}$$

$$C = \sum_{i=1}^{T} \alpha_i \cdot X_i \tag{6.10}$$

The Hybrid CNN-GRU model integrates the attention-weighted context vector with the output of both a CNN and a GRU. The attended feature map ($\hat{Z}$) is obtained through:

$$\hat{Z}_{i,j} = \sum_{m=1}^{H} \sum_{n=1}^{W} \alpha_{m,n} \cdot Z_{i+m,j+n} \tag{6.11}$$

6.5.5 Soft Voting Ensemble Model

The Soft Voting Ensemble Model, as shown in Figure 6.1, is a sophisticated ensemble learning approach that combines the predictive outputs of multiple base models, aiming to enhance overall predictive accuracy. In this context, it integrates the results of the Multi-head Attention model and the Hybrid CNN-GRU with Attention model using a soft voting mechanism.

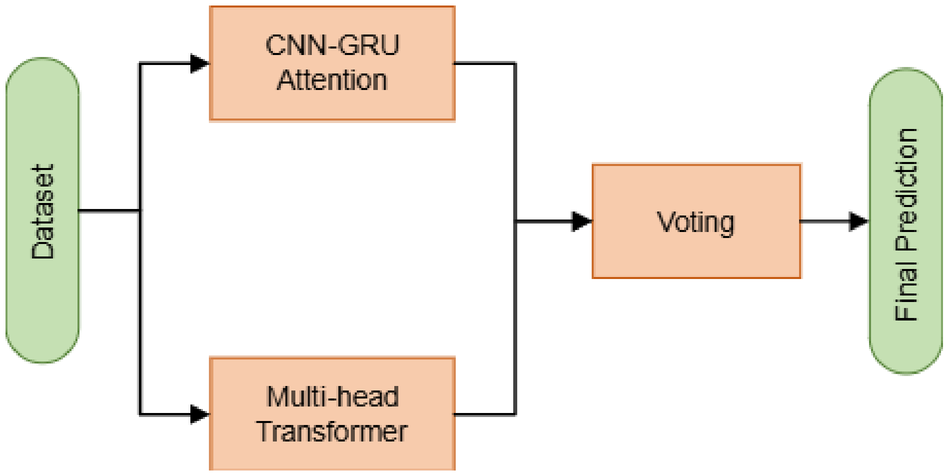


Figure 6.1: Soft voting ensemble model

Multi-Head Attention Model

The Multi-Head Attention model is adept at capturing intricate patterns and dependencies within time-series data. Utilizing parallel attention heads, each focused on different aspects of the input sequence, it produces diverse and informative representations of the data.

Hybrid CNN-GRU with Attention Model

The Hybrid CNN-GRU with Attention model seamlessly integrates the Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU), and Attention mechanisms. This combination allows for effective feature extraction and context-aware processing, contributing to the model's ability to capture both spatial and temporal dependencies.

Soft Voting Mechanism

The Soft Voting Mechanism combines the outputs of the Multi-head Attention and Hybrid CNN-GRU with Attention models by assigning weights (α_i) to each model's output. The Softmax function is then applied to the weighted sum to obtain the final ensemble prediction as presented in equation 6.12.

$$\text{Soft Voting Ensemble Output} = \text{Softmax}\left(\sum_{i=1}^{N} \alpha_i \cdot \text{Output}_i\right) \tag{6.12}$$

here, N is the number of base models (2 in this case), α_i represents the weight or confidence assigned to each base model's output, and Output_i is the output of the i-th base model.

6.5.6 Stacking Ensemble Model

The Stacking Ensemble Model, which is presented in Figure 6.2 is an advanced ensemble learning technique designed to enhance predictive performance by combining the outputs of multiple base models. In this configuration, it integrates the predictive capabilities of the Multi-Head Attention model and the Hybrid CNN-GRU with Attention model, creating a powerful and accurate ensemble. combines the predictions of the Multi-head Attention and Hybrid CNN-GRU with Attention models in a final Dense layer. The outputs of both models serve as input features for the Dense layer, which learns to weigh and combine these features to produce the ensemble prediction.

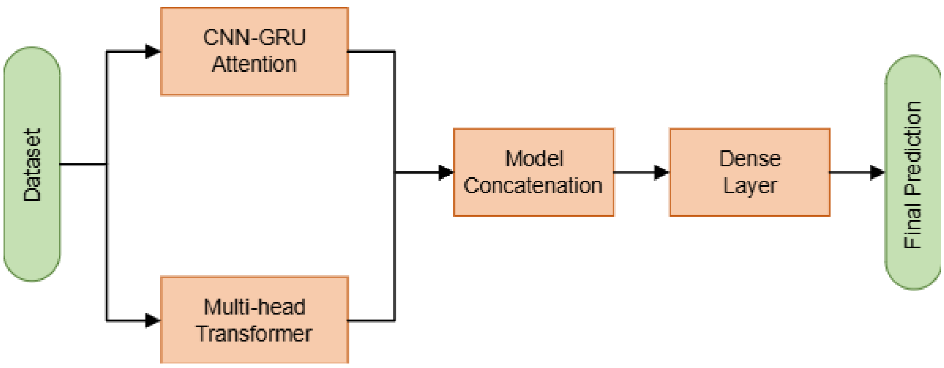


Figure 6.2: Stacking ensemble model

$$Ensemble\ Output = \text{Dense}\left(\left[\text{Output}_{\text{Multi-head Attention}}, \text{Output}_{\text{Hybrid CNN-GRU}}\right]\right) \tag{6.13}$$

6.5.7 Optimized Multi-Head Attention model

The optimized version of the Multi-Head Attention model [221] is an extension of the original Transformer architecture using the Optuna optimization algorithm, allowing the model to jointly attend to different positions or patterns in the input sequence effectively. Using multiple attention heads, it captures diverse information simultaneously. Below are the key components and mathematical equations for a Multihead Transformer: The Advanced Transformer model leverages the foundational architecture of the Transformer while incorporating advanced parameterizations for enhanced adaptability. The scaled dot-product attention mechanism is defined as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \cdot V \tag{6.14}$$

For multihead attention, the model concatenates individual attention heads, each computed as:

$$\text{Multihead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W_O \tag{6.15}$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{6.16}$$

where Q, K, V are the Query, the Key and the Value matrices, W_i^Q, W_i^K, W_i^V are learned weight matrices specific to each head and W^O is a learnable weight matrix for the output.

The position-wise feedforward (FFN) networks are given by:

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \tag{6.17}$$

where W_1, W_2, b_1, and b_2 are learnable weight matrices and biases.

Layer normalization and residual connections are applied through the following:

$$\text{LayerNorm}(x + \text{Sublayer}(x)) \tag{6.18}$$

As mentioned above, this model is introduced as optimized, using the Optuna hyperparameter optimization framework, which is further elaborated in subsection 4.7. The optimization parameters are the Dimension of the model, the Number of attention heads, and the Number of transformer layers, and the goal is to find the appropriate combination of these parameters that can minimize the optimization function of the specific optimization algorithm, which is the validation loss of Mean Squared Error.

6.5.8 Benchmark Perceptron model

The Benchmark Perceptron model constitutes a fundamental, yet powerful architecture commonly used as a baseline in various machine learning tasks. This model typically involves a single Dense layer, also known as a fully connected layer, which serves as the primary component for learning and making predictions. The Dense layer accepts input data, processes it through a set of interconnected neurons, each connected to every neuron in the subsequent layer, and outputs the final predictions or classifications.

The Perceptron model is a foundational neural network architecture, forming the basis for more complex neural networks. It consists of a single layer with input nodes, each associated with a weight, and an activation function that determines the output. Mathematically, the output ($\hat{y}$) of the Perceptron for input features $x_1, x_2, \ldots, x_n$ and weights $w_1, w_2, \ldots, w_n$ is calculated as:

$$\hat{y} = \text{activation}\left(\sum_{i=1}^{n} x_i \cdot w_i + b\right) \tag{6.19}$$

where b is the bias term.

Completing the description of the architectures of all models, a brief review of both the Explainability Process Technique and the Optuna Optimization Framework, on which the optimization of each model was conducted, is presented.

6.5.9 Explainability Process

In order to enhance interpretability, the current work introduces an explainability process, shedding light on the impact of variables within the models. The specific equations for this process may vary based on the chosen explainability technique.

The integration of explainability in electricity price forecasting holds paramount significance, as it enhances the transparency and interpretability of predictive models. In an era dominated by complex machine learning algorithms and sophisticated models, understanding the rationale behind price predictions becomes imperative for stakeholders such as energy traders, grid operators, and policymakers. Explainability not only fosters trust in forecasting models but also facilitates the identification and mitigation of potential biases or inaccuracies. By shedding light on the intricate relationships between input variables and predicted outcomes, explainable models empower industry professionals to navigate the intricate terrain of electricity markets with greater confidence and precision, ultimately contributing to more robust and reliable forecasts.

For this reason, SHAP (SHapley Additive exPlanations) [222] is utilized. SHAP operates by decomposing the model output into sums of individual feature impacts. It computes values representing each feature's contribution to the overall model outcome, enabling a nuanced understanding of feature importance. However, time series data's temporal dependencies necessitate additional preprocessing or the use of models that inherently capture sequential patterns. SHAP values, which indicate the impact of features on predictions, can be interpreted through visualization, with positive and negative values denoting positive and negative contributions, respectively. It is crucial to consider dynamic features in the analysis.

6.5.10 Optimization Framework

Optuna [223] is a Python framework that automates hyperparameter optimization, making the task of finding optimal hyperparameter values to achieve superior model performance more convenient. It employs a trial-and-error methodology to autonomously explore and pinpoint the best hyperparameter configurations. This algorithm uses a historical record of previous trials to guide its decision-making regarding which hyperparameter values to investigate next. These data inform its selection of promising areas within the hyperparameter search space, guiding subsequent trials in those regions. As new results become available, Optuna refines its identification of even more promising regions. This iterative process is based on the knowledge accumulated from previous trials. At the core of this optimization process is the utilization of a Bayesian optimization algorithm called the Tree-structured Parzen Estimator.

6.6 Exploratory Data Analysis and Features Creation

In this section, the behavior of time series of electricity prices over time is presented and analyzed in detail. Initially, reference was made to the selection of the dataset, the separation of the entire data set into training and test sets. Subsequently, the variability of the time series is extensively presented, including the seasonal patterns it exhibits, and its overall behavior. This analysis initially includes the visualization and statistical features of the price timeseries, followed by the analysis of all variables under study and the correlations they exhibit with the price.

6.6.1 Dataset selection

For this study, the German electricity market is selected for three main reasons. Germany has more than 83 million residents and borders nine European countries; Denmark to the north; the Netherlands, Belgium, Luxembourg, and France to the west; Switzerland and Austria to the south; and the Czech Republic and Poland to the east. Additionally, in Germany is presented a diversity of energy sources used at the power plants for electricity production from conventional energy sources like lignite, coal, gases, and nuclear power to renewable sources like wind turbines onshore and offshore, photovoltaics, and sewage gases. The selected period to retrieve these data was from 1st of October 2018 to 30th of August 2023.

The actual electricity consumption, electricity generation from all energy plants, and energy price data were retrieved from the Bundesnetzagentur's electricity market information platform SMARD [224]. The resolution of the data is one time step per hour and does not contain missing values.

Furthermore, the climate condition for the under investigation period is retrieved from NASA's Modern Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) [142]. The parameters retrieved from the M2I1NXASM database are air temperature and air-specific humidity at 2-meters in units K and kg/kg, respectively. The resolution of the data is one time step per hour. The dataset presents missing values that are filled with mean value of the same timestep of other years in the dataset.

The final dataset is composed of the primary time series of sources mentioned above which are processed and combined. Regarding data partitioning, the temporal period from October 2018 to February 2023 is selected as the training data set for the problem studied. The next six months of the data set are then designated as the testing set.

6.6.2 Electricity price variation

Initially, a crucial consideration involves evaluating the time series for stable seasonality. To achieve this, Figure 6.3 illustrates both the price variations from 2018 to 2023 and their corresponding weekly average prices. The figure depicted in the following chart confirms substantial fluctuations in electricity price volatility throughout the observed period.

More specifically, it is observed that until November 2021, both the price and the average price exhibit stability, without extreme values. From November onward, a significant increase is observed in both time series, with the average hourly price reaching 600€/MWh and the

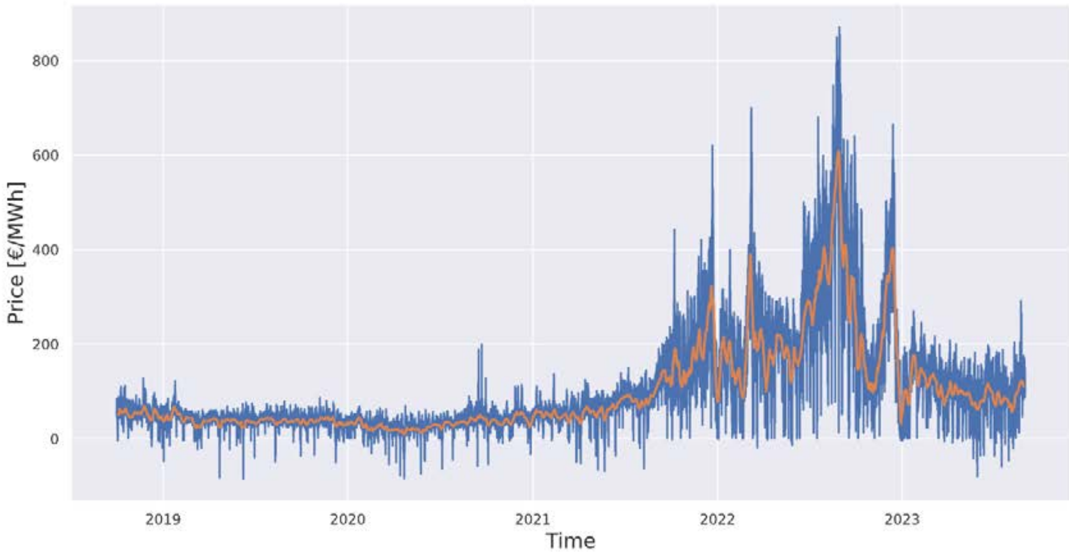


Figure 6.3: Electricity price variations from 2018 to 2023

corresponding mean price reaching 40€/MWh. Subsequently, from the beginning of 2023 onward, the price variation appears to stabilize, and extreme values are eliminated. These extreme fluctuations create imbalances in the electricity market, making price prediction in wholesale trading an exceptionally challenging process. Political-economic factors caused imbalances in wholesaler transactions, resulting in the market losing its stability with respect to price fluctuations. Therefore, it becomes evident that the accurate prediction of the hourly price is a difficult task, but simultaneously a significant challenge for deep learning models.

In contrast, Table 6.2 illustrates the annual electricity price statistics. The electricity price data spanning from 2018 to 2022 are utilized as training data, January and February 2023 serving as validation data, and the final six months of our sample designated for testing purposes. Upon examination of the data, notable patterns come to light. The standout year is 2022, marked by substantial shifts, particularly in maximum prices. A significant surge compared to preceding years renders it the most volatile year within our dataset.

Therefore, from Table 6.2, it is evident that the maximum electricity price increases exponentially between 2020 and 2022. Specifically, there is a 772% increase between 2020 and 2022 in mean electricity values, highlighting the exceptionally challenging nature of accurately predicting the electricity price. This fact gives special value to the problem, as it is a challenge.

Table 6.2: Descriptive statistics for electricity price.

Year	Number of samples	Mean value (€)	Standard deviation (€)	Minimum value (€)	Maximum value (€)	Percentile 25% (€)	Percentile 50% (€)	Percentile 75% (€)
2018	2208	52.606	18.59	-19.43	128.26	44.305	51.83	64.207
2019	8760	37.674	15.468	-85.04	121.46	31.057	38.06	46.270
2020	8784	30.469	17.499	-83.94	200.04	21.750	30.99	40.242
2021	8760	96.855	73.676	-69.00	620.00	53.007	75.48	112.032
2022	8760	235.446	142.808	-19.04	871.00	134.197	208.34	310.080
2023	5832	99.705	44.918	-80.69	291.93	80.090	100.515	126.067

6.6.3 Electricity Price Descriptive Analysis

In this subsection, a comprehensive analysis of the electricity price time series is presented, starting from its fluctuation at the quarter, monthly, and ending at the hour level for each average month.

Quarterly Density Variation

The regular hourly pattern persists consistently throughout each month throughout the course of the study. In Figure 6.4 the quarterly analysis reveals that Q1 and Q4 notably display the highest occurrences of electricity price outliers, with June emerging as the month characterized by the highest frequency.

Furthermore, in all study years, July, August, September and December consistently manifest the highest price levels. These observed trends can be attributed to diverse factors, including seasonal fluctuations in demand influenced by weather conditions and concurrent upswings in economic activity, among other contributing elements.

Monthly Price Variation

Figure 6.5 presents the violin plot of the average monthly price variation, taking into account all the years from 2018 to 2023. We observe that all months show negative electricity prices. While, August and September have the highest prices compared to the other months, January, February, April, May, October, and November show peaks which represent high frequency and probability to exhibit values nearly the median.

Daily Price Variation

Figure 6.6 presents the volatility of the average price over a period of 24 hours (1-24 hours). It is noted that peak demand occurs during the hours of 09:00 and 20:00 throughout the day.

Daily Per Month Price Variation

Figure 6.7 visualizes the average hourly price fluctuation of electrical energy for each of the months of the year. Specifically, during peak hours (specifically at 8 and 19 o'clock), the price rises during the daily peak. It should be noted that the fluctuation of the average hourly price exhibits common peak periods for each month of the year. Within the twenty-four-hour period, two peak intervals are observed. The first includes the period between 7-9 hours, and the second includes hours 18-20. Therefore, it is evident that despite the time series exhibiting abrupt and irregular changes from hour to hour, the average fluctuation follows a more stable behavior. The presentation of exogenous variables follows, along with their correlations with the time series of prices, to guide us in the final configuration of input data for model training.

6.6.4 Feature Engineering and Variables Selection

The strategy followed in the current work, regarding the data used as input variables in the created models, initially includes the collection of the relevant time series of data. Regarding the data collected, they included the following variables:

- Price [€/MWh]
- Actual Load
- Fossil Brown coal/Lignite
- Fossil Gas
- Fossil Hard coal
- Hydro Pumped Storage
- Hydro Run-of-river and poundage
- Hydro Water Reservoir

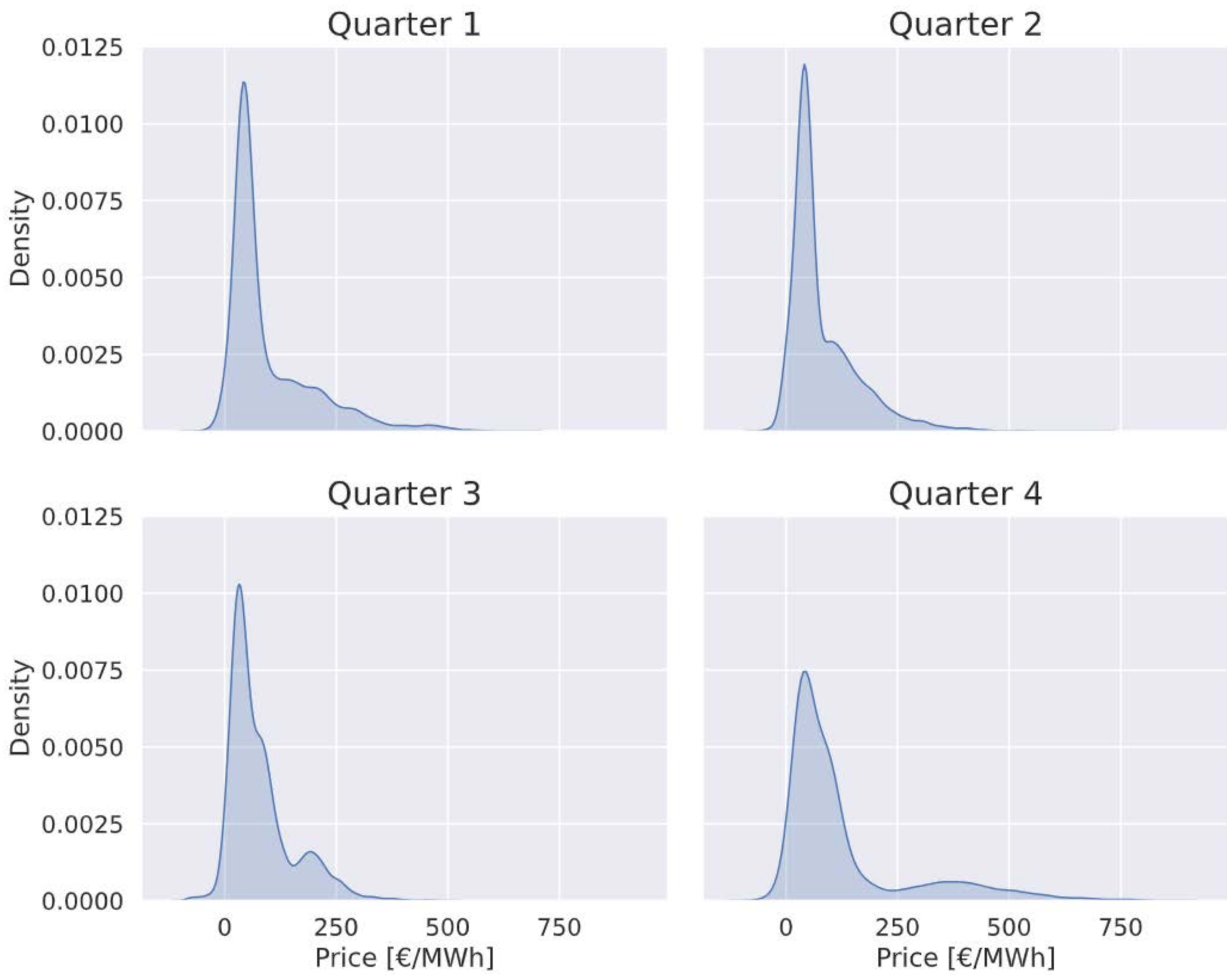


Figure 6.4: Quarterly Density Variation

- Nuclear
- Solar
- Wind Onshore
- temperature
- humidity

Then, we used the one-hot encoding technique [225] to create cyclical time variables, which can be used in short-term electricity price forecasting. This particular method enhances the ability of Deep Learning algorithms to capture the behavior of the time series at the weekly, daily, and hourly levels. Specifically, the following are detailed:

- day_of_week_sin

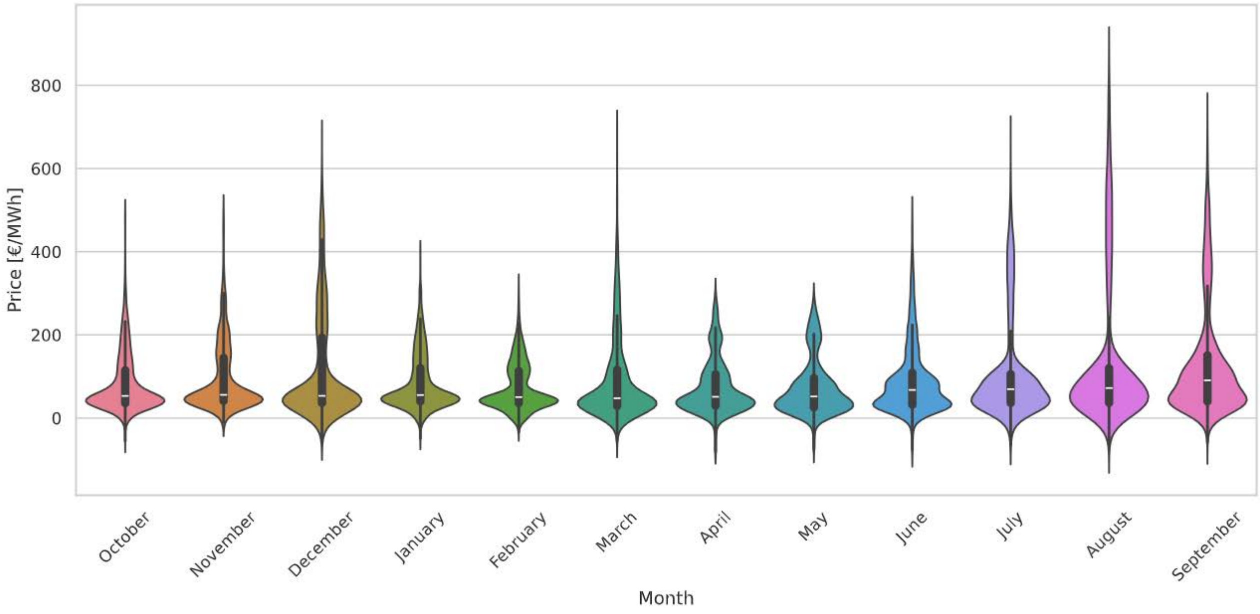


Figure 6.5: Violin plot of average monthly price variation for each month for all years

- day_of_week_cos
- Month of the Year Sin
- Month of the Year Cos
- IsWeekend
- hour_sin
- hour_cos
- quarter

Time series refer to systematically organized datasets arranged in chronological order and collected at regular intervals, such as hourly, daily, or weekly observations. The inherent structure of these sequential datasets introduces potential interconnections among response variables, which poses significant challenges for machine learning (ML) and deep learning (DL) algorithms, especially those commonly used in regression tasks. These challenges manifest mainly as issues of multi-collinearity and non-stationarity. In addressing the prediction problem at hand, we have a total of 21 variables available for integration into our models. Generally, short-term forecasting of electricity prices is a complex, multidimensional problem that requires in-depth analyses of how each variable influences the final price prediction.

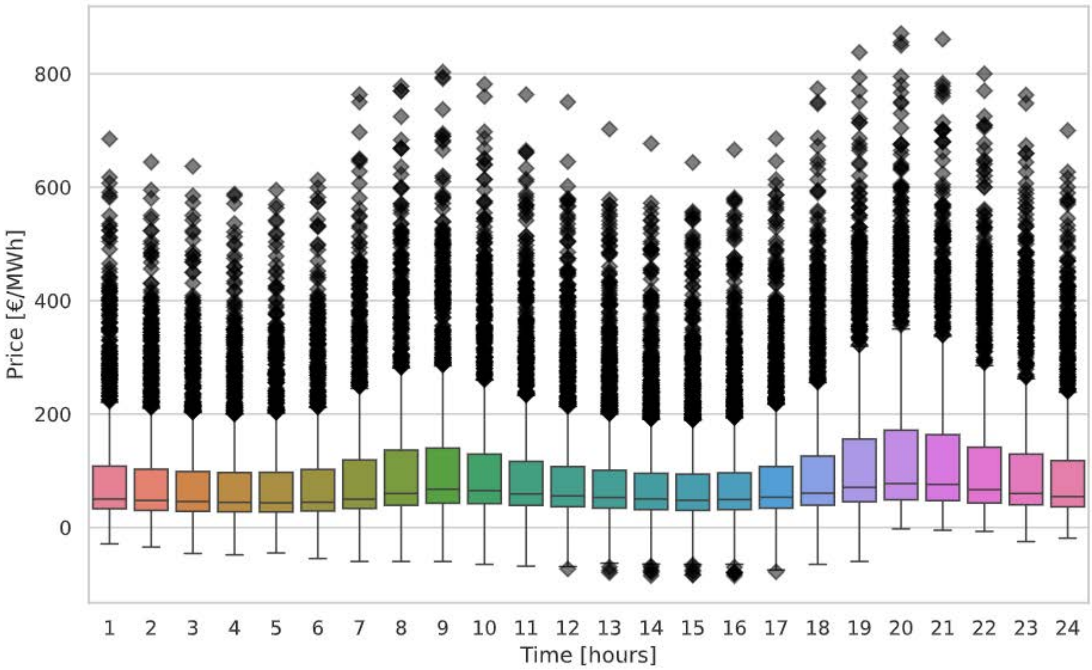


Figure 6.6: Boxplot of Hourly Price Variation Per Month

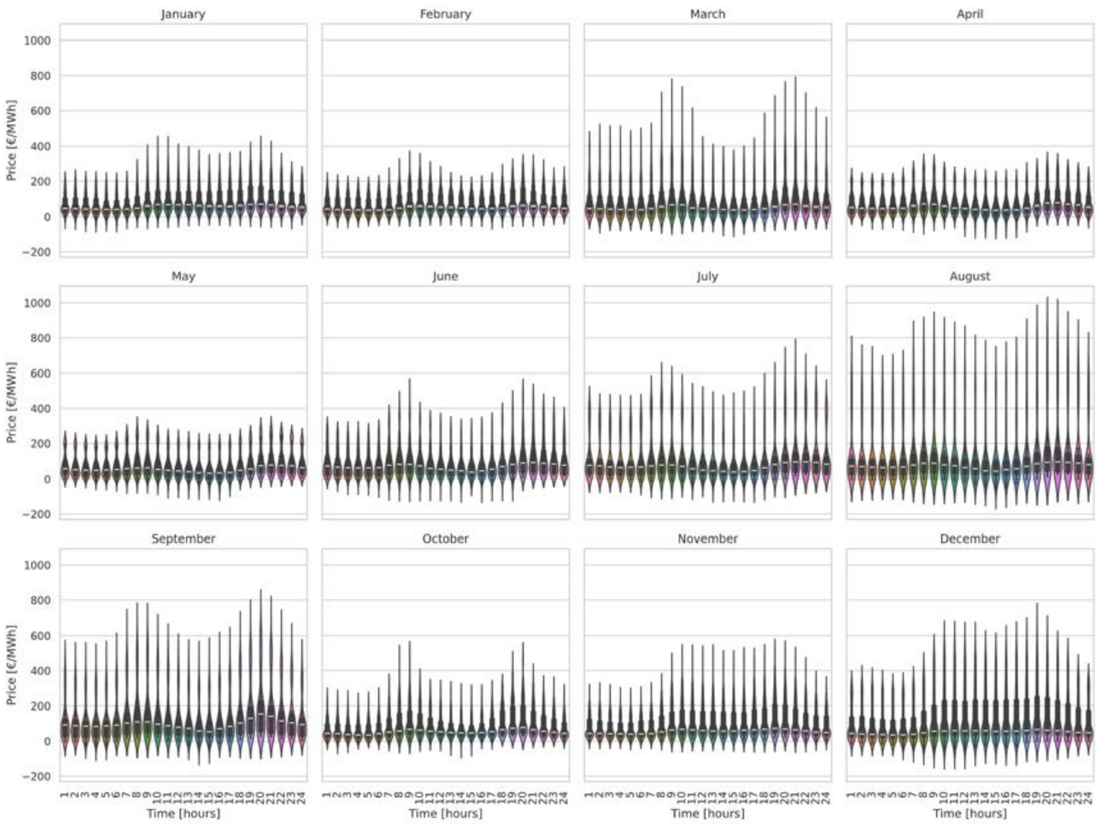


Figure 6.7: Boxplot of Daily Per Month Price Variation

Forecasting time series problems and tasks requiring the application of artificial intelligence models demand a comprehensive analysis of all correlations among variables to identify the most fitting ones for each specific problem.

To curate a dataset suitable for the forecasting models employed, the time series undergoes transformation into a reshaped dataset featuring nontime-dependent input features (X) and output variables (Y). The input features take the form of [samples, lags, features], while the output variables maintain a structure of [samples, horizon], aligning with the model architecture outlined in this work. This transformation involves implementing the sliding-window technique, particularly advantageous for predicting time series data. In this context, a 24-hour forecasting sliding window with electricity prices, as well as the exogenous variables, is selected for each forecasting hour. These choices result from a thorough analysis of time series tools and meticulous model tuning, aligning with the model's objective of providing short-term predictions for electricity prices. The application of this technique is visually represented.

6.6.5 Pearson's and Spearman's Correlation

In this subsection, the correlations among all relevant time series are presented with the aim of arriving at the optimal combination of input parameters for the algorithms. For this purpose, the general correlation heatmap and the Pearson and Spearman correlations are calculated and presented in Figures 6.8 and 6.9, respectively. In the realm of statistical analysis, the Pearson correlation constitutes a parametric metric utilized to evaluate the magnitude and direction of linear associations between two continuous variables. Based on the assumption of an approximate normal distribution of variables, Pearson's (r) computes the correlation through normalization of covariance by the product of standard deviations. In contrast, the Spearman rank correlation, denoted as Spearman's ρ, emerges as a nonparametric measure specifically designed to scrutinize the monotonic relationship between two variables. Not reliant on linearity assumptions, Spearman's correlation is apt for ordinal or ranked data and, crucially, exhibits heightened resistance to outliers in comparison to the Pearson correlation. Both metrics assume pivotal roles in systematically examining and comprehending the interrelationships among variables within various research domains.

Based on the features that were studied and the correlations they exhibit, numerous experiments were conducted to identify the optimal variable combinations, aiming to enhance the

Figure 6.8: Correlation Heatmap

models' performance in forecasting electricity prices. For this purpose, tests were performed with models that included both high-variability and single-variable models. Extensive experiments were carried out using historical data from the time series of prices in combination with cyclical temporal variables, as presented above. Concerning the high variability scenario, tests were performed both with independent or exogenous variables in combination with the mentioned temporal variables and without them. On the basis of both the following diagrams and the experimental approaches conducted, we arrived at the optimal combination of input parameters for each of our models as follows. Based on experimental approaches, we arrived at the optimal combination of exogenous variables, from which the most accurate prediction results were derived. It was observed that our models perform better in understanding the patterns of the electricity price time series when parameters are selected that

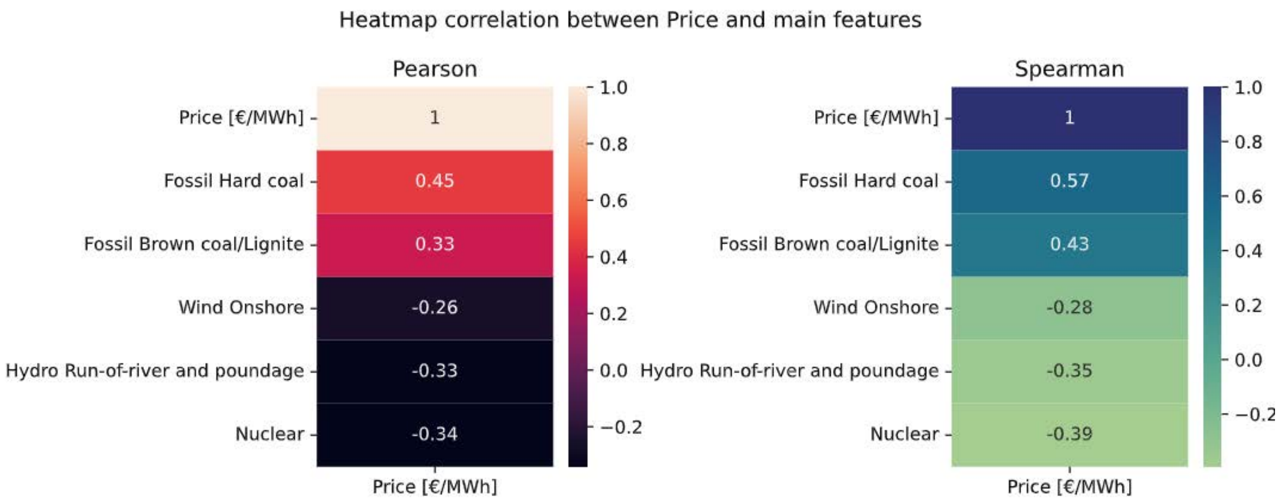


Figure 6.9: Pearson and Spearman's Correlation

not only exhibit high average correlation but also capture both negative and positive values of the price. Based on Figure 6.9, significant conclusions emerge regarding the formation of electricity prices in the Germany-Luxembourg interconnection and, more broadly, in the wholesale market.

The first reasonable conclusion drawn is that conventional energy sources exhibit a positive correlation, both in Pearson and Spearman analyses. Furthermore, renewable energy sources (RES) show negative correlations with price fluctuations. Remarkably, the variable 'Hydro Run of River and Poundage,' despite accounting for only 14% of the average production compared to 'Fossil Brown coal/Lignite,' demonstrates scientifically acceptable correlations of -0.33 and -0.35 with the price, underscoring its consideration in price forecasting.

Therefore, upon a more comprehensive examination, it is discerned that the price time series manifests substantial fluctuations in polarity, encompassing both notably negative and positive values. This intrinsic characteristic introduces complexity for a model to adapt to the time series, given the absence of stable patterns typically encountered in load forecasting. A thorough investigation into the origins of this phenomenon reveals that during periods characterized by extremely negative prices, there is a significant surge in RES production, particularly marked by a rapid increase in Wind Onshore generation. Simultaneously, conventional modes of production, such as 'Fossil Brown coal/Lignite' and 'Fossil Hard coal', witness substantial decreases in their output. Consequently, an excess of electrical energy occurs, surpassing demand during these intervals, leading to exceptionally negative price spikes. Figure 6.10 analyzes and confirms the above interactions between the examined features.

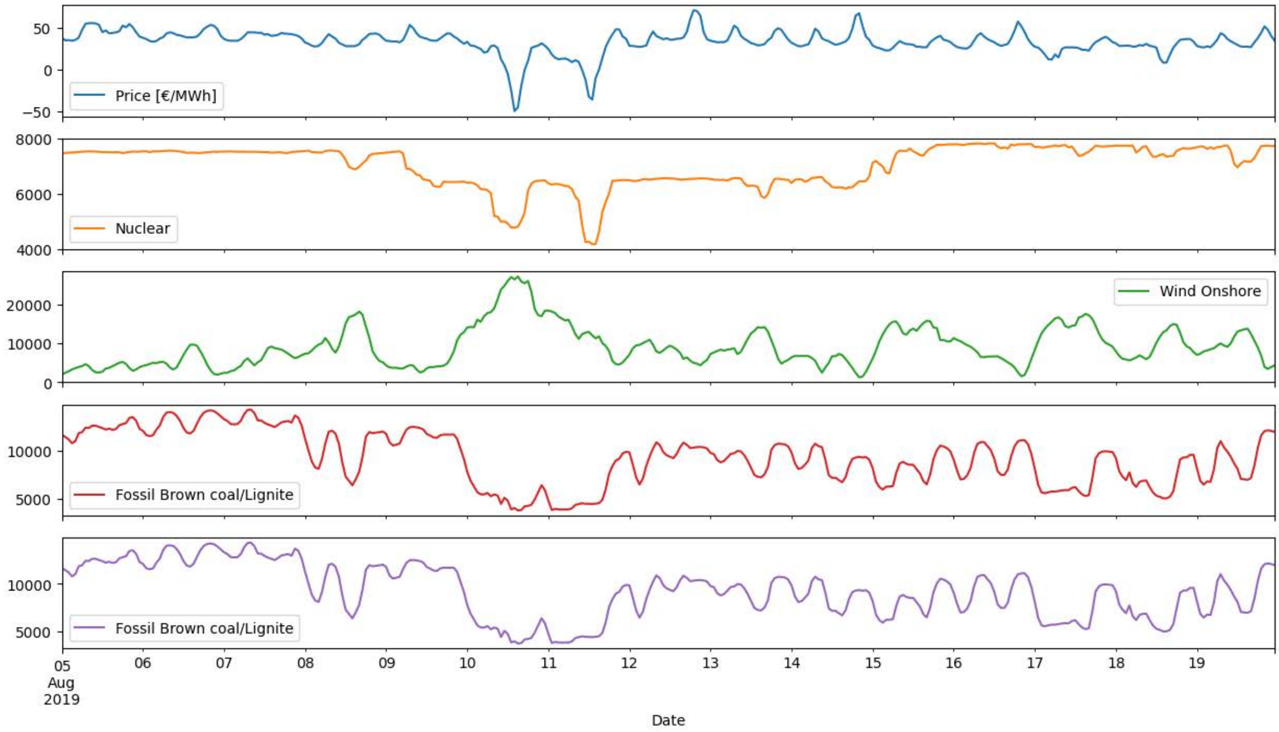


Figure 6.10: Variation of the features emphasizing the negative values of the price

More generally, in recent years the phenomenon of zero or even negative electricity prices has been observed in the European Energy Market for some hours of the day. In particular, the combination of RES and reduced demand seems to be shaking the European electricity market. The problem arises when the increase in RES production, in periods of favorable weather conditions, is combined with a limited demand. Given that RES are prioritized in the system, on days when there is excess production along with reduced demand, there is no room left in the system for thermal units. They do not function as reserves either, leaving the country's system uncontrollable in the event of a sudden change in weather conditions, which will abruptly drop RES production. Negative prices are an indication of what will happen to European electricity markets if the glut of planned renewable energy generation is not met by demand shifting. The hope is that eventually bigger fleets of electric cars, smarter grids, and better battery technology will fill the gap, but for now the mismatch is causing concern for governments and companies.

6.6.6 Formation and Analysis of the Independent Variables

Based on the above analysis and after several trials conducted, which include both tests only with the price variable along with the cyclical temporal variables formed by One-Hot

Encoding, as well as combinations of energy production variables with and without these cyclical variables, we have concluded that for an optimal model performance a timeseries for each of the 24 preceding hours is used for the following six input variables: electricity price and the values of energy produced from the exogenous variables of ’Fossil Brown coal/Lignite’, ’Fossil Hard coal’, ’Nuclear’, ’Wind Onshore’ and ’Hydro Run-of-river and poundage’.

In Figure 6.11, the overall time series of values is presented, which represents the count of values for each hour. It appears that most of the values in the time series fall within the range of 50-100€/MWh. However, simultaneously, the timeseries shows outliers, as a significant number of values exceed 300€, even reaching 500€. It is observed that the price distribution does not follow normality, with the majority of values less than 100€/MWh. Additionally, the combination of high values along with negative values creates a time series that does not follow consistent patterns in hourly fluctuations, unlike load time series, which typically exhibit stable seasonality and variations.

Moving to the variables used as input to the models, Figure 6.12 illustrates the graphical representation of the distribution of the six input variables. Specifically, it is noteworthy that the time series of ”Fossil Hard Coal” and ”Wind Onshore” exhibit a uniform distribution, a pattern also observed between the variables ”Hydro Run-of-river and poundage” and ”Fossil Brown coal/Lignite”. Notably, the distribution of ”Nuclear” presents an intriguing pattern, containing fluctuations and not adhering to a specific distribution approach, unlike the other forms of energy.

6.6.7 Graphical illustration of feature relations

Figure 6.13 visualizes the relationship between each of the exogenous variables and the variable of electricity price. Contours and solid color map represent the estimated probability density function of the joint distribution.

Based on Figure 6.13, it is observed that the features ”Fossil Brown coal/Lignite” and ”Hydro Run-of-river and poundage” exhibit a smooth correlation with the price of electrical energy, a fact which is proved by the intense color that extends almost across their entire value range. Additionally, concerning the variables ”Fossil Hard coal” and ”Wind onshore,” a similar smoothness of prices is noted with a narrower range of price density concentration. Finally, regarding nuclear energy production, it is noted that the variable ”Nuclear” demon-

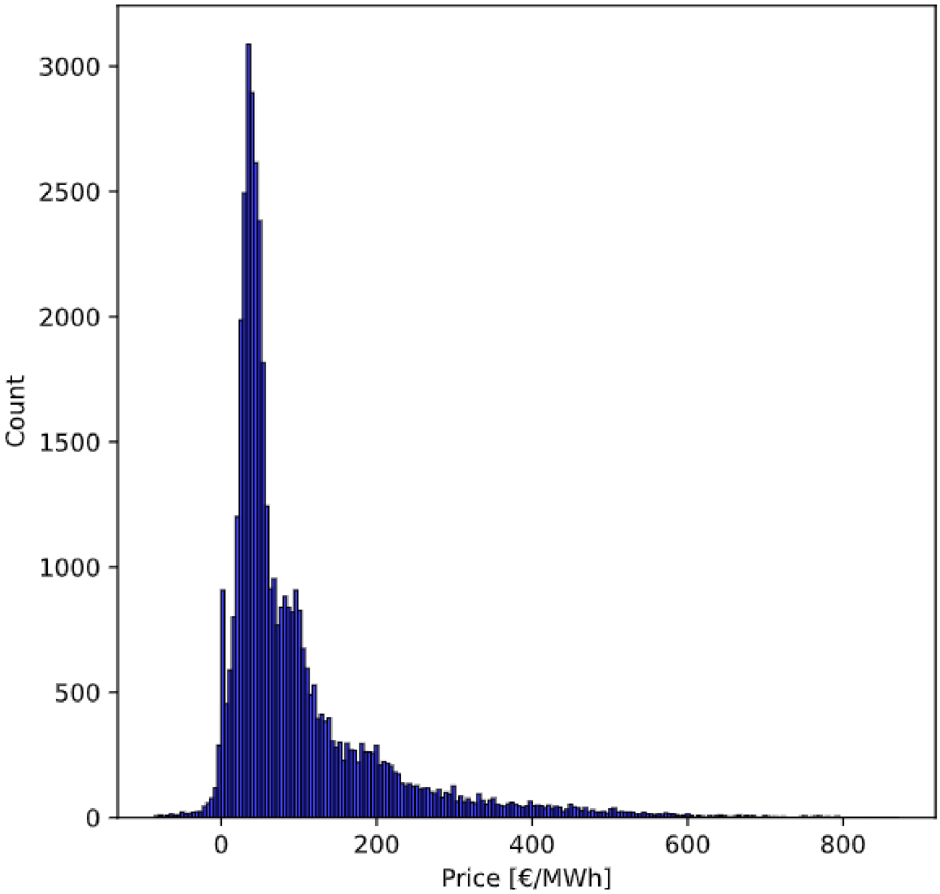


Figure 6.11: Electricity price countplot

strates higher density at higher values, a particularly interesting fact indicating its significant influence on shaping prices at higher nuclear energy production values.

6.6.8 Sliding Window Technique

For the training of our models and the formation of the 3D input, the sliding-window technique will be employed. For each forecasted hour, a 24-hour sample of each of the six variables provided as inputs will be utilized, in order to make forecasts suitable for the Intra-Day and Balancing Market. In more detail, the methodology involves utilizing the preceding 24 values of electricity prices, the corresponding 24 historical values of conventional generation units, namely 'Fossil Hard Coal' and 'Fossil Brown coal/Lignite', as well as the historical 24 values of Renewable Energy Sources such as 'Wind Onshore' and 'Hydro Run-of-river and poundage'. Finally, the historical values of nuclear energy for the variable 'Nuclear' will be incorporated.

6.6.9 Software Environment

Python 3.9 programming language, along with the Tensorflow, sklearn, Keras, Pandas and Numpy libraries, was used for the simulations in this study. All computational resources for these simulations were exclusively cloud-based, specifically utilizing Google's Colab infrastructure. The daily forecasts for all models also differed depending on the size and complexity of each forecasting model. More intricate and computationally demanding models, such as deep learning structures, required extensive virtual GPU resources. The run times for these complex models ranged from several hours to several days, and they were executed in batches spanning three to five months. The adoption of cloud-based resources facilitated efficient and scalable computation of simulations, enabling concurrent execution of multiple models and datasets. The varying duration of each simulation run underscores the computational requirements of the forecasting models and emphasizes the importance of suitable resources to ensure accurate and timely results. These findings offer valuable information on the practical considerations when implementing energy demand forecasting models in real-world applications.

6.6.10 Evaluation Strategy

To evaluate the effectiveness of the various forecasting models and techniques used in this investigation, we assessed their results using a set of metrics. The following metrics were used for this purpose, offering valuable insights into the precision and dependability of the forecasting models. Specifically, the mean square error (MSE) is given by equation 6.20 where y_i is the true value and $\hat{y}_i$ is the predicted value.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6.20}$$

The Root Mean Squared Error (RMSE) is given by equation 6.21 where y_i is the true value and $\hat{y}_i$ is the predicted value.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6.21}$$

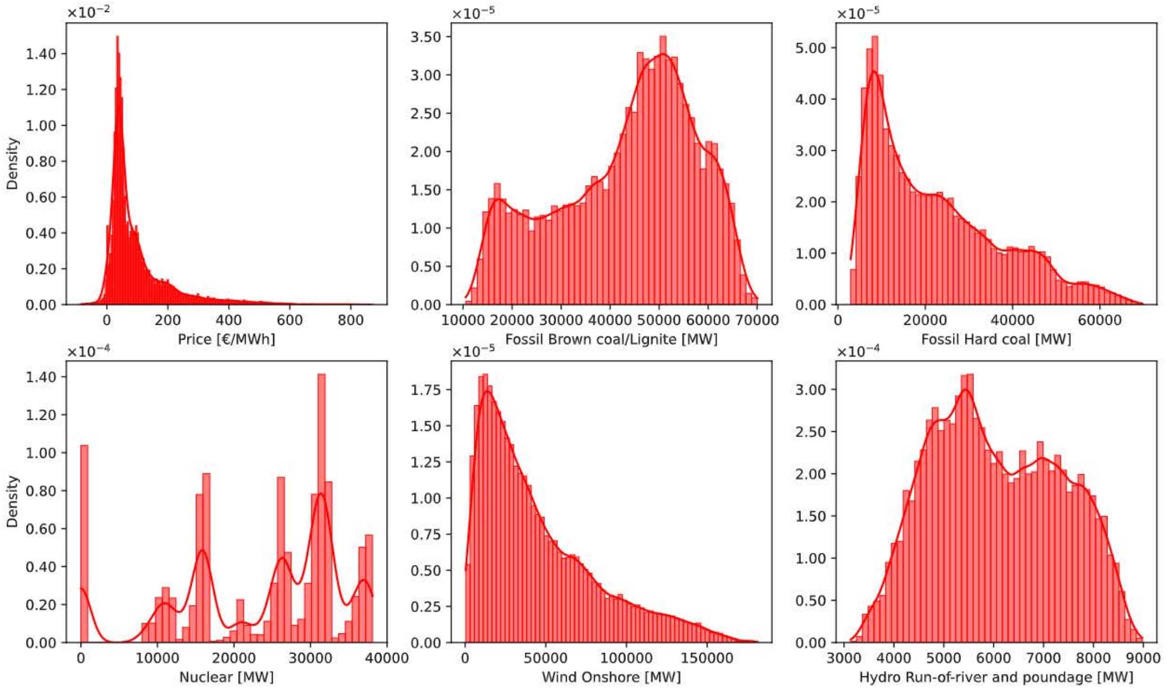


Figure 6.12: Distribution of the input variables of the prediction models

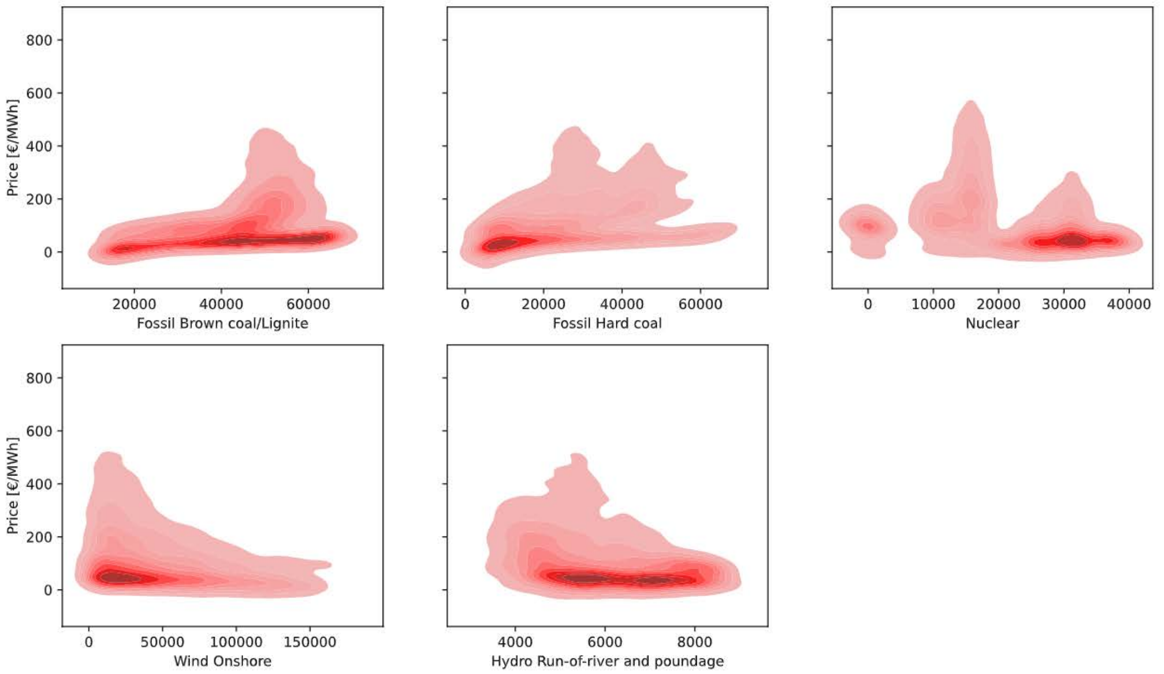


Figure 6.13: Price relation with exogenous features: probability density function of the joint distribution

The Mean Absolute Percentage Error (MAPE) is given by equation 6.22 where y_i is the true value and $\hat{y}_i$ is the predicted value.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \tag{6.22}$$

6.7 Analysis and Discussion of the Results

In this section, the forecasting results of each model are presented. Based on the results of Table 6.3, it is evident that initially the Hybrid CNN-GRU with Attention Mechanism and the Multi-Head Attention models achieve great prediction accuracy. Subsequently, the Soft Voting Ensemble model achieves the highest performance in each of the four metrics in Table 6.3. These outcomes are quite satisfactory and encouraging, as in most previous studies on time-series prediction, models incorporating Attention mechanisms and Transformers showed low performances.

Table 6.3: Performance comparison of each model.

Model	**MSE (€/MWh)^2**	**RMSE (€/MWh)**	**MAE (€/MWh)**	**MAPE (%)**
CNN-Attention	163.111	12.771	9.849	10.699
Multi-Head Attention	109.920	10.484	6.342	6.889
CNN-GRU-Attention	110.738	10.523	5.832	6.333
Stacking Ensemble	128.389	11.330	8.377	9.100
Soft Voting Ensemble	108.185	10.401	5.639	6.125
Benchmark: Perceptron	229.706	15.156	12.240	13.296

6.7.1 Performance of the models

The models' performance is evaluated based on Mean Absolute Percentage Error (MAPE) metric, providing a clear and intuitive understanding of the prediction accuracy of the models. In particular, the novel CNN-GRU with Attention mechanism model demonstrates a high accuracy of prediction, with a MAPE of 6.333%. This is a notable achievement, especially considering the historical challenges associated with attention mechanisms in transformer models. Also, the Multi-head Attention model achieves very high performance, reaching a MAPE of 6.889%. The CNN-Attention model is observed to exhibit lower prediction results compared to the other Deep Learning models, indicating that Convolutional mechanisms may not efficiently collaborate with the Attention Mechanism to achieve accurate prediction results.

The performance evaluation of the Soft Voting Ensemble model showcases its efficacy in electricity price prediction. Metrics such as MAPE demonstrate the high accuracy of the model, with a notable value of 6. 125%. Simultaneously, achieving MSE 108.185€/MWh, RMSE 10.401€/MWh, and MAE 5.639€/MWh, this model leverages the strengths of both the Multi-head Attention and the Hybrid CNN-GRU with Attention models. The weighting mechanism allows the ensemble to adaptively give more influence to the model with higher confidence, resulting in improved predictive accuracy and robustness. Regarding the Stacking Ensemble model, this algorithm achieves a MAPE of 9.100%, higher than Soft Voting for 2.975%. This difference may be possibly attributed to the final Dense Layer, which introduces additional trainable parameters to the final model, thereby making it more challenging to converge to the optimal result during the training period.

Regarding the benchmark Perceptron model, it is well-suited for linearly separable problems and binary classification tasks. Its simplicity makes it computationally efficient and easy to interpret. However, its limitations include the inability to handle non-linear relationships in data, which leads it in this manner to achieve the lowest performance compared to the other models, with an MAPE of 13.296%.

Additionally, concurrently with the performance of the models, a significant factor, especially for real-time conditions, is the training time of the models. As emphasized above, a virtual GPU was utilized for faster and more efficient training for all models. For the three deep learning models, CNN-attention, CNN-GRU with attention, and Multi-head-Attention, the training time, once the optimal hyperparameters are found, does not exceed 20 minutes. In particular, the Multi-head-Attention model, due to the multiple number of heads, is the slowest compared to the other two. Regarding the Stacking Ensemble model, the training time exceeds that of the simple ones, and is approximately 30 minutes. The Soft Voting Ensemble model, due to initially relying on the training of 2 separate models, requires approximately 20 minutes of training time. The Perceptron model is considerably simpler than the others, with its training time estimated at 10 minutes.

The minimal disadvantages of modeling techniques are based on the fact that the proposed models, which all use attention mechanisms, are quite complex and have multiple trainable parameters, with the result that their training is quite difficult compared to simpler models, such as the benchmark Perceptron model that is used.

Additionally, Figure 6.14, the residuals between actual and predicted values are plotted, so it can clearly observe the performance of each model. The specific plot shows that all the models behave approximately the same, they present common prediction characteristics and prediction residual outliers.

Finally, Figure 6.15 illustrates the plot diagram of the actual and predicted values for the first seven days of each month in the testing period. From the plots most of the models present exceptional predictability except for the Base model and CNN with attention model as described above.

6.7.2 Explainability Process

In this subsection, the Explainability process is presented. We will try to interpret the impact of the independent exogenous variables used in our models as input parameters. For this purpose, we will use the benchmark model, and through the SHAP library, we will delve deeper into its functionality. According to Figure 6.16, it is observed that for the longest period of observations, the price, as well as the variable ‘Fossil Brown coal/Lignite’ has the highest effect on price prediction. More specifically, the price achieves score -0.52, the ‘Fossil Brown coal/Lignite’ 0.55, the ‘Fossil Hard coal’ 0.3, the ’Wind onshore’ -0.73 and the variable ‘Hydro Run-of-river and poundage’ -0.55, respectively.

Following the explainability analysis, Figure 6.17 presents the effect of the input variables for a period in which negative electricity prices are observed. It is noteworthy that for the specific interval the effect of Renewable Energy Sources is very high and presents negative values, a fact that shows a negative correlation between RES variation and the price of electricity, as was analyzed above. Finally, it seems that the effect of conventional energy sources for these cases is quite small, with the variable ’Fossil Hard coal’ showing a value of 0.3.

6.8 Summary and Future Work

In this study, a comprehensive and integrated investigation was conducted with the aim of developing novel high-accuracy deep learning algorithms to predict the price of electricity on

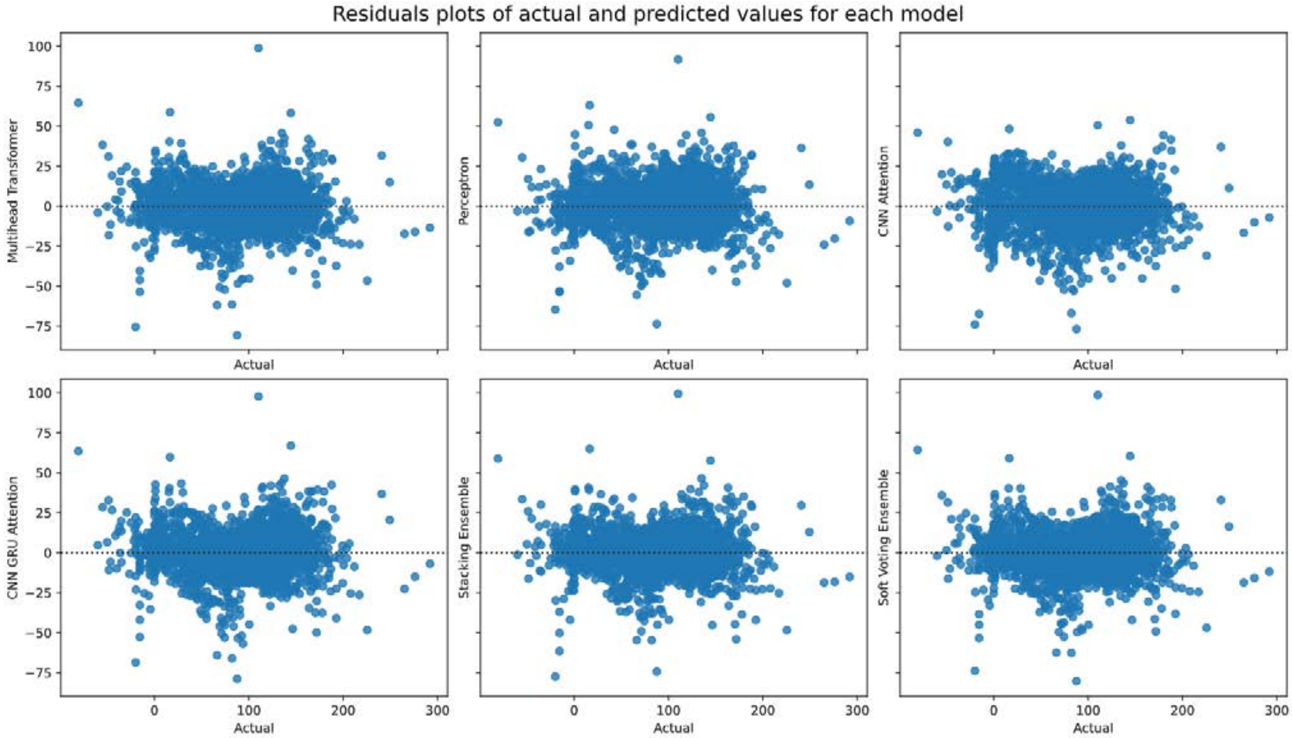


Figure 6.14: Residuals plots of actual and predicted values for each model

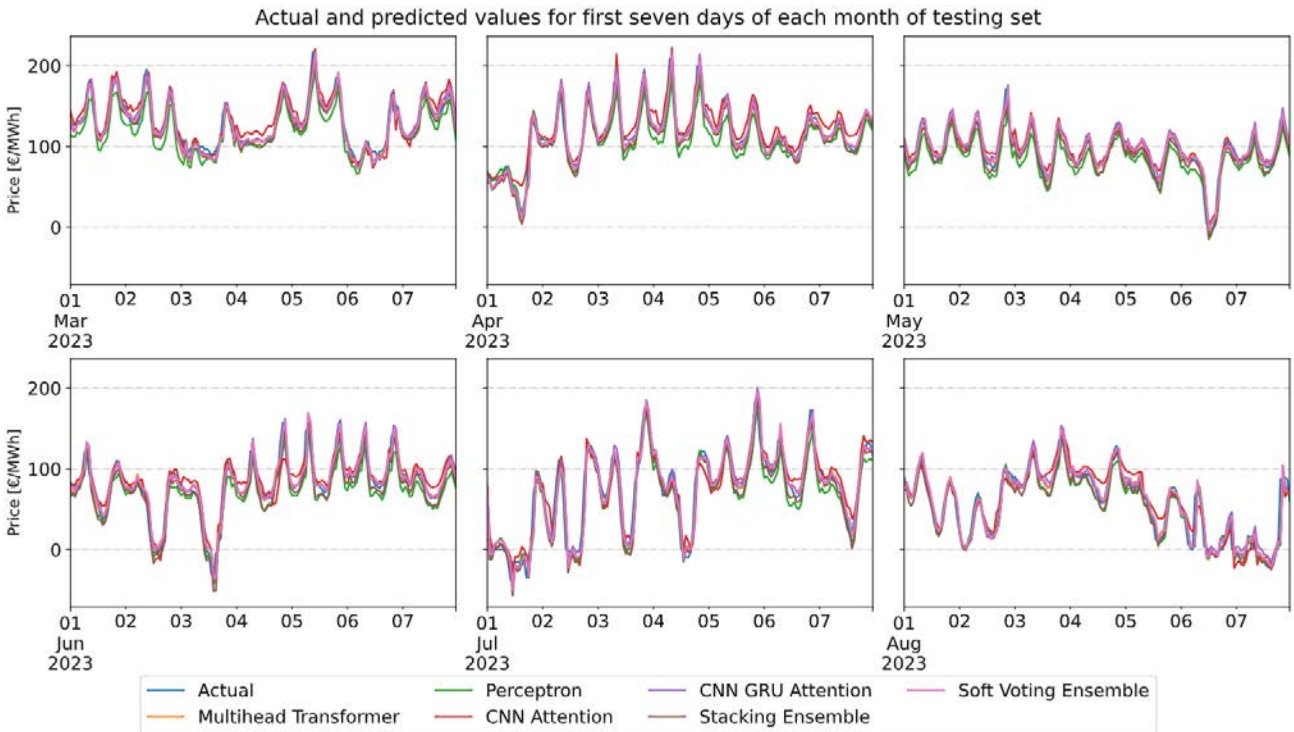


Figure 6.15: The actual and predicted values for the first seven days of each month in the testing set

Figure 6.16: SHAP values of impact on model output

Figure 6.17: SHAP values of impact on model output

a short-term basis. Its main objective is two-fold. Firstly, the development and proposal of two novel deep learning models, a hybrid CNN-GRU with Attention mechanism and an ensemble learning model that performs with the highest accuracy. Second, the creation of advanced Deep Learning models with Attention Mechanisms and testing them on the electricity price data of the Germany-Luxembourg interconnection. For this purpose, an optimized multihead transformer, a CNN-Attention model, and another Ensemble Learning category - a stacking ensemble - are investigated. A Machine Learning model, the Perceptron, was selected as a benchmark reference. Among these models, the hybrid CNN-GRU with attention mechanism stands out as a recommended model, as it has not been used in the literature for electricity price prediction problems until now, a fact that instills particular optimism for future studies and applications. Moreover, despite the general performance issues of most Transformer algorithms in time-series prediction problems, in our case, the Multi-head Transformer model predicts with exceptional accuracy on unknown data. Finally, the integration of two models from the Ensemble Learning domain and their successful results is both satisfying and challenging in terms of their selection.

A significant observation through the analysis is that the time-series of electricity prices does not follow a specific pattern of formation because their values are influenced by factors such as political conditions prevailing in neighboring countries. These factors are extremely difficult to simulate because of the uncertainty and reshuffling they cause in the wholesale energy market. Also, one more important observation is that the creation of advanced Deep Learning algorithms with integrated Attention Mechanisms could play a very important role in the field of wholesale electricity market. In combination with the use of the Optuna optimization algorithm, the hyperparameters of the models are optimized, acquiring the most suitable architecture, resulting in the achievement of the best results. Accurate price prediction could effectively contribute, in turn, to better production planning for producers, thus reducing deviations with undesirable consequences for system operation.

Regarding future work, the initial focus will be on delving into attention-based algorithms for tackling more intricate challenges in predicting electricity prices, which stands as a significant endeavor within the realm of Artificial Intelligence. Additionally, there is a prospect to assess the implementation of the algorithms elaborated in this study on short-term load prediction tasks for the flexibility market and day-ahead forecasting. Moreover, the Soft Voting Ensemble Model underscores the potential of amalgamating diverse deep learning structures to enhance timeseries forecasting. Subsequent research could further investigate alternative model combinations, ensemble tactics, and optimization of weight assignments to push the boundaries of excellence in the field. It is worth noting that the models introduced, alongside their hyperparameter optimization techniques, could be explored in various other domains, including industrial, financial, and biomedical sectors for predictive purposes.

Finally, based on the Explainability process presented above, future research could concentrate on Uncertainty Quantification, based on which methods to quantify and explain uncertainties associated with electricity price predictions could be explored. Understanding the level of confidence or uncertainty in forecasts is crucial for decision makers in the energy industry and allows them to make more informed choices regarding resource allocation, risk management, and strategic planning. Furthermore, by incorporating uncertainty quantification into decision processes, stakeholders can respond proactively to potential variations in electricity prices, leading to more resilient and adaptive energy strategies.

Chapter 7

Transfer Learning in Energy Market

Transfer learning is a subfield of machine learning that leverages knowledge gained from one domain (the source domain) to improve performance in another related domain (the target domain). This approach is particularly valuable in the energy market, where data can be sparse or highly variable, and models must adapt to changes in supply, demand, and external conditions. Instead of starting the learning process from scratch, a model leverages the features, representations, or parameters acquired from a pre-trained model on a large dataset, often in a domain where data is abundant, and then adapts this learned information to a new target domain that may have limited data. This approach is especially beneficial in scenarios where data collection is expensive or time-consuming, as it allows for more rapid model development and often results in improved performance. The underlying assumption is that many tasks share commonalities—such as visual features in images or semantic patterns in text—that can be effectively reused, reducing the amount of effort required to train a model for the new task. Consequently, transfer learning has become a powerful tool in fields like computer vision, natural language processing, and beyond, where pre-trained models on extensive datasets have paved the way for breakthroughs in related, data-scarce applications [226].

In practice, transfer learning works by repurposing parts of a pre-trained model to serve as a foundation for the new task. One common method is to use the pre-trained model as a feature extractor, where its layers—typically the early ones that capture generic features—are frozen, and only the later layers are fine-tuned on the new dataset to learn task-specific representations. Another technique involves initializing a new model with the weights of a pre-trained model and then fine-tuning the entire network to adapt to the nuances of the target domain. This process may involve domain adaptation strategies that adjust the model to account for differences between the source and target data distributions, ensuring that the transferred knowledge remains relevant. By doing so, the model not only benefits from a robust starting point but also learns to refine its parameters in a way that suits the specific characteristics of the new problem, ultimately leading to better performance with less training data and reduced computational cost.

Fundamentals of Transfer Learning

Traditional machine learning models typically assume that training and testing data share the same distribution. However, in the context of the energy market, this assumption often does not hold due to factors such as seasonal changes, regional differences, and evolving market dynamics. Transfer learning addresses these challenges by:

- **Feature Transfer:** Reusing learned features or representations from a well-resourced source domain to enhance the learning process in a target domain with limited data.
- **Parameter Transfer:** Initializing target models with parameters obtained from a source model, thereby speeding up convergence and improving overall performance.
- **Domain Adaptation:** Adjusting the source model to better align with the target domain's data distribution, reducing the risk of negative transfer.

Applications in the Energy Market

The energy market benefits from transfer learning in several key areas:

1. **Load Forecasting:** Accurate demand forecasting is essential for grid management and operational planning. Transfer learning can integrate data from different regions or historical periods to refine predictive models, even when local data is limited [227].
2. **Renewable Energy Integration:** The intermittent nature of renewable energy sources such as solar and wind makes it challenging to predict generation capacity. Transfer learning models can incorporate weather patterns and historical performance data from diverse geographical areas to improve reliability [227].
3. **Anomaly Detection and Fault Diagnosis:** Detecting irregularities and potential faults in energy systems is crucial for maintaining system stability. By transferring knowledge

from simulated environments or similar operational systems, models can more quickly and accurately identify anomalies in real-world scenarios [228].

4. **Market Analysis and Price Forecasting:** Energy prices are influenced by a complex interplay of factors including supply-demand imbalances, regulatory changes, and macroeconomic trends. Transfer learning can help build more robust forecasting models by synthesizing information from related markets or time periods [229].

7.1 Enhanced Sequence-to-Sequence Deep Transfer Learning for Day-Ahead Electricity Load Forecasting

7.1.1 Problem Description

Electricity load forecasting is a crucial undertaking within all the deregulated markets globally. Among the research challenges on a global scale, the investigation of Deep Transfer Learning (DTL) in the field of electricity load forecasting represents a fundamental effort that imparts generality to Artificial Intelligence applications. In this research a comprehensive study is conducted for day-ahead electricity load forecasting. For this purpose, three Sequence-to-Sequence (Seq2seq) Deep Learning (DL) models are used, namely the Multilayer Perceptron (MLP), the Convolutional Neural Network (CNN) and the Ensemble Learning Model (ELM), which consists of the weighted combination of the outputs of MLP and CNN models. Also, the study focuses on the development of different forecasting strategies based on DTL and emphasizing the way the datasets are trained and fine-tuning for higher forecasting accuracy. In order to implement the forecasting strategies using Deep Learning models, load datasets from three Greek islands, Rhodes, Lesvos, and Chios, are used. The main purpose is to apply DTL for day-ahead predictions (1-24 hours) for each month of the year for Chios dataset after training and fine-tuning the models using the datasets of the three islands in various combinations. Four DTL strategies are illustrated. In the first strategy (DTL Case 1), each of the three DL models is trained using only Lesvos dataset, while fine-tuning is performed on the dataset of Chios island, in order to create day-ahead predictions for Chios load. In the second strategy (DTL Case 2), data from both Lesvos and Rhodes concurrently are used for the DL model training period, and fine-tuning is performed on the data from Chios. The third DTL strategy (DTL Case 3) contains the training of the DL models using

Lesvos dataset, and the testing period performed directly on the Chios dataset without fine-tuning. The fourth strategy is a Multi-task Deep Learning (MTDL) approach, which has been extensively studied in recent years. In MTDL, the three DL models are trained simultaneously on all three datasets and the final predictions are made on the unknown part of the dataset of Chios. The results demonstrated that DTL can be applied with high efficiency for day-ahead load forecasting. Specifically, DTL Case 1 and 2 outperformed MTDL in terms of load prediction accuracy. Regarding the DL models, all three exhibit very high prediction accuracy, especially in the two cases with fine-tuning. The ELM excels compared to the single models. More specifically, for conducting day-ahead predictions, it has been concluded that the MLP model presents best monthly forecasts with a MAPE of 6.24% and 6.01% for the first two cases, the CNN model presents best monthly forecasts with a MAPE of 5.57% and 5.60% respectively and the ELM model achieves best monthly forecasts with a MAPE of 5.29% and 5.31%, respectively, indicating the very high accuracy it can achieve.

Deep Transfer Learning (DTL) refers to the use of Deep Neural Networks (DNNs) in the domain of Transfer Learning. DTL utilizes the knowledge learned from one task to improve the performance of another related task. Usually, in DTL a pre-trained DNN model is fine-tuned for a different task. This domain improves data efficiency, reduces training time, and allows models to generalize well, capturing the underlying patterns in electricity consumption. The ability to adapt to dynamic conditions and the improved accuracy stemming from pre-trained models make transfer learning a valuable tool for addressing the challenges of forecasting in the energy sector. With limited and heterogeneous datasets, transfer learning enables models trained in one domain to be adapted to another, addressing the challenges of varying temporal scales and spatial characteristics. This approach proves essential for optimizing electricity forecasting models, particularly in the face of emerging technologies, changing infrastructures, and the need for resource-efficient solutions. By reusing pre-trained models and enhancing adaptability, transfer learning contributes significantly to robust and accurate predictions, ultimately supporting more effective energy management in the dynamically evolving landscape of the electricity sector.

7.1.2 Related Studies

Research interest in the electricity sector has been growing for several key reasons. First, there is a global shift toward using sustainable and renewable energy sources like solar and wind. This shift has led researchers to find ways to better incorporate these technologies into existing power grids. Second, the increasing need for electricity due to population growth and industrialization requires new and efficient solutions for transmission, distribution, and consumption of electricity. Third, the development of smart grids and improvements in energy storage technologies have opened up opportunities to improve the resilience, reliability, and responsiveness of power grids. Additionally, concerns about the environment and climate change have motivated researchers to explore cleaner and more eco-friendly energy options. The digitization of the electricity sector, thanks to advances in data analytics, machine learning, and Internet of Things (IoT) technologies, has also spurred research to create smarter and more efficient energy systems. In general, these factors have created a dynamic landscape, urging researchers to explore new approaches and technologies to tackle the changing challenges and opportunities in the electricity sector.

More generally, in recent years, Transfer Learning is gathering increasing scientific interest. For this purpose, an extensive literature review follows, aiming to present the most comprehensive publications that investigate the specific field. Meng et al. in [230] propose a transfer learning based method for abnormal electricity consumption detection, where a pre-trained model is fine-tuned using a small amount of data from the target domain. Antoniadis et al in [231] discuss the use of transfer learning techniques for electricity load forecasting, specifically in the context of leveraging information from finer scales to improve forecasts at wider scales. Yang et al. in [232] discuss the implementation of a transfer learning strategy to address the multi-parameter coupling problem in the design of water flow enabled generators. Dong et. al. in [233] propose a transfer learning model based on Xception neural network for electrical load prediction, which is trained using pre-trained models and fine-tuned in the training process. Li et al. in [234] explore a transfer learning scheme for non-intrusive load monitoring (NILM) in smart buildings, which involves transferring a well-trained model to estimate power consumption in another data set for all appliances. Peirelinck et al. in [235] discuss the use of transfer learning techniques in the context of demand response in the electricity sector. It shows that transfer learning can improve performance by over 30% in various tasks. Laitsos et al. in [148] use Transfer Learning Technique with several Deep Learning models to predict energy consumption of Greek Islands and the pre-trained model demonstrates outstanding flexibility when adapting to a new and unknown dataset. Wu et al. in [236] propose an attentive transfer framework for efficient residential electric load forecasting using transfer learning and graph neural networks. Kamalov et al. in [237] introduce a NBEATS model in order to test its results in medium-term electricity forecasting for zero-shot transfer learning. Syed et al. in [238] propose a reliable inductive transfer learning (ITL) method for load forecasting in electrical networks, which uses knowledge from existing deep learning models to develop accurate ITL models at other distribution nodes. Laitsos et al. in [81] propose an automated Deep Learning application for electricity load forecasting. Santos et al. in [239] propose a novel methodology that combines transfer learning and deep learning techniques to enhance short-term load forecasting for buildings with limited electricity data. Arvanitidis et al. in [240] propose clustering MLP models for short-term load forecasting. Luo et al. in [241] discuss the use of transfer learning techniques for load, solar, and wind power predictions, but it does not specifically mention the application of transfer learning to load prediction. Li et al. in [242] discuss a short-term load forecasting framework that adopts Transfer Learning. Transfer learning is used to train learnable parameters based on trend components and then transfer them to the load forecasting model. Chan et al. in [243] introduce a hybridized modelling approach, using a Convolutional Neural Network (CNN) and a Support Vector Machine (SVM), for short-term load forecasting. Gontijo et al. in [244] examine the hourly power generation data in Brazil from 2018 to 2020, categorized based on the different electrical subsystems and their corresponding energy sources. The aim is to assess the precision of key methods for combining and splitting forecasts generated by the Autoregressive Integrated Moving Average (ARIMA) and Error, Trend, Seasonal (ETS) models.

Jung et al. in [245] propose a monthly electricity load forecasting framework for smart cities, using transfer learning techniques. Collect data from multiple districts, select similar data based on correlation coefficients, and fine-tune the model using target data. Al-Hajj et al. in [246] present a survey of transfer learning in renewable energy systems, specifically in the prediction of solar and wind power, the prediction of load, and the diagnosis of faults. Nivarthi et al. in [247] discuss the use of transfer learning in renewable energy systems, specifically in power forecasting and anomaly detection. It proposes a transfer learning framework and a feature embedding approach to handle missing sensor data. Miraftabzadeh et al. in [248] present a framework based on transfer learning and deep neural networks for the prediction of day-ahead photovoltaic power. akovic et al. in [249] conduct an extensive review of machine learning applications aimed at addressing energy-related issues through the examination of various energy types and opportunities for energy reduction. Vontzos et al. in [149] propose a data-driven short-term forecasting for electricity consumption in airport. Yang et al. in [250] propose an innovative monthly DNN approach, for load forecasting in urban and regional areas. In order to draw more secure conclusions, an extended comparison with other Machine Learning models is conducted. Li et al. in [251] propose a building electricity load forecasting method based on Maximum Mean Discrepancy (MMD) and an improved TrAdaBoost algorithm (iTrAdaBoost). Gontijo et al. in [252] introduce the Dynamic Time Scan Forecasting (DTSF) technique as a novel approach to predict hourly energy prices in Brazil. By identifying similarity patterns in time series data, DTSF has demonstrated competitive advantages over traditional forecasting methods in prior research.

7.1.3 Research Contribution

The relentless growth of electricity demand, coupled with the dynamic and often unpredictable nature of energy consumption patterns, requires advanced forecasting methods for effective grid management. Contrary to the aforementioned studies, which investigate Transfer Learning only within a few hours or so, the forecasting strategies developed in this work target day-ahead prediction, thereby enhancing the effectiveness of the models. Additionally, none of the aforementioned studies explore Ensemble Deep Learning, as is done in this study. This Deep Learning domain includes models that gather significant scientific interest due to their high performance, as well as the potential they offer for further investigations and improvements. So, in this context, transfer learning emerges as a promising paradigm to address the challenges associated with limited and disparate data sources. This research work delves into the application of deep transfer learning techniques in the domain of electricity forecasting, with the aim of exploiting the knowledge gained from a source domain to improve predictive accuracy in a target domain. By leveraging preexisting models trained on related datasets or domains, transfer learning seeks to enhance the adaptability and robustness of forecasting models, ultimately contributing to more accurate and reliable predictions in the complex and ever-evolving landscape of electricity demand. This research explores the theoretical foundations, methodologies and practical implications of transfer learning in the specific context of electricity forecasting, shedding light on its potential to revolutionize the field and pave the way for more resilient and efficient energy management systems.

With respect to the contribution of the current work, the following points are emphasized:

- For the first time, a high accuracy results implementation of Sequence to Sequence (Seq2seq) Ensemble Deep Transfer Learning for day-ahead (1-24 hours) forecasting is conducted on three distinct datasets from islands of the Greek power system. Although the Rhodes training data set exhibits somewhat different behavior compared to the other two datasets, the proposed algorithms provide very satisfactory results. This fact further enhances the performance of the proposed strategies and models. The characteristics of the Rhodes dataset lead to a more robust and comprehensive evaluation of the models, as it introduces variability and challenges that may not be present in the other datasets. This diversity in behavior across datasets provides a more realistic and thorough assessment of the models' capabilities.

- The results obtained indicate that Deep Transfer Learning (DTL) could provide particular value to both Transmission System Operators (TSO) and Distribution System Operators (DSO), within various regions of the Greek system.

- The application of the models is performed on actual load data with minimal data preprocessing, a fact that creates optimistic conclusions regarding their applicability under real-time conditions.

The current research is organized as follows. First, in Section 7.2, Exploratory Dataset Analysis and Feature Creation is used. In Section 7.3, the forecasting strategies are analyzed. In Section 7.4, the emergent results for each algorithm are presented, along with a discussion about their performance. Finally, in Section 7.5, the main conclusions and future study proposals are analyzed.

7.2 Materials and Methods

7.2.1 Dataset Analysis

In this section, all features, behaviors, and correlations of the three datasets used are investigated and analyzed. Initially, the three timeseries under study are presented, and then emphasis is given to their monthly and daily average fluctuations.

Figure 7.1 illustrates the three power timeseries fluctuations for each of the three datasets in hourly resolution. What is noteworthy is that Rhodes, especially during the summer months,

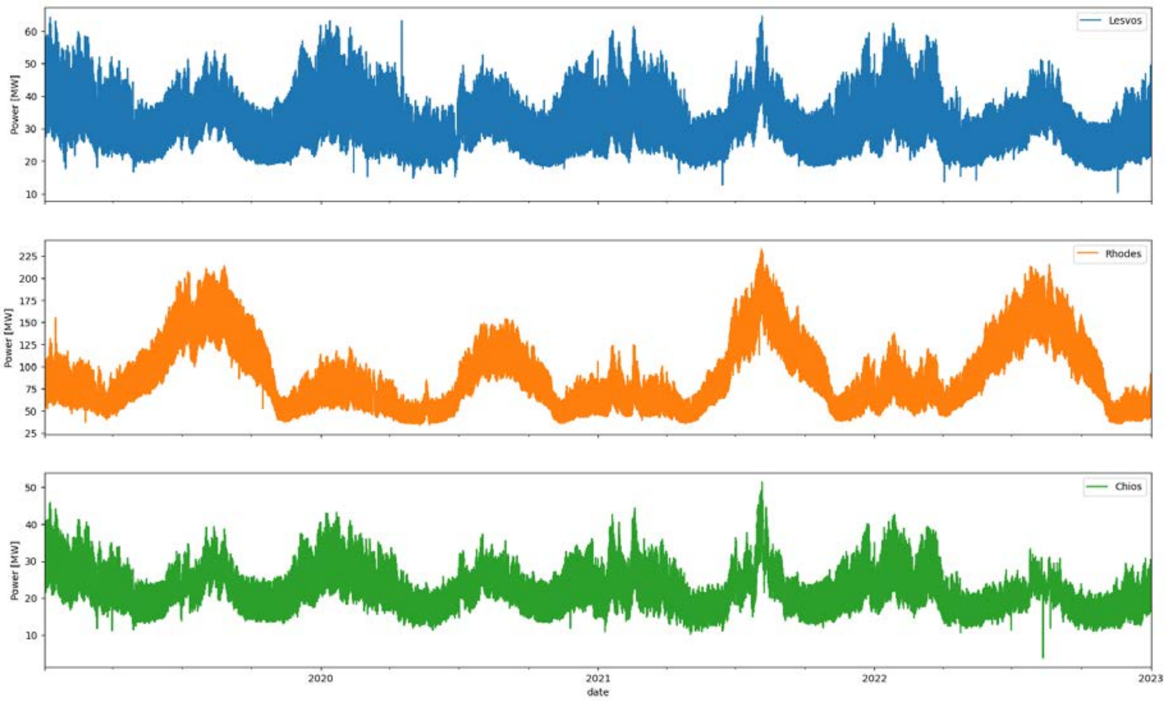


Figure 7.1: Hourly consumption for Lesvos, Rhodes and Chios

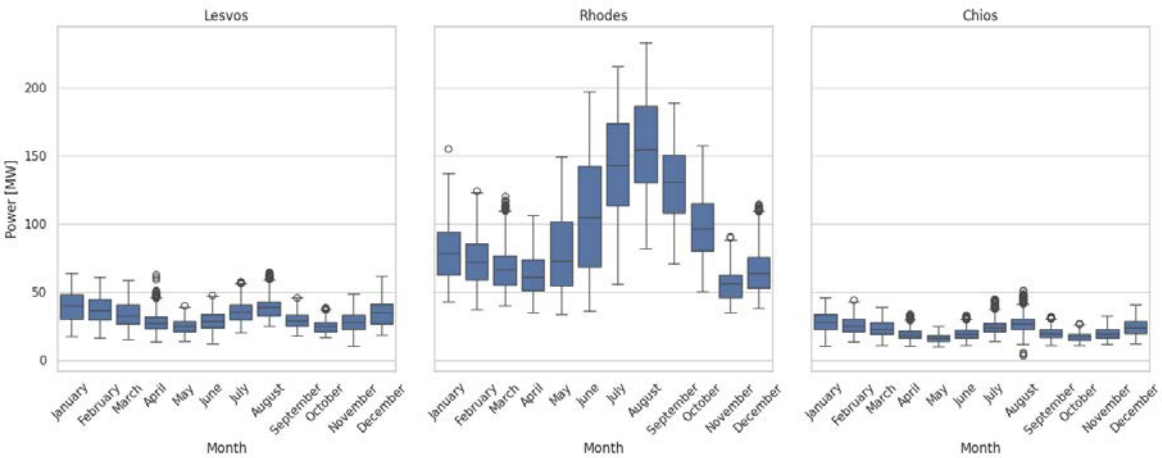


Figure 7.2: Monthly average consumption for every island

experiences a substantial increase in demand. The other two islands exhibit several similarities between them, with both the average and extreme values behaving relatively similarly. This consistency in behavior across the two datasets suggests common characteristics or patterns in the energy-related dynamics of these islands. The shared trends in both average and extreme values contribute to a more coherent and comparable analysis between the two data sets, helping to develop and evaluate models for these specific island environments.

Figure 7.2 presents the monthly boxplots for each of the three datasets in hourly resolution.

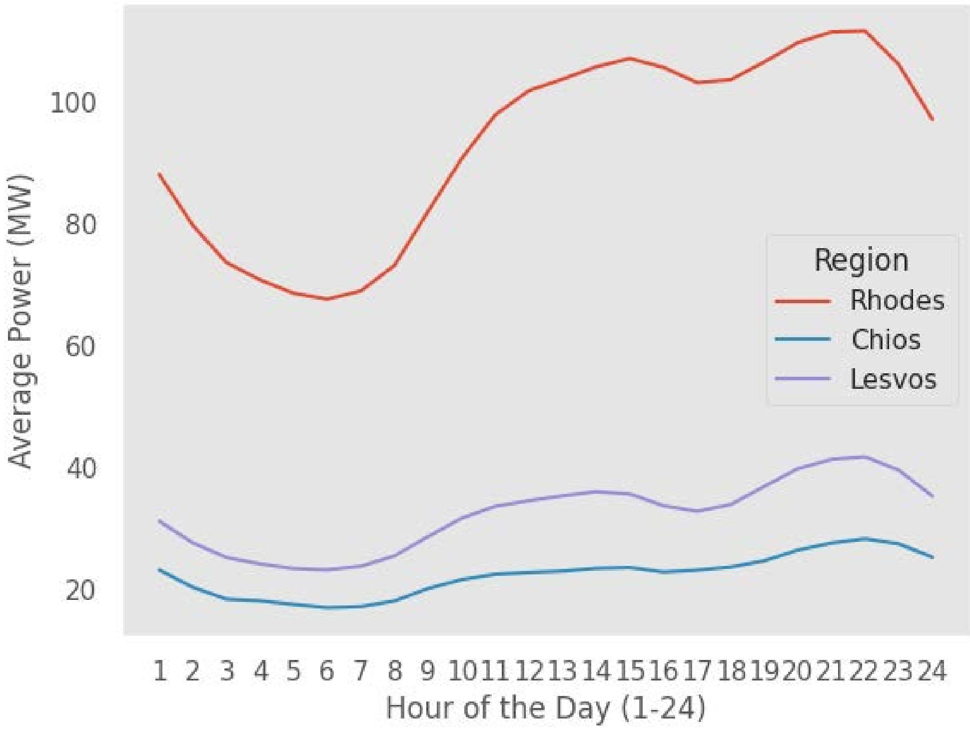


Figure 7.3: Hourly average consumption per day for each island

In addition, Figure 7.3 visualizes the average daily electricity consumption (1-24 hours) for each island.

This figure clearly illustrates that the average hourly values for Rhodes are significantly higher than those of the other two islands. However, it should be noted that the three patterns exhibit high similarities between them, a fact demonstrated by common peak and off-peak demand hours among the three islands.

7.2.2 Data preprocessing

In order to shape the raw data into a suitable format, capable of being used as input in deep learning models, the preprocess involves the following four stages:

1. Anomaly Detection: Anomaly detection in timeseries involves establishing a baseline of normal behavior through statistical methods or machine learning algorithms, extracting relevant features, and training a model on labeled data to distinguish normal patterns from anomalies. Due to instances of zero consumption during specific hours, probably caused by network faults, these particular values were set equal to the corresponding values from one week prior. This adjustment was made to address the chal-

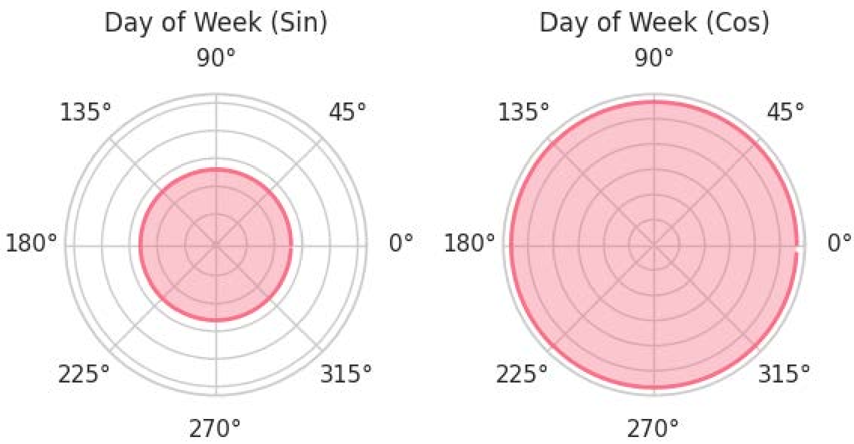


Figure 7.4: Day of the week in sine and cosine formulation

lenge of unexpected situations in the data, as algorithms may struggle to account for such anomalies. The goal is to ensure optimal training for each model by handling these irregularities in the data set.

2. Filling missing values: Filling missing values in timeseries data is a crucial preprocessing step for anomaly detection. Since anomalies are often identified based on patterns and trends in the data, it is essential to address gaps caused by missing values.

3. Min-Max Scaling: This preprocessing method applied to all datasets in this work, involves Min-Max Scaling, which normalizes data points to a range between 0 and 1. To achieve this, two distinct scalers were employed—one for input and another for output datasets. The primary rationale behind utilizing Min-Max Scaling is its ability to enhance the efficiency of training deep learning models during the training phase, facilitating faster convergence to the optimal solution of the loss function.

4. One-Hot Encoding: With this process, numerical data are transformed to cyclical, through trigonometric equations. In this study, day of the week, hour of the day and month of the year were converted to sin and cosine formulation. Figure 7.4 represents the day of the week transformed in sin and cosine format. The periodicity that appears with this method helps the models to better understand the patterns presented by each timeseries studied.

7.2.3 Feature Creation

In this subsection, the parameters used as inputs for the models are being generated. For this reason, several input characteristics were studied and evaluated in order to understand the most significant to predict electricity demand. After several trials, we ended up with the best combination of features, which gave us the most accurate prediction results. So, eight input characteristics were selected and investigated to forecast the day-ahead electricity demand. More specifically, variables 2-8 concern the methodology of One-Hot Encoding, which was introduced earlier and helps the DL models better understand and adapt to timeseries exhibiting seasonal patterns, such as those studied in the current work. The input variables used for all Deep Learning Models remain consistent and are presented in detail below.

1. Power in hourly resolution: The sequence of 168 hours of load values for 7 days/one week.
2. Cos of Day of Week: The sequence of 168 values of Day of the Week (0-6) converted by One Hot Encoding to cosine type.
3. Sin of Day of Week: The sequence of 168 values of Day of the Week (0-6) converted by One Hot Encoding to sin type.
4. Cos of Hour of Day: The sequence of 168 values of Hour of Day (0-23), converted by One Hot Encoding to cosine type.
5. Sin of Hour of Day: The sequence of 168 values of Hour of Day (0-23), converted by One Hot Encoding to sin type.
6. Cos of Month of Year: The sequence of 168 values of Month of the Year (1-12) converted by One- Hot Encoding to cosine type.
7. Sin of Month of Year: The sequence of 168 values of Month of the Year (1-12) converted by One- Hot Encoding to sin type.
8. IsWeekend: The sequence of 168 values of a dummy variable named “Is Weekend”, with value equal to 0 for working days and 1 for weekends and holidays.

Our target is to utilize a historical sequence of 168 hours from the aforementioned 8 features, and create day-ahead predictions for the load, i.e., a sequence of 24 values of the Power. Figure 7.5 visualizes the Seq2seq forecasting technique.

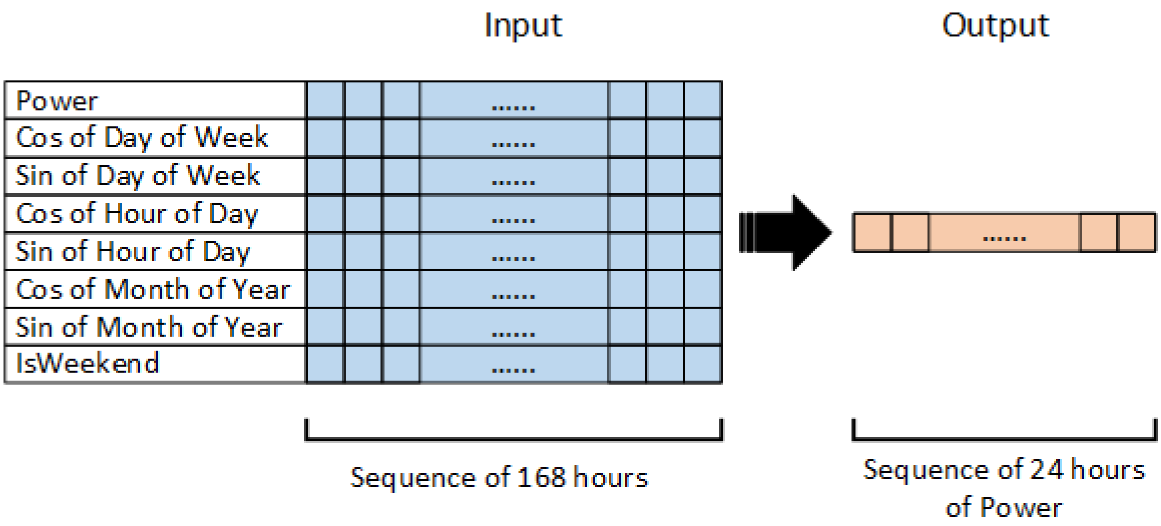


Figure 7.5: Sequence-to-sequence forecasting technique

Finally, the correlation heatmap is presented in Figure 7.6, in order to highlight the relationships between the power timeseries for each of the three islands.

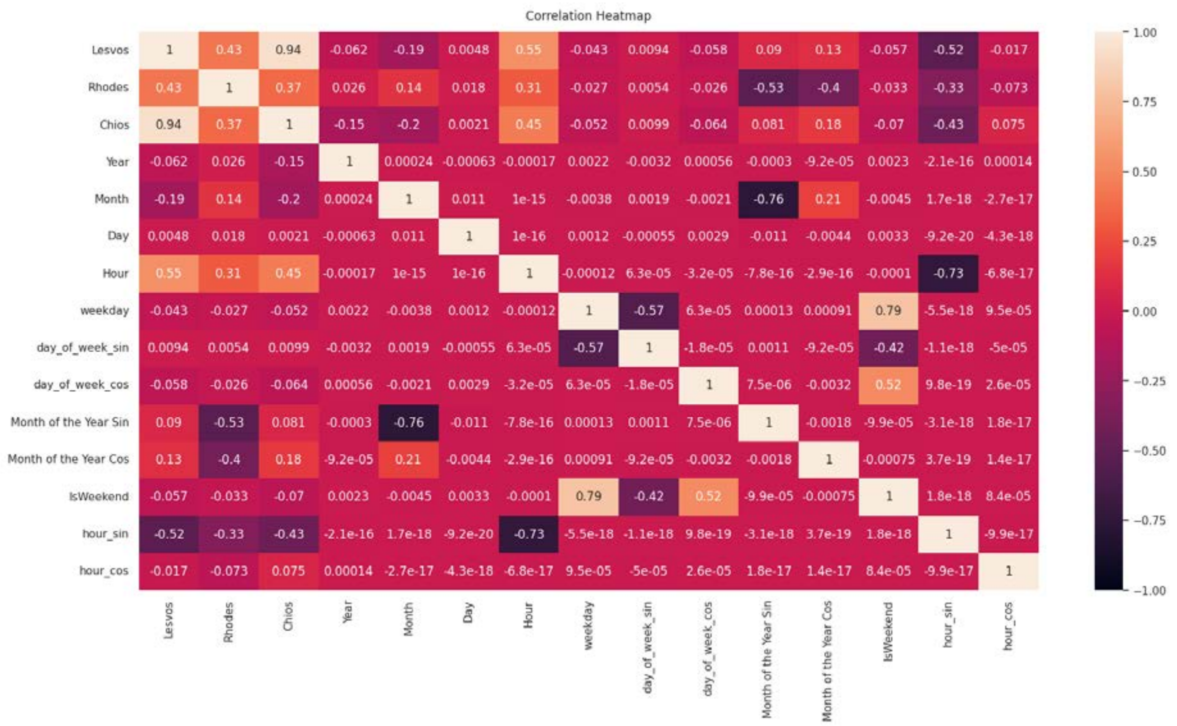


Figure 7.6: Correlation heatmap

What is noteworthy is that the target timeseries of Chios shows a correlation of 0.37 with that of Rhodes and 0.94 with that of Lesvos. This particular characteristic further enhances the reliability of the implemented applications, demonstrating generality and robustness through the results that will be presented below. The correlation between the target timeseries of different islands suggests some level of interdependence or shared patterns, which can contribute to the generalization and effectiveness of the models developed for these islands.

7.3 Methodology

In this section, the fundamental methodology followed in the current work is presented. Initially, an introduction of the three DL models is presented. Then, the forecasting strategies are analyzed in depth, emphasizing all the details of each one, highlighting the way datasets are utilized for each strategy. Subsequently, a reference was made to the Optimization framework, which has been used for enhancing the performance of the DL models. Additionally, the evaluation metrics are presented, which were used to evaluate and compare the performance of the algorithms of each strategy. Finally, the software tools that were used are analyzed.

7.3.1 Deep Learning Models

In this subsection, the functionality and architectures of the three DL models used in the ongoing research, MLP, CNN, and ELM, are analyzed. The main reason why the specific models are chosen is both because they are very powerful and advanced architectures of Deep Learning domain in the field of timeseries forecasting, and because they are a field of high scientific interest compared to simpler Machine Learning algorithms.

Multilayer Perceptron

A Multilayer Perceptron (MLP), which is presented in Figure 7.7, is a versatile artificial neural network architecture employed for learning and modeling complex relationships within data. Comprising an input layer (X), one or more hidden layers (H_i), and an output layer (Y), the MLP is characterized by its capacity to capture intricate non-linear patterns. During forward propagation, the input data is transformed through weighted connections and activation functions (σ) in the hidden layers, generating progressively abstract representations. The hidden layer outputs (H_i) can be mathematically expressed as:

$$H_i = \sigma(W_i H_{i-1} + b_i) \tag{7.1}$$

where W_i denotes the weight matrix connecting layer $i-1$ to i, H_{i-1} is the output of the previous layer, and b_i represents the bias term for layer i. The activation function introduces non-linearity, enabling the network to capture complex mappings.

The final output (Y) is computed through similar operations in the output layer:

$$Y = \sigma(W_{\text{out}} H_{\text{last}} + b_{\text{out}}) \tag{7.2}$$

During training, the network adjusts its weights to minimize a defined loss function (L), which quantifies the disparity between the predicted output and actual target values. The weights are updated using an optimization algorithm, typically gradient descent, with the weight update rule expressed as:

$$W_{\text{new}} = W_{\text{old}} - \eta \frac{\partial L}{\partial W} \tag{7.3}$$

where η is the learning rate. This iterative process, known as back propagation, refines the model's weights to improve its predictive accuracy.

In the domain of timeseries forecasting, MLPs exhibit efficiency owing to their inherent ability to capture temporal dependencies and non-linear patterns. The adaptability of the model enables it to discern and model various temporal structures, including trends and seasonality. The hidden layers serve as dynamic feature extractors, automatically learning relevant temporal features from the timeseries data. This feature learning capability, coupled with the tunability of model parameters, positions MLPs as robust and effective tools for a wide array of timeseries forecasting tasks.

Convolutional Neural Network

Convolutional Neural Networks (CNNs), the architecture of which is presented in Figure 7.8, constitute a class of sophisticated deep learning architectures specifically designed for the analysis and processing of visual data. The principal structure of CNNs encompasses multiple layers, notably including convolutional layers, pooling layers, and fully connected layers. Convolutional layers assume a pivotal role in feature extraction from input images through the application of convolutional operations utilizing trainable filters. These filters adeptly identify patterns and features at various spatial scales, enabling the network to discern intricate details within the data. Accompanying pooling layers serve to diminish the spatial

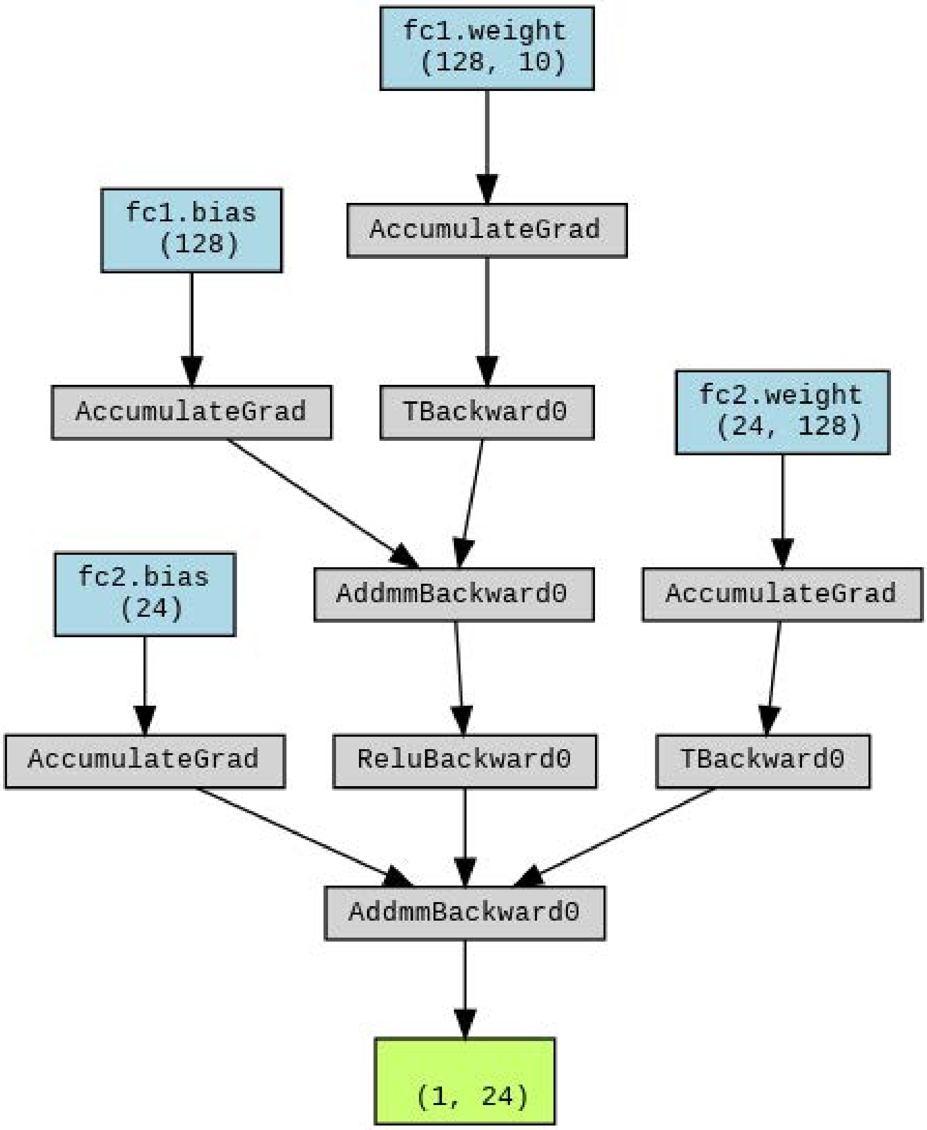


Figure 7.7: Multilayer Perceptron Architecture

dimensions of feature maps, thereby diminishing computational complexity and augmenting the model's capacity for generalization. The culmination of these operations transpires in fully connected layers positioned at the conclusion of the network, where the amalgamated features facilitate conclusive predictions. The applicability of CNNs extends across diverse computer vision domains, manifesting notable success in tasks such as image classification, object detection, and image segmentation.

In timeseries forecasting, CNNs adapt to sequential data using 1D convolutional layers. These layers analyze temporal patterns, aided by pooling layers for downsizing. CNNs efficiently capture short and long-term dependencies, making them valuable for tasks such as stock price prediction and weather forecasting, showcasing their versatility across diverse

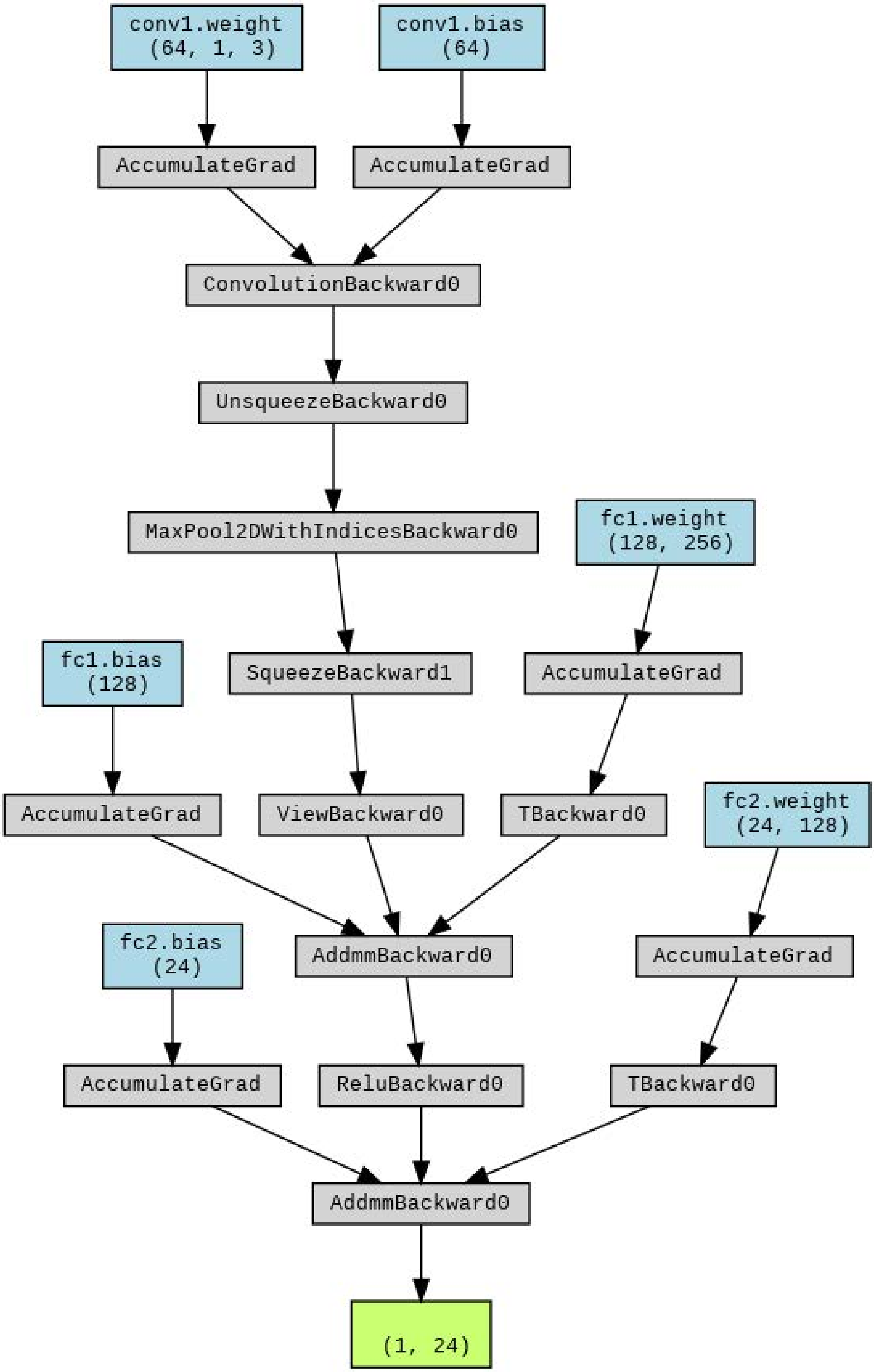


Figure 7.8: CNN model architecture

data types.

Ensemble Learning Model

The Ensemble Learning Model (ELM), comprising a Multilayer Perceptron (MLP) and a Convolutional Neural Network (CNN), operates by independently training both models on a given dataset and then combining their predictions through weighted averaging. The MLP focuses on learning non-linear relationships, while the CNN excels at extracting hierarchical features. The weights assigned to each model in the ensemble are determined based on their performance, enhancing the contribution of the better-performing model. The final prediction is generated by summing the weighted predictions, aiming to capitalize on the complementary strengths of the MLP and CNN for improved predictive accuracy and robustness across diverse data patterns. Figure 7.9 visualizes the main body of ELM created in this section.

7.3.2 Deep Transfer Learning Forecasting Strategies

The basic idea on which the four implemented strategies are based is the extended application and exploration of Deep Transfer Learning in three different datasets, finding the best method, in combination with the creation of powerful DTL forecasting strategies and tools with robust generalization capabilities. The first forecasting strategy, named Deep Transfer Learning Case 1 (DTL Case 1), involves training each of the three DL models exclusively on the Lesvos dataset, with fine-tuning carried out using the Chios dataset. In the second strategy, Deep Transfer Learning Case 2 (DTL Case 2), both Lesvos and Rhodes datasets are used concurrently during the DL model training phase, followed by fine-tuning using the Chios dataset. The third strategy, Deep Transfer Learning Case 3 (DTL Case 3), entails training the DL models solely on the Lesvos dataset, with the testing phase conducted directly on the Chios dataset, without any fine-tuning. Lastly, in the Multi-task Deep Learning application strategy (MTDL), each of the three DL models are trained simultaneously on all three datasets, with final predictions made on the unused portion of the Chios dataset.

The way in which the available data were used in order to implement the DTL strategies and the MTDL strategy is described below:

- For DTL Case 1, only the dataset of Lesvos was used for the first training phase of the models, specifically for the time period from 2019-01-01 01:00:00 to 2021-12-31

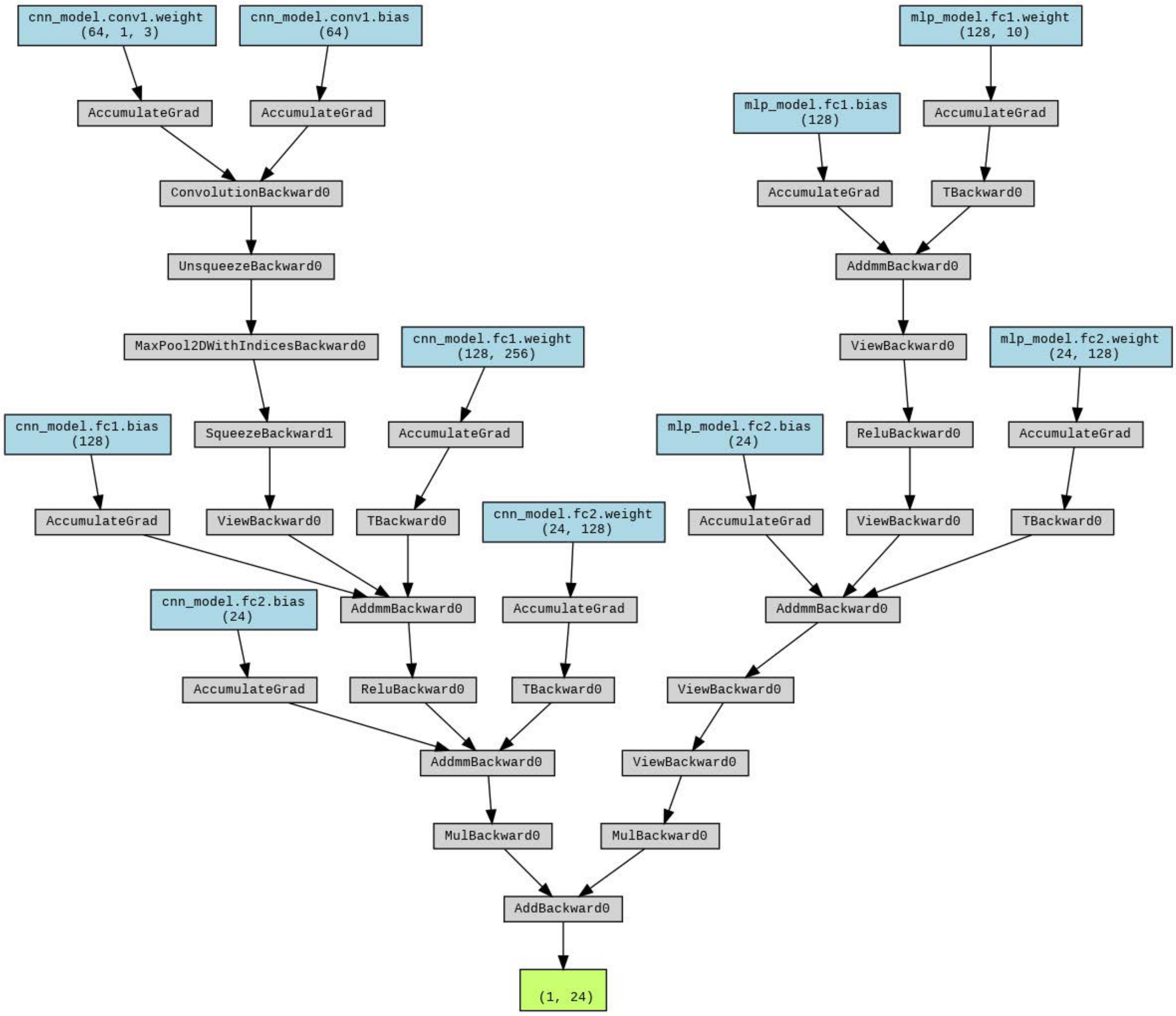


Figure 7.9: Ensemble Learning Model Architecture

23:00:00. Then, for the second phase, i.e., fine-tuning, the time period from 2019-01-01 01:00:00 to 2021-12-31 23:00:00 of the Chios dataset is utilized.

- For DTL Case 2, the first training phase of the models was based on the datasets from Lesvos and Rhodes, and more specifically, for the time period from 2019-01-01 01:00:00 to 2021-12-31 23:00:00 for each dataset. Then, for the second phase, i.e., fine-tuning, the same time period as in DTL Case 1 is used, from 2019-01-01 01:00:00 to 2021-12-31 23:00:00 of the Chios dataset.

- For DTL Case 3, the training dataset that was used covers the time period from 2019-01-01 01:00:00 to 2022-12-31 23:00:00 from the Lesvos dataset without fine tuning. The choice of Lesvos alone was made because it exhibits a higher correlation and similarity with the corresponding dataset of the Chios timeseries, compared to the Rhodes island.

- For MTDL, the datasets from all three islands were used simultaneously for training. Specifically, the datasets of Rhodes and Lesvos for the time period 2019-01-01 01:00:00 to 2022-12-31 23:00:00, and for Chios dataset, for the time period 2019-01-01 01:00:00 to 2021-12-31 23:00:00 were used.

Deep Transfer Learning is a field of Transfer Learning that entails utilizing knowledge gained from solving one task to make predictions on another related task, employed Deep Neural Networks. This field often consists of pre-training a neural network on a source task with abundant labeled data and subsequently applying the acquired knowledge to a target task characterized by a foreign dataset. Two prevalent scenarios in transfer learning include domain adaptation, where the source and target tasks have the same input space but differ in output spaces, and feature extraction, where the source and target tasks share similar input and output spaces.

In the realm of timeseries approaches, is valuable for several reasons. Timeseries data often exhibits complex patterns, trends, and seasonality, and acquiring labeled data for training deep models can be challenging due to limited availability. It allows a model pre-trained on a source timeseries task to capture generic temporal features and representations that can be beneficial for a target task. The learned features can serve as a useful initialization for the target task, reducing the need for extensive training data and potentially enhancing the model's ability to generalize to new patterns. Additionally, transfer learning is particularly advantageous when the source and target tasks share similar temporal characteristics, enabling the model to transfer knowledge effectively and improve its performance on the target task. This approach is especially relevant in situations where collecting large amounts of labeled data for every specific task is impractical or costly.

In this investigation, DTL strategies with and without fine-tuning are developed. Initially, the two cases of DTL with fine-tuning are examined (DTL Case 1 and 2) and are analyzed below. Subsequently, the scenario where the pre-trained model, as configured without fine-tuning, is examined (DTL Case 3). Finally, the MTDL strategy, involving the simultaneous training of each DL model on the datasets of the three islands, is presented. The four proposed forecasting strategies implemented are analyzed and presented in detail below.

Deep Transfer Learning Case 1

In this methodology a DNN pre-trained on a source task is adjusted to perform a related target task. Initially trained on a large dataset for a general task, such as laod forecasting, the pre-trained model captures broad features. This knowledge is then transferred to a target task, which is a foreign dataset. During fine-tuning, the model's weights, especially in the deeper layers, are adjusted based on the target task's data, allowing the model to adapt its learned representations to task-specific characteristics. This approach is particularly advantageous when the target task has also limited labeled data, enabling the model to leverage the knowledge gained from the source task and enhance its performance on the target task.

In DTL Case 1, the testing dataset of Chios is used, and the training dataset consists only of the timeseries of Lesvos island. This approach is followed due to higher correlation between Chios and Lesvos compared to Rhodes, influencing the selection of training data for better model performance. After, the training period, fine-tuning is performed on the training parts of each model on the dataset of Chios, creating the final fine-tuned models. Finally, these models are used for Chios day-ahead load forecasting. Graphically, Figure 7.10 visualizes the DTL Case 1 strategy.

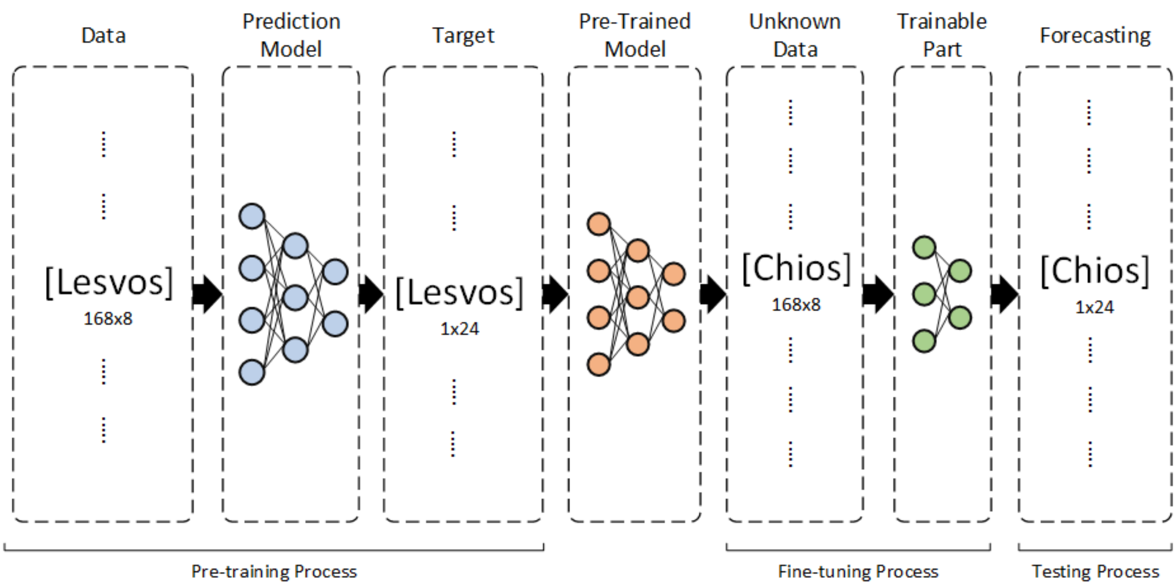


Figure 7.10: Deep Transfer Learning Case 1

Deep Transfer Learning Case 2

Similarly with DTL Case 1, for DTL Case 2, the two datasets from Rhodes and Lesvos are merged, and the three models are trained on the combined training dataset. The trainable part of the pre-trained model is fine-tuned in the dataset of Chios. Finally, predictions are made on the unused dataset from Chios island. Figure 7.11 visualizes the DTL Case 2 strategy.

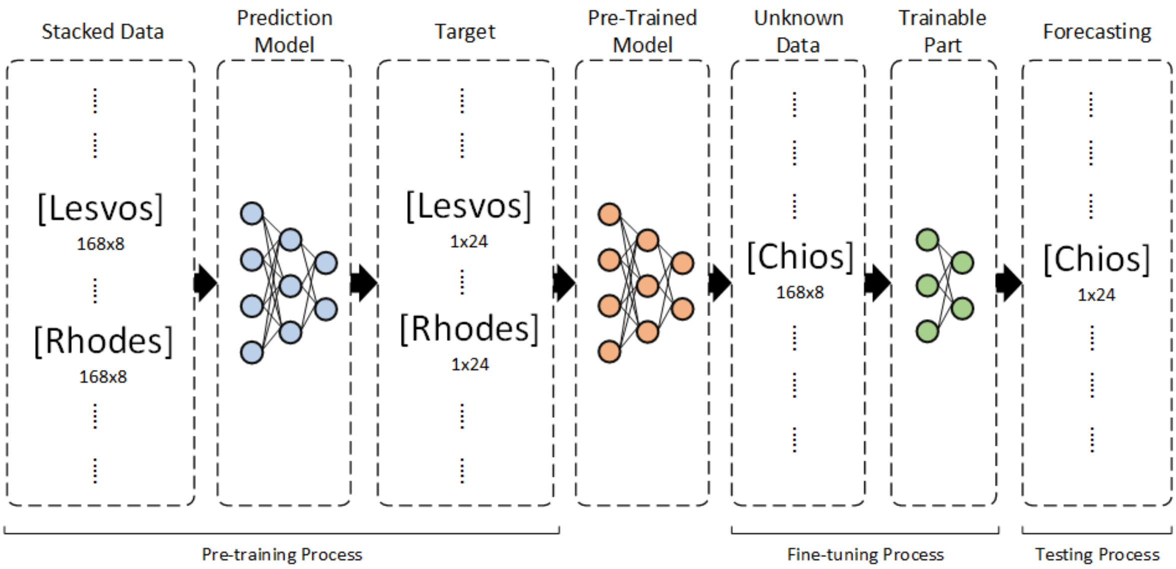


Figure 7.11: Deep Transfer Learning Case 2

Deep Transfer Learning Case 3

Deep Transfer learning without fine-tuning involves a two-step process. Initially, a Deep Neural Network is trained on a source task using a substantial dataset, learning hierarchical features relevant to that task. Subsequently, in the transfer phase, the pre-trained model is utilized with the exact same structure as it was formed during the training period, in order to create predictions for a target task. The learned features are extracted without further adjusting the model's weights, and these fixed representations serve as input to a new task-specific regressor trained on the target task's dataset.

This approach proves advantageous when the target task possesses limited labeled data, as it allows for knowledge transfer from a source task without the computational overhead of fine-tuning the entire model. By utilizing the pre-trained model as a feature extractor, the knowledge encapsulated in the generic representations can be harnessed for tasks that share similar low-level features and structures, promoting effective knowledge transfer across related tasks while mitigating the need for task-specific fine-tuning.

In DTL Case 3 strategy, the three DL models are trained on the dataset of Lesvos, and subsequently, predictions are made directly on the Chios day-ahead load, without fine-tuning process. This strategy is used in order to draw secure conclusions regarding the approaches with and without fine-tuning. The DTL Case 3 strategy is presented in Figure 7.12.

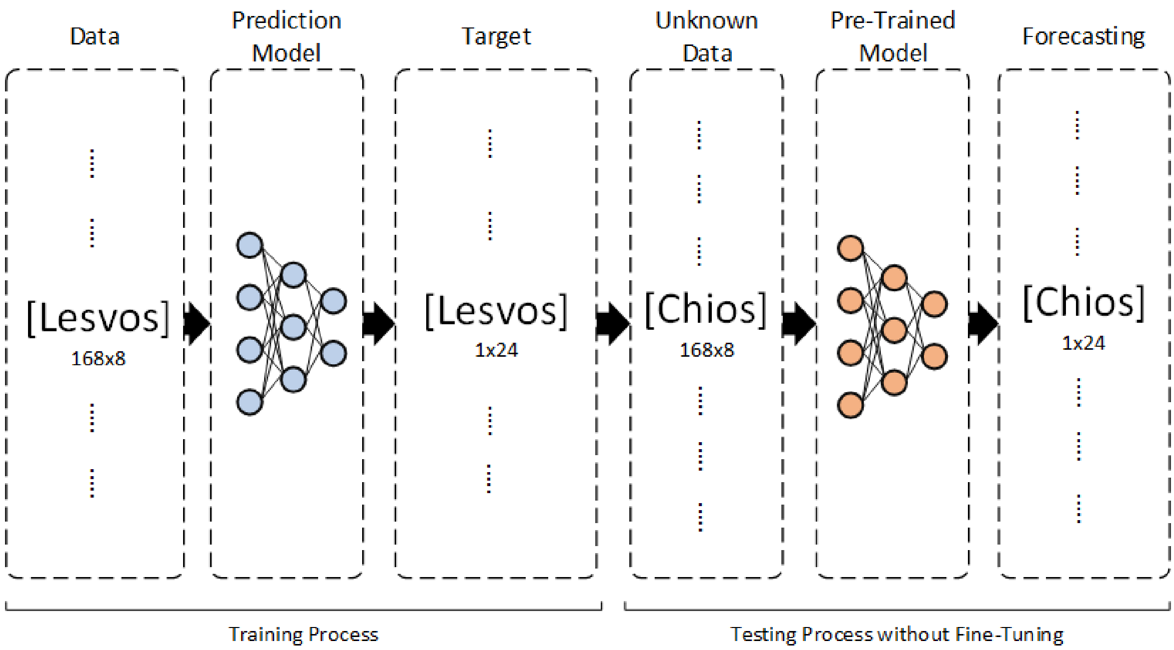


Figure 7.12: Deep Transfer Learning Case 3

Multi-task Deep Learning

Multi-task Deep Learning (MTDL) is a methodology in Deep Neural Networks (DNN), where a DNN model is trained simultaneously on multiple datasets, leveraging the shared knowledge across the different datasets to improve the model's generalization performance. The underlying principle of MTDL is to exploit the relationships and commonalities among distinct but related tasks, allowing the model to learn a shared representation that captures the inherent structure present in the data. Essentially, a unified dataset is created which is a concatenation of different datasets, where each of them corresponds to a specific output. During training, the model optimizes its parameters by jointly minimizing the loss across all tasks. The shared representation learned across tasks facilitates the transfer of knowledge between them, leading to enhanced generalization performance, particularly in scenarios where individual tasks lack sufficient data for robust learning. The success of MTDL lies in its ability to induce a form of regularization, encouraging the model to discover and focus on relevant features that are beneficial for multiple tasks simultaneously.

Each of the training tasks has its own objective function, and the model learns to jointly optimize these objective functions. The general idea is to share information between tasks to improve overall performance. Mathematically, the following applies:

- N is the total number of tasks.
- X are the input data.
- Y_i is the output for task i.
- θ are the parameters of the neural network model.

For each task i, there is an associated loss function $L_i(\theta)$ that measures the error between the predicted output and the true output for that task. The overall loss function for all tasks can be defined as a combination of the individual task loss functions, often using a weighted sum.

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} \alpha_i L_i(\theta) \tag{7.4}$$

where α_i are optional weighting factors and Li represents the loss for the $i - th$ task.

The goal is to minimize this overall loss function with respect to the model parameters θ. The minimal value of the loss function, θ^*, is given below:

$$\theta^* = \arg\min_{\theta} \mathcal{L}(\theta) \tag{7.5}$$

The model parameters are then updated using gradient descent or other optimization techniques to minimize this combined loss. The shared representation in the intermediate layers allows the model to discover commonalities and relationships among tasks, promoting a more generalized feature extraction process. By training on diverse datasets simultaneously, MDTL facilitates the development of a model that not only excels in individual tasks but also demonstrates improved performance on new, foreign data.

In the strategy employed in this work, which is presented in Figure 7.13, training is conducted simultaneously on all three different datasets of the islands, Rhodes, Chios and Lesvos. The objective is for the model to acquire high generalization capabilities and make predictions on the testing dataset selected of Chios. Due to the distinct variations in the three timeseries, this approach proves more robust than cases involving singular training, imparting generality in performance across the models.

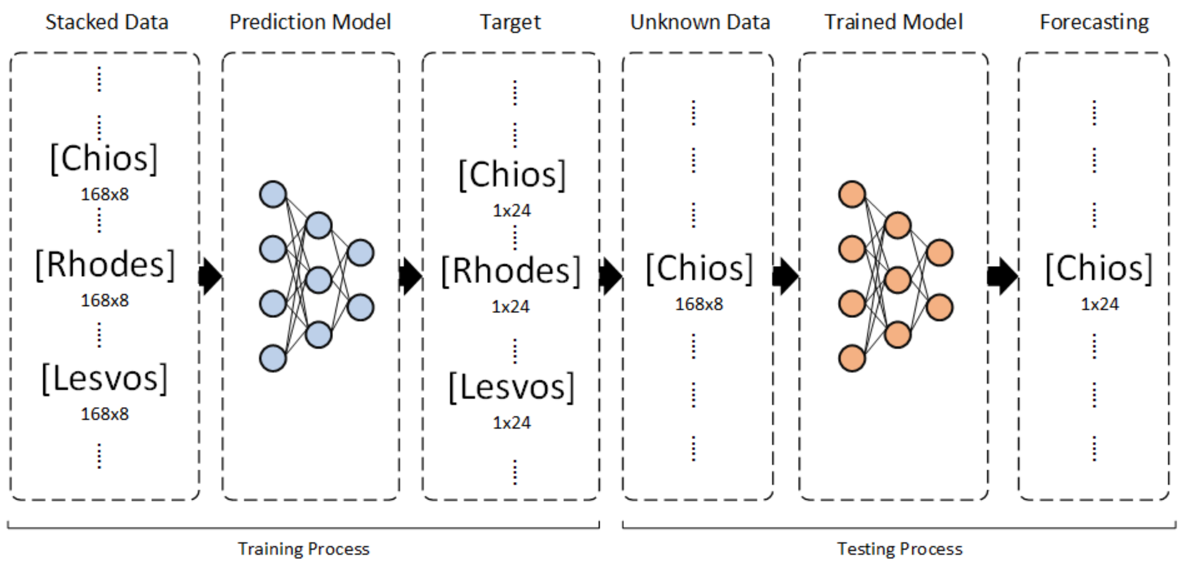


Figure 7.13: Multi-task Deep Learning

7.3.3 Optimization Framework

The Bayesian Optimization Algorithm (BOA) is utilized for each training period and for each model. It is a probabilistic optimization approach designed to tackle complex and computationally expensive objective functions. Central to BOA is the use of a surrogate model, typically a Gaussian Process, which provides a probabilistic representation of the unknown objective function. This surrogate model captures both the mean and uncertainty associated with the objective function across the parameter space. The algorithm iteratively refines its understanding of the objective function by selecting points for evaluation based on an acquisition function that balances exploration and exploitation. The chosen points are then used to update the surrogate model through Bayesian inference, adjusting the model's predictions in light of the new information. This iterative process allows BOA to systematically explore the parameter space, adapt to the underlying structure of the objective function, and efficiently converge towards optimal solutions.

BOA excels in making informed decisions using the uncertainty measured by its surrogate model. An acquisition function guides the algorithm to explore areas where the objective function is uncertain or is likely to have optimal values. As the optimization progresses, the surrogate model of BOA improves continuously, enhancing its understanding of the objective function and focusing the search on regions most likely to contain the global optimum. This principled approach makes it particularly well-suited for optimization problems in scientific and engineering domains where objective function evaluations are resource-intensive or subject to noise. It efficiently identifies optimal parameter configurations in such scenarios. In this study, the architecture of each DL model was optimized and configured through BOA. For each case, each model has been saved in h5 format. Each DL model can operate using online data. Regarding the hyperparameters of each DL model, Table 7.1 presents in detail each hyperparameter, the value of which was optimized in order to achieve the most accurate predictions.

Table 7.1: Models optimization hyperparameters.

Type	**Deep Learning Model**	**Input Sequence Length**	**Model Hyperparameters**	**Optimization Function**
DL	MLP	168 h	• Dense layer with search space of Neurons: min_value = 16, max_value = 512, step = 16 • Dropout layer with search space: min_value = 0, max_value = 0.25, step = 0.05 • Optimizer = Adam • Learning rate with search space: min_value = 0.0010, max_value = 0.010, sampling = “log”	Validation loss of MSE

Continued on next page...

Table 7.1 – continued from previous page.

Type	**Deep Learning Model**	**Input Sequence Length**	**Model Hyperparameters**	**Optimization Function**
DL	CNN	168 h	• Filters of CNN with range: min_value = 64, max_value = 128, step = 16 • CNN kernel size with search space: min_value = 4, max_value = 8, step = 2 • Neurons of Dense Layer with search space: min_value = 24, max_value = 120, step = 12 • Optimizer = Adam • Learning rate with search space: min_value = 0.0010, max_value = 0.010, sampling = “log”	Validation loss of MSE

Continued on next page...

Table 7.1 – continued from previous page.

Type	Deep Learning Model	Input Sequence Length	Model Hyperparameters	Optimization Function
DL	ELM	168 h	• Dense layer with search space of Neurons: min_value = 16, max_value = 512, step = 16 • Dropout layer with search space: min_value = 0, max_value = 0.25, step = 0.05 • Filters of CNN with range: min_value = 64, max_value = 128, step = 16 • CNN kernel size with search space: min_value = 4, max_value = 8, step = 2 • Neurons of Final Dense Layer with search space: min_value = 24, max_value = 48, step = 4 • Optimizer = Adam • Learning rate with search space: min_value = 0.0010, max_value = 0.010, sampling = "log"	Validation loss of MSE

7.3.4 Evaluation Metrics

For this work, the following four error evaluation metrics are used.

Mean Absolute Error (MAE): In this metric, the average of the absolute differences between the forecasted and true values is calculated.

Root Mean Squared Error (RMSE): This metric calculates the square root of the average of the squared differences between the forecasted and true values.

Mean Absolute Percentage Error (MAPE): This metric computes the average of the absolute percentage differences between the predicted and actual values.

R-squared (R^2): Is a statistical metric that measures how the independent variable(s) in a forecasting model explain the variation of the dependent variable. It takes values between 0 and 1. 1 belongs to a satisfactory fit, meaning all variation in the dependent variable is explained by the independent variable(s). 0 demonstrates zero connection between the variables.

The above metrics are defined as follows:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{7.6}$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - x_i)^2}{n}} \tag{7.7}$$

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} |\frac{y_i - x_i}{x_i}| \tag{7.8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{x}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{7.9}$$

where x_i, y_i, $\hat{x}_i$ and $\bar{y}$ are the forecasting, the actual, the mean of the forecasting values and the mean of the actual values, respectively.

7.3.5 Software environment

All the algorithms in this work were developed using the Python 3.10 programming language. The open source software library Tensorflow 2.15.0 and the Keras 2.15.0 high-level API were used to train and test the deep learning algorithms. Furthermore, the Pandas 2.1.0 and Numpy 1.26.0 libraries were used for data analysis. For visualization purposes in exploratory analysis and prediction results, the Seaborn, Plotly, Matplotlib, graphviz and torchviz libraries were incorporated. Also, the official Calendar library was used, in order to identify the weekends and the Greek holidays. The research was carried out on the Google Colab Pro environment. For the MLP model the GPU Tesla T4 with specific characteristics: NVIDIA-SMI 535.104.05 Driver Version: 535.104.05 CUDA Version: 12.2, RAM: 12.7 GB, and disk space: 78.2 GB was used. For the CNN and Ensemble model a cloud TPU 28.6 GB RAM abd 107.7 GB disc space was employed.

7.4 Results Analysis

In this section, the experimental results obtained are presented. First, only the mean variance of the Mean Absolute Error (MAE) on a monthly basis for the MLP, CNN, and ELM, respectively, is presented for the economy of space. Then, MAPE (Mean Absolute Percentage Error), RMSE (Root Mean Square Error) and the aggregated results are visualized, analyzed, and compared for each strategy and each model, on a monthly basis. Finally, the tables detailing R^2 are presented in the aggregated results.

7.4.1 Multilayer Perceptron Results

Figure 7.14 visualizes the variation of MAE, on a monthly basis, for each of the four strategies, for the MLP model. It is observed that May and September exhibit better prediction results, while January, July, and August are the months with poorer performance.

7.4.2 Convolutional Neural Networks Results

Figure 7.15 presents the variation of MAE, on a monthly basis, for each of the four strategies, for CNN model. It clearly seems that April and May exhibit better average prediction results, while June, July, and August are the months with the worst performance.

7.4.3 Ensemble Learning Model Results

Figure 7.16 visualizes the variation of MAE, on a monthly basis, for each of the four strategies, for the Ensemble Learning model. It is observed that June, July, and August exhibit

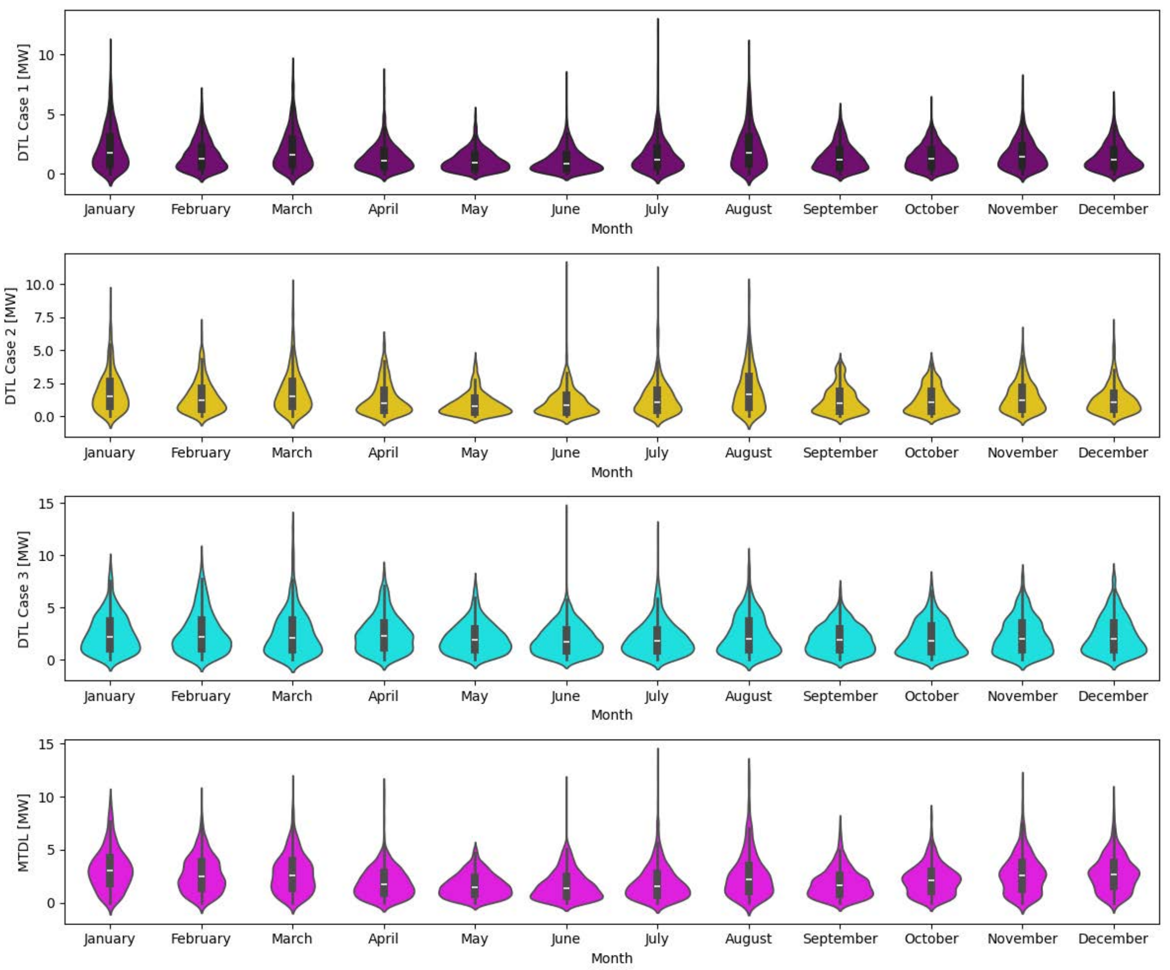


Figure 7.14: Monthly Variation of MAE for MLP model

better average prediction results, while May, September, and October are the months with the lowest prediction accuracy.

7.4.4 Results Comparison

In order to make an evaluation of the performance of each model, the results are consolidated and compared for each of the four strategies that were pursued in the following radial graph. This particular graph was chosen because it constitutes an illustrated and clear approach in presenting results for each of the twelve months of the year. In other words, the closer a point is to the center of the circle, the lower the prediction error achieved.

Variation of MAE for each strategy

Figure 7.17 presents the comparison month by month of Mean Absolute Error (MAE) of the DL models for each strategy.

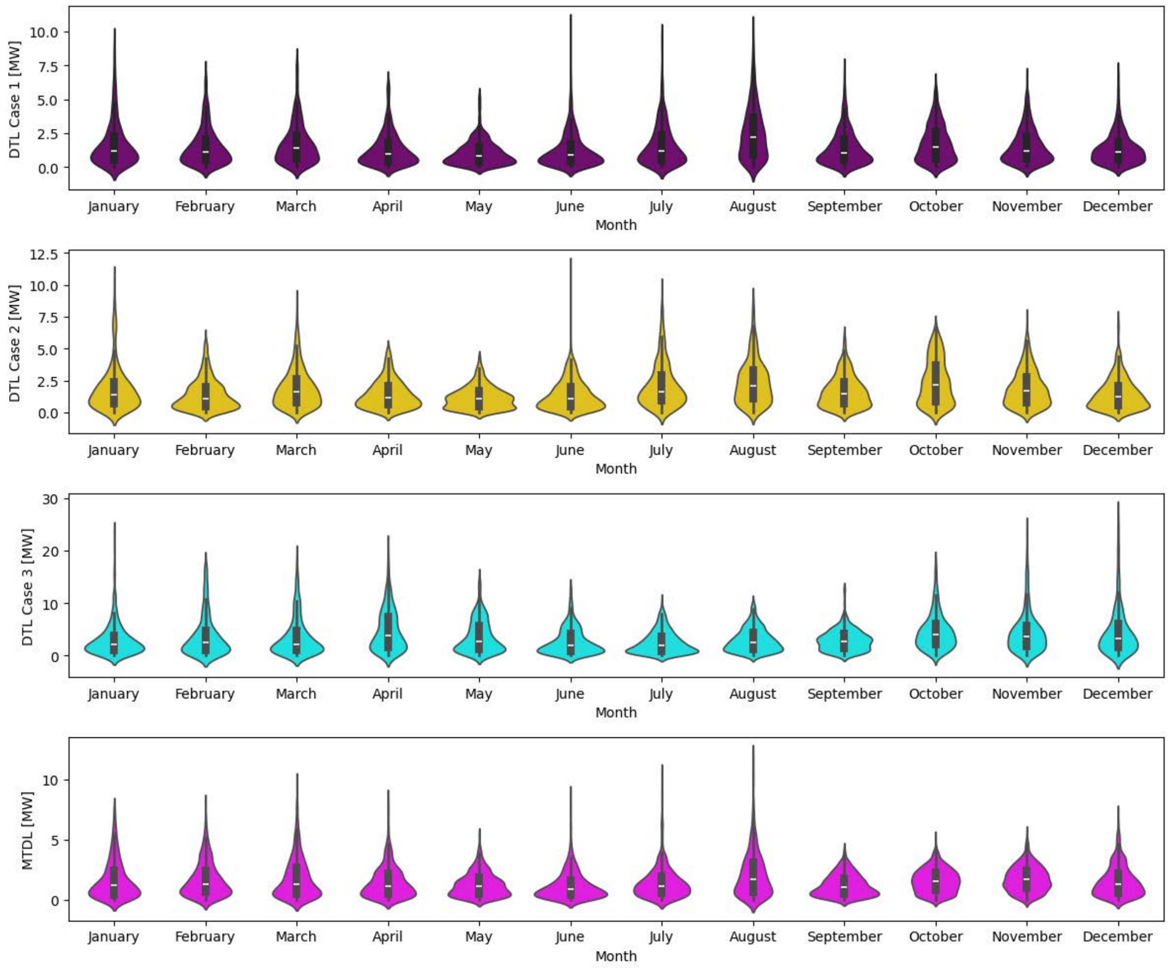


Figure 7.15: Monthly Variation of MAE for CNN model

Based on the comparative results, the following observations emerge:

- For the DTL strategy with fine-tuning and training data from the Lesvos timeseries, DTL Case 1, it is observed that the MLP and Ensemble models outperform the CNN model.

- In the forecasting strategy with fine-tuning and training data from both the Rhodes and Lesvos timeseries, DTL Case 2, it seems that the MLP and Ensemble models exhibit comparable performance, except for October, where MLP outperforms.

- Regarding for the use of pre-trained models without fine-tuning, DTL Case 3, the CNN significantly lags behind the other two models, with MLP consistently exhibiting the highest prediction accuracy.

- Finally, regarding the Multi-task Deep Learning application strategy, MTDL, it is observed that the MLP model consistently shows inferior results compared to the other

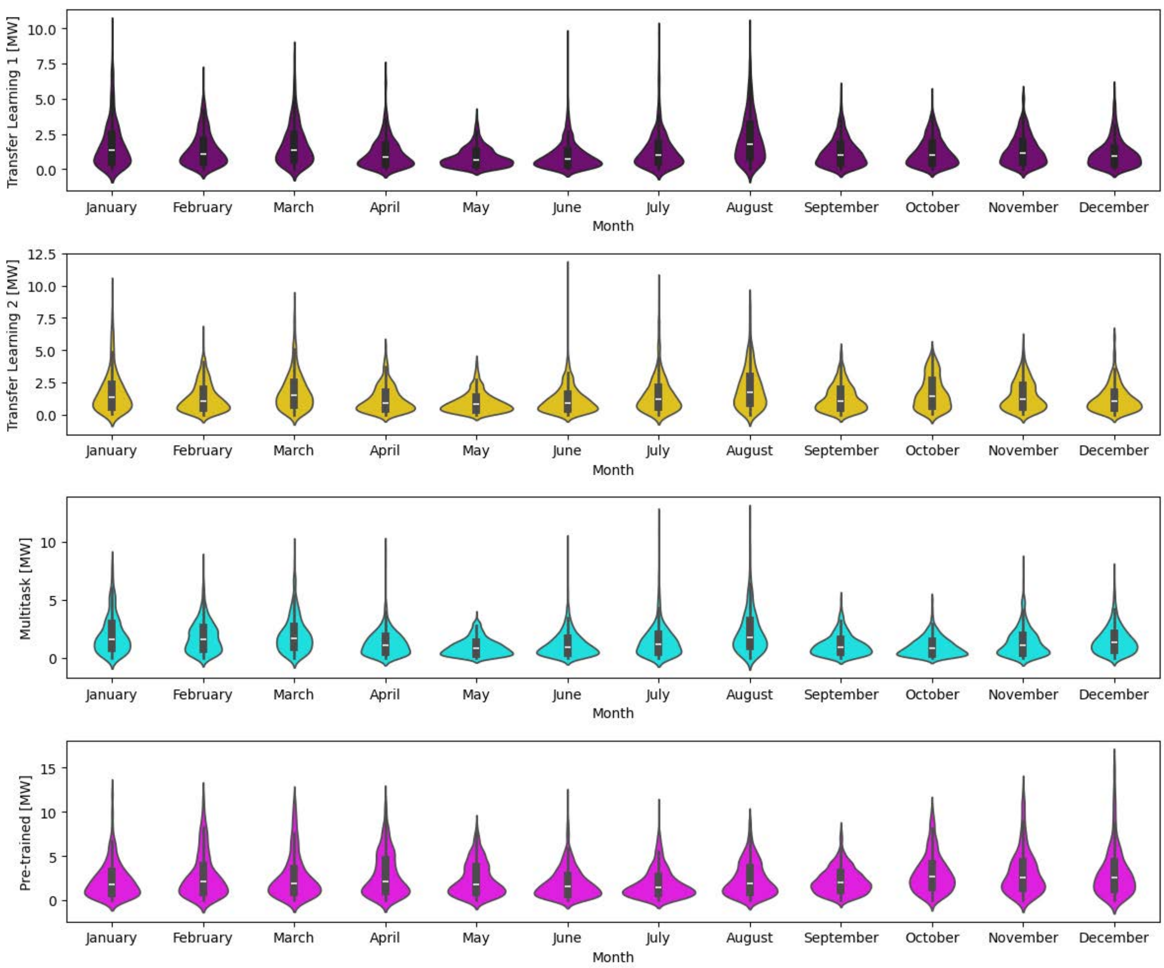


Figure 7.16: Monthly Variation of MAE for ELM model

two models for all months. Here, CNN and Ensemble achieve similar accuracy.

For additional analysis and understanding of the behavior of the algorithms, Figures 7.18 and 7.19 plot the variation of the Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) for each month, respectively.

Variation of MAPE for each strategy

Figure 7.18 presents the comparison of Mean Absolute Percentage Error (MAPE) for each model and strategy, on a monthly basis.

It is clearly observed that for the DTL Cases 1 and 2, June exhibits the best prediction accuracy, with the Ensemble model achieving a MAPE of 5.29% for Case 1 and the MLP achieving a MAPE of 6.01% for Case 2. In the case of MTDL, the best predictions are observed in the months of June and July, with the ELM presenting predictions of 6.33% and 6.86%, respectively. Finally, for the DTL Case 3, the best prediction is observed in January,

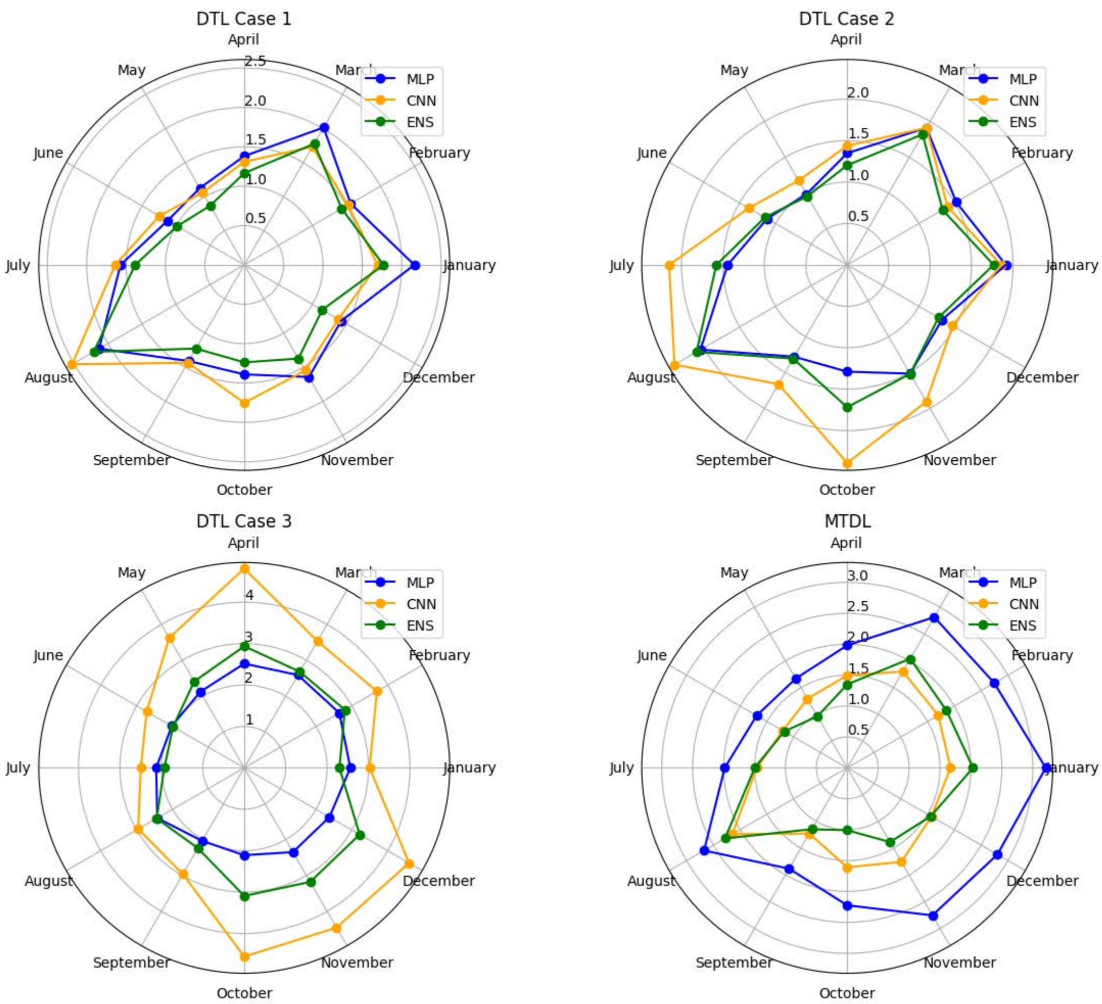


Figure 7.17: Comparison of monthly MAE of each model for each forecasting strategy

corresponding to a MAPE of 7.85%, achieved by the ELM model.

Variation of RMSE for each strategy

Figure 7.19 presents the comparison of the root mean square error (RMSE) for each model and strategy, monthly.

7.4.5 Aggregated Results

This subsection provides a analyzes the results obtained for each forecasting strategy and each DL model. Tables 7.2, 7.3, 7.4, and 7.5 present all aggregated results for each forecasting strategy, to provide a complete view of the performance of each algorithm and for each forecasting month. MAE and RMSE metrics are given in MWs, MAPE is given in percent (%), and R^2 takes values between 0 and 1.

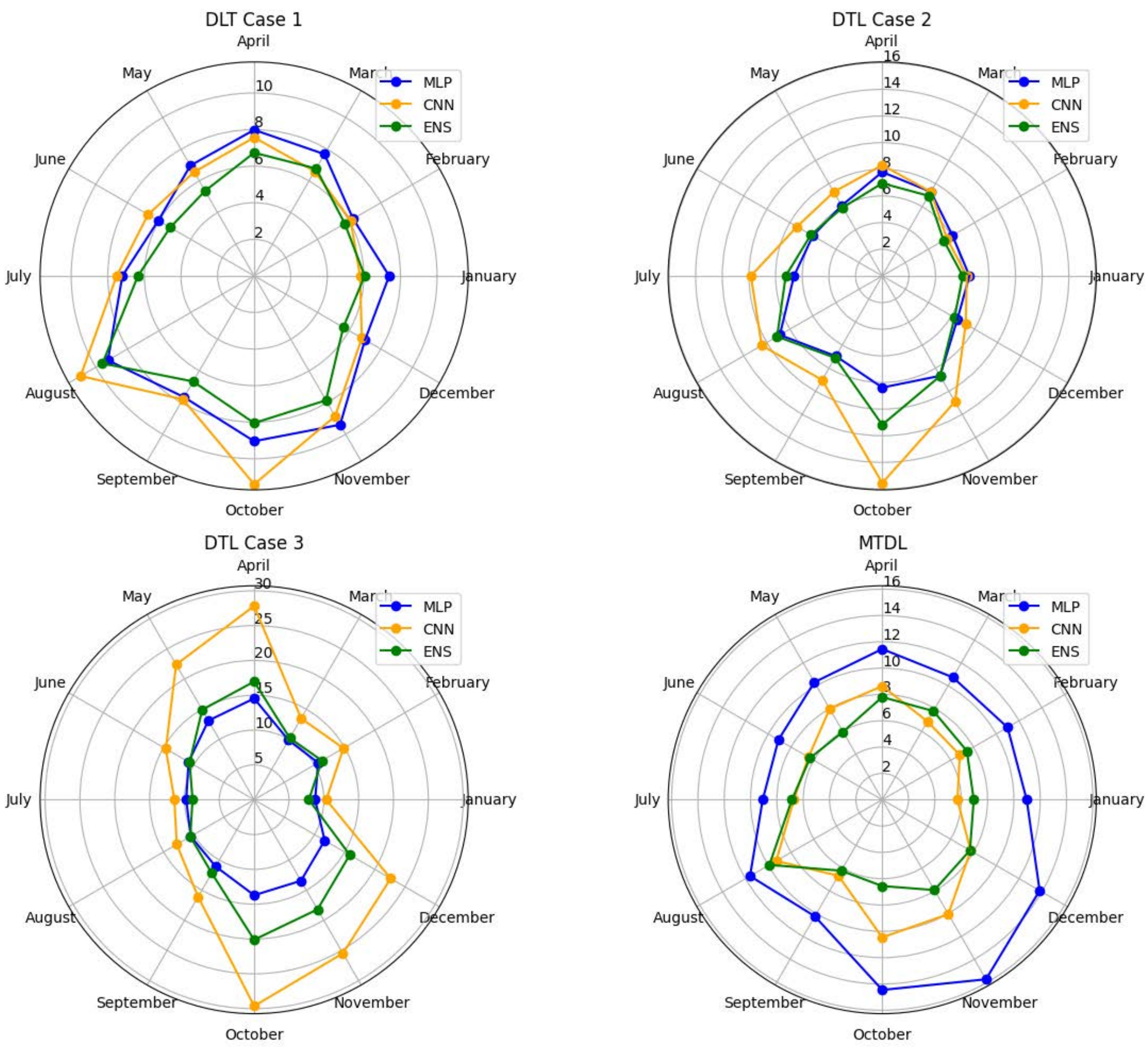


Figure 7.18: Comparison of monthly MAPE of each model for each forecasting strategy

Table 7.2: Deep Transfer Learning Case 1

Model	Metric	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
MLP	MAE	2.16	1.56	2.01	1.38	1.12	1.12	1.57	2.12	1.40	1.39	1.64	1.42
	MAPE	7.38	6.24	7.68	7.98	6.97	6.05	7.01	9.23	7.66	9.04	9.39	7.00
	RMSE	2.79	1.98	2.59	1.78	1.41	1.48	2.15	2.76	1.78	1.70	2.06	1.79
	R^2	0.83	0.87	0.82	0.67	0.73	0.74	0.63	0.08	0.75	0.53	0.70	0.82
CNN	MAE	1.69	1.52	1.73	1.31	1.06	1.24	1.63	2.52	1.43	1.75	1.54	1.38
	MAPE	5.57	6.07	6.60	7.57	6.55	6.69	7.46	10.95	7.79	11.39	8.85	6.79
	RMSE	2.37	2.03	2.28	1.78	1.38	1.74	2.25	3.19	1.90	2.20	1.98	1.79
	R^2	0.88	0.87	0.86	0.68	0.74	0.64	0.59	0.11	0.70	0.21	0.73	0.82
ELM	MAE	1.76	1.42	1.77	1.16	0.86	0.98	1.38	2.20	1.22	1.23	1.37	1.14
	MAPE	5.87	5.72	6.77	6.73	5.36	5.29	6.24	9.57	6.67	8.04	7.82	5.62
	RMSE	2.42	1.84	2.31	1.57	1.12	1.36	1.91	2.80	1.58	1.57	1.73	1.51
	R^2	0.88	0.89	0.86	0.75	0.83	0.78	0.71	6.25	0.80	0.61	0.80	0.87

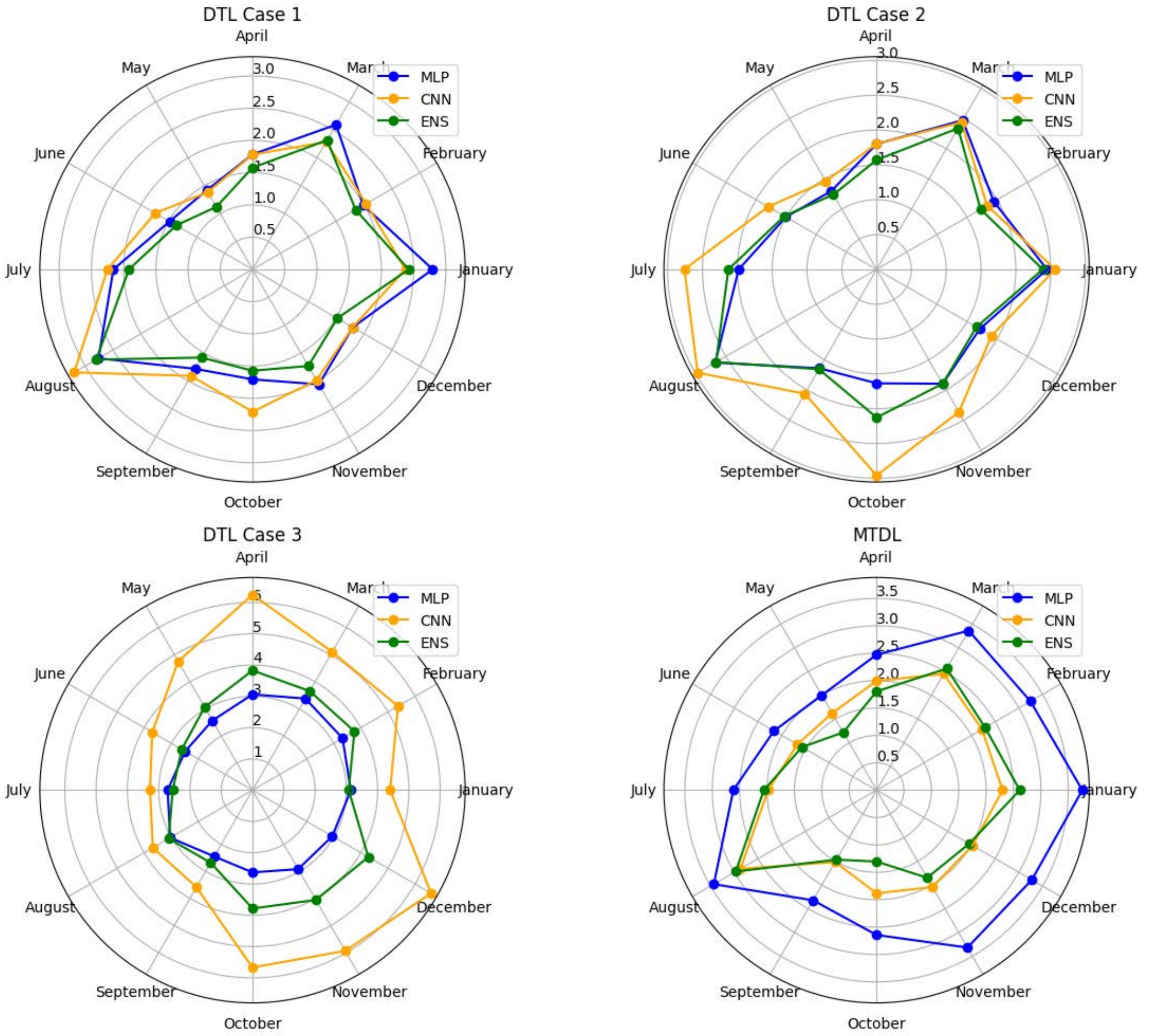


Figure 7.19: Comparison of monthly RMSE of each model for each forecasting strategy

Table 7.3: Deep Transfer Learning Case 2

Model	Metric	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
MLP	**MAE**	1.91	1.51	1.91	1.34	0.98	1.11	1.44	2.04	1.28	1.28	1.51	1.32
	MAPE	6.50	6.07	7.29	7.79	6.24	6.01	6.61	8.87	6.95	8.38	8.66	6.51
	RMSE	2.44	1.94	2.47	1.80	1.29	1.50	1.98	2.67	1.64	1.63	1.89	1.71
	R^2	0.87	0.88	0.84	0.67	0.77	0.73	0.69	0.15	0.78	0.57	0.75	0.84
CNN	**MAE**	1.85	1.40	1.91	1.43	1.17	1.37	2.14	2.40	1.66	2.39	1.90	1.46
	MAPE	6.26	5.60	7.29	8.30	7.27	7.38	9.85	10.44	9.02	15.51	10.90	7.23
	RMSE	2.56	1.83	2.43	1.80	1.46	1.79	2.76	2.96	2.06	2.96	2.36	1.90
	R^2	0.86	0.89	0.84	0.67	0.71	0.62	0.39	0.02	0.66	0.08	0.61	0.80
ELM	**MAE**	1.76	1.32	1.82	1.20	0.95	1.14	1.57	2.10	1.30	1.71	1.51	1.27
	MAPE	6.03	5.31	6.94	6.95	5.92	6.17	7.23	9.10	7.09	11.17	8.68	6.26
	RMSE	2.38	1.72	2.33	1.57	1.23	1.52	2.12	2.66	1.65	2.12	1.89	1.65
	R^2	0.88	0.91	0.85	0.75	0.80	0.72	0.64	0.16	0.78	0.27	0.75	0.85

Table 7.4: Deep Transfer Learning Case 3

Model	Metric	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
MLP	**MAE**	2.56	2.65	2.58	2.51	2.12	2.01	2.12	2.42	2.03	2.10	2.34	2.37
	MAPE	8.75	10.60	9.87	14.54	13.12	10.86	9.74	10.51	11.05	13.71	13.43	11.69
	RMSE	3.14	3.32	3.38	3.06	2.56	2.48	2.71	3.03	2.44	2.62	2.91	2.94
	R^2	0.79	0.64	0.69	0.05	0.10	0.27	0.41	-0.10	0.52	-0.11	0.41	0.51
CNN	**MAE**	3.02	3.69	3.53	4.81	3.62	2.71	2.49	2.95	2.96	4.55	4.45	4.59
	MAPE	10.34	14.79	13.48	27.82	22.40	14.63	11.45	12.82	16.09	29.63	25.47	22.62
	RMSE	4.39	5.37	5.10	6.25	4.75	3.70	3.27	3.67	3.58	5.66	5.93	6.59
	R^2	0.58	0.05	0.29	-2.96	-2.08	-0.64	0.14	-0.61	-0.05	-4.18	-1.45	-1.45
ELM	**MAE**	2.29	2.80	2.68	2.93	2.41	1.98	1.92	2.44	2.23	3.09	3.18	3.21
	MAPE	7.85	11.22	10.22	16.96	14.91	10.68	8.84	10.58	12.16	20.10	18.18	15.84
	RMSE	3.07	3.76	3.65	3.82	3.06	2.61	2.55	3.09	2.67	3.78	4.05	4.27
	R^2	0.80	0.54	0.64	-0.48	-0.28	0.19	0.48	-0.14	0.42	-1.30	-0.14	-0.03

Table 7.5: Multi-task Deep Learning

Model	Metric	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
MLP	**MAE**	3.21	2.74	2.81	1.98	1.66	1.69	1.98	2.67	1.88	2.22	2.75	2.80
	MAPE	10.97	10.99	10.73	11.45	10.31	9.10	9.06	11.50	10.23	14.45	15.75	13.80
	RMSE	3.75	3.24	3.36	2.47	2.00	2.16	2.61	3.43	2.32	2.64	3.31	3.27
	R^2	0.70	0.66	0.69	0.38	0.46	0.44	0.45	-0.40	0.56	-0.12	0.24	0.40
CNN	**MAE**	1.67	1.69	0.80	1.49	1.28	1.20	1.46	2.14	1.22	1.60	1.75	1.57
	MAPE	5.62	6.78	6.86	8.63	7.97	6.47	6.58	9.30	6.66	10.46	10.03	7.75
	RMSE	2.30	2.22	2.45	1.99	1.62	1.67	1.98	2.87	1.51	1.88	2.04	2.02
	R^2	0.89	0.84	0.84	0.60	0.64	0.67	0.69	0.01	0.82	0.43	0.70	0.77
ELM	**MAE**	2.03	1.85	2.03	1.35	0.96	1.17	1.49	2.28	1.14	1.00	1.38	1.56
	MAPE	6.86	7.42	7.77	7.80	5.96	6.33	6.86	9.89	6.22	6.56	7.92	7.69
	RMSE	2.61	2.28	2.56	1.80	1.21	1.57	2.05	2.96	1.46	1.30	1.84	1.95
	R^2	0.85	0.83	0.82	0.67	0.80	0.71	0.66	-0.05	0.83	0.73	0.77	0.79

Based on the results presented in the above tables, the following summarizations apply:

- In DTL Case 1, it is observed that the Ensemble model achieves the best prediction for the month of June, presenting an MAPE of 5.29%.
- In DTL Case 2, again, the Ensemble model achieves the best prediction, which pertains to the month of February and exhibits an MAPE of 5.31%.

- Regarding the DTL Case 3, the Ensemble model achieves the best prediction in January with an MAPE of 7.85%.

- In MTDL, the CNN model manages the best prediction in January, corresponding to an MAPE of 5.62%.

Generally, based on Tables 7.2, 7.3, 7.4, and 7.5, and in Figures 7.17, 7.18, and 7.19, a notable difference is observed between the best and worst prediction months. The best corresponds to the ELM prediction for May, with MAPE of 5.36% and the worst, corresponds to the CNN model prediction for December, with MAPE of 22.62%, resulting in a precision difference of 17.26%. This fact further reinforces the high performance of ELM in cases with fine-tuning, which is influenced by the predictions of the two distinct models. On the other hand, CNN, as evidenced and in the case of the worst-performing month, exhibits general instability in the case without fine-tuning, which is related to the inability of its trainable parameters to adapt to foreign data. This fact may indeed reveal the general difficulty of the models using convolution mechanisms to adapt to unknown datasets.

7.4.6 Discussion of Results

Based on the above results, it becomes evident that the application of Deep Learning algorithms in the domain of Deep Transfer Learning (DTL) can yield satisfactory outcomes, reducing computational power requirements and model training times, due to the fact that, after the initial training of the model, only the fine-tuning needs to take place each time the DL model is applied in a different area. The time required for a DL model to be trained during the fine-tuning period is significantly shorter compared to the time needed for direct training, as in the case of MTDL. Also, the variation in results for each month indicates that the ELM improves predictions for the majority of forecasted months.

In general, it is observed that the two strategies of Deep Transfer Learning with fine-tuning (DTL Case 1 and 2) significantly outperform DTL Case 3 and MTDL. Specifically, in the comparison between fine-tuning strategies and Multi-task Deep Learning, the differences suggest that the utilized models can adapt better when trained separately on different datasets, as opposed to parallel and simultaneous training on multiple datasets together. Both of these cases involve efforts to create models capable of efficiently generalizing to unknown and differently behaving timeseries.

Additionally, in the case of the direct use of a pre-trained model (DTL Case 3), a poorer performance is achieved compared to other cases. For the ELM, which is influenced by both the MLP and CNN models, the poor performance of the CNN negatively affects the accuracy for most months, with exceptions in January and July.

The variation in results clearly demonstrates that the employment of more than one model in an ensemble combination significantly improves the performance compared to individual algorithms. The reason behind this improvement lies in the weighted average learning, which takes into account the best predictions from both models, MLP and CNN separately. As a result, the final day-ahead load prediction is considerably improved, demonstrating the effectiveness of combining multiple models.

Finally, it is worth noting that the adaptability of the algorithms to the three examined timeseries relies on both the trainable parameters of each model and the different features and patterns exhibited by each case. Seasonality, peak demand periods, and average values of each dataset are some of the characteristics that influence algorithmic functionalities.

7.5 Summary and Future Work

In this study, an extensive investigation is conducted regarding Seq2Seq Deep Transfer Learning on timeseries data. For this reason, a case study of a month-to-month approach was employed with the aim of day-ahead forecasting of electricity load in three islands of the Greek power system. The results obtained provided us with valuable information regarding the application of such methods and their effectiveness. The first major conclusion is that transfer learning outperforms simple learning, even in the case of Multi-task Deep Learning, which is utilized for better model generalization.

Furthermore, another conclusion is that Deep Transfer Learning using Ensemble models outperforms simple DL models, as evidenced by the results obtained. More specifically, in the strategies DTL Case 1, DTL Case 2, as well as MTDL, it is observed that, for the majority of months, the ELM enhances the predictions achieved by the two individual DL models. This fact creates particularly optimistic conclusions regarding further exploration of Ensemble Models in the field of prediction in power systems.

DTL strategies are cost-effective, requiring significantly less computational power and time compared to simple prediction methods, due to the fact that the DL models are training once in a source dataset, after they are saved in appropriate format, and subsequently only their final part is fine-tuned for each specific task. Therefore, DTL minimizes the computational resources required and speeds up training for a specific target task, such as forecasting the day-ahead electrical load. By leveraging the knowledge stored in pre-trained models, DTL efficiently utilizes resources, facilitating swift deployment of effective models across different domains. Thus, it becomes evident that the application of DTL on real-time data can be achieved with high performance, due to the aforementioned conditions, as well as the flexibility provided by this specific domain compared to simple Transfer Learning applications.

More specifically, deploying the models studied in this work in real-time conditions is really challenging. Adapting to fast-changing data streams, maintaining low latency, and balancing model complexity with real-time requirements are key hurdles. Continuous updates of the model to address concept drift and ensure interpretability in real-time settings are also critical. Success requires tailored optimization and infrastructure to handle real-time processing demands effectively.

In the context of future study proposals, it should be emphasized initially that DTL could be applied to other branches of power systems, such as fault prediction in electric power transmission and distribution networks. Furthermore, beyond energy systems, it could be highlighted that a significant challenge lies in the application of DTL to areas like healthcare, where numerous research studies are conducted globally.

As already mentioned, the study of DTL using ELM, both in the field of power systems and in other domains, can prove capable of further improving the results of Transfer Learning.

Finally, the combination of DTL with Reinforcement Learning holds promise for future research and offers potential advancements. It could be explored to improve the efficiency of demand forecasting and load management systems. For instance, a pre-trained Reinforcement Learning agent could learn general patterns and behaviors from historical data across different regions or time periods. This pre-trained agent could then be fine-tuned on a specific locality or timeframe, in order to adapt to unique characteristics and changes in electricity demand. This approach may lead to more accurate and adaptable models for load prediction, contributing to improved resource planning and energy efficiency in the electricity grid.

Chapter 8

Electric Vehicles Load Forecasting

The automotive sector has evolved into a key force in both the global economy and the sphere of research and development. Continuous technological breakthroughs have led to vehicles now being outfitted with advanced safety systems that protect not only passengers but also pedestrians. This wave of innovation has contributed to a growing number of vehicles on our roads, delivering the benefits of rapid and comfortable travel. However, this progress has also ushered in significant environmental challenges. Cities are experiencing increased levels of harmful pollutants—such as sulfur dioxide, nitrogen oxides, carbon monoxide, and particulate matter—underscoring the complex balance between technological advancement and environmental stewardship. While these innovations have vastly improved transportation and quality of life, they have simultaneously accelerated environmental degradation, calling for urgent solutions to offset the negative impacts.

On a larger scale, the transportation industry is recognized as one of the major contributors to global energy consumption and greenhouse gas emissions, accounting for over a quarter of worldwide energy use, with road transport alone responsible for more than 70% of the sector's emissions. This reality has intensified concerns over carbon emissions and the sustainability of oil supplies, prompting a shift towards the concept of sustainable transportation. Electric vehicles (EVs) are at the forefront of this movement, offering significantly higher "tank-to-wheel" efficiency compared to traditional internal combustion engine vehicles, along with reduced noise and vibrations. As nations strive to address climate change and energy stability, there has been a notable increase in EV adoption across regions such as the United States, Europe, and China. Despite these encouraging trends, challenges such as underdeveloped battery technology, limited driving range, lengthy charging times, high initial costs, and insufficient charging infrastructure remain significant hurdles. Overcoming these obstacles is essential for fully integrating EVs into future smart cities, where enhanced battery performance and expanded charging networks will be critical in creating a sustainable, interconnected transportation ecosystem [253].

8.1 Smart Charging Working Principle

Smart EV charging is an innovative approach to powering electric vehicles, where the charging process is integrated with the grid and connected devices like smartphones and smart meters. This system adjusts the charging schedule based on variables such as time of day, electricity prices, and the battery level of the vehicle. Essentially, it enables you to charge your car efficiently by allowing remote control, ensuring it charges during off-peak hours or when electricity rates are most affordable, without the need for constant manual intervention required by traditional chargers.

Also, it operates using specialized smart chargers, like the Ohme ePod, which incorporate algorithms to align the charging process with your personal preferences—such as the desired battery level or a specific time by which the car should be fully charged. Once these settings are configured, the charger automatically manages the power flow, pausing and resuming the charge based on optimal pricing conditions. Moreover, these chargers can be programmed to harness renewable energy sources, such as solar or wind power, storing excess energy for later use. Typically, the entire process can be controlled remotely via an app or web interface, ensuring convenient management as long as you're connected to the internet [254].

8.2 Benefits of Smart EV Charging

Smart EV charging provides a variety of advantages. It enhances grid performance by providing load balancing, grid stability, and renewable energy integration. It helps distribute electricity demand efficiently by scheduling EV charging to smooth out peak loads or adjusting charging rates in real time to maintain system balance. Additionally, it improves grid stability by preventing congestion, overloading, and voltage fluctuations, reducing the need for costly infrastructure upgrades. Moreover, smart charging supports the integration of renewable energy by synchronizing charging sessions with renewable generation, promoting more sustainable and eco-friendly energy consumption.

The most immediate benefit is cost reduction. By charging during off-peak hours when electricity rates are lower, you can significantly reduce your energy expenses.

In addition to saving money, smart charging offers several other benefits [255]:

- Enhanced Convenience: Smart chargers allow you to schedule charging sessions according to your personal timetable. Once your preferences are set, the charger automatically manages the process, freeing you from manual adjustments.

- Improved Reliability: These chargers supply real-time updates on the charging process, ensuring your vehicle receives the necessary power precisely when needed.

- Reduced Emissions: By optimizing the charging schedule and taking advantage of renewable energy sources when available, smart charging helps lower emissions.

- Grid Stability: Smart chargers can contribute to stabilizing the electrical grid by monitoring demand and avoiding charging during peak periods. This minimizes the need for expensive grid upgrades and improves overall system efficiency.

- Optimized EV Performance: By maintaining the battery charge at the manufacturer's recommended level, smart charging can enhance the overall range and performance of your electric vehicle.

8.3 Smart Charging and EV Load Forecasting

Smart charging technology represents a significant advancement in managing the charging of electric vehicles (EVs), as it leverages real-time communication and advanced algorithms to optimize charging sessions. A fundamental component of this system is EV load forecasting, which predicts future energy demand based on historical data, time-of-day patterns, and user behavior. Accurate forecasting enables smart charging systems to schedule charging during off-peak hours and when renewable energy is abundant, thereby enhancing grid stability and reducing energy costs. By anticipating fluctuations in EV demand, utilities can avoid overloading the grid during peak times, ensure efficient distribution of power, and better integrate intermittent renewable resources.

The precision of EV load forecasting is essential for maximizing the benefits of smart charging. With reliable demand predictions, charging systems can proactively adjust charging rates and schedules, mitigating the risks of grid congestion and inefficiencies. This predictive capability not only supports the effective integration of renewable energy sources but also contributes to lowering emissions by minimizing reliance on fossil fuels during high-demand periods. Ultimately, the synergy between smart charging and accurate EV load forecasting is critical for creating a resilient, cost-effective, and environmentally sustainable energy ecosystem.

8.4 Impact of Electric Vehicle Load on Wholesale Energy Market Operations

The increasing penetration of electric vehicles (EVs) introduces new challenges and opportunities in wholesale energy market operations. The load from electric vehicles, especially from large fleets, plays an important role in the system's operation since it is difficult to predict. Consider the case at an MT/HT substation: how much will it be overloaded if a fleet of electric vehicles starts charging simultaneously? What will happen in the network, and to what extent is it at risk of going out of operation?

So, the transition toward electrified transportation is altering electricity consumption patterns. As EV adoption rises, the additional demand has significant implications for wholesale electricity markets. Key aspects include shifts in peak demand, price volatility, and potential grid congestion. Understanding these effects is essential for market participants, regulators, and policymakers.

EV charging demand is influenced by factors such as consumer behavior, charging infrastructure, and price signals, resulting in different impacts on wholesale electricity markets depending on the charging strategy employed. Uncontrolled charging, where vehicles begin charging immediately upon plugging in, typically aligns with peak demand periods, potentially exacerbating grid stress. In contrast, time-managed charging enables users to operate during off-peak hours, thereby taking advantage of lower electricity prices and reducing the load during critical times. Additionally, vehicle-to-grid (V2G) operations allow EVs to serve as flexible energy resources by discharging stored electricity back into the grid, which can help balance demand and enhance overall system stability [256].

Some of the areas of smart energy grids affected by the integration of EVs are described below [257]:

- **Impact on Energy Demand:** EV charging requires substantial power, and when many vehicles begin charging at the same time, substations can experience sudden demand surges. These peaks can stress the system beyond its designed capacity, leading to issues such as voltage drops and potential service interruptions. The unpredictable nature of charging behavior further complicates load forecasting and grid management.

- **Infrastructure Challenges:** Traditional grid systems are optimized for relatively stable and predictable demand profiles. The dynamic load introduced by EV charging, especially when uncontrolled, can exacerbate stress on both distribution and transmission networks. To cope with these challenges, it may be necessary to upgrade grid infrastructure and integrate advanced monitoring systems capable of responding to rapid load changes.

- **Mitigation Strategies:**

  - **Smart Charging Approaches:** Implementing smart charging techniques can help distribute the EV load more evenly across time. By scheduling charging during off-peak periods or in response to real-time pricing signals, grid operators can reduce the risk of overloads and better manage energy demand. This approach not only alleviates strain on the grid but also promotes more efficient energy usage.
  - **Vehicle-to-Grid (V2G) Systems:** V2G technology offers a dual benefit by enabling EVs to return stored energy back to the grid during periods of high demand. This bidirectional flow of electricity can act as a buffer, helping to balance load fluctuations and improve overall grid stability. Integrating V2G into the energy system is seen as a promising measure to mitigate the impact of peak charging events.

8.5 Comparison of Deep Learning Approaches for Electric Vehicle Load Curves Forecasting

8.5.1 Summary of the work

Electric Vehicles (EVs) are becoming increasingly prevalent in modern transportation systems. Their fast penetration is followed by a need for accurate forecasting of EV load curves, a process particularly important for market entities like electricity suppliers and/or electric vehicle aggregators that participate in wholesale electricity markets. In this work, two deep learning forecasting models, which have been proven effective for time-series forecasting, the Temporal Convolutional Networks and the Temporal Fusion Transformer are developed and evaluated for EV load curves forecasting. Their forecasting performance is compared with the Persistence baseline model using common forecasting error metrics. Furthermore, an ensemble methodology is proposed, which combines the deep learning models and the baseline model to improve the overall performance of the predictions. The performance of the examined forecasting models is evaluated on a publicly available EV dataset and results are presented and discussed.

8.6 Related Publications

Electric Vehicles (EVs) are gaining popularity in modern transportation networks. The mass adoption of EVs necessitates precise forecasting of EV Load Curves (EVLCs), an important process for wholesale energy market participants such as electricity suppliers and/or electric vehicle aggregators [258]. The EVLCs must be forecasted on time and accurately to drive the optimal submission of competitive bids in the energy market and/or in the ancillary services markets. The need for accurate short-term forecasting of EVLCs is particularly useful for short-term electricity markets, such as the Day-Ahead Market and/or the Intra-day Markets [259], where market participants are required to submit energy bids (MWh) with a small lead time of a few hours. Moreover, accurate forecasting of the EVLCs is useful for a many other applications, e.g. power grid management, efficient integration of renewable energy resources, optimal charging schedule of EVs and policy making.

There are various methodologies used for EVLC forecasting, the performance of which is based on the type of available data and their temporal variation. A common approach is to use statistical models, such as time series analysis, to predict the load for electricity based solely on historical data. Another approach is to use machine learning algorithms, such as neural networks or support vector machines, to analyze data from multiple sources, including weather patterns, commuting behavior, and vehicle charging patterns.

Other techniques, such as clustering and pattern recognition, can also be used to identify trends and patterns in the data. The effectiveness of these methodologies depends on the quality and quantity of available datasets, as well as on the ability of each model to adapt to each dataset. All in all, the goal is to provide accurate and reliable forecasts that can support utilities and grid operators to manage the demand for electricity in the power system effectively.

Two types of Deep Learning (DL) models, the Temporal Convolutional Networks (TCNs) and Temporal Fusion Transformers (TFTs) seem very promising for use in time-series forecasting. Although they have been used in a variety of applications, their exploitation in EVLC forecasting is very limited. Regarding the TCN model, in [260] it is used for short-term load forecasting in North American New England Control Area and in [261] it is used for load forecasting in Toronto City. Additionally, the authors in [262] proposed a hybrid Gated Recurrent Network-TCN model for short term load forecasting of a southwest province in China. The most relevant to EVLC forecasting application can be found in [263], where a TCN model is compared with a hybrid Multi-Channel CNN-TCN model in order to predict EVLC in the urban area of a city in northern China. Regarding the TFT model, the authors in [264] developed a TFT-based interpretable multi-horizon time-series forecasting using a variety of real-world datasets from the electricity, traffic, and retail sector, while the authors in [265] used a TFT-based model to forecast short-term load, using real data collected from a university campus. Finally, in [265], a TFT-based forecasting model for wind speed prediction is presented.

8.7 Structure of the work

Following the very limited application of TCN and TFT models in EVLC forecasting, the current work evaluates their performance on a real-world, publicly available, EV dataset. The evaluation period covers a year in hourly time steps (8760 hours). The forecasting is executed on a rolling manner, once per day at midnight and the forecasting horizon expands from 1 up to 24 hours-ahead. The performance of the TCN and TFT models is evaluated in comparison with the Persistence baseline model and two ensemble models. The first ensemble model combines the two DL models, while the second ensemble model combines the two DL models and the baseline model. Finally, the value of weekly versus monthly model re-training is evaluated. To the best of the authors' knowledge, this attempt is the first that performs such a systematic evaluation of TCN and TFT models on EVLC forecasting and the first that creates an ensemble EVLC forecasting model combining these two DL models.

8.8 Methodology

The methodology applied in this study, according to which the data are collected, preprocessed into an appropriate form, and input into the deep learning models, includes the following steps:

- Firstly, the EV charging data are collected. The dataset used is publicly available and contains information for each EV charging session. The key records of each EV session used for this research work are: a) the plug-in/out time of each EV, b) the active charging time of each EV, and c) the total energy consumed by each EV.
- Secondly, the EV charging data are preprocessed. This step includes data cleaning, feature selection, and data transformation, converting the raw data into suitable time-series vectors, used as input for the developed models.
- Thirdly, five (5) forecasting models are created. Two (2) DL (i.e., TFT and TCN), one (1) Persistence baseline model, and two (2) ensemble models. The first ensemble model uses the average of the predictions of the two DL models, while the second one considers the average of the three models (the DLs and the Persistence).
- Following the preparation of input and output vectors for each model, the process of Hyperparameter Optimization follows, where the optimal hyperparameters for each model are adjusted. This step defines the fundamental architecture of the models.
- In the next step, the models' training process takes place. Each forecasting model is trained on the training dataset, and the trainable parameters are determined. Then each model is ready to make predictions on the separate testing dataset.
- Finally, the prediction process takes place, by feeding the testing dataset into each model, providing the required predictions. The predicted values are then compared to the actual values of the testing data, and various metrics such as MAE, RMSE and R-Squared (R^2) are calculated to evaluate each model's performance.

The above methodology is applied in two different cases, depending on the re-training frequency of each model:

Case 1: Each DL model (TFT, TCN) is re-trained on a monthly basis. At the end of each month, the models are re-trained, including the data of the previous month in the training set and the forecasting execution continues sequentially, as above. The forecasting is executed sequentially, once per day at midnight, and the forecasting horizon expands from 1 up to 24 hours-ahead, for each hour of the next day.

Case 2: Each DL model (TFT, TCN) is re-trained on a weekly basis. Thus, after the end of each week, the models are re-trained, including the data of the previous week in the training set. The forecasting process is the same as the one described in Case 1.

The Persistence model does not need to be trained since it makes instant inferences using historical data, as is further explained in section "Forecasting Models".

By applying the above methodology, conclusions can be drawn regarding the model's forecasting performance, the variation of forecasting accuracy throughout the year, and the impact of model re-training frequency on the forecasting performance.

8.9 Dataset Description

The dataset used in the study is titled *"Electric Vehicle Charging Station Use (July 2011 - December 2020)"* and is provided publicly by the city of Palo Alto through its Open Data Portal [266]. The data consists of 259,415 EV charging sessions. Each session includes records for the energy consumed (kWh), the charging duration, and the duration each EV is plugged-in into the socket.

8.10 Data Pre-processing

The data pre-processing methodology follows the steps below:

- Initially, the abnormal sessions were removed, i.e., sessions with a charging start time later than the charging end time.
- Then, the assumption that each EV charges at a constant power during the whole session was made. Following this assumption, the instant power value per second (kW) was calculated by dividing the energy consumed (kWh) per session by the charging duration (secs) per session. Thus, the EVLC for the whole dataset is formulated by summing the energy per second (kWh/sec) time series of all sessions of the dataset,

starting from the first available second of the first session until the last available second of the latest session. The kWh/sec EVLC is then re-sampled to hourly time step (kWh/hour), using averaging. The time series produced represent the hourly EVLC of the dataset. The hourly EVLC for the first 15 days of April 2019, is presented in Figure 8.1.

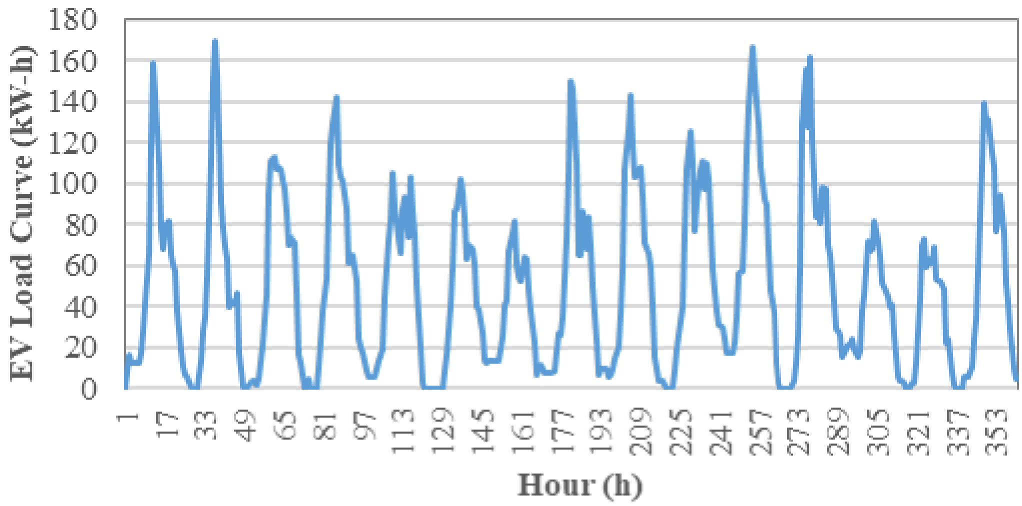


Figure 8.1: The EVLC for the first 15 days of April 2019

8.11 Forecasting Models

8.11.1 Temporal Fusion Transformer

TFT is a network architecture that combines multi-horizon forecasting with interpretable insights into temporal dynamics. TFT employs specialized building blocks to choose key features and omit unnecessary components, hence enabling great performance across a wide range of workloads. Their primary constituents are: a) Gated Residual Network techniques. They enable to elimination of any unneeded architectural components and avoid non-linear processing if it is not required, b) Variable Selection Networks with variable selection to choose significant input features at each time step, c) Static covariate encoders that integrate static factors into the network in order to condition temporal dynamics, d) Temporal processing to learn both long- and short-term temporal associations from time-varying inputs that are both observed and known, e) Prediction intervals via quantile predictions to determine the range of likely target values at each prediction horizon. TFT enhances the interpretability of time series forecasting by identifying globally relevant variables, enduring temporal patterns, and significant events for the prediction problem [267].

8.11.2 Temporal Convolutional Networks

Temporal Convolutional Networks (TCNs) are a type of neural network that can be used for time-series forecasting tasks. They are similar to traditional Convolutional Neural Networks (CNNs) but designed to handle sequences of variable length. In a TCN model, input sequences are first passed through a series of convolutional layers, which apply a set of filters to the input at different time scales. These filters can be thought of as windows of different widths that slide over the input sequence, allowing the model to capture patterns at different levels of granularity. The output of the convolutional layers is then passed through a series of dilated causal convolutional layers, which help the model capture long-range dependencies in the input sequence. Finally, the output of the dilated causal convolutional layers is passed through a set of fully connected layers to produce the final prediction [268].

8.11.3 Persistence Baseline Model

The Persistence Baseline model is a simple time series forecasting technique that is often used for comparison with more sophisticated forecasting methods. The algorithmic approach of the Persistence model involves computing the forecast for each target step as the average value of the corresponding hours from the two most recent weeks. However, if any of the corresponding hours from the prior weeks fall on a holiday, the algorithm skips those hours recursively until an hour in a non-holiday is found. Additionally, if the target hour falls on a weekend or holiday, the corresponding hour from the previous closest holiday or weekend is considered as the forecast. Despite its simplicity and inherent limitations, the Persistence method can serve as an effective initial step for short-term time series forecasting.

8.11.4 Ensemble Models

Ensemble methodology is based on the premise that by combining the predictions of different forecasting models, the errors and biases of individual models are reduced, resulting in a more reliable, robust, and accurate forecast [269]. There are several approaches to building ensemble models for time series forecasting, including:

- **Model averaging:** This involves taking the average of the predictions from multiple models. The idea is that by combining the predictions of multiple models, the impact of any individual model's biases or errors is reduced.

- **Model stacking:** This involves using the predictions from multiple models as input to a second-level model that produces the final forecast.

- **Bagging and boosting:** These are techniques from the field of machine learning that involve generating multiple models from different samples of the training data (bagging) or modifying the weights of the training data (boosting).

The Ensemble modeling has been found to be particularly effective in time series forecasting, where it is often challenging to obtain accurate forecasts due to the complex and dynamic nature of the real-world time series. In the current work, two ensemble models are used: The first ensemble model computes the average of the two deep learning models predictions, whereas the second ensemble incorporates the average of all three models, the two deep learning models and the Persistence Baseline model.

8.12 Evaluation Metrics

To compare the aforementioned algorithms and techniques, their results were evaluated using a series of metrics: Mean Absolute Error, Root Mean Squared Error, and R-squared (R^2). Their mathematical formulations and brief descriptions follow, where f_i is the predicted value, y_i is the actual value, $\bar{y}$ is the mean value of all y_i values, and N is the total number of data points.

8.12.1 Mean Absolute Error (MAE)

MAE is a statistical metric used to measure the average absolute difference between predicted values and actual values in a regression problem. It measures how well a regression line (or curve) fits the data points.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - f_i| \tag{8.1}$$

8.12.2 Root Mean Squared Error (RMSE)

RMSE is a statistical metric used to measure the square root of the average squared difference between the predicted values and the actual values in a regression problem, being more sensitive to large forecast deviations. It is a measure of how well a regression line (or curve) fits the data points.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - f_i)^2} \tag{8.2}$$

8.12.3 R-squared (R^2)

The coefficient of determination (R2) constitutes the comparison of the variance of the errors to the variance of the data which is to be modeled. It can be calculated as follows:

$$R^2 = 1 - \frac{\sum(y_i - f_i)^2}{\sum(y_i - \bar{y})^2} \tag{8.3}$$

8.13 Results and Discussion

This section describes the forecasting results of the study. As stated above, the forecasting is executed sequentially, once per day at midnight, and the forecasting horizon expands from 1 up to 24 hours-ahead, for each hour of the next day, creating a continuous forecasted time-series with no overlapping predictions. The forecasting period extends to the whole year 2019. The training period includes the whole year 2018 and extends by month for Case 1 and by week in Case 2, in every retrain process, as described in the Methodology section.

The model's performance is compared, both for Case 1 and Case 2 (i.e., monthly, and weekly re-training). The detailed results for all models and for Cases 1 and 2 are presented in Tables 8.2, 8.3, 8.4 and 8.5. It is noted that the MAE, RMSE and R2 scores of the Persistence model are the same between Case 1 and Case 2, since they are independent of the re-training frequency.

8.13.1 TFT versus TCN forecasting performance

The average of the performance scores for the 12 months period for Cases 1 and 2, are presented in Table 8.1. MAE and RMSE scores are referred to kWh. The lower the score the better the performance. Focusing on Case 1, the yearly average MAE for the TFT and the TCN model is 12.52 and 13.39, respectively, while the Persistence baseline model, has an average MAE of 14.58. In addition, the yearly RMSE for the TFT and the TCN model are 18.29 and 18.79, respectively, while the baseline model has an average RMSE of 20.26. Similar results are recorded for Case 2, where the MAE for the TFT model is lower than the TCN model (i.e., 12.74 and 13.39 respectively) and the RMSE for TFT model is lower than the TCN model (i.e., 18.40 and 18.78 respectively).

These scores indicate that both DL models improve the forecasting accuracy of the EVLC, compared to the baseline model. In addition, the overall performance of the TFT-based forecasting model outperforms the TCN model. The main reason is considered that the TFT employs the self-attention mechanism, which enables the model to focus on the most pertinent portions of the sequence for forecasting while TCN extracts features from the input sequence using fixed convolutional filters, which may not be as successful at capturing complicated temporal patterns. Additionally, TFT employs a multiscale design that enables the model to capture time series data patterns at various granularities. In contrast, TCN employs just fixed-size convolutional filters, which may not be as good at capturing patterns of varying scales. Finally, in addition to the dynamic aspects of the time series data, TFT can also combine static features that may be pertinent to the forecasting assignment, while TCN is often employed just for modeling the temporal characteristics of data. Compared to TCN, TFT's additional characteristics make it a more effective model for forecasting time series data.

8.13.2 Monthly versus weekly re-training frequency

The impact of the re-training frequency on the predictions' accuracy can be evaluated by comparing the error metrics of Case 1 (i.e., monthly re-training) and Case 2 (weekly re-training). For the TCN model, the MAE and the RSME are similar between monthly and weekly re-training, thus the allocation of computational resources for weekly model re-training is unnecessary. For the TFT model the performance under monthly re-training frequency is slightly better than the weekly re-training frequency, since MAE is 12.52 in Case 1 and 12.74 in Case 2 while the RMSE is 18.29 in Case 1 and 18.40 in Case 2. The results indicate that the performance of the TFT forecasting model slightly decreases by increasing the model re-training frequency from monthly to weekly. Thus, similarly to the TCN model, the weekly re-training of the TFT model does not seem to improve the overall performance.

8.13.3 Ensemble models performance

The performance of the Ensemble models is presented in Table 8.1. The first Ensemble TCN-TFT model, combines the results of the two DL models, taking their mean value for each hour as the final prediction value, leading to higher prediction accuracy compared to TFT model only or TCN model only, for all error metrics. In addition, the second Ensemble model combines the results of the two DL models and the Persistence model, taking their mean value for each hour as the final prediction value, leading to higher prediction accuracy compared to TFT model only or TCN model only, for all error metrics. In Case 1 the 3-models Ensemble, achieved a MAE equal to 11.64 while the Ensemble, TCN-TFT model achieved MAE equal to 12.1. Additionally, the RMSE in Case 1 is 16.11 and 17.15 for the 3-models Ensemble and the TCN-TFT Ensemble model, respectively. In Case 2 the 3-models Ensemble achieved a MAE equal to 11.79 while the Ensemble, TCN-TFT model a MAE equal to 12.03. Additionally, the RMSE in Case 2 is 16.35 and 17.11 for the 3-models Ensemble and the TCN-TFT Ensemble model, respectively. In both Cases, the combination of the three (3) forecasting models (Ensemble, 3 Models) produce better forecasting results than the combination of the two DL models.

Table 8.1: Yearly Evaluation Metrics For The Compared Models

Model / Frequency	**MAE**		**RMSE**		**R^2**	
	Case 1	**Case 2**	**Case 1**	**Case 2**	**Case 1**	**Case 2**
TFT	12.52	12.74	18.29	18.40	0.847	0.845
TCN	13.39	13.39	18.79	18.78	0.837	0.837
Persistence	14.58	14.58	20.26	20.26	0.811	0.811
Ensemble, TCN-TFT	12.10	12.03	17.15	17.11	0.865	0.866
Ensemble, 3 Models	11.64	11.79	16.10	16.35	0.880	0.880

8.13.4 Forecasting performance per month

Figure 8.2 and Figure 8.3 present the MAE (kWh) score per month, for all the examined models for Case 1 and Case 2, respectively. The monthly MAE scores in Case 1 (Figure 8.2), indicate that the TFT forecasting model outperforms the TCN forecasting model for nine of the twelve months. Additionally, the 3-models Ensemble outperforms both the TCN model

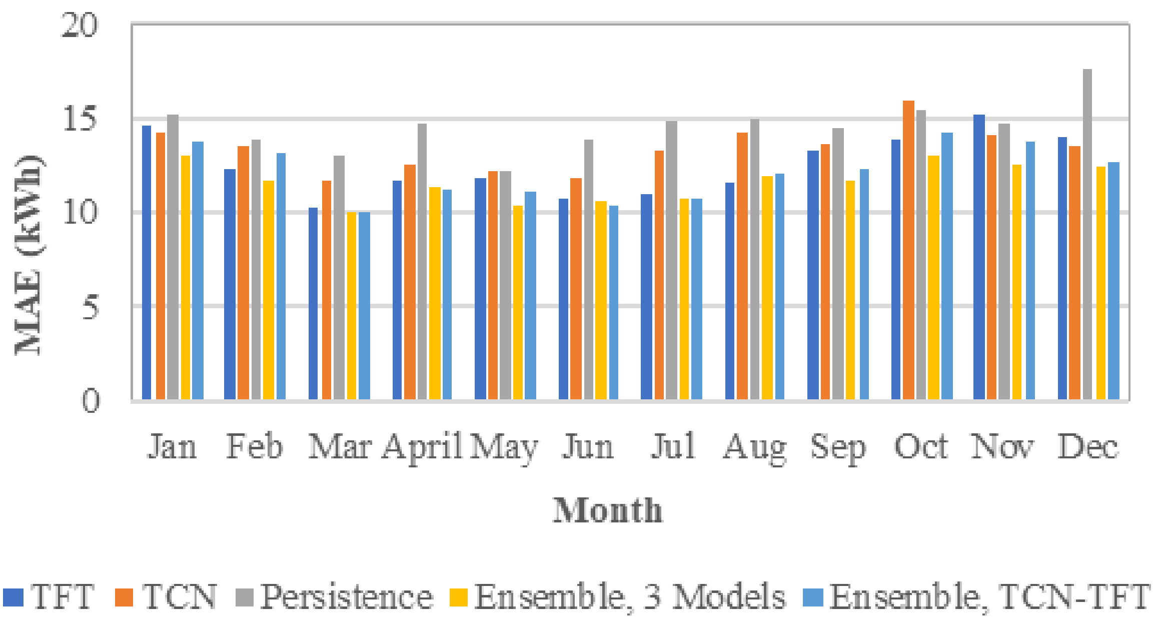


Figure 8.2: MAE comparison between the examined models for Case 1 (monthly re-training frequency)

and the TFT model for every month. The monthly MAE scores per month for Case 2 (Figure 8.3) indicate that the TFT model outperforms the TCN model for nine of the twelve months (the same months as Case 1). Additionally, the Ensemble, 3 Models model outperforms the TCN model every month and the TFT model for eleven months. Regarding the comparison between the two Ensemble models, in Case 1, the Ensemble, 3 Models model outperforms the Ensemble, TCN-TFT model for eight months, while for the rest four months their performance is similar. In Case 2, the Ensemble, 3 Models model outperforms the Ensemble, TCN-TFT model only for seven months, indicating that for the weekly re-training frequency their performance is close. In conclusion, the two Ensemble models outperform the other models, showing relative stability throughout the year.

As a representative example of the hourly forecasting performance, Figure 8.4 presents the forecasting results for the first week of April 2019, for three models, for Case 1.

Finally, Tables 8.2, 8.3, 8.4 and 8.5 present, in a monthly base, the forecasting results for each model for monthly and weekly re-training frequency.

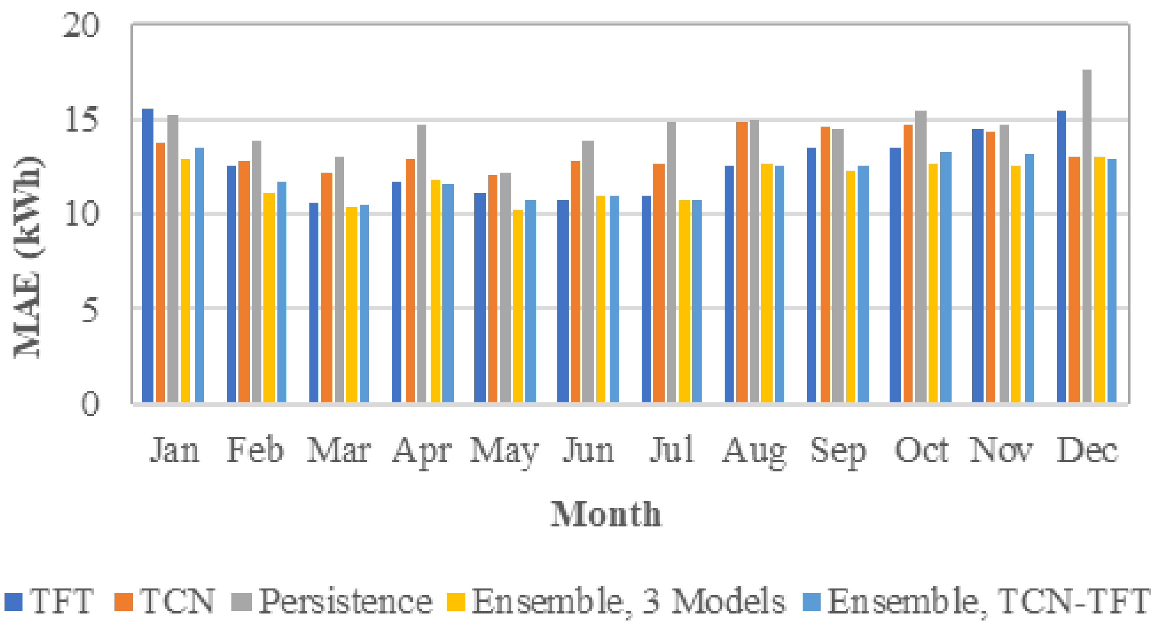


Figure 8.3: MAE comparison between the examined models for Case 2 (weekly re-training frequency)

Table 8.2: Monthly re-training frequency: Performance of TFT and TCN models

Metric	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Avg
TFT Model													
MAE	14.60	12.30	10.31	11.70	11.82	10.69	10.92	11.57	13.33	13.84	15.21	13.96	12.52
RMSE	21.90	16.57	14.17	16.09	19.36	14.15	16.62	15.95	19.18	19.50	22.87	20.57	18.08
R^2	0.78	0.88	0.90	0.87	0.83	0.90	0.85	0.88	0.82	0.84	0.79	0.82	0.85
TCN Model													
MAE	14.20	13.52	11.75	12.59	12.18	11.79	13.26	14.24	13.66	15.94	14.11	13.53	13.40
RMSE	20.14	18.57	16.50	17.47	18.10	16.47	19.33	19.66	18.60	21.31	19.74	18.93	18.73
R^2	0.81	0.85	0.87	0.85	0.85	0.87	0.80	0.82	0.83	0.81	0.84	0.85	0.84

Table 8.3: Monthly re-training frequency: Performance of Persistence and Ensemble models

Metric	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Avg
Persistence Model													
MAE	15.27	13.87	12.98	14.73	12.19	13.92	14.89	14.94	14.47	15.41	14.77	17.61	14.59
RMSE	22.27	18.53	17.89	19.62	17.55	19.04	20.32	21.20	19.08	20.58	19.62	25.95	20.14
R^2	0.77	0.85	0.85	0.81	0.86	0.82	0.78	0.79	0.83	0.82	0.85	0.72	0.81
Ensemble (3 Models)													
MAE	13.07	11.68	10.06	11.34	10.40	10.64	10.79	11.92	11.75	13.07	12.53	12.46	12.11
RMSE	18.47	15.96	13.82	15.43	15.16	14.41	15.06	16.39	15.40	17.85	17.26	17.94	16.67
R^2	0.84	0.89	0.91	0.88	0.89	0.90	0.88	0.87	0.89	0.87	0.88	0.86	0.87
Ensemble (TCN-TFT)													
MAE	13.81	13.16	9.97	11.19	11.08	10.33	10.77	12.01	12.32	14.30	13.74	12.64	12.11
RMSE	19.49	18.25	13.64	15.46	17.15	14.21	16.14	16.44	16.68	19.36	19.72	17.98	17.04
R^2	0.82	0.86	0.91	0.88	0.87	0.90	0.86	0.87	0.87	0.84	0.84	0.86	0.87

Table 8.4: Weekly re-training frequency: Performance of TFT, TCN, and Persistence models

Metric	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Avg
TFT Model													
MAE	15.63	12.53	10.65	11.70	11.15	10.70	11.04	12.51	13.55	13.52	14.52	15.44	12.74
RMSE	22.52	17.17	14.62	15.60	18.29	14.50	16.77	17.17	18.93	19.27	21.32	22.29	18.20
R^2	0.76	0.87	0.90	0.88	0.85	0.90	0.85	0.86	0.83	0.85	0.82	0.79	0.85
TCN Model													
MAE	13.72	12.77	12.17	12.97	12.04	12.80	12.67	14.85	14.63	14.74	14.33	13.06	13.40
RMSE	19.23	17.10	17.08	18.12	18.76	17.86	17.82	20.79	19.16	20.49	19.82	18.68	18.74
R^2	0.83	0.87	0.86	0.84	0.84	0.84	0.83	0.80	0.82	0.83	0.84	0.85	0.84
Persistence Model													
MAE	15.27	13.87	12.98	14.73	12.19	13.92	14.89	14.94	14.47	15.41	14.77	17.61	14.59
RMSE	22.27	18.53	17.89	19.62	17.55	19.04	20.32	21.20	19.08	20.58	19.62	25.95	20.14
R^2	0.77	0.85	0.85	0.81	0.86	0.82	0.78	0.79	0.83	0.82	0.85	0.72	0.81

Table 8.5: Weekly re-training frequency: Performance of the two Ensemble models

Metric	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Avg
Ensemble (3 Models)													
MAE	12.92	11.15	10.36	11.78	10.28	11.00	10.77	12.61	12.28	12.71	12.53	12.99	12.16
RMSE	18.41	14.95	14.44	15.87	15.31	15.14	14.90	17.48	15.72	17.74	17.22	19.02	16.74
R^2	0.84	0.90	0.90	0.87	0.89	0.88	0.88	0.85	0.88	0.87	0.88	0.84	0.87
Ensemble (TCN-TFT)													
MAE	13.57	11.74	10.51	11.55	10.72	10.95	10.75	12.57	12.59	13.33	13.16	12.97	12.03
RMSE	19.41	15.92	14.53	15.63	17.24	15.16	15.80	17.46	16.53	18.74	19.19	18.80	17.03
R^2	0.82	0.89	0.90	0.88	0.86	0.89	0.87	0.86	0.87	0.85	0.85	0.85	0.87

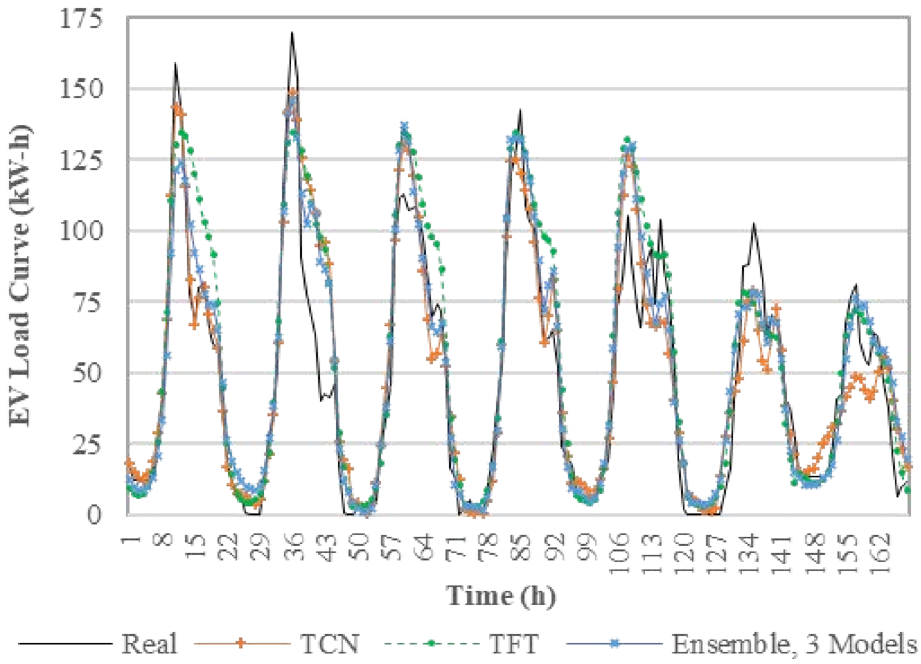


Figure 8.4: Performance of the different algorithms for the first week of April

8.14 Summary and Future Work

In this research investigation two deep learning models, the Temporal Convolutional Networks (TCN) and the Temporal Fusion Transformer (TFT), are developed and evaluated for the forecasting of EV load curves. Accurate forecasting of EV load curves is important for market entities such as electricity suppliers and electric vehicle aggregators participating in wholesale electricity markets. The performance of the two models is evaluated and compared with the Persistence baseline model, as well as with two ensemble models, where the aforementioned models are combined. The study is based on a real-world, publicly available, EV dataset and the common error metrics MAE, RMSE and R2 are used. The methodology consists of six steps: data collection, data pre-processing, model creation, hyperparameter optimization, training, and prediction. The study concludes that the deep learning models outperform the Persistence model in forecasting the evaluated EV load curve, while the Ensemble models demonstrate superior forecasting performance compared to all individual models. The proposed methodology is applied in two separate Cases, one with monthly re-training frequency and one with weekly re-training frequency. Based on the results, it is concluded that the deep learning models exhibit high adaptability to EV load curves, and the increased models' re-training frequency does not necessarily lead to improved prediction results.

The research findings provide practical insights on the EV load curve forecasting performance which could be useful for the relevant stakeholders to promote a more sustainable and efficient future of electric vehicle industry.

Chapter 9

General Conclusions and Future Study Proposals

Based on the analyses presented in the previous chapters and the overall work carried out throughout the entire PhD dissertation, the current chapter is divided into two parts: first, the conclusions that emerge for each area analyzed, and secondly, the proposals for future study that arise are presented. The reason for organizing them by area is to clearly highlight the dissertation's contributions and findings across all examined dimensions.

9.1 General Conclusions

Based on the scientific analyses and approaches developed throughout the dissertation, the primary conclusion is clear: Deep Learning and all its branches serve as the cornerstone for the advancement of Artificial Intelligence, the design of computing systems, and ongoing research. The energy sector, through the accurate forecasting provided by Deep Learning, is empowered and gains increasing reliability. The functioning of the market achieves greater stability and trust among participants, as data-driven decision-making systems lead to faster and more accurate decisions. More specifically, the following are the conclusions that emerged for each of the fields investigated in the comparative dissertation.

9.1.1 Forecasting Approaches

Based on the models that were studied, as well as those developed across all implementations of this dissertation, it is evident that Deep Learning significantly outperforms models belonging to other categories. The flexibility provided by these algorithms — in terms of data handling, preprocessing, and their ability to deliver highly accurate forecasts — makes them an exceptionally powerful tool with a wide range of scientific applications.

Special mention should be made of Transformers and Ensemble models. The former are perhaps the most powerful yet complex algorithmic models, due to their large number of parameters and their architecture, which lead to outstanding results in time series tasks. The latter, owing to the wide range of possible combinations they offer, often enhance the final performance of the individual models under study.

9.1.2 Electricity Load Forecasting

The developments that have contributed to significant changes in the energy sector have introduced new challenges, requiring efficient solutions to ensure the reliable operation of the power grid. Accurate demand forecasting plays a crucial role in maintaining the stability and robustness of modern power systems. In particular, short-term load forecasting is essential for market flexibility, as it helps balance supply and consumption in real time. To address these needs, the ELF tasks presented in this PhD dissertation, and more specifically in Chapter 3, contain an innovative and automated deep learning forecasting model for the energy sector. The forecasting performance of five advanced models—CNN-1D, LSTM, CNN-LSTM, TCN, and MLP—is evaluated and compared with a proposed hybrid model, the ConvLSTM Encoder-Decoder. The results indicate that the ConvLSTM Encoder-Decoder outperforms the other models, achieving a MAPE of 2.15%. It is followed by LSTM at 2.25%, CNN-LSTM at 2.31%, MLP at 2.38%, TCN at 2.40%, and CNN-1D at 2.66%. The differences in MAPE values among the models are relatively small, likely due to the extensive optimization applied to each. A key contribution of this research is the development of an automated deep learning algorithm capable of identifying the most suitable model for predicting future time series data across different applications. While this study focuses on load forecasting, the proposed approach could also be applied to predicting variables in systems incorporating renewable energy sources, such as wind and solar power. Additionally, the methodology has potential applications in forecasting building energy consumption, particularly in commercial, industrial, and military facilities, where factors like climate conditions, indoor environment, and occupant behavior influence energy use. Beyond introducing an effective forecasting model, this study presents a novel deep learning approach that achieves the highest prediction accuracy among the tested models. The proposed hybrid model integrates a ConvLSTM component to encode input time series data, which is then processed through an Encoder Vector and passed to an LSTM model responsible for forecasting future values. Given these contributions, it is evident that the developed methodology sets itself apart from existing approaches in the literature. While previous studies primarily compare results from different statistical and machine learning models, this work takes a different direction by employing an automated model selection process. This enhances flexibility and usability for potential users of the developed application. Furthermore, the introduction of the ConvLSTM Encoder-Decoder, a state-of-the-art model that has not been widely applied to load forecasting, adds another valuable dimension to this research.

Additionally, in Chapter 3 we aim to forecast the electricity consumption of households on Skiathos Island using hourly data from 15 households. However, having such a small number of households makes it challenging to build a reliable prediction model. Moreover, household electricity use varies throughout the month, week, and day due to the often random nature of individual consumption patterns. In general, as the number of consumers increases, the overall load curve tends to stabilize, which in turn improves the accuracy of the predictions.

Our findings show that deep learning models are better than simple statistical models, like the naive model we used, at capturing consumer behavior patterns. Additionally, we found that when the sample size is limited, weather factors—such as temperature, humidity, and wind speed—do not have a significant impact on hourly consumption variations, as indicated by the low correlation between these variables in our study.

9.1.3 Wholesale Electricity Price Forecasting

Regarding EPF, a thorough and integrated study was carried out in Chapter 4, in order to develop high-accuracy deep learning algorithms for short-term electricity price forecasting. The study had two main objectives. First, we aimed to design and propose two novel deep learning models: a hybrid CNN-GRU with an attention mechanism and an ensemble learning model that achieves the highest accuracy. Second, we explored advanced deep learning models incorporating attention mechanisms and tested them using electricity price data from the Germany–Luxembourg interconnection. Specifically, we investigated an optimized Multi-Head Transformer, a CNN attention model, and a stacking ensemble learning model. A perceptron-based machine learning model was chosen as a benchmark reference. Among the models tested, the hybrid CNN-GRU with an attention mechanism emerged as a promising approach, as it has not previously been applied to electricity price prediction. This fact offers strong potential for future research and applications. Additionally, despite the general challenges transformer-based algorithms face in time series forecasting, the Multi-Head Transformer model delivered highly accurate predictions on unseen data. Furthermore, the successful implementation of two ensemble learning models highlights both the effectiveness and the complexity of selecting suitable approaches. A key observation from the analysis is that electricity price time series do not follow a consistent pattern, as prices are influenced by external factors such as political conditions in neighboring countries. These factors introduce uncertainty and fluctuations in the wholesale energy market, making them difficult to model accurately. Another significant finding is that advanced deep learning algorithms with integrated attention mechanisms could play a crucial role in improving price forecasting in the wholesale electricity market. By utilizing the Optuna optimization algorithm, model hyperparameters can be fine-tuned to achieve the most suitable architecture, leading to optimal performance. More accurate price predictions, in turn, can support better production planning for electricity producers, minimizing deviations that could negatively impact system operation.

9.1.4 Transfer Learning

Regarding Transfer Learning, the study conducted in Chapter 5 is an in-depth investigation into Seq2Seq deep transfer learning for time series data. A month-to-month approach was applied as a case study, focusing on day-ahead electricity load forecasting for three islands connected to the Greek power system. The results offer valuable insights into the application and effectiveness of these methods. The first key finding is that transfer learning outperforms conventional learning approaches, even when multi-task deep learning is used to improve model generalization. Additionally, deep transfer learning with ensemble models achieves better performance than standalone deep learning models, as demonstrated by the results. Specifically, in DTL Case 1, DTL Case 2, and MTDL strategies, the ensemble learning model (ELM) consistently enhances prediction accuracy over the two individual deep learning models in most months. This finding underscores the potential of ensemble models for further exploration in power system forecasting. DTL strategies offer a cost-effective solution by significantly reducing computational time and power requirements. Since deep learning models are pre-trained on a source dataset, saved in an optimized format, and later fine-tuned only for specific tasks, DTL minimizes resource consumption while accelerating training. This efficiency makes it well-suited for day-ahead electricity load forecasting, as it allows pre-trained models to be adapted swiftly to new tasks with minimal overhead. The ability to leverage stored knowledge from pre-trained models enhances resource utilization and facilitates the rapid deployment of effective models across different applications. Deploying these models in real-time presents several challenges. Adapting to rapidly changing data streams, maintaining low latency, and balancing model complexity with real-time performance requirements are critical hurdles. Continuous model updates are necessary to address concept drift and ensure interpretability in dynamic environments. Achieving success in real-time forecasting requires careful optimization and an infrastructure capable of handling the demands of real-time processing efficiently.

9.1.5 Market Functionalities and Demand Response

Based on the implementations carried out in Chapter 2 in the field of Demand Response, it is observed that DR will play a decisive role in the future operation of the modern electricity market. Although the simulations do not fully capture consumer behavior—as we know, customer actions generally follow a stochastic path—experimental approaches under real conditions undoubtedly lead to more reliable conclusions. However, executing them requires extensive behavioral analysis, research, and adequate incentives. From the algorithmic analyses conducted, we conclude that if the incentives are sufficient, even if they do not provide a direct financial benefit to consumers, they respond to signals from the network operator and limit their demand when necessary.

These programs contribute to the smoother operation of the market, offer dynamic interaction among all stakeholders, and significantly help reduce costly investments. They also protect energy systems from overloading and mitigate peak demand periods, thus contributing to the balancing of the generation-demand curve and the overall stabilization of power flow as much as possible.

9.1.6 Electric Vehicle Load Curve Forecasting

The study presented in Chapter 6 examined the performance of nine forecasting models, showing that relatively simple Machine Learning (ML) models like XGBoost and MLP outperformed more complex Deep Learning (DL) approaches in predicting the Electric Vehicle Load Curve (EVLC) for Day-Ahead Market (DAM) participation. The models were evaluated using four datasets—”Palo Alto,” ”Crowdcharge,” ”Greenflux,” and ”ACN”—and assessed with multiple metrics, including nMAE, nRMSE, and R^2. Across all datasets, the ML models consistently outperformed the baseline Persistencewo,sc model, demonstrating their ability to capture EVLC trends despite fluctuations and data irregularities.

Although more complex DL models were expected to yield better results, simpler ML approaches proved more effective across different datasets. The study also highlighted the impact of dataset size and characteristics on model performance. For instance, in smaller datasets like ”ACN,” DL models tended to overfit, emphasizing the need for model selection based on data attributes. Additionally, an hourly-level analysis revealed that all models struggled with forecasting during peak load periods, reinforcing the need for further refinements to improve accuracy in these critical time frames.

9.2 Future Study Proposals

Regarding the suggestions for future research arising, it is emphasized that, more broadly, Deep Learning models will represent the cutting edge of Artificial Intelligence and computational science. In all the areas studied, these specific algorithms are the subject of ongoing research and experimentation, due to the flexibility and effectiveness they offer in handling large volumes of data.

Furthermore, the energy market is fully aligning with the advancement of computational systems, aiming to gain flexibility and become fairer for the participants. More specifically, similar to the conclusions drawn, the proposed future research directions are focused on the following areas.

9.2.1 Forecasting Approaches

So, in terms of future study proposals in the forecasting approaches, firstly, managing with non-stationary and nonlinear load data is a challenge, as conventional STLF models typically assume that load data remains stationary and linear. Nonetheless, load data can exhibit non-stationarity and nonlinearity as a result of shifts in consumer behavior, the emergence of new technologies, and various other factors. Future investigations might focus on devising STLF models capable of managing non-stationary and nonlinear load data, potentially utilizing advanced ML models or more adaptable statistical frameworks.

Secondly, the incorporation of online learning is another avenue for improvement. Online learning, a branch of ML, adapts to real-time changes in the energy system. These algorithms continuously update their learning based on new data, allowing them to refine forecasts in real-time. Future studies could aim to develop STLF models that leverage online learning algorithms, thereby improving the accuracy and timeliness of forecasts.

Thirdly, the development of advanced models and the evaluation of their results using cross-validation techniques or not, represents a notable direction for future research, which will provide us with meaningful comparative results and conclusions, as cross-validation ensures a reliable estimate of model performance on future, unseen data.

Fourth, the development of models for DERs, like standalone batteries, is becoming more crucial with the growing adoption of technologies such as rooftop solar panels and integrated energy storage systems. These advancements pose new challenges for STLF models. Future research could focus on creating STLF models that integrate DERs, either by forecasting renewable energy generation or by assessing their impact on electricity demand.

Finally, the investigation of behavior analysis in the field of ELF is crucial, as discussed above. By understanding the patterns and habits of consumers, especially residential users, it is possible to refine load forecasting models. Human behavior significantly influences energy consumption, such as variations in demand based on time of day, weather conditions, or lifestyle choices. Integrating behavior analysis could improve the accuracy of forecasting models by capturing these dynamic factors and could be particularly useful in predicting consumption patterns during peak periods, contributing to more efficient energy management for flexibility and demand response programs.

9.2.2 Electricity Load Forecasting

In the field of ELF, looking ahead several promising directions emerge for future research based on the work have done. The novel automated model introduced could be further improved by integrating more efficient architectures, such as the Exponential Smoothing LSTM (ESLSTM) and advanced Multi-Head Transformers. A key challenge involves extending these models to medium- and long-term multi-step forecasting tasks, in order to assess their robustness over longer prediction horizons. Another important aspect is testing the model's performance on a wider variety of time series datasets, which would help evaluate its generalizability. Additionally, combining the current models with custom activation functions and conducting a comparative analysis of their predictive accuracy could provide further insights. Finally, applying these methods to Demand Side Management and Demand Response problems presents a significant challenge, particularly due to the precision required in short-term forecasting scenarios.

9.2.3 Wholesale Electricity Price Forecasting

Regarding future work in the field of EPF, the initial focus will be on delving into attention-based algorithms for tackling more intricate challenges in predicting electricity prices, which stands as a significant endeavor within the realm of Artificial Intelligence. Additionally, there is a prospect to assess the implementation of the algorithms elaborated in this study on short-term load prediction tasks for the flexibility market and day-ahead forecasting. Moreover, the Soft Voting Ensemble Model underscores the potential of amalgamating diverse deep learning structures to enhance timeseries forecasting. Subsequent research could further investigate alternative model combinations, ensemble tactics, and optimization of weight assignments to push the boundaries of excellence in the field. It is worth noting that the models introduced, alongside their hyperparameter optimization techniques, could be explored in various other domains, including industrial, financial, and biomedical sectors for predictive purposes.

Finally, based on the Explainability process that was carried out, future research could concentrate on Uncertainty Quantification, based on which methods to quantify and explain uncertainties associated with electricity price predictions could be explored. Understanding the level of confidence or uncertainty in forecasts is crucial for decision makers in the energy industry and allows them to make more informed choices regarding resource allocation, risk management, and strategic planning. Furthermore, by incorporating uncertainty quantification into decision processes, stakeholders can respond proactively to potential variations in electricity prices, leading to more resilient and adaptive energy strategies.

9.2.4 Transfer Learning

In the context of future study proposals for Transfer Learning area, it should be emphasized initially that Deep Transfer Learning could be applied to other branches of power systems, such as fault prediction in electric power transmission and distribution networks. Furthermore, beyond energy systems, it could be highlighted that a significant challenge lies in the application of DTL to areas like healthcare, where numerous research studies are conducted globally. As already mentioned, the study of DTL using ELM, both in the field of power systems and in other domains, can help to further improve the results of transfer learning.

Finally, the combination of DTL with reinforcement learning holds promise for future research and offers potential advancements. This could be explored to improve the efficiency of demand forecasting and load management systems. For instance, a pre-trained reinforcement learning agent could learn general patterns and behaviors from historical data across different regions or time periods. This pre-trained agent could then be fine-tuned on a specific locality or timeframe in order to adapt to unique characteristics and changes in electricity demand. This approach may lead to more accurate and adaptable models for load prediction, contributing to improved resource planning and energy efficiency in the electricity grid.

9.2.5 Market Functionalities and Demand Response

Future research in the area of Energy Market Functionalities and Demand Response should aim to explore more dynamic, decentralized, and consumer-centric frameworks that leverage emerging technologies such as blockchain, artificial intelligence, and IoT-enabled smart devices. As energy systems evolve toward greater integration of renewable sources and prosumers, there is a growing need to enhance market mechanisms that support real-time pricing, peer-to-peer energy trading, and automated demand-side flexibility. Future studies could investigate the role of intelligent demand response strategies that optimize both consumer utility and grid stability, particularly under conditions of high renewable penetration and fluctuating demand. Additionally, the design and simulation of market structures that incentivize active participation from residential, commercial, and industrial consumers—while ensuring fairness, transparency, and interoperability—remains a key research challenge. These advancements will be crucial in transitioning to a more resilient, efficient, and sustainable energy ecosystem.

9.2.6 Electric Vehicle Load Curve Forecasting

A promising direction for future research in Electric Vehicle (EV) load forecasting involves the development and refinement of advanced machine learning and deep learning models that can accurately predict EV charging demand at various spatial and temporal resolutions. As EV adoption increases, the variability and uncertainty in charging behavior pose significant challenges to grid stability and energy management. Future studies could focus on integrating real-time data sources—such as traffic patterns, weather conditions, user mobility profiles, and dynamic electricity pricing—into predictive models to enhance their accuracy and adaptability. Moreover, incorporating federated learning or privacy-preserving techniques can allow utilities and researchers to access decentralized data without compromising user privacy. Emphasis should also be placed on developing models that can generalize across different geographic regions and adapt to evolving EV usage trends, enabling smarter grid planning and improved demand-side management strategies.

References

[1] Pablo Ponce, Cristiana Oliveira, Viviana Álvarez, and María de la Cruz del Río-Rama. The liberalization of the internal energy market in the european union: Evidence of its influence on reducing environmental pollution. *Energies*, 13(22), 2020.

[2] What does liberalization and unbundling of energy markets mean? https://www.next-kraftwerke.com/knowledge/liberalization-energy-markets. Access Date: 14-01-2025.

[3] Global energy review: Co2 emissions in 2021. https://www.iea.org/reports/global-energy-review-co2-emissions-in-20212. Access Date: 15-02-2025.

[4] Net zero by 2050 a roadmap for the global energy sector. https://iea.blob.core.windows.net/assets/deebef5d-0c34-4539-9d0c-10b13d840027/NetZeroby2050-ARoadmapfortheGlobalEnergySector_CORR.pdf. Access Date: 26-03-2025.

[5] Reducing carbon emissions: Eu targets and policies. https://www.europarl.europa.eu/topics/en/article/20180305STO99003/reducing-carbon-emissions-eu-targets-and-policies. Access Date: 17-04-2025.

[6] Greenhouse gas emissions by country and sector (infographic). https://www.europarl.europa.eu/topics/en/article/20180301STO98928/greenhouse-gas-emissions-by-country-and-sector-infographic. Access Date: 19-12-2024.

[7] European climate law. https://climate.ec.europa.eu/eu-action/european-climate-law_en. Access Date: 21-02-2025.

[8] Delivering the european green deal on the path to a climate-neutral europe by 2050. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal/delivering-european-green-deal_en#transforming-our-economy-and-societies. Access Date: 23-02-2025.

[9] Fit for 55. https://www.consilium.europa.eu/en/policies/fit-for-55/. Access Date: 25-03-2025.

[10] Commission welcomes completion of key 'fit for 55' legislation, putting eu on track to exceed 2030 targets. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_4754. Access Date: 20-03-2025.

[11] Energy efficiency targets. https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficiency-targets-directive-and-rules/energy-efficiency-targets_en. Access Date: 26-03-2025.

[12] Thorsten Lenck Ing. Konstantin Lenz and Frank Peter. The liberalisation of electricity markets in germany – history, development and current status. In *Agora Energiewende*, Dec. 2019.

[13] Internal energy market. Ημερομηνία πρόσβασης: 12-01-2025.

[14] Jemma J. Makrygiorgou, Christos-Spyridon Karavas, Christos Dikaiakos, and Ioannis P. Moraitis. The electricity market in greece: Current status, identified challenges, and arranged reforms. *Sustainability*, 15(4), 2023.

[15] Ipto. greek balancing market. https://www.admie.gr/en/market/general/description. Access Date: 13-04-2025.

[16] Chalkiadaki Ilektra. The european target model in energy markets. objectives, challenges and prospects. Master's thesis, Department of International and European Studies, 2021.

[17] Entsoe. single intraday coupling (sidc). https://www.entsoe.eu/network_codes/cacm/implementation/sdac/. Access Date: 22-03-2025.

[18] Vasileios Laitsos, Georgios Vontzos, Paschalis Paraschoudis, Eleftherios Tsampasis, Dimitrios Bargiotas, and Lefteri H. Tsoukalas. The state of the art electricity load and price forecasting for the modern wholesale electricity market. *Energies*, 17(22), 2024.

[19] Ipto. technical limits for bidding prices and clearing prices in the balancing market. https://www.admie.gr/sites/default/files/users/dda/KAE/TECHNICAL%20LIMITS_EN.pdf. Access Date: 12-03-2025.

[20] Nikolaos Koltsaklis and Athanasios Dagoumas. Policy implications of power exchanges on operational scheduling: Evaluating euphemia's market products in case of greece. *Energies*, 11(10), 2018.

[21] Filippos Ioannidis, Kyriaki Kosmidou, Kostas Andriosopoulos, and Antigoni Everkiadi. Assessment of the target model implementation in the wholesale electricity market of greece. *Energies*, 14(19), 2021.

[22] Merit order. https://nanoenergies.eu/knowledge-base/merit-order. Access Date: 16-03-2025.

[23] Stelios Loumakis, Eugenia Giannini, and Zacharias Maroulis. Merit order effect modeling: The case of the hellenic electricity market. *Energies*, 12(20), 2019.

[24] Chris Johnathon, Ashish Prakash Agalgaonkar, Joel Kennedy, and Chayne Planiden. Analyzing electricity markets with increasing penetration of large-scale renewable power generation. *Energies*, 14(22), 2021.

[25] What is the merit order curve in the power system? https://www.next-kraftwerke.be/knowledge-hub/merit-order-curve. Access Date: 18-03-2025.

[26] Ferc. an introductory guide for participation in pjm processes. https://www.ferc.gov/introductory-guide-participation-pjm-processes?utm_source=chatgpt.com. Access Date: 27-11-2024.

[27] Nyiso. manual 11 day-ahead scheduling manual. https://www.nyiso.com/documents/20142/2923301/dayahd_schd_mnl.pdf/0024bc71-4dd9-fa80-a816-f9f3e26ea53a?utm_source=chatgpt.com. Access Date: 10-10-2024.

[28] Caiso. day-ahead market enhancements and extended day ahead market business practice summary. https://www.caiso.com/documents/business-practice-summary-extended-day-ahead-market.pdf?utm_source=chatgpt.com. Access Date: 10-10-2024.

[29] Fariba Mousavi, Morteza Nazari-Heris, Behnam Mohammadi-Ivatloo, and Somayeh Asadi. Chapter 1 - energy market fundamentals and overview. In Behnam Mohammadi-Ivatloo, Amin Mohammadpour Shotorbani, and Amjad Anvari-Moghaddam, editors, *Energy Storage in Energy Markets*, pages 1–21. Academic Press, 2021.

[30] Deniz Kenan Kılıç, Peter Nielsen, and Amila Thibbotuwawa. Intraday electricity price forecasting via lstm and trading strategy for the power market: A case study of the west denmark dk1 grid region. *Energies*, 17(12), 2024.

[31] Entsoe. european cross-border intraday (xbid) solution trades exceed 2.5 million since go-live. https://www.entsoe.eu/news/2018/09/06/european-cross-border-intraday-xbid-solution-trades-exceed-2-5-million-since-go-live/. Access Date: 24-04-2025.

[32] Single intraday coupling (sidc). https://www.entsoe.eu/network_codes/cacm/implementation/sidc/. Access Date: 12-01-2025.

[33] Gerald Schweiger, Lisa V. Eckerstorfer, Irene Hafner, Andreas Fleischhacker, Johannes Radl, Barbara Glock, Matthias Wastian, Matthias Rößler, Georg Lettner, Niki Popper, and Katja Corcoran. Active consumer participation in smart energy systems. *Energy and Buildings*, 227:110359, 2020.

[34] Mohammad Mehdi Amiri, Mohammad Taghi Ameli, Goran Strbac, Danny Pudjianto, and Hossein Ameli. The role of flexibility in the integrated operation of low-carbon gas and electricity systems: A review. *Energies*, 17(9), 2024.

[35] next. what is flexibility in the electricity sector? https://www.next-kraftwerke.be/knowledge-hub/flexibility-electricity-sector. Access Date: 21-02-2025.

[36] Vasileios M. Laitsos, Dimitrios Bargiotas, Aspassia Daskalopulu, Athanasios Ioannis Arvanitidis, and Lefteri H. Tsoukalas. An incentive-based implementation of demand side management in power systems. *Energies*, 14(23), 2021.

[37] Diana Khripko, Heiko Dunkelberg, Daniel Summerbell, and Jens Hesselbach. Energy efficiency and demand side management: A case study of a holistic energy concept in polymer processing. *Procedia Manufacturing*, 21:702–709, 01 2018.

[38] Ankit Kumar Sharma and Akash Saxena. A demand side management control strategy using whale optimization algorithm. *Springer Nature*, 1, 2019.

[39] Amit Shewale, Anil Mokhade, Nitesh Funde, and Neeraj Dhanraj Bokde. A survey of efficient demand-side management techniques for the residential appliance scheduling problem in smart homes. *Energies*, 15(8), 2022.

[40] Sven Koenig and Yury V. Smirnov. Graph learning with a nearest neighbor approach. In *Proc. of the 9th Annual Conf. on Computational Learning Theory, COLT '96, Desenzano del Garda, Italy, June 28-July 1, 1996*, pages 19–28. ACM, 1996.

[41] Christian Finck, Paul Beagon, John Clauß, Thibault Péan, Pierre Vogler-Finck, Kun Zhang, and Hussain Kazmi. Review of applied and tested control possibilities for energy flexibility in buildings - a technical report from iea ebc annex 67 energy flexible buildings, 05 2018.

[42] Miadreza Shafie-khah, Ehsan Heydarian-Forushani, Gerardo Osório, Fabio Gil, J. Aghaei, Mostafa Barani, and João Catalão. Optimal behavior of electric vehicle parking lots as demand response aggregation agents. *IEEE Transactions on Smart Grid*, 7:1–12, 11 2015.

[43] Peter Palensky and Dietmar Dietrich. Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE Transactions on Industrial Informatics*, 7(3):381–388, 2011.

[44] Chenyi Xia, Hongjun Lu, Beng Chin Ooi, and Jin Hu. Gorder: An efficient method for KNN join processing. In *Proc. of the Thirtieth Int. Conf. on Very Large Data Bases, VLDB '04, Toronto, Canada, August 31 - September 3 2004*, pages 756–767. Morgan Kaufmann, 2004.

[45] Huu Vu Lam Cao, Trong Nhan Phan, Minh Quang Tran, Thanh Luan Hong, and Minh Nhat Quang Truong. Processing all k-nearest neighbor query on large multidimensional data. In *Proc. of the 2016 Int. Conf. on Advanced Computing and Applications, ACOMP '16, Can Tho City, Vietnam, November 23-25, 2016*, pages 11–17. IEEE, 2016.

[46] Nikolaos G. Paterakis, Ozan Erdinç, and João P.S. Catalão. An overview of demand response: Key-elements and international experience. *Renewable and Sustainable Energy Reviews*, 69:871–891, 2017.

[47] Mohammad Hossein Fouladfar, Abdolah Loni, Mahsa Bagheri Tookanlou, Mousa Marzband, Radu Godina, Ameena Al-Sumaiti, and Edris Pouresmaeil. The impact of demand response programs on reducing the emissions and cost of a neighborhood home microgrid. *Applied Sciences*, 9(10), 2019.

[48] M.H. Albadi and E.F. El-Saadany. A summary of demand response in electricity markets. *Electric Power Systems Research*, 78(11):1989–1996, 2008.

[49] Cui Yu, Bin Cui, Shuguang Wang, and Jianwen Su. Efficient index-based KNN join processing for high-dimensional data. *Information and Software Technology*, 49(4):332–344, 2007.

[50] Waqas Haider Bangyal, Kashif Nisar, Tariq Rahim Soomro, Ag Asri Ag Ibrahim, Ghulam Ali Mallah, Nafees Ul Hassan, and Najeeb Ur Rehman. An improved particle swarm optimization algorithm for data classification. *Applied Sciences*, 13(1), 2023.

[51] Mohammad Reza Bonyadi and Zbigniew Michalewicz. Particle swarm optimization for single objective continuous space problems: A review. *Evolutionary Computation*, 25(1):1–54, 03 2017.

[52] Henerica Tazvinga, Bing Zhu, and Xiaohua Xia. Optimal management of renewable energy sources considering split-diesel and dump energy. *Solar Energy*, 108:412–420, 2014.

[53] Sajjad Asefi, Hossein Shayeghi, Elnaz Shahryari, and rashid dadkhah. Optimal management of renewable energy sources considering split-diesel and dump energy. *International Journal on "Technical and Physical Problems of Engineering" (IJTPE)*, 10:34–40, 03 2018.

[54] James F. Manwell, Jon G. McGowan, and Anthony L. Rogers. *Wind Energy Explained: Theory, Design and Application*. John Wiley & Sons, 2nd edition, 2009.

[55] M. Fahrioglu and F.L. Alvarado. Designing incentive compatible contracts for effective demand management. *IEEE Transactions on Power Systems*, 15(4):1255–1260, 2000.

[56] Bert Willems and Juulia Zhou. The clean energy package and demand response: Setting correct incentives. *Energies*, 13(21), 2020.

[57] Nnamdi I. Nwulu and Xiaohua Xia. Optimal dispatch for a microgrid incorporating renewables and demand response. *Renewable Energy*, 101:16–28, 2017.

[58] Ppc. greek balancing market. `https://www.dei.gr/el/gia-to-spiti/revma/`. Access Date: 15-04-2021.

[59] Anish Tiwari and Anandita Chowdhury. Impact of nl and ev penetration on ieee 13 node test feeder and q-mszs bess charger. In *2024 IEEE 3rd International Conference on Electrical Power and Energy Systems (ICEPES)*, pages 1–5, 2024.

[60] Om Prakash Mahela, Jaya Sharma, Bipul Kumar, Baseem Khan, and Hassan Haes Alhelou. An algorithm for the protection of distribution feeders using the stockwell and hilbert transforms supported features. *CSEE Journal of Power and Energy Systems*, 7(6):1278–1288, 2021.

[61] Gridlab-d technical support document: Residential end-use module version 1.0. `https://www.pnnl.gov/main/publications/external/technical_reports/PNNL-17694.pdf`. Access Date: 14-03-2025.

[62] Amir H. Gandomi, Fang Chen, and Laith Abualigah. Big data analytics using artificial intelligence. *Electronics*, 12(4), 2023.

[63] Flink vs. kafka and their role in the streaming data pipeline. `https://www.redpanda.com/guides/event-stream-processing-flink-vs-kafka`. Access Date: 17-03-2025.

[64] Ana Almeida, Susana Brás, Susana Sargento, and Filipe Cabral, Pinto. Time series big data: a survey on data stream frameworks, analysis and algorithms. *Journal of Big Data*, 10, 2023.

[65] Yulia Kumar, Jose Marchena, Ardalan H. Awlla, J. Jenny Li, and Hemn Barzan Abdalla. The ai-powered evolution of big data. *Applied Sciences*, 14(22), 2024.

[66] History and evolution of machine learning: A timeline. `hhttps://www.techtarget.com/whatis/feature/History-and-evolution-of-machine-learning-A-timeline`. Access Date: 13-07-2024.

[67] Mohammad Mahdi Forootan, Iman Larki, Rahim Zahedi, and Abolfazl Ahmadi. Machine learning and deep learning in energy systems: A review. *Sustainability*, 14(8), 2022.

[68] Mehak Usmani, Zulfiqar Ali Memon, Adil Zulfiqar, and Rizwan Qureshi. Preptimize: Automation of time series data preprocessing and forecasting. *Algorithms*, 17(8), 2024.

[69] Amal Tawakuli, Bastian Havers, Vincenzo Gulisano, Daniel Kaiser, and Thomas Engel. Survey:time-series data preprocessing: A survey and an empirical analysis. *Journal of Engineering Research*, 2024.

[70] Ranjai Baidya and Sang-Woong Lee. Addressing the non-stationarity and complexity of time series data for long-term forecasts. *Applied Sciences*, 14(11), 2024.

[71] Milan Čistý, Michal Danko, Silvia Kohnová, Barbora Považanová, and Andrej Trizna. Machine learning enhanced by feature engineering for estimating snow water equivalent. *Water*, 16(16), 2024.

[72] Raydonal Ospina, João A. M. Gondim, Víctor Leiva, and Cecilia Castro. An overview of forecast analysis with arima models during the covid-19 pandemic: Methodology and case study in brazil. *Mathematics*, 11(14), 2023.

[73] Talal H. Noor, Abdulqader M. Almars, Majed Alwateer, Malik Almaliki, Ibrahim Gad, and El-Sayed Atlam. Sarima: A seasonal autoregressive integrated moving average model for crime analysis in saudi arabia. *Electronics*, 11(23), 2022.

[74] Tryggvi Jónsson, Pierre Pinson, Henrik Aa. Nielsen, and Henrik Madsen. Exponential smoothing approaches for prediction in real-time electricity markets. *Energies*, 7(6):3710–3732, 2014.

[75] Rhenan G. S. Queiroz and Sergio A. David. Performance of the realized-garch model against other garch types in predicting cryptocurrency volatility. *Risks*, 11(12), 2023.

[76] Rosita Guido, Stefania Ferrisi, Danilo Lofaro, and Domenico Conforti. An overview on the advancements of support vector machine models in healthcare applications: A review. *Information*, 15(4), 2024.

[77] Soo-Jin Kim, Seung-Jong Bae, and Min-Won Jang. Linear regression machine learning algorithms for estimating reference evapotranspiration using limited climate data. *Sustainability*, 14(18), 2022.

[78] Jianting Zhang, Ruifei Wang, Ailin Jia, and Naichao Feng. Optimization and application of xgboost logging prediction model for porosity and permeability based on k-means method. *Applied Sciences*, 14(10), 2024.

[79] Shantaram B. Nadkarni, G. S. Vijay, and Raghavendra C. Kamath. Comparative study of random forest and gradient boosting algorithms to predict airfoil self-noise. *Engineering Proceedings*, 59(1), 2023.

[80] Ana Lazcano, Miguel A. Jaramillo-Morán, and Julio E. Sandubete. Back to basics: The power of the multilayer perceptron in financial time series forecasting. *Mathematics*, 12(12), 2024.

[81] Vasileios Laitsos, Georgios Vontzos, Dimitrios Bargiotas, Aspassia Daskalopulu, and Lefteri H Tsoukalas. Enhanced automated deep learning application for short-term load forecasting. *Mathematics*, 11(13):2912, 2023.

[82] Pedro Lara-Benítez, Manuel Carranza-García, José M. Luna-Romera, and José C. Riquelme. Temporal convolutional networks applied to energy-related time series forecasting. *Applied Sciences*, 10(7), 2020.

[83] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.

[84] Da Huang, Zhaoguo Liu, and Dan Wu. Research on ensemble learning-based feature selection method for time-series prediction. *Applied Sciences*, 14(1), 2024.

[85] Arnaud De Myttenaere, Boris Golden, Bénédicte Le Grand, and Fabrice Rossi. Mean absolute percentage error for regression models. *Neurocomputing*, 192:38–48, 2016.

[86] A Colin Cameron and Frank AG Windmeijer. An r-squared measure of goodness of fit for some common nonlinear regression models. *Journal of econometrics*, 77(2):329–342, 1997.

[87] Cort J Willmott and Kenji Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.

[88] Ibrahim Salem Jahan, Vaclav Snasel, and Stanislav Misak. Intelligent systems for power load forecasting: A study review. *Energies*, 13(22), 2020.

[89] Rajat Sethi and Jan Kleissl. Comparison of short-term load forecasting techniques. In *2020 IEEE Conference on Technologies for Sustainability (SusTech)*, pages 1–6. IEEE, 2020.

[90] Jui-Sheng Chou and Duc-Son Tran. Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy*, 165:709–726, 2018.

[91] S. Hadri, Y. Naitmalek, M. Najib, M. Bakhouya, Y. Fakhri, and M. Elaroussi. A comparative study of predictive approaches for load forecasting in smart buildings. *Procedia Computer Science*, 160:173–180, 2019. The 10th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2019) / The 9th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2019) / Affiliated Workshops.

[92] Petra Vrablecová, Anna Bou Ezzeddine, Viera Rozinajová, Slavomír Šárik, and Arun Kumar Sangaiah. Smart grid load forecasting using online support vector regression. *Computers & Electrical Engineering*, 65:102–117, 2018.

[93] Paweł Pełka. Analysis and forecasting of monthly electricity demand time series using pattern-based statistical methods. *Energies*, 16(2), 2023.

[94] Niematallah Elamin and Mototsugu Fukushige. Modeling and forecasting hourly electricity demand by sarimax with interactions. *Energy*, 165:257–268, 2018.

[95] Ismit Mado, Antonius Rajagukguk, Aris Triwiyatno, and Arif Fadllullah. Short-term electricity load forecasting model based dsarima. *International Journal of Electrical, Energy and Power System Engineering*, 5(1):6–11, 2022.

[96] Mobarak Abumohsen, Amani Yousef Owda, and Majdi Owda. Electrical load forecasting using lstm, gru, and rnn algorithms. *Energies*, 16(5), 2023.

[97] Songjiang Li, Wenxin Zhang, and Peng Wang. Ts2arcformer: A multi-dimensional time series forecasting framework for short-term load prediction. *Energies*, 16(15), 2023.

[98] Mesbaholdin Salami, Farzad Movahedi Sobhani, and Mohammad Sadegh Ghazizadeh. Short-term forecasting of electricity supply and demand by using the wavelet-pso-nns-so technique for searching in big data of iran's electricity market. *Data*, 3(4), 2018.

[99] Xin Gao, Xiaobing Li, Bing Zhao, Weijia Ji, Xiao Jing, and Yang He. Short-term electricity load forecasting model based on emd-gru with feature selection. *Energies*, 12(6), 2019.

[100] Bin Yuan, Biao He, Jiang Yan, Jianwen Jiang, Zetao Wei, and Xiaodong Shen. Short-term electricity consumption forecasting method based on empirical mode decomposition of long-short term memory network. *IOP Conference Series: Earth and Environmental Science*, 983(1):012004, feb 2022.

[101] Fangze Zhou, Hui Zhou, Zhaoyan Li, and Kai Zhao. Multi-step ahead short-term electricity load forecasting using vmd-tcn and error correction strategy. *Energies*, 15(15), 2022.

[102] Umar Javed, Khalid Ijaz, Muhammad Jawad, Ejaz A. Ansari, Noman Shabbir, Lauri Kütt, and Oleksandr Husev. Exploratory data analysis based short-term electrical load forecasting: A comprehensive analysis. *Energies*, 14(17), 2021.

[103] Feifei He, Jianzhong Zhou, Zhong kai Feng, Guangbiao Liu, and Yuqi Yang. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with bayesian optimization algorithm. *Applied Energy*, 237:103–116, 2019.

[104] Wan He. Load forecasting via deep neural networks. *Procedia Computer Science*, 122:308–314, 2017. 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.

[105] Heng Shi, Minghao Xu, and Ran Li. Deep learning for household load forecasting—a novel pooling deep rnn. *IEEE Transactions on Smart Grid*, 9(5):5271–5280, 2018.

[106] Kasun Amarasinghe, Daniel L. Marino, and Milos Manic. Deep neural networks for energy load forecasting. In *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, pages 1483–1488, 2017.

[107] Meng Wang, Junqi Yu, Meng Zhou, Wei Quan, and Renyin Cheng. Joint forecasting model for the hourly cooling load and fluctuation range of a large public building based on ga-svm and ig-svm. *Sustainability*, 15(24), 2023.

[108] Michel Bessani, Julio A.D. Massignan, Talysson M.O. Santos, João B.A. London, and Carlos D. Maciel. Multiple households very short-term load forecasting using bayesian networks. *Electric Power Systems Research*, 189:106733, 2020.

[109] Joan Sebastian Caicedo-Vivas and Wilfredo Alfonso-Morales. Short-term load forecasting using an lstm neural network for a grid operator. *Energies*, 16(23), 2023.

[110] Weicong Kong, Zhao Yang Dong, Youwei Jia, David J. Hill, Yan Xu, and Yuan Zhang. Short-term residential load forecasting based on lstm recurrent neural network. *IEEE Transactions on Smart Grid*, 10(1):841–851, 2019.

[111] Maryam Imani and Hassan Ghassemian. Electrical load forecasting using customers clustering and smart meters in internet of things. In *2018 9th International Symposium on Telecommunications (IST)*, pages 113–117, 2018.

[112] M. Pratapa Raju and A. Jaya Laxmi. Iot based online load forecasting using machine learning algorithms. *Procedia Computer Science*, 171:551–560, 2020. Third International Conference on Computing and Network Communications (CoCoNet'19).

[113] Lechen Li, Christoph J. Meinrenken, Vijay Modi, and Patricia J. Culligan. Short-term apartment-level load forecasting using a modified neural network with selected autoregressive features. *Applied Energy*, 287:116509, 2021.

[114] Mohammad Navid Fekri, Harsh Patel, Katarina Grolinger, and Vinay Sharma. Deep learning for load forecasting with smart meter data: Online adaptive recurrent neural network. *Applied Energy*, 282:116177, 2021.

[115] Nikos Andriopoulos, Aristeidis Magklaras, Alexios Birbas, Alex Papalexopoulos, Christos Valouxis, Sophia Daskalaki, Michael Birbas, Efthymios Housos, and George P. Papaioannou. Short term electric load forecasting based on data transformation and statistical machine learning. *Applied Sciences*, 11(1), 2021.

[116] Meshari D. Alanazi, Ahmad Saeed, Muhammad Islam, Shabana Habib, Hammad I. Sherazi, Sheroz Khan, and Mohammad Munawar Shees. Enhancing short-term electrical load forecasting for sustainable energy management in low-carbon buildings. *Sustainability*, 15(24), 2023.

[117] Seyed Mohammad Jafar Jalali, Sajad Ahmadian, Abbas Khosravi, Miadreza Shafie-khah, Saeid Nahavandi, and João P. S. Catalão. A novel evolutionary-based deep convolutional neural network model for intelligent load forecasting. *IEEE Transactions on Industrial Informatics*, 17(12):8243–8253, 2021.

[118] Gopal Chitalia, Manisa Pipattanasomporn, Vishal Garg, and Saifur Rahman. Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks. *Applied Energy*, 278:115410, 2020.

[119] William L Hamilton. *Graph representation learning*. Morgan & Claypool Publishers, 2020.

[120] Georgios Vontzos, Vasileios Laitsos, Avraam Charakopoulos, Dimitrios Bargiotas, and Theodoros E Karakasidis. Estimating spatio-temporal building power consumption based on graph convolution network method. *Dynamics*, 4(2):337–356, 2024.

[121] Yuqing Hu, Xiaoyuan Cheng, Suhang Wang, Jianli Chen, Tianxiang Zhao, and Enyan Dai. Times series forecasting for urban building energy consumption based on graph convolutional network. *Applied Energy*, 307:118231, 2022.

[122] Minjie Wei, Mi Wen, and Yi Zhang. A novel spatial electric load forecasting method based on ldtw and gcn. *IET Generation, Transmission & Distribution*, 18(3):491–505, 2024.

[123] Xiaoqun Liao, Nanlan Cao, Ma Li, and Xiaofan Kang. Research on short-term load forecasting using xgboost based on similar days. In *2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, pages 675–678, 2019.

[124] Zhenhai Guo, Weigang Zhao, Haiyan Lu, and Jianzhou Wang. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. *Renewable Energy*, 37(1):241–249, 2012.

[125] Mohammad H. Alobaidi, Fateh Chebana, and Mohamed A. Meguid. Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. *Applied Energy*, 212:997–1012, 2018.

[126] Pei Du, Yu Jin, and Kequan Zhang. A hybrid multi-step rolling forecasting model based on ssa and simulated annealing—adaptive particle swarm optimization for wind speed. *Sustainability*, 8(8), 2016.

[127] Xuejiao Ma, Yu Jin, and Qingli Dong. A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. *Applied Soft Computing*, 54:296–312, 2017.

[128] Hossein Javedani Sadaei, Petrônio Cândido de Lima e Silva, Frederico Gadelha Guimarães, and Muhammad Hisyam Lee. Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. *Energy*, 175:365–377, 2019.

[129] Shafiul Hasan Rafi, . Nahid-Al-Masood, Shohana Rahman Deeba, and Eklas Hossain. A short-term load forecasting method using integrated cnn and lstm network. *IEEE Access*, 9:32436–32448, 2021.

[130] Jiakang Wang, Hui Liu, Guangji Zheng, Ye Li, and Shi Yin. Short-term load forecasting based on outlier correction, decomposition, and ensemble reinforcement learning. *Energies*, 16(11), 2023.

[131] Mahmoud A. Hammad, Borut Jereb, Bojan Rosi, and Dejan Dragan. Methods and models for electric load forecasting: A comprehensive review. *Logistics, Supply Chain, Sustainability and Global Challenges*, 11(1):51–76, 2020.

[132] Tao Hong and Shu Fan. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3):914–938, 2016.

[133] Ozgur Cemal Ozerdem, Ebenezer O. Olaniyi, and Oyebade K. Oyedotun. Short term load forecasting using particle swarm optimization neural network. *Procedia Computer Science*, 120:382–393, 2017. 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, 22-23 August 2017, Budapest, Hungary.

[134] John D. Hobby and Gabriel H. Tucci. Analysis of the residential, commercial and industrial electricity consumption. In *2011 IEEE PES Innovative Smart Grid Technologies*, pages 1–7, 2011.

[135] Baxter Williams, Daniel Bishop, Patricio Gallardo, and J. Geoffrey Chase. Demand side management in industrial, commercial, and residential sectors: A review of constraints and considerations. *Energies*, 16(13), 2023.

[136] Wenzhi Cao, Houdun Liu, Xiangzhi Zhang, and Yangyan Zeng. Residential load forecasting based on long short-term memory, considering temporal local attention. *Sustainability*, 16(24), 2024.

[137] Guozhen Xia, Fubin Zhang, Cheng Wang, and Chen Zhou. Ed-convlstm: A novel global ionospheric total electron content medium-term forecast model. *Space Weather*, 20(8):e2021SW002959, 2022.

[138] Si Wang, Lin Mu, and Darong Liu. A hybrid approach for el niño prediction based on empirical mode decomposition and convolutional lstm encoder-decoder. *Computers & Geosciences*, 149:104695, 2021.

[139] Huixin Tian and Jianhua Chen. Deep learning with spatial attention-based conv-lstm for soc estimation of lithium-ion batteries. *Processes*, 10(11):2185, 2022.

[140] Jason Brownlee. How does attention work in encoder-decoder recurrent neural networks, 2019.

[141] Metar - wikipedia. `https://en.wikipedia.org/wiki/METAR`. Access Date: 12-01-2023.

[142] Global Modeling and Assimilation Office (GMAO), Goddard Earth Sciences Data and Information Services Center (GES DISC). MERRA-2 inst1_2d_asm_Nx: 2d, 1-hourly, instantaneous, single-level,assimilation, single-level diagnostics 0.625 x 0.5 degree v5.12.4 (m2i1nxasm) at ges disc, 2015.

[143] Publication of nii daily energy planning data | hedno. `https://deddie.gr/en/themata-tou-diaxeiristi-mi-diasundedemenwn-nisiwn/leitourgia-mdn/dimosieusi-imerisiou-energeiakou-programmatismou/`. Accessed: (2023-01-10).

[144] Claude Sammut and Geoffrey I. Webb, editors. *Encyclopedia of Machine Learning*, pages 652–653. Springer US, Boston, MA, 2010.

[145] Arnaud de Myttenaere, Boris Golden, Bénédicte Le Grand, and Fabrice Rossi. Mean Absolute Percentage Error for regression models. *Neurocomputing*, 192:38–48, jun 2016.

[146] Davide Chicco, Matthijs J Warrens, and Giuseppe Jurman. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, 7:e623, 2021.

[147] Alberto Gasparin, Slobodan Lukovic, and Cesare Alippi. Deep learning for time series forecasting: The electric load case. *CAAI Transactions on Intelligence Technology*, 7(1):1–25, 2022.

[148] Vasileios Laitsos, Georgios Vontzos, and Dimitrios Bargiotas. Investigation of transfer learning for electricity load forecasting. In *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pages 1–7. IEEE, 2023.

[149] Georgios Vontzos, Vasileios Laitsos, and Dimitrios Bargiotas. Data-driven airport multi-step very short-term load forecasting. In *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pages 1–6. IEEE, 2023.

[150] Almendra Awerkin, Humberto Verdejo, Cristhian Becker, and Karina Barbosa. Statistical parametric techniques for power residential demand forecasting. In *2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, pages 1–7, 2017.

[151] Chinthaka Dinesh, Stephen Makonin, and Ivan V. Bajić. Residential power forecasting based on affinity aggregation spectral clustering. *IEEE Access*, 8:99431–99444, 2020.

[152] M.A. Rafe Biswas, Melvin D. Robinson, and Nelson Fumo. Prediction of residential building energy consumption: A neural network approach. *Energy*, 117:84–92, 2016.

[153] Tae-Young Kim and Sung-Bae Cho. Predicting residential energy consumption using cnn-lstm neural networks. *Energy*, 182:72–81, 2019.

[154] Fath U Min Ullah, Amin Ullah, Ijaz Ul Haq, Seungmin Rho, and Sung Wook Baik. Short-term prediction of residential power energy consumption via cnn and multi-layer bi-directional lstm networks. *IEEE Access*, 8:123369–123380, 2020.

[155] Hedno. `https://deddie.gr/en/`. Access Date: 10-12-2023.

[156] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

[157] Robert Skopal. Short-term hourly price forward curve prediction using neural network and hybrid arima-nn model. *2015 International Conference on Information and Digital Technologies*, pages 335–338, 2015.

[158] Heping Liu and Jing Shi. Applying arma–garch approaches to forecasting short-term electricity prices. *Energy Economics*, 37:152–166, 2013.

[159] Preethi Rajan and K.R.M Vijaya Chandrakala. Statistical model approach of electricity price forecasting for indian electricity market. In *2021 IEEE Madras Section Conference (MASCON)*, pages 1–5, 2021.

[160] Daniel Bissing, Michael T. Klein, Radhakrishnan Angamuthu Chinnathambi, Daisy Flora Selvaraj, and Prakash Ranganathan. A hybrid regression model for day-ahead energy price forecasting. *IEEE Access*, 7:36833–36842, 2019.

[161] Grzegorz Marcjasz, Tomasz Serafin, and Rafał Weron. Selection of calibration windows for day-ahead electricity price forecasting. *Energies*, 11(9), 2018.

[162] Agustín A. Sánchez de la Nieta, Virginia González, and Javier Contreras. Portfolio decision of short-term electricity forecasted prices through stochastic programming. *Energies*, 9(12), 2016.

[163] Mustafa Abunofal, Nikhilkumar Poshiya, Ramiz Qussous, and Anke Weidlich. Comparative analysis of electricity market prices based on different forecasting methods. In *2021 IEEE Madrid PowerTech*, pages 1–6. IEEE, 2021.

[164] Sergio Beltrán, Sergio Beltrán, Alain Castro, Alain Castro, I. Irizar, I. Irizar, Gorka Naveran, and Imanol Yeregui. Framework for collaborative intelligence in forecasting day-ahead electricity price. *Applied Energy*, 2022.

[165] Aurora Poggi, Luca Di Persio, and Matthias Ehrhardt. Electricity price forecasting via statistical and deep learning approaches: The german case. *AppliedMath*, 3(2):316–342, 2023.

[166] Mustafa Abunofal, Nikhilkumar Poshiya, Ramiz Qussous, and Anke Weidlich. Comparative analysis of electricity market prices based on different forecasting methods. In *2021 IEEE Madrid PowerTech*, pages 1–6, 2021.

[167] Behrouz Banitalebi, Md. Erfanul Hoque, Srimantoorao S. Appadoo, and Aerambamoorthy Thavaneswaran. Regularized probabilistic forecasting of electricity wholesale price and demand. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 28–35, 2020.

[168] Anna Shikhina, Alexei Kochengin, George Chrysostomou, and Vladimir Shikhin. Investigation of autoregressive forecasting models for market electricity price. In *2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON)*, pages 570–575, 2020.

[169] Fatin Khairunnisa Khairuddin, Farah Anishah Zaini, Mohamad Fani Sulaima, Nur Hazahsha Shamsudin, and Mohd Shahrin Abu Hanifah. A single-buyer model of imbalance cost pass-through pricing forecasting in the malaysian electricity supply industry. *Electricity*, 5(2):295–312, 2024.

[170] Negin Entezari and José Alberto Fuinhas. Quantifying the impact of risk on market volatility and price: Evidence from the wholesale electricity market in portugal. *Sustainability*, 16(7), 2024.

[171] Nikola Mišnić, Bojan Pejović, Jelena Jovović, Sunčica Rogić, and Vladimir Đurišić. The economic viability of pv power plant based on a neural network model of electricity prices forecast: A case of a developing market. *Energies*, 15(17), 2022.

[172] Virginia Gonzalez, Javier Contreras, and Derek W. Bunn. Forecasting power prices using a hybrid fundamental-econometric model. *IEEE Transactions on Power Systems*, 27(1):363–372, 2012.

[173] Sajjad Khan, Shahzad Aslam, Iqra Mustafa, and Sheraz Aslam. Short-term electricity price forecasting by employing ensemble empirical mode decomposition and extreme learning machine. *Forecasting*, 3(3):28, 2021.

[174] Huixin Liu, Xiaodong Shen, Xisheng Tang, and Junyong Liu. Day-ahead electricity price probabilistic forecasting based on shap feature selection and lstnet quantile regression. *Energies*, 16(13):5152, 2023.

[175] Milutin Pavićević and Tomo Popović. Forecasting day-ahead electricity metrics with artificial neural networks. *Sensors*, 22(3):1051, 2022.

[176] Biao Zhang, Chao Song, Xuchu Jiang, and Ying Li. Electricity price forecast based on the stl-tcn-nbeats model. *Heliyon*, 9(1), 2023.

[177] Mughees Ali, Zahoor Ali Khan, Sana Mujeeb, Shahid Abbas, and Nadeem Javaid. Short-term electricity price and load forecasting using enhanced support vector machine and k-nearest neighbor. In *2019 Sixth HCT Information Technology Trends (ITT)*, pages 79–83. IEEE, 2019.

[178] Ali Najem Alkawaz, Abdallah Abdellatif, Jeevan Kanesan, Anis Salwa Mohd Khairuddin, and Hassan Muwafaq Gheni. Day-ahead electricity price forecasting based on hybrid regression model. *IEEE Access*, 10:108021–108033, 2022.

[179] Athanasios Ioannis Arvanitidis, Dimitrios Bargiotas, Dimitrios Kontogiannis, Athanasios Fevgas, and Miltiadis Alamaniotis. Optimized data-driven models for short-term electricity price forecasting based on signal decomposition and clustering techniques. *Energies*, 15(21):7929, 2022.

[180] Mahmood Hosseini Imani, Ettore Bompard, Pietro Colella, and Tao Huang. Predictive methods of electricity price: an application to the italian electricity market. In *2020 IEEE*

International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC I&CPS Europe), pages 1–6, 2020.

[181] Hamza Mubarak, Shameem Ahmad, Al Amin Hossain, Ben Horan, Abdallah Abdellatif, Saad Mekhilef, Mehdi Seyedmahmoudian, Alex Stojcevski, Hazlie Mokhlis, Jeevan Kanesan, and Mohamed Becherif. Short-term electricity price forecasting using interpretable hybrid machine learning models. In *2023 IEEE IAS Global Conference on Renewable Energy and Hydrogen Technologies (GlobConHT)*, pages 1–6, 2023.

[182] Alberto Menéndez Medina and José Antonio Heredia Álvaro. Using generative pretrained transformers (gpt) for electricity price trend forecasting in the spanish market. *Energies*, 17(10), 2024.

[183] Prasenjit Chatterjee, Morteza Yazdani, Francisco Fernández-Navarro, and Javier Pérez-Rodríguez. CRC Press, 2023.

[184] N Krishna Prakash, Jai Govind Singh, et al. Electricity price forecasting using hybrid deep learned networks. *Journal of Forecasting*, 42(7):1750–1771, 2023.

[185] Wenjie Zhang, Farwa Cheema, and Dipti Srinivasan. Forecasting of electricity prices using deep learning networks. In *2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pages 451–456, 2018.

[186] Siyu Zhou, Lin Zhou, Mingxuan Mao, Heng-Ming Tai, and Yihao Wan. An optimized heterogeneous structure lstm network for electricity price forecasting. *IEEE Access*, 7:108161–108173, 2019.

[187] Hanhong Shi, Lei Wang, Rafał Scherer, Marcin Woźniak, Pengchao Zhang, and Wei Wei. Short-term load forecasting based on adabelief optimized temporal convolutional network and gated recurrent unit hybrid neural network. *IEEE Access*, 9:66965–66981, 2021.

[188] Ekaterina Popovska and Galya Georgieva-Tsaneva. Day-ahead electricity price forecasting using long-short term memory recurrent neural network. *Innovative STEM Education*, 4:139–148, 2022.

[189] Nazia Daood, Zahoor Ali Khan, Ashrafullah, Muhammad Jaffar Khan, Muhammad Adil, and Nadeem Javaid. *Electricity Price Forecasting Based on Enhanced Convolutional Neural Network in Smart Grid*, pages 1212–1226. Springer, 2020.

[190] Alessandro Brusaferri, Matteo Matteucci, Pietro Portolani, and Andrea Vitali. Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. *Applied Energy*, 250:1158–1175, 2019.

[191] Shan Zhang, Powei Chen, and Ziqiang Yang. Electricity price prediction based on a new hybrid model. In *2022 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*, pages 995–999. IEEE, 2022.

[192] Bo Wang, Wenting Wei, and Wang Su. Short-term electricity price forecasting based on data mining. In *2022 2nd International Conference on Algorithms, High Performance Computing and Artificial Intelligence (AHPCAI)*, pages 394–397. IEEE, 2022.

[193] Nafise Rezaei, Roozbeh Rajabi, and Abouzar Estebsari. Electricity price forecasting model based on gated recurrent units. In *2022 IEEE International Conference on Environment and Electrical Engineering and 2022 IEEE Industrial and Commercial Power Systems Europe (EEEICI&CPS Europe)*, pages 1–5. IEEE, 2022.

[194] Dimitrios Kontogiannis, Dimitrios Bargiotas, Aspassia Daskalopulu, Athanasios Ioannis Arvanitidis, and Lefteri H Tsoukalas. Error compensation enhanced day-ahead electricity price forecasting. *Energies*, 15(4):1466, 2022.

[195] Thiago Conte and Roberto Oliveira. Comparative analysis between intelligent machine committees and hybrid deep learning with genetic algorithms in energy sector forecasting: A case study on electricity price and wind speed in the brazilian market. *Energies*, 17(4), 2024.

[196] Vasileios Laitsos, Georgios Vontzos, Dimitrios Bargiotas, Aspassia Daskalopulu, and Lefteri H. Tsoukalas. Data-driven techniques for short-term electricity price forecasting through novel deep learning approaches with attention mechanisms. *Energies*, 17(7), 2024.

[197] Arim Jin, Dahan Lee, Jong-Bae Park, and Jae Hyung Roh. Day-ahead electricity market price forecasting considering the components of the electricity market price; using

demand decomposition, fuel cost, and the kernel density estimation. *Energies*, 16(7), 2023.

[198] Denis E. Baskan, Daniel Meyer, Sebastian Mieck, Leonhard Faubel, Benjamin Klöpper, Nika Strem, Johannes A. Wagner, and Jan J. Koltermann. A scenario-based model comparison for short-term day-ahead electricity prices in times of economic and political tension. *Algorithms*, 16(4), 2023.

[199] Miguel Pinhão, Miguel Fonseca, and Ricardo Covas. Electricity spot price forecast by modelling supply and demand curve. *Mathematics*, 10(12), 2022.

[200] Leena Heistrene, Juri Belikov, Dmitry Baimel, Liran Katzir, Ram Machlev, Kfir Levy, Shie Mannor, and Yoash Levron. An improved and explainable electricity price forecasting model via shap-based error compensation approach. *IEEE Transactions on Artificial Intelligence*, pages 1–11, 2024.

[201] Leena Heistrene, Ram Machlev, Michael Perl, Juri Belikov, Dmitry Baimel, Kfir Levy, Shie Mannor, and Yoash Levron. Explainability-based trust algorithm for electricity price forecasting models. *Energy and AI*, 14:100259, 2023.

[202] Jun Dong, Xihao Dou, Aruhan Bao, Yaoyu Zhang, and Dongran Liu. Day-ahead spot market price forecast based on a hybrid extreme learning machine technique: A case study in china. *Sustainability*, 14(13), 2022.

[203] Mustafa Saglam, Xiaojing Lv, Catalina Spataru, and Omer Ali Karaman. Instantaneous electricity peak load forecasting using optimization and machine learning. *Energies*, 17(4), 2024.

[204] Kushagra Bhatia, Rajat Mittal, Jyothi Varanasi, and M.M. Tripathi. An ensemble approach for electricity price forecasting in markets with renewable energy resources. *Utilities Policy*, 70:101185, 2021.

[205] Weishang Guo and Zhenyu Zhao. A novel hybrid bnd-foa-lssvm model for electricity price forecasting. *Information*, 8(4), 2017.

[206] Laiqing Yan, Zhenwen Li, Ning Ma, and Ran Li. Electricity market price prediction based on quadratic hybrid decomposition and thpo algorithm. *Energies*, 16(13):5098–5098, 2023.

[207] Khalid Rashid Chughatta and Shaikh Saaqib Haroon. Short-term electricity price forecasting using eemd and gru-nn. *International Journal of Advanced Natural Sciences and Engineering Researches*, 7(4):420–427, 2023.

[208] Haolin Yang and Kristen R. Schell. Ghtnet: Tri-branch deep learning network for real-time electricity price forecasting. *Energy*, 238:122052, 2022.

[209] Hamidreza Jahangir, Hanif Tayarani, Sina Baghali, Ali Ahmadian, Ali Elkamel, Masoud Aliakbar Golkar, and Miguel Castilla. A novel electricity price forecasting approach based on dimension reduction strategy and rough artificial neural networks. *IEEE Transactions on Industrial Informatics*, 16(4):2369–2381, 2020.

[210] Morteza Gholipour Khajeh, Akbar Maleki, Marc A. Rosen, and Mohammad Hossein Ahmadi. Electricity price forecasting using neural networks with an improved iterative training algorithm. *International Journal of Ambient Energy*, 39:147 – 158, 2018.

[211] Jinliang Zhang, Zhongfu Tan, and Shuxia Yang. Day-ahead electricity price forecasting by a new hybrid method. *Computers & Industrial Engineering*, 63(3):695–701, 2012. Energy Management and Economics.

[212] Daniel Manfre Jaimes, Manuel Zamudio López, Hamidreza Zareipour, and Mike Quashie. A hybrid model for multi-day-ahead electricity price forecasting considering price spikes. *Forecasting*, 5(3):499–521, 2023.

[213] Vasileios Laitsos, Georgios Vontzos, Dimitrios Bargiotas, Aspassia Daskalopulu, and Lefteri H. Tsoukalas. Enhanced automated deep learning application for short-term load forecasting. *Mathematics*, 11(13), 2023.

[214] Athanasios Ioannis Arvanitidis, Dimitrios Kontogiannis, Georgios Vontzos, Vasileios Laitsos, and Dimitrios Bargiotas. Stochastic heuristic optimization of machine learning estimators for short-term wind power forecasting. In *2022 57th International Universities Power Engineering Conference (UPEC)*, pages 1–6. IEEE, 2022.

[215] Deqiang Gan, Donghan Feng, and Jun Xie. *Electricity markets and power system economics*. CRC Press, 2013.

[216] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon,

U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[217] Zhihe Yang, Jiandun Li, Haitao Wang, and Chang Liu. An informer model for very short-term power load forecasting. *Energies*, 18(5), 2025.

[218] Bulent Ayhan, Erik P. Vargo, and Huang Tang. On the exploration of temporal fusion transformers for anomaly detection with multivariate aviation time-series data. *Aerospace*, 11(8), 2024.

[219] Zhu Baozhou, Peter Hofstee, Jinho Lee, and Zaid Al-Ars. An attention module for convolutional neural networks. *arXiv preprint arXiv:2108.08205*, 2021.

[220] Jiyan Chen and Zijiang Yang. Revolutionizing time series data preprocessing with a novel cycling layer in self-attention mechanisms. *Applied Sciences*, 14(19), 2024.

[221] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[222] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017.

[223] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[224] German electricity market. `https://www.smard.de/en`. Accessed: (2023-09-02).

[225] scikit-learn. `https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html`. Accessed: (2023-09-04).

[226] Mohammadreza Iman, Hamid Reza Arabnia, and Khaled Rasheed. A review of deep transfer learning and recent advancements. *Technologies*, 11(2), 2023.

[227] Zhuoqun Xing, Yiqun Pan, Yiting Yang, Xiaolei Yuan, Yumin Liang, and Zhizhong Huang. Transfer learning integrating similarity analysis for short-term and long-term building energy consumption prediction. *Applied Energy*, 365:123276, 2024.

[228] Gabriel Michau and Olga Fink. Unsupervised transfer learning for anomaly detection: Application to complementary operating condition transfer. *Knowledge-Based Systems*, 216:106816, 2021.

[229] Rami Al-Hajj, Ali Assi, Bilel Neji, Raymond Ghandour, and Zaher Al Barakeh. Transfer learning for renewable energy systems: A survey. *Sustainability*, 15(11), 2023.

[230] Songping Meng, Chengdong Li, Chongyi Tian, Wei Peng, and Chenlu Tian. Transfer learning based graph convolutional network with self-attention mechanism for abnormal electricity consumption detection. *Energy Reports*, 9:5647–5658, 2023.

[231] Anestis Antoniadis, Solenne Gaucher, and Yannig Goude. Hierarchical transfer learning with applications to electricity load forecasting. *International Journal of Forecasting*, 2023.

[232] Ce Yang, Haiyan Wang, Jiaxin Bai, Tiancheng He, Huhu Cheng, Tianlei Guang, Houze Yao, and Liangti Qu. Transfer learning enhanced water-enabled electricity generation in highly oriented graphene oxide nanochannels. *Nature Communications*, 13(1):6819, 2022.

[233] Yunxuan Dong and Ling Xiao. A transfer learning based deep model for electrical load prediction. In *2022 IEEE 8th International Conference on Computer and Communications (ICCC)*, pages 2251–2255, 2022.

[234] Dandan Li, Jiangfeng Li, Xin Zeng, Vladimir Stankovic, Lina Stankovic, Changjiang Xiao, and Qingjiang Shi. Transfer learning for multi-objective non-intrusive load monitoring in smart building. *Applied Energy*, 329:120223, 2023.

[235] Thijs Peirelinck, Hussain Kazmi, Brida V Mbuwir, Chris Hermans, Fred Spiessens, Johan Suykens, and Geert Deconinck. Transfer learning in demand response: A review of algorithms for data-efficient modelling and control. *Energy and AI*, 7:100126, 2022.

[236] Di Wu and Weixuan Lin. Efficient residential electric load forecasting via transfer learning and graph neural networks. *IEEE Transactions on Smart Grid*, 14(3):2423–2431, 2023.

[237] Firuz Kamalov, Hana Sulieman, Sherif Moussa, Jorge Avante Reyes, and Murodbek Safaraliev. Powering electricity forecasting with transfer learning. *Energies*, 17(3), 2024.

[238] Dabeeruddin Syed, Ameema Zainab, Shady S. Refaat, Haitham Abu-Rub, Othmane Bouhali, Ali Ghrayeb, Mahdi Houchati, and Santiago Bañales. Inductive transfer and deep neural network learning-based cross-model method for short-term load forecasting in smarts grids. *IEEE Canadian Journal of Electrical and Computer Engineering*, 46(2):157–169, 2023.

[239] Miguel López Santos, Saúl Díaz García, Xela García-Santiago, Ana Ogando-Martínez, Fernando Echevarría Camarero, Gonzalo Blázquez Gil, and Pablo Carrasco Ortega. Deep learning and transfer learning techniques applied to short-term load forecasting of data-poor buildings in local energy communities. *Energy and Buildings*, 292:113164, 2023.

[240] Athanasios Ioannis Arvanitidis, Dimitrios Bargiotas, Aspassia Daskalopulu, Dimitrios Kontogiannis, Ioannis P Panapakidis, and Lefteri H Tsoukalas. Clustering informed mlp models for fast and accurate short-term load forecasting. *Energies*, 15(4):1295, 2022.

[241] Tongtong Luo, Zhiyuan Tang, Junyong Liu, and Bo Zhou. A review of transfer learning approaches for load, solar and wind power predictions. In *2023 Panda Forum on Power and Energy (PandaFPE)*, pages 1580–1584, 2023.

[242] Shiwei Li, Hongbin Wu, Xiaoming Wang, Bin Xu, Long Yang, and Rui Bi. Short-term load forecasting based on am-cif-lstm method adopting transfer learning. *Frontiers in Energy Research*, 11:1162040, 2023.

[243] S Chan, I Oktavianti, and V Puspita. A deep learning cnn and ai-tuned svm for electricity consumption forecasting: Multivariate time series data. In *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 0488–0494. IEEE, 2019.

[244] Tiago Silveira Gontijo and Marcelo Azevedo Costa. Forecasting hierarchical time series in power generation. *Energies*, 13(14), 2020.

[245] Seung-Min Jung, Sungwoo Park, Seung-Won Jung, and Eenjun Hwang. Monthly electric load forecasting using transfer learning for smart cities. *Sustainability*, 12(16):6364, 2020.

[246] Rami Al-Hajj, Ali Assi, Bilel Neji, Raymond Ghandour, and Zaher Al Barakeh. Transfer learning for renewable energy systems: A survey. *Sustainability*, 15(11):9131, 2023.

[247] Chandana Priya Nivarthi. Transfer learning as an essential tool for digital twins in renewable energy systems. *arXiv preprint arXiv:2203.05026*, 2022.

[248] Seyed Mahdi Miraftabzadeh, Cristian Giovanni Colombo, Michela Longo, and Federica Foiadelli. A day-ahead photovoltaic power prediction via transfer learning and deep neural networks. *Forecasting*, 5(1):213–228, 2023.

[249] Damir Đaković, Miroslav Kljajić, Nikola Milivojević, Đorđije Doder, and Aleksandar S. Anđelković. Review of energy-related machine learning applications in drying processes. *Energies*, 17(1), 2024.

[250] Mingzhi Yang, Yue Liu, and Quanlong Liu. Nonintrusive residential electricity load decomposition based on transfer learning. *Sustainability*, 13(12):6546, 2021.

[251] Kangji Li, Borui Wei, Qianqian Tang, and Yufei Liu. A data-efficient building electricity load forecasting method based on maximum mean discrepancy and improved tradaboost algorithm. *Energies*, 15(23):8780, 2022.

[252] T. Silveira Gontijo, R. Barbosa de Santis, and M. Azevedo Costa. Application of a data-driven dtsf and benchmark models for the prediction of electricity prices in brazil: A time-series case. *Journal of Renewable and Sustainable Energy*, 2023.

[253] Fayez Alanazi. Electric vehicles: Benefits, challenges, and potential solutions for widespread adaptation. *Applied Sciences*, 13(10), 2023.

[254] Geraldo L. Maia, Caio C. L. Santos, Paulo R. M. Nunes, José F. C. Castro, Davidson C. Marques, Luiz H. A. De Medeiros, Leonardo R. Limongi, Márcio E. C. Brito, Nicolau K. L. Dantas, Antônio V. M. L. Filho, Amanda L. Fernandes, Jiyong Chai, and Chenxin Zhang. Ev smart-charging strategy for power management in distribution grid with high penetration of distributed generation. *Energies*, 17(21), 2024.

[255] Julia Hildermeier, Christos Kolokathis, Jan Rosenow, Michael Hogan, Catharina Wiese, and Andreas Jahn. Smart ev charging: A global review of promising practices. *World Electric Vehicle Journal*, 10(4), 2019.

[256] Megan Zielke, Adria Brooks, and Gregory Nemet. The impacts of electric vehicle growth on wholesale electricity prices in wisconsin. *World Electric Vehicle Journal*, 11(2), 2020.

[257] Wattcrop. the impact of electric vehicles on the power grid. `https://wattcrop.com/the-impact-of-electric-vehicles-on-the-power-grid/`. Access Date: 17-04-2025.

[258] Takuya Yokoyama, Yoshiharu Ishikawa, and Yu Suzuki. Processing all k-nearest neighbor queries in hadoop. In *Proc. of the 13th Int. Conf. on Web-Age Information Management, WAIM '12, Harbin, China, August 18-20, 2012*, volume 7418 of *LNCS*, pages 346–351. Springer, 2012.

[259] Nikolaos Nodarakis, Evaggelia Pitoura, Spyros Sioutas, Athanasios K. Tsakalidis, Dimitrios Tsoumakos, and Giannis Tzimas. kdann+: A rapid aknn classifier for big data. *Transactions on Large-Scale Data- and Knowledge-Centered Systems*, 24:139–168, 2016.

[260] Christian Böhm and Florian Krebs. The *k*-nearest neighbour join: Turbo charging the KDD process. *Knowledge and Information Systems*, 6:728–749, 2004.

[261] Wei Lu, Yanyan Shen, Su Chen, and Beng Chin Ooi. Efficient processing of k nearest neighbor joins using mapreduce. *Proc. of the VLDB Endowment*, 5(10):1016–1027, 2012.

[262] Yun Chen and Jignesh M. Patel. Efficient evaluation of all-nearest-neighbor queries. In *Proc. of the 23rd Int. Conf. on Data Engineering, ICDE '07, The Marmara Hotel, Istanbul, Turkey, April 15-20, 2007*, pages 1056–1065. IEEE, 2007.

[263] Tobias Emrich, Franz Graf, Hans-Peter Kriegel, Matthias Schubert, and Marisa Thoma. Optimizing all-nearest-neighbor queries with trigonometric pruning. In *Proc. of the 22nd Int. Conf. on the Scientific and Statistical Database Management, SSDBM '10, Heidelberg, Germany, June 30 - July 2, 2010*, volume 6187 of *LNCS*, pages 501–518. Springer, 2010.

[264] Ahmed Eldawy and Mohamed F. Mokbel. The era of big spatial data. *Proc. of the VLDB Endowment*, 10(12):1992–1995, 2017.

[265] Binrong Wu, Lin Wang, and Yu-Rong Zeng. Interpretable wind speed prediction with multivariate time series and temporal fusion transformers. *Energy*, 252:123990, 2022.

[266] Daniel J. Eisenstein and Piet Hut. Hop: A new group-finding algorithm for n-body simulations. *The Astrophysical Journal*, 498(1):137–142, 1998.

[267] Bryan Lim, Sercan Ö. Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021.

[268] Mingping Liu, Xihao Sun, Qingnian Wang, and Suhui Deng. Short-term load forecasting using emd with feature selection and tcn-based deep learning model. *Energies*, 15(19), 2022.

[269] Nick Roussopoulos, Stephen Kelley, and Frédéic Vincent. Nearest neighbor queries. In *Proc. of the 1995 ACM SIGMOD Int. Conf. on Management of Data, San Jose, California, USA, May 22-25, 1995*, pages 71–79. ACM, 1995.