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

# Short-Term Load Forecasting Using Artificial Neural Networks

MSc Thesis

**Arvanitidis Athanasios Ioannis**

**Supervisor:** Bargiotas Dimitrios

June 2021

UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

# Short-Term Load Forecasting Using Artificial Neural Networks

MSc Thesis

**Arvanitidis Athanasios Ioannis**

**Supervisor:** Bargiotas Dimitrios

June 2021

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

# Βραχυπρόθεσμη Πρόβλεψη Φορτίου με τη Χρήση Τεχνητών Νευρωνικών Δικτύων

Μεταπτυχιακή Διπλωματική Εργασία

**Αρβανιτίδης Αθανάσιος Ιωάννης**

**Επιβλέπων:** Μπαργιώτας Δημήτριος

Ιούνιος 2021

Approved by the Examination Committee:

Supervisor **Bargiotas Dimitrios**
Associate Professor, Department of Electrical and Computer Engineering,
University of Thessaly

Member **Daskalopulu Aspassia**
Assistant Professor, Department of Electrical and Computer Engineering,
University of Thessaly

Member **Panapakidis Ioannis**
Assistant Professor, Department of Electrical and Computer Engineering,
University of Thessaly

# DISCLAIMER ON ACADEMIC ETHICS AND INTELLECTUAL PROPERTY RIGHTS

Being fully aware of the implications of copyright laws, I expressly state that this MSc thesis, as well as the electronic files and source codes developed or modified in the course of this thesis, 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 thesis or part of it does not belong to me because it is a product of plagiarism.

The declarant

Arvanitidis Athanasios Ioannis

# Acknowledgements

I would first like to thank my supervisor, Dr. Dimitrios Bargiotas, whose expertise and guidance were invaluable in accomplishing this research. His insightful feedback pushed me to sharpen my thinking and to embrace science. I warmly thank him for his cooperation.

I would also like to thank the examination committee, Dr. Daskalopulu Aspassia and Dr. Panapakidis Ioannis, for their valuable guidance throughout my studies. You provided me with the tools that I needed to choose the right direction and successfully complete my thesis.

Finally I want to extend my sincere gratitude to my parents for their wholehearted support and abundant intellectual resources during my postgraduate studies.

MSc Thesis

**Short-Term Load Forecasting Using Artificial Neural Networks**

**Arvanitidis Athanasios Ioannis**

# Abstract

The growing needs for additional electricity has emphasized the importance of modernizing and optimizing the existing power systems. An optimal and modernized power system will provide more controllable power electronic equipment to allow the most use of current circuits, ensure flexibility and optimal power system efficiency, and make it easier to integrate renewable energy resources at all voltage levels. The current revolution in communication technologies, fuelled primarily by the internet, provides the opportunity for even greater supervision and regulation in the power grid, resulting in more reliable, efficient, and cost-effective services. One of the most critical aspects of efficient power system operation is the ability to predict energy load requirements Load forecasting is one of the most distinctive topics that the science community has extensively considered as it is essential for balancing demand and supply and for deciding electricity prices. In this thesis, a novel data processing strategy is proposed that emphasizes the importance of specific input data by applying neural networks to load forecast. This innovative method is implemented using consumption data from the Greek interconnected power system leading to improved forecasted values.

Μεταπτυχιακή Διπλωματική Εργασία

**Βραχυπρόθεσμη Πρόβλεψη Φορτίου με τη Χρήση Τεχνητών Νευρωνικών Δικτύων**

**Αρβανιτίδης Αθανάσιος Ιωάννης**

# Περίληψη

Σήμερα, η αυξανόμενη ανάγκη για επιπλέον ηλεκτρική ενέργεια έχει τονίσει την ανάγκη για εκσυγχρονισμό και για βελτιστοποίηση των υπαρχόντων συστημάτων ενέργειας. Ένα βέλτιστο και εκσυγχρονισμένο σύστημα ισχύος θα διασφαλίζει την ευελιξία και τη βέλτιστη απόδοση του συστήματος ισχύος και θα διευκολύνει την ενσωμάτωση ανανεώσιμων πηγών ενέργειας σε όλα τα επίπεδα τάσης. Η τρέχουσα επανάσταση στο τομέα των επικοινωνιών, που αφορά κυρίως το Διαδίκτυο, παρέχει την ευκαιρία για ακόμη μεγαλύτερη επίβλεψη και ρύθμιση στο δίκτυο ισχύος, με αποτέλεσμα πιο αξιόπιστες, αποδοτικές και οικονομικά ωφέλιμες υπηρεσίες. Μία από τις πιο κρίσιμες πτυχές της αποτελεσματικής λειτουργίας του συστήματος ισχύος είναι η ικανότητα πρόβλεψης φορτίου. Η πρόβλεψη φορτίου είναι ένα ιδιαίτερο θέμα που η επιστημονική κοινότητα έχει εξετάσει εκτενώς, καθώς είναι απαραίτητο για την εξισορρόπηση της ζήτησης και της προσφοράς και για τη λήψη τιμών ηλεκτρικής ενέργειας. Σε αυτή την εργασία, προτείνεται μια νέα στρατηγική επεξεργασίας δεδομένων που υπογραμμίζει τη σημασία συγκεκριμένων δεδομένων εισόδου κατά την εφαρμογή των νευρωνικών δικτύων στο ζήτημα της πρόβλεψης φορτίου. Αυτή η καινοτόμος μέθοδος εφαρμόζεται σε τρεις μελέτες που βασίζονται σε δεδομένα από το ελληνικό διασυνδεδεμένο σύστημα παράγοντας καλύτερα αποτελέσματα πρόβλεψης.

# Table of contents

# List of figures

# List of tables

# Abbreviations

| | |
|---|---|
| ICTs | Information and Communication Technologies |
| STLF | Short-Term Load Forecasting |
| MTLF | Medium-Term Load Forecasting |
| LTLF | Long-Term Load Forecasting |
| ISOs | Independent System Operators |
| RTOs | Regional Transmission Organizations |
| GNP | Gross National Product |
| GDP | Gross Domestic Product |
| DGs | Distributed Generation |
| RES | Renewable Energy Sources |
| UC | Unit Commitment |
| ANNs | Artificial Neural Networks |
| MLPs | Multilayer Perceptrons |
| RBF | Radial Basis Function |
| MSE | Mean Squared Error |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| DRBFN | Decay Radial Basis Function Networks |
| SVR | Support Vector Regression |
| ELM | Extreme Learning Machine |
| ErrCor | Error Correction algorithm |
| WFNN | Wavelet Fuzzy Neural Networks |
| IMF | Intrinsic Mode Function |
| PSO | Particle Swarm Optimization |
| EKF | Extended Kalman Filter |

| | |
|---|---|
| KELM | Extreme Learning Machine with Kernel |

# Chapter 1

# Introduction

An electric power system, or simply power system, is a network of electrical components deployed to generate, transfer, distribute, supply, and use electric power to consumers or customers. The generation, transmission, and distribution of electricity at the points of demand, i.e. at the points of customer access to the electricity network, is the primary role of an electricity grid.

An energy system must be built and controlled in such a way that it is safe, reliable, environmentally friendly, and supplies high quality electricity at the lowest possible price in order to be practical. Reliability applies not only to the quantitative coverage of consumers' overall needs, but also to the fulfilment of temporal and local fluctuations in load. The term "quality" refers to the observance of known limits for voltage and frequency variations, which are usually 5% and 0.5%, respectively. To be operational, a well-designed and structured electricity grid must fulfil the following minimum requirements:

- The device must be able to respond to changes in demand for active and reactive power on a continuous basis.
- The system should supply power at a low cost while having a low environmental impact.
- The power supplied by the electricity grid must be of high quality and is dependent on voltage stability, network frequency, and system reliability.

The management of the electricity systems and the implementation of different methods of optimizing their operations must be constant and uninterrupted in order to be able to satisfy the above-described requirements. System optimization adds significant economic benefit, with major utilities saving hundreds of millions of dollars each year in fuel costs, increased operating efficiency, and system security. As a result, in terms of fuel cost reduction and environmental protection, optimization has been critical for the service of power grid utilities. Different optimisation problems such as economic transmission, unit engagement, hydrothermal scheduling, optimum power flow, maintenance scheduling etc. were included in power system operation [1].

To solve power system optimization challenges, a variety of strategies have been used, including conventional and artificial intelligence techniques. Due to their non-linearity, optimization problems are complex, and dealing with them using traditional approaches is a time-consuming procedure. Recognizing that the rapid growth of the Internet and computing power presents significant opportunities to modernize the operation of electrical networks has coincided with a realization that the power sector can only be evolved, developed, and de-carbonized at a reasonable cost if it is effectively monitored and controlled. Economic dispatch, unit commitment, hydrothermal scheduling, optimal power flow, maintenance scheduling, and other optimization concerns have all been examined in the operation of power systems. The optimization issue in this thesis is short-term load forecasting using artificial neural networks.

The purpose of the master's thesis is to investigate the effects of various input data preprocessing approaches on the outcomes of short-term load forecasting using neural networks in order to improve the accuracy of the forecasted values. Specifically, Chapter 2 develops the needs that led to the creation of Smart Grids and what problems they are called to solve, as well as emphasizes the importance of the implementation of load forecasting. Chapter 3 presents the theoretical analysis of neural networks, while Chapter 4 is an overview of the application of neural networks in the issue of short-term load forecasting. Finally, Chapter 5 provides a novel data processing strategy that varies from previous work in that it emphasizes the importance of specific input data by applying neural networks to load prediction. While proposing an improved approach to the issue, various case studies for short-term load forecasting in the Greek Power System are reviewed. In comparison to the current literature based on data from the Greek interconnected system, this method produces better prediction results.

# Chapter 2

# Application of Smart Grids in Existing Electricity Systems

## 2.1 Introduction

This chapter presents the features of Smart Grids and the vital importance of load forecasting for the smooth and uninterrupted operation of electrical power systems. The correlation between these two issues is strong and directly affects the modernization of the power grid.

## 2.2 Smart Grids

The power grid developed steadily, in their current known form, since early 1900s in many areas of the world (e.g. the United States of America and most European countries), with the transmission and supply equipment being constructed at that moment going past its construction life and in need of substitution. The capital costs of like-for-like replacement would be extremely high, and it is also debatable if the requisite power equipment manufacturing capability and trained personnel are already sufficient. The need to rehabilitate transmission and supply circuits presents an obvious incentive to experiment with innovative designs and operational procedures. As a result, some of the existing power transmission and distribution lines are nearing capacity, and some renewable energy cannot be attached. This necessitates more intelligent methods of dynamically raising the power transmission capability of circuits and rerouting power flows into less loaded circuits.

Climate experts are unanimous in their belief that man-made greenhouse gases are causing harmful climate change. As a result, methods of using energy more efficiently and producing electricity without or reduced emission of $CO_2$ must be created. Accurate knowledge is required for efficient load management and the elimination of losses and unused energy, while the use of significant volumes of renewable generation necessitates the inclusion of the load in the operation of the power grid to better balance supply and demand. For this reason, there has been a resurgence of interest in linking generation to the distribution network since about 1990.

Overvoltages and repeated variations in power system's frequency values may result from integration from distributed generation. Every disturbance is expressed as a frequency variation from its nominal value (50 or 60 Hz) or abnormal flows in the tie lines connecting the various regions of large power systems. System controllers ensures the system's operation within strict guidelines and wherever is needed respond in order to bring it back within operational limits. As more and more vital loads are connected, modern society needs an extremely dependable energy supply. The conventional solution to increasing redundancy was to install additional redundant circuits, which came at a high capital expense and had a negative environmental effect.

An optimal and modernized power system will provide more controllable power electronic equipment to allow the most use of current circuits, ensure flexibility and optimal power system efficiency, and make it easier to integrate renewable energy resources at all voltage levels. To a large degree, all of these capabilities are supported by genuine and advanced power networks known as Smart Grids (Figure 2.1).

In addition to these specifications, the current revolution in communication technologies, fuelled primarily by the internet, provides the opportunity for even greater supervision and

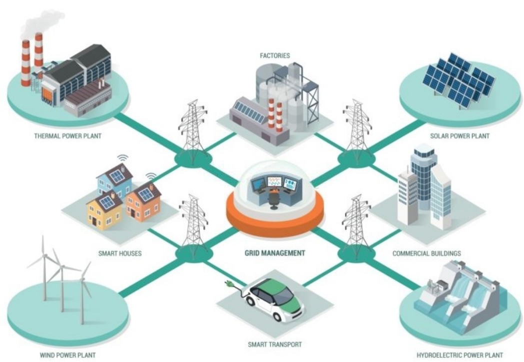


Figure 2.1: Typical interface of a smart grid.

regulation in the power grid, resulting in more reliable, efficient, and cost-effective services. The Smart Grid represents an opportunity to revolutionize the electrical power infrastructure by using modern Information and Communication Technologies (ICTs). The sharing of information in a Smart Grid is highly reliant on two-way contact. Real-time data must be transmitted to and from massive central generators, substations, consumer loads, and distributed generators. Power grid communication services are currently limited to central generation and transmission systems, with some coverage of high voltage distribution networks.

The extension of communication in the distribution system and the establishment of two-way connections with consumers through a smart meter or smart interfacing unit is a critical development of the Smart Grid. Smart meters are an essential component of the Smart Grid and they can provide information about the demands and, as a result, the power flows across the network. As all of the components of the power system are controlled, the system's condition becomes visible, and several control options arise. Obtaining information on customers' loads, on the other hand, may be of concern to unauthorised parties and may infringe on customers' privacy. The opportunity to obtain access to energy usage data and consumer account numbers opens up a plethora of opportunities for theft. Smart Grids necessitates the timely and safe exchange of information [2].

Smart Grids have many benefits over traditional and outdated power networks. Any of the characteristics that highlight their significance are as follows:

- They allow demand response and demand side management through the introduction of smart meters, smart appliances and market loads, micro-generation, and power storage (such as electric vehicles), as well as by providing consumers with energy usage and price information. Customers are expected to be given knowledge and incentives to change their consumption habits in order to alleviate any of the power system's restrictions.

- They accommodate and promote both renewable energy sources, distributed generation, domestic micro-generation, and storage solutions, reducing the overall environmental footprint of the power market and even provide aggregation.

- They optimize and effectively manage properties by intelligent distribution system service.

- They ensure and increase supply stability and protection by being resilient to disruptions, threats, and natural disasters, forecasting and adapting to device fluctuations (predictive maintenance and self-healing), and improving supply security by improved transfer capacities.

- Increased transmission routes, aggregated supply and demand response programs, and ancillary content provisions improve consumer access.

In the future, renewable energy sources, electric vehicles, and heat pumps will be increasingly connected to the distribution network through the Smart Grid. More flexible loads can be required to sustain the grid by accepting growing renewable energy supplies and monitoring demand peaks. The level of supply would be critical for sensitive loads such as computers and high-value processing plants. As a result, visibility, controllability, and stability will be critical aspects of the future power grid, with power electronics playing an important part.

Nevertheless, converting traditional power systems to intelligent power systems is a time-consuming procedure that necessitates extensive research and a thorough knowledge of the issues at hand. To schedule the effective operation and sustainable capital extension of an electric power distribution system, the system owner must be able to predict the need for power sypply - how much power is required, and when and where it will be required. Load forecasting is one of the most distinctive topics that the science community has extensively considered.

## 2.3 Load Forecasting

There is much more demand from electricity suppliers to slash production cost than they have previously had. Modern power supply providers also undergo the same regulatory paradigm, but have many different financial pressures to regulate their planning goals such that they frequently have to invest even more than the conventional paradigm would have warranted.

The idea that the system response closely matches the load specifications is one of the most critical aspects of power system operation. As the system load increases or decreases, the power generation increases or decreases accordingly. This on-demand power generation necessitates having an adequate amount of generation capacity available. As a result, knowing the load parameters in advance allows the electric utility operator to manage grid capacity optimally [3].

The forecasting of electricity loads is critical for the management of power systems. To manage the generation and distribution of electrical power, system operators need reliable load forecasts. Load forecasts are essential for balancing demand and supply, but they also play an important role in deciding electricity prices. The effective modelling methodology is determined by the implementation and the predicted horizon [4].

One of the most critical aspects of efficient power system operation is the ability to predict energy load requirements. The accuracy of forecasts has a significant effect on the economic feasibility and dependability of any electricity utility. Many critical operational decisions, such as power generation scheduling, fuel purchase scheduling, maintenance scheduling, and electricity transaction preparation, are dependent on electric load forecasting. For the reasons mentioned above, we typically divide load forecasting approaches into three categories: short-term load forecasting (STLF), medium-term load forecasting (MTLF), and long-term load forecasting (LTLF). Long-term forecasts are generally necessary for the scheduling of power systems, medium-term predictions are required for maintenance and planning of fuel supply while short-term predictions are required for daily operations of the power system (Figure 2.2). In a deregulated environment, all parties concerned must conduct load forecasting on a regular basis. Load forecasts are used by generation providers, transmission companies, independent system operators (ISOs), and regional transmission organizations (RTOs) to prepare, negotiate, and operate [5].

Therefore, it is necessary to list the parameters that directly affect the issue of Load Forecasting and are used appropriately by researchers in their various prediction models. Some

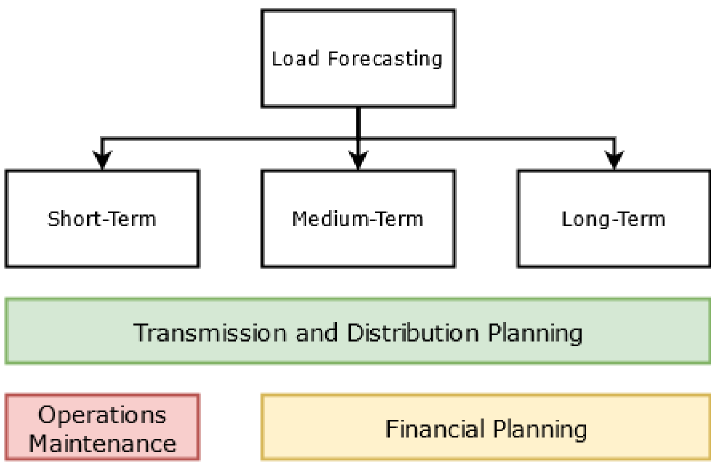


Figure 2.2: Categories into which load forecasting is divided.

of the parameters affecting the forecasted load of future are [6]:

- Time factors such as hours of the day (day or night), day of the week (week day or weekend), time of the year (season or month).
- Weather conditions (temperature and humidity).
- Class of customers (residential, commercial, industrial, agricultural, public, etc).
- Special events (TV programmes, public holidays, etc).
- Population.
- Economic indicators (per capita income, Gross National Product (GNP), Gross Domestic Product (GDP), etc).
- Trends in using new technologies.
- Electricity price.

The capability of power systems to fulfil load specifications instantly and at all occasions is perhaps the most difficult feature of their service. Because of the large load variations during the day, it is important for the device operator to be mindful of the demand that will be anticipated in the coming hours so that adequate preparation can be done. Generators, especially fossil fuel generators, require a significant amount of time to be synchronized to the network if they are initially decommitted for several hours.

The operation of today's power systems is based on short-term electric load forecasting. Previously, experienced system operators may forecast electric load specifications within reasonable ranges (based on their experience with the particular power system). This is not so easy these days. Load size, infrastructure specifications, tighter power efficiency requirements, and deregulation have all necessitated the invention of innovative load forecasting techniques. Short-term load forecasting is so critical that no electric utility can run in a cost-effective, safe, and stable manner without it. In addition, it has a number of key elements that ensure power systems' reliability, security and economic operation; such as the negotiation between utility and regional transmission operators of bilateral agreements, studies including economic dispatch, unit commitment, load flow analysis and safety study and operations such as committing or decommitting generating scheduling.

Electricity must be transmitted from the generating resources to the end users in the centralized power station model. The majority of these generators are located hundreds of miles away from the actual users. This necessitates long-distance transmission of electricity. This centralized power plant system has numerous drawbacks. In addition to the issue of long-distance transmission, this system with large-scale power plants contributes to greenhouse gas emissions, the production of nuclear waste, inefficiencies and power loss over the lengthy transmission lines, environmental impact near the area where the power lines are constructed, and security-related issues. Many of these concerns can be avoided with a distributed generation model. The challenges associated with the building of transmission lines are rendered obsolete by placing the energy source near or at the end user location.

For a variety of reasons, there is increased interest around the world in increasing the percentage of electricity generated by Distributed Generators (DGs), notably renewable energy sources (RESs). Globally, there is a general trend toward steady increase of renewable energy sources, particularly wind and solar resources. Because of their traits of variability, controllability, partial unpredictability, and locational dependency, these resources provide significant challenges for system operation. Therefore, RESs as DG a resource pose the same challenges while the growing proportion of RESs in DG resources may exacerbate the operational problems. As a result, DGs are expected to expand rapidly in the near future. Currently, the emphasis is on implementing rules to define technical standards for DG connections to distribution networks.

With the introduction of DG, many distribution networks are becoming energy harvesting systems, with significantly higher variability and bidirectional power flows where DG penetration is significant. This presents numerous new issues for utilities and necessitates the operation of distribution networks as active distribution systems, with real-time control and optimization of multiple dispersed energy supplies. The issue of load forecasting is intimately related to real-time control and optimization of power systems. This optimization also entails improving the existing network's reliability coefficients.

Distribution system reliability is a critical issue in power engineering for both utilities and customers. The probability of a distribution system providing continuous power without failure for a specified period of time is defined as distribution system reliability. Because of the growing demand for more dependable service with lower interruption frequency and duration, the reliability assessment of future distribution networks is an essential topic. Future distribution systems are likely to be incorporated into the system in order to monitor, control, and operate it. As a result, the reliability of future grids is projected to become a more difficult issue in the near future, as system configurations become more intricate and the penetration of small-scale units increases. Therefore, the urgent application of STLF-related approaches is deemed essential for the modernisation and optimization of existing networks [7].

Furthermore, economic dispatch and unit commitment are two additional categories that necessitate the adoption of load forecasting. The aim of real power economic dispatch is to minimize the generator's fuel consumption or the overall system operating cost by determining the power output of each generating unit under the constraint condition of system load demands.

Since generators cannot be switched on and produce power instantly, unit commitment (UC) must be planned ahead of time to ensure that enough generation is always available to handle system demand with an adequate reserve margin in the event that generators or transmission lines fail or load demand increases. UC manages a power system's unit generation schedule in order to minimize operating costs while meeting prevailing restrictions such as load demand and system reserve requirements throughout a series of time periods. The classical UC problem, which belongs to a class of combinatorial optimization problems [8], is concerned with setting the start-up and shutdown schedules of energy generator units in order to fulfil predicted demand during specific time periods (24 h to 1 week).

The STLF question is one of the most significant problems of optimization, which is why a number of research teams have addressed it. Optimization problems, including nonlinear objective functions and limitations of nonlinear equality and inequalities, are generally nonlinear. Historical charging statistics are used to extrapolate trend techniques to previous charged development trends. Simulation-based load predicting attempts to replicate or to model the load growth process itself to predict where, when and how the load develops, and to describe some of the reasons behind their growth. The prediction's accuracy and speed depend upon the degree of detail used in simulation, the selection of relevant contributing factors (such as social and weather trends) and the level of testing conducted by various methods. Although the concepts used in each forecaster's development should be used mostly for all power systems, the majority of approaches do not work well when the weight of various influencing factors is generalised.

Seasonal input variables, such as load fluctuations induced by air conditioning and heating systems, weather forecast variables, such as temperature, humidity, wind, and cloud cover, and historical data, such as hourly loads for the previous hour, previous date, and same day of the previous week, are the three key types of input variables used for STLF. It should be remembered that distinguishing weekdays from weekends and holidays requires extra care, as the load pattern differs significantly on each form of day. The projected average load for each hour of the day, the daily peak load, and the daily or weekly energy production are standard outputs of short-term forecasts.

Short-term load forecasting has traditionally been performed using approaches such as time series models, regression-based approaches, and Kalman filtering. These approaches are often mixed with the operator's knowledge to arrive at assumptions about effective generation scheduling. Methods based on artificial intelligence have recently become popular for solving optimization problems. These techniques have the advantage of being able to deal with complex challenges that traditional methods cannot overcome. Furthermore, because of their basic mathematical nature, these methods are simple to implement and easy to integrate with other methods to create hybrid solutions that combine the strengths of each individual process.

There are two types of short-term load forecasting techniques: traditional or classical methods and computational intelligence-based techniques. Methods in the first group include time series models, regression models, and Kalman filtering-based techniques. Expert algorithms, artificial neural networks, deep neural networks, fuzzy inference and fuzzy-neural structures, and evolutionary programming are examples of computational intelligence-based techniques.

# Chapter 3

# Introduction to Artificial Neural Networks

## 3.1 Introduction

Artificial Neural Networks (ANNs) are widely use deep learning methods that mimic the learning process inspired by biological organisms. Neural networks were created to emulate the human nervous system for machine learning activities by treating the computational units in a learning model similarly to human neurons, since they have proved to be capable of learning any mathematical function given enough training data. The expanded data availability and processing efficiency of modern computers has exceeded the limitations of conventional machine learning algorithms, which reflects a major part of neural networks' recent performance [9].

In this chapter, we discuss single-layer, multi-layer, and radial basis function neural networks. A series of inputs is directly transferred to an output in a single layer network using a simplified version of a linear function. The perceptron is another name for this basic neural network instantiation. The neurons in multi-layer neural networks are organized in a layered manner, with the input and output layers divided by a set of hidden layers. The neural network's layer-wise architecture is also known as a feed-forward network.

Short-term load forecasting has traditionally been achieved using approaches such as time series equations, regression-based approaches, and Kalman filtering. These approaches are often mixed with the operator's knowledge to arrive at assumptions about correct generation scheduling. Artificial neural network approaches, as well as other computational intelligence techniques, have emerged as potentially useful instruments in electric load forecasting in recent years [10].

The aim of this work is to provide a solution to the problem of short-term load forecasting using neural networks. As a result, the theoretical context and mathematical analysis of the major neural networks used mainly in STLF should be presented thoroughly. The key metrics for measuring their performance are presented in depth at the end of this chapter.

## 3.2 Perceptrons

The perceptron is the simplest neural network and was introduced in 1957 by Frank Rosenblatt. This neural network has one input layer and one output node. Figure 3.1 depicts the perceptron's simple design.

The input layer contains n nodes that transmit the n features $X = [x_1 \dots x_n]$ and each input connection is associated with a weight $W = [w_1 \dots w_n]$. Firstly, a weighted sum $z$ of its inputs is computed at the output node as follows:

$$z = x_1 \cdot w_1 + x_2 \cdot w_2 + \cdots + x_n \cdot w_n = \sum_{i=1}^{n} x_i \cdot w_1 \tag{3.1}$$

Subsequently, the sign of this real value is used in order to predict the dependent variable

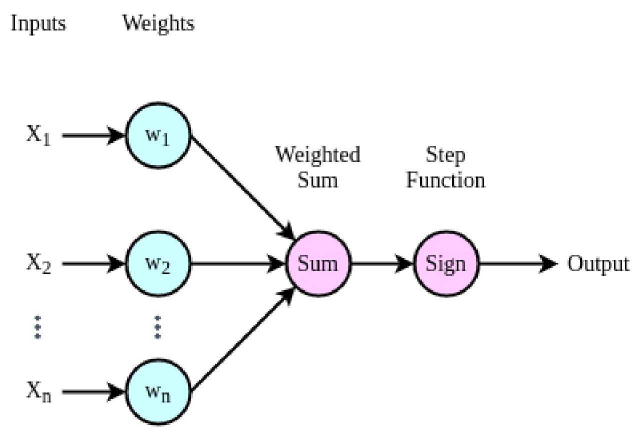


Figure 3.1: The most common form of a perceptron.

of $X$. Therefore, the prediction $y$ is computed as follows:

$$y = sign(z) = sign(x_1 \cdot w_1 + x_2 \cdot w_2 + \cdots + x_n \cdot w_n) = sign(\sum_{i=1}^{n} x_i \cdot w_1) \tag{3.2}$$

where $sign(z)$ is:

$$sign(z) = \begin{cases} -1, & \text{if } z < 0. \\ 0, & \text{if } z = 0. \\ 1, & \text{if } z > 0. \end{cases} \tag{3.3}$$

The sign function maps a real value to either +1 or $-1$, which is appropriate for binary classification. The error of the prediction is therefore:

$$Error(X) = y_{\text{real}} - y_{\text{predicted}} \tag{3.4}$$

When the error value $Error(X)$ is nonzero, the neural network weights must be modified in the negative direction of the error gradient. In this case, training a perceptron entails determining the appropriate values for $w_0$, $w_1$, and $w_2$. Despite the perceptron's similarities to conventional machine learning models, its representation as a computational device is extremely useful because it helps one to combine several units to generate much more efficient models than are possible in traditional machine learning.

## 3.3 Multi-layer Perceptrons

There are several computational layers in multilayer perceptron (MLPs) neural networks. The perceptron has an input and an output layer, with the output layer being the only one that performs computation. The data is transmitted from the input layer to the output layer, and all computations are entirely transparent to the user. MLPs, as opposed to Perceptrons, are made up of the following layers:

- Input layer
- Hidden layer
- Output layer

All of the data that you want the system to learn from is fed into the input layer. Hidden layers attempt to examine various variations of the input layer to determine which of them is relevant and how much weight should be assigned to them. They accomplish this with the assistance of weights. As a result, the hidden layers accept weighted input. When all of the computation has completed, the output layer computes all of the program's outputs and calculates the results (Figure 3.2).

MLPs are suitable for regression tasks. If you want to forecast a single value (e.g., the load of the next day) with many of its characteristics, you only need a single output neuron: its output is the predicted value. For multivariate regression (predicting several values at once), one output neuron is needed for each output dimension.

The principles of forward feeding and backpropagation are used by neural networks to function. Training the network entails the system attempting to discover all possible trends in the data and then learning them. The advantage of training is that when you provide the system a new set of data, it attempts to adapt the previously learned patterns to the new set. If the patterns fit, a determination is taken based on what was done to the training data after this pattern was found.

Each node is computed by performing some operation on the node before it. As a result, the hidden nodes are derived from the input nodes, and the output nodes are derived from

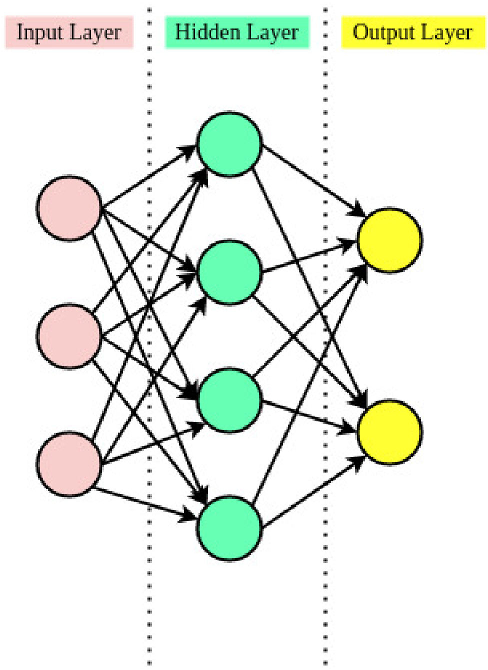


Figure 3.2: Three-layer MLP is widely used as a computational method.

the hidden nodes. Forward propagation refers to the mechanism by which the output node is computed by the operations of previous nodes and the information supplied by the input nodes is transferred directly to the output node. When the output nodes have been generated and their values calculated using the forward propagation method, the gradient must be computed, which necessitates information flowing backward from the output node to the input node. This is known as backpropagation. In neural networks, calculating gradients using the backpropagation technique is critical because it helps reduce the cost function, resulting in much smoother and more reliable predictions.

The forecast output is compared to the training case, and the derivative of the loss function with respect to the output is calculated. This loss' derivative must now be calculated with respect to the weights in all layers in the backwards process. The primary aim of the backward step is to learn the gradient of the loss function with respect to the various weights using differential calculus' chain law. The weights are modified using these gradients. The aim of backpropagation is to find a point at the bottom of the curve where the magnitude of the loss function is as low as possible. As the value decreases, you must determine the values of the variables that contributed to the decrease. These factors are known as weights and prejudice in neural networks. The backpropagation algorithm employs differential calculus' chain law, which computes error gradients as summations of local-gradient products over various paths from a node to the output [11].

Each neuron in the hidden or output layer has a unique activation mechanism. This aids in determining whether the performance of a specific neuron is essential or not. The system's trained weights are multiplied by the input neuron value. The bias attribute is then applied. The activation function, which decides the significance, determines the output value. When an activation function is applied to an output neuron, it takes the output of all prior neurons where the activation function was applied and provides a final response by computing the weighted sum. There are several categories of activation functions, some of which are as Sigmoid, Tanh, Softmax, ReLU and Leaky ReLU.

The curve of a sigmoid activation feature is s-shaped. It has a value spectrum of 0 to 1. Since its higher and lower bounds are 0 and 1, it is most often found in binary classification problems. The following equation gives the sigmoid activation function:

$$Sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{3.5}$$

Tanh is another name for the hyperbolic tangent function. This function ranges from -1 to

+1. This function is used when you want to consider negative outputs. In general, using Tanh in the hidden layers is advised because it causes the quality of various layers to be negative as well. The following equation gives the hyperbolic tangent function:

$$Tanh(x) = \frac{2}{1 + e^{-2x}} \tag{3.6}$$

For dealing with binary classification problems, sigmoid functions are used. However, if you have several classes, you can instead use softmax activation functions. A softmax function's contribution is the probability of each class over all classes. The formula is given by the next equation:

$$Softmax(x) = \frac{e^i}{\sum e^i} \tag{3.7}$$

The lower limit of ReLU activation functions is 0, but there is no upper limit. If the weighted sum is an integer or whole number, the same value will be returned as output. However, if the output is less than zero, it will be translated to zero. It can be represented by the following formula:

$$ReLU(x) = max(0, x) \tag{3.8}$$

Leaky ReLU is the same as ReLU, only that instead of having a lower limit of exactly 0, values will be less than 0 by taking a value and multiplying it by the initial value. It can be represented by the following formula:

$$LeakyReLU(x) = \begin{cases} x, & \text{if } x > 0. \\ a \cdot x, & otherwise. \end{cases} \tag{3.9}$$

Increasing the strength of several layers requires the use of nonlinear activation functions. A multilayer network can be seen to simplify to linear regression if all layers use an identity activation mechanism. A network with a single hidden layer of nonlinear units and a single (linear) output layer has been shown to compute almost any element. As a consequence, neural networks are often referred to as universal function approximators, despite the fact that this theoretical assertion is not always easily translated into functional usefulness. The key problem is that the number of secret units used to accomplish this is very high, which raises the number of parameters to be studied. As a consequence, training the network with a small number of data presents realistic challenges. Deeper networks, in particular, are often chosen because they minimize the number of hidden units in each layer as well as the total number of parameters.

Generally, you don't want to use any stimulation for the output neurons when creating an MLP for regression, so that all values are free to output. You should use the ReLU activation feature in the output layer if you want to ensure that the output will still be optimistic. In conclusion, you can use the logistics function or the hyperbolic tangent if you want to guarantee that the forecasts fall within the range of values given and then scale labels into the required ranges of 0 to 1 of the logistical function and -1 to 1 of the hyperbolic tangents.

## 3.4 Radial Basis Function

Radial basis function (RBF) networks have a radically different architecture from the previous chapters. Both of the preceding chapters make use of a feed-forward network, in which the inputs are transferred forward from layer to layer in a similar manner to provide the final outputs. The nonlinearity in a feed-forward network is usually provided by the repeated composition of activation functions. An RBF network, on the other hand, usually consists of a single input layer, a single hidden layer (with a special form of action specified by RBF functions), and an output layer (Figure 3.3). Although the output layer can be replaced with several feed-forward layers, as in a traditional network, the resulting network is still very shallow, and its behaviour is heavily affected by the existence of the unique hidden layer.

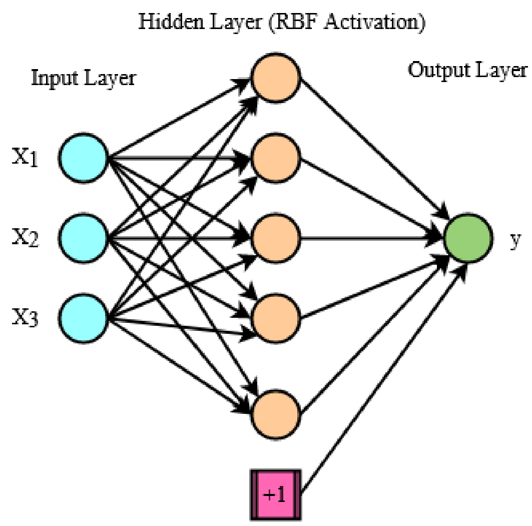


Figure 3.3: A typical form of an RBF neural network.

The nature of the calculations in hidden layers differs greatly from the feed forward networks we have seen to date. The structure and computations performed by the special hidden layer are the secret to the RBF network's strength. The RBF network is a generalization of kernel classification and regression, much as the perceptron is a version of the linear support vector machine. The RBF network's layers are constructed as follows:

- Simply transmit the input layer to the hidden layers from the input functions. The number of input devices is therefore exactly the same as the dimension d of the data. As with feed forward networks, input layers are not calculated. The input units are entirely attached to the hidden units, as with all feed-forward networks, and take their input forward.

- In the hidden layers, computations are dependent on comparisons with prototype vectors. A d-dimensional prototype vector is present in each hidden unit. Let $\mu_i$ represent the prototype vector of the i$^{\text{th}}$ secret unit. Furthermore, the i$^{\text{th}}$ secret unit has a bandwidth denoted by $\sigma_i$. The activation $\Phi_i(X)$ of the i$^{\text{th}}$ hidden unit is then defined for any input training point $X$ as follows:

$$h_i = \Phi_i(X) = exp(-\frac{\|X - \mu_i\|^2}{2 \cdot \sigma_i^2}), \quad i = 1 \ldots m. \tag{3.10}$$

  where, $m$ represents the cumulative number of hidden units. Each of these $m$ units is intended to exert a significant impact on the cluster of points nearest to its prototype vector. As a result, $m$ can be thought of as the number of clusters used for simulation, and it is a significant hyper-parameter open to the algorithm.

- Let $h_i$ be the output of the i$^{\text{th}}$ hidden unit for any given training point $X$, and the weights of the links from the hidden to the output nodes be set to $w_i$. The RBF network's prediction $y$ in the output layer is then described as follows:

$$\hat{y} = \sum_{i=1}^{m} w_i \cdot h_i = \sum_{i=1}^{m} w_i \cdot \Phi_i(X) = \sum_{i=1}^{m} w_i \cdot exp(-\frac{\|X - \mu_i\|^2}{2 \cdot \sigma_i^2}) \tag{3.11}$$

The RBF neural network model has been shown to be a good approximator of any function. The ability of the RBF structure to recognize whether an input is near the training set or in an untrained area of the input space gives it a considerable advantage over the MLP structure. RBF networks can also be conditioned more quickly. Because of its shorter training period, the RBF neural network architecture is found to be a feasible option.

## 3.5 Testing and Validating

The only way to determine how well a model can adopt to new cases is to test it on new cases. A smarter choice is to divide the data into two sets: training and testing. As the names suggest, you prepare the model with the training set and compare it with the test set. The error rate on new cases is known as the generalization error, and you can approximate it by testing your model on the test sample. This attribute indicates how well the model can do in situations it has never seen before.

Despite neural networks' formidable credibility as universal feature approximators, significant challenges remain in training neural networks to provide this level of accuracy. These difficulties are largely due to a number of realistic training issues, the most significant of which are overfitting and underfitting. Overfitting occurs when the model is trained for an excessively long period of time, while underfitting occurs when the model is not trained for an insufficiently long period of time. In other words, overfitting may occur if the model is trained to the point where error begins to increase. Underfitting can occur if you stop training the model when the error is high but can still be minimized.

The data used to create the model is just a subset of all the data available in the universe. The data can be described as incomplete and noisy. As a result, when the model is practiced, it attempts to learn how well it generalizes to new data. In other words, if the model is able to correctly extend the principles it has applied to new data is referred to as generalization. An overfitting problem occurs when the model is applied to the data too well. Often there are so many specifics in the data, as well as a lot of needless material. If the model learns from this highly specific data, especially the details and the extra noise, it could lead to overfitting. This negatively impacts the performance. Therefore, in this scenario, the model performs really well in the training data but not on the unseen data.

It has long been assumed that neural networks are technically capable of approximating any function. However, a lack of data access will lead to low results, which is one of the reasons neural networks have only recently gained popularity. Underfitting occurs where the model is unable to learn from the data and thus cannot perform well on new unseen data. In general, neural networks must be carefully designed to avoid the negative consequences of overfitting and underfitting.

## 3.6 Data preprocessing

There are some critical problems involved with neural network setup, preprocessing and initialization. The aim of data preprocessing is to make raw data more amenable to neural networks. This covers vectorization, normalization, missing-value management, and feature extraction. The importance of feature preprocessing and initialization cannot be overstated. As opposed to other machine learning algorithms, neural networks have greater parameter spaces, which magnifies the effect of preprocessing and initialization in a variety of ways.

The feature processing approaches employed in neural network training are not dissimilar to those employed in other machine learning algorithms. Normalization is commonly accomplished by dividing each function value by its standard deviation. The data is said to be normalized when this form of function scaling is paired with mean-centering. The basic principle is that each function is assumed to come from a regular normal distribution with a zero mean and unit variance. Where the data must be scaled in the range [0, 1], the other form of function normalization is useful. Let $Min_j$ and $Max_j$ be the $j^{th}$ attribute's minimal and maximum values. Then, using min-max normalization, each function value $x^{ij}$ for the $j^{th}$ dimension of the $i^{th}$ point is scaled as follows:

$$y_{\mathrm{ij}} = \frac{x_{\mathrm{ij}} - min_j}{max_j - min_j} \tag{3.12}$$

It is known as Min-Max and commonly used for short term load forecasting in the literature. The normalization of features also guarantees improved results because relative values of features are commonly different in more than one order of magnitude. In such situations, the learning parameter faces the issue of malconditioning, in which the loss function is more susceptible than other parameters.

## 3.7 Neural Network Performance Metrics

As expected, performance measures are used to select the better network model in terms of architecture, training algorithm, training parameters, etc. This section examines the inconsistencies of the three most commonly used performance measures for selecting the best possible network in terms of training parameter-tolerance [12]. For the problem of short-term load forecasting, the metrics most frequently mentioned in the literature and which this thesis deals with are Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute

Percentage Error (MAPE).

In statistics, the MSE of an estimator is defined as the sum of the squares of the errors, or the average squared difference between the expected and actual values. MSE is almost often purely positive due to randomness or that the estimator fails to account for facts that might yield a more reliable result. MSE is calculated with the help of the following mathematical formula:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \tag{3.13}$$

where, $x_i$ is the actual value for the $i^{th}$ observation and $y_i$ is the predicted value.

Furthermore, MAE is a measure of the difference in errors between paired experiments describing the same phenomenon. MAE is calculated as follows:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \tag{3.14}$$

where, $x_i$ is the actual value for the $i^{th}$ observation and $y_i$ is the predicted value and $n$ is the number of samples.

In mathematics, the MAPE is an indicator of a forecasting method's prediction accuracy. It normally communicates accuracy as a formula-defined ratio:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}|\frac{A_i - F_i}{A_i}| \tag{3.15}$$

where, $A_t$ is the actual value, $F_t$ is the forecast value and $n$ is the number of samples.

The literature, as can be seen in subsequent chapters, focuses primarily on the presentation of the findings and the assessment of the respective forecasting models in MAPE. All three measurement criteria will be used to test the model presented in this master's thesis. However, the comparison of the MAPE seen in the numerical results with the other models in the current literature would highlight its efficiency.

# Chapter 4

# Review of Short-Term Load Forecasting Methods Using Artificial Neural Networks

## 4.1 Introduction

In recent decades, a variety of Artificial Intelligence algorithms, Deep Learning and Neuro-Fuzzy methods seem to find application and evolve in the field of electricity systems. Some of their main applications concern the optimal operation and management of the power system, load forecasting and electricity price forecasting.

For instance, Miltiadis Alamaniotis, Dimitrios Bargiotas and Lefteri H. Tsoukalas propose a Gaussian process regression (GPR) and a relevance vector regression (RVR) for approaching load forecasting issue based on historical New England's power system load data [13]. In parallel, Dimitrios Kontogiannis et al. [14], proposes the design of a fuzzy control system that will use environmental data, such as those of the weather, to achieve the minimum energy consumption in buildings. This fuzzy control system relies on the parallel application of XGBoost and decision tree metrics to determine the importance of data's features.

In a similar effort to promote AI methods based on RVRs, Miltiadis Alamaniotis et al., re-use the historical data of New England's power system to forecast the price of electricity the next day [15]. In a later attempt [16], the author proposes a novel hybrid methodology to address the same issue. Initially, it uses RVRs to determine the price of the next day's electricity. It then uses these prediction results in conjunction with a micro-genetic algorithm to enhance the prediction and determine the final value.

Despite the importance of these techniques, this chapter provides an overview of existing STLF-based neural network methods.

## 4.2 Short term load forecasting using Multi-layer Perceptrons

The first attempts to address the issue of hourly load prediction using neural networks appeared in the early 1990s. One of the first approaches to STLF was the use of multi-layered perceptron. In [17], D.C. Park et al., cites theoretical analysis and mathematical background for the application of three-layer perceptrons to load prediction. Using historical load data and temperature data their proposed system managed to produce three different forecast variables (peak load, total daily load and hourly load) with MAPE values less than 3%.

In an effort to develop existing techniques, Kun-Long Ho et al., applied an adaptive learning algorithm to MLPs to enhance the ability to make more accurate predictions [18]. A different approach is presented in the paper [19], where a minimum distance measurement is used to find the correlations of the data used as neuronal inputs. The differences appear in the fact that as input data are used prior days total load and maximum and minimum temperature as well as the predicted maximum and minimum temperature for the forecast day and with an enhanced learning algorithm, better prediction results are noted.

A first attempt for short-term load forecasting based on the data of the Greek power system is presented in [20]. A fully connected three-layer feedforward ANN consisting of 63 input neurons, 24 hidden neurons and 24 output neurons was proposed from Bakirtzis et al. in order to predict the hourly values of the next day's loads. Initially the ANN was trained via back-propagation algorithm using 365 input/output training patterns from the previous year. However, improved performance of the model was observed when the ANN parameters were updated every month and even better results were obtained when the model parameters were updated on a daily basis. This selection of training data set gave satisfactory forecasts for regular days but resulted in high forecast errors for holidays. To eliminate these errors in STLF an improved holiday forecasting ANN was proposed. This model trained on a special - for each holiday - data set reduced the forecast errors of consecutive holidays and days following a holiday.

P. Mandal et al., based on a range of similar days to the predicted day, use a solid method for several hours ahead energy price and load forecasting [21]. They measured the correlation coefficient between the data and then integrated them into a three-layer MLP, which was trained with the backpropagation algorithm, using historical half-hourly data for the Victoria electricity sector. This study, without taking into account variables such as weather and special days, offered a more accurate approach to the STLF issue compared to the previous simple methods.

In a more sophisticated approach, N. Kandil et al, used a simple MLP for short term load forecasting [22]. In this paper, they focused on increasing the important and highly correlated data used as neural network inputs. Thus, the authors used as variable inputs an hour indicator, a day indicator, the estimated temperature at hour k, k-1 and k-2. Special emphasis is given to the fact that historical loads are not used as inputs.

Another work related to STLF for the Greek Intercontinental Power System is presented in [23]. In their work, G. J. Tsekouras et al., compared various neural network training algorithms to predict hourly load demand by measuring the MAPE of each separately. The proposed neural network creation does not differ significantly in the structure and input data. The differentiation of researchers is in the normalization of input data where they follow a function of their own to pre-process the data.

In addition, in another attempt to approach the STLF issue Alamaniotis and Tsoukalas [24] presented a data-driven method for minutely active power forecasting based on Gaussian processes, highlighting therefore the importance of minute predictions while Kontogiannis et al. [25] presented a baseline performance comparison of neural network models for minutely active power forecasts derived from residential data.

## 4.3 Short term load forecasting using Radial Basis Functions

Another category of neural networks that has occupied the research community in recent decades and is directly applicable to load prediction is that of Radial Basis Function Networks.

An initial analysis of the application of RBFNs (Radial Basis Function Networks) to short-term load forecasting is presented in [26]. First, the mathematical background of their operation is analyzed in parallel with those of ANFIS and the necessary comparisons are made. Then a modified ANFIS with the possibility for load adjustment and a different mathematical formulation to load prediction is proposed. The proposed model is applied to STLF issue with the Load data from January to August 2004 of an Australian region giving satisfactory results compared to the referenced literature.

In [27], Zbigniew Gontar et al, emphasizing the special importance and usefulness of RBFNs, recorded the results obtained using as input the data of the loads of Crete. In order to be able to identify the seasonality of the data, they proposed the creation of four neural networks, one for each time of year. A fifth neural network was used to predict the loads on the special days and the weekend, as well. The data used as neuronal inputs were historical load data, mainly the data of previous hours and the previous day, the time of day, the initial network forecast, the maximum and minimum value of the day temperature, the cumulative density function and the variances between predicted value and previous actual load value. In conclusion, the authors observed that with the use of RBFNs the next hour load forecast predicted a significant improvement over conventional MLPs, as MAPE recorded a value of 2.01%.

A comparative study using RBFNs is described in [28]. The authors, trying to achieve better generalization of data, faster execution time and less error in prediction, consider the application of various algorithms for the training of neural networks. To evaluate the different learning techniques, they compare the prediction results obtained from RBFNs during their training with Decay Radial Basis Function Networks (DRBFN), Support Vector Regression (SVR), Extreme Learning Machine (ELM), Improved Second Order algorithm (ISO) and Error Correction algorithm respectively (ErrCor). The dataset introduced consists of New England hourly loads, hourly temperature data for the period 2004-2011, an appropriate feature that returns the time of day, the day of the week and a variation on whether the day in question is special day or not. The result extracted from the neural network is the hourly load prediction for the next day. The numerical results showed that the ErrCorr modified algorithm showed the best prediction since the value of MAPE says a value less than 2 percent.

## 4.4 Short term load forecasting using Hybrid Models

A first approach to the issue of STLF with the application of hybrid models is presented in [29]. The researchers suggest using wavelet fuzzy neural networks (WFNN) and modified fuzzy neural networks (FNCI) to predict the next hour's load. The researchers used load data from the Northern Region Load Dispatch Center in Delhi, India, as well as temperature, wind speed and humidity data to evaluate their proposed hybrid models. After introducing these data into various models and comparing the results, they concluded that the technique they proposed yielded better prediction results than the traditional ANFIS model, which has been used extensively in the literature.

In [30], Ioannis P. Panapakidis proposes a robust hybrid model for forecasting day-ahead and hour-ahead load predictions by using hourly load values of 10 buses of the Greek Power System located in the area of Thessaloniki, North Greece. The hybrid model described is based on the combination of historical load and temperature data clustering and embedding in an MLP neural network. Specifically, the author recommends using the minCEntropy clustering algorithm on the training set in order to formulate k clusters. A different ANN is used for each subset. As a result, the data from the corresponding clusters is used to train k ANNs respectively. The Euclidean distance is used to relate each pattern in the test set to k centroids, and the results are fed into the corresponding ANN.

Following the pattern of previous researchers, an innovative approach to load prediction comes from [31]. Xishuang Dong et al., Proposed the implementation of a convolutional neural network (CNN) enhanced with K-means clustering in order to achieve higher scalability in the data and thus reduce the error rate in the forecast. As a first step, the data is cleaned of noise, so that they show greater self-correlation with each other. The K-means clustering algorithm is then applied to the denoised data. The now grouped data is entered into a CNN for hourly load forecasting. The researchers measured the MAPE of the proposed hybrid model and concluded that it offers lower error values compared to other hybrid techniques.

Another hybrid load prediction system is described in detail in [32]. In their work, the authors emphasize the importance of preprocessing load data and propose an improved neural network learning algorithm. The data entered in the modification model refer to historical load data per hour and exogenous data that directly affect the load behaviour, such as temperature and humidity. The Min - Max normalization is now applied to them so that their values range between 0 and 1. Then, the data that show greater autocorrelation are entered into an MLP neural network, which is trained with a modified harmony search (MHS) algorithm. This proposed model is evaluated based on data from the Pennsylvania - New Jersey - Maryland (PJM) power system. The researchers conclude that compared to the other models analyzed in the literature, their models show less MAPE values thanks to both the preprocessing of the data and the modified MHS learning algorithm.

Remaining in the category of hybrid models, L. Ekonomou et al., proposed a load forecasting model based on the combination of MLP neural networks and wavelet analysis [33]. The researchers manipulated historical load data from the Bulgarian power system grid as time series and applied a wavelet denoising algorithm to remove their noise and split them into signals with different frequencies. Subsequently, they inserted these denoised data into an improved ANN significantly increasing the success rate of the prediction. Despite the detailed work of the researchers, the model does not take into consideration the crucial for the STLF issue holiday factor.

In [34], K-shape is proposed as a new clustering technique for categorizing consumers based on their load consumption behaviour. At the same time, various ways for network-level load forecasting are proposed and the characteristics used as inputs of the neural network are pointed out. More specifically, the inputs include historical load data of the previous and the same day, the corresponding temperature data, an innovative approach to data cyclicity and a discriminating between working days and non-working days. The researchers applied the proposed algorithm to a series of Deep-Learning technical forecasts for the next day's load, concluding that the lowest value of MAPE is 2.15%.

Another hybrid approach is described in the paper [35]. Researchers use various machine learning algorithms to optimize the data they use in short-term load forecasting. Initially, they use the load data as time series data and decompose it based on the Intrinsic Mode Function (IMF) technique. Then, with the help of the Particle Swarm Optimization (PSO) algorithm, they are filtered and used by the Extended Kalman Filter (EKF), Extreme Learning Machine with Kernel (KELM) for STLF. The authors conclude that this approach yields acceptable forecasting accuracy and time performance. In a similar addition, the author of [36] uses day or week ahead load data for creating clusters which he feeds to an ANN. The error resulting from comparing this data with the actual ones is fed to a WNN where the final prediction and the various error metrics are calculated. This process is repeated for various machine learning approaches. The results of these techniques are compared with the MAPE and the nRMSE where the approach with the least error is chosen as preferable.

The above papers focused on the different forms of neural networks that can be used for short-term load prediction. The bibliography is very extensive and is constantly increasing in size as new deep learning methods approach the subject of STLF. The effects of the various forms of data normalization, the various activation functions, and the various morphologies of neural networks are presented in [37, 38, 39].

Following a thorough review of the literature, it was find out that a complete inestigation of the various scaling strategies of the input data to the neural networks and the influence they have on the prediction outcomes could be analyzed and improved.

# Chapter 5

# Proposed Approach and Implementation for Short-Term Load Forecasting Methods

## 5.1 Proposed approach for STLF

Following a thorough review of the literature, an innovative data processing technique is suggested, which differs from earlier work, highlighting the values of specific input data. The two primary preprocessing strategies suggested concentrate on the gravity of particular neural network input variables in relation to output variables, resulting in superior prediction outcomes than traditional methods in the existing literature. As a result, numerical findings highlight the significance of this work.

As mentioned in previous chapters, the aim of this master's thesis is to examine the impacts of multiple input data preprocessing approaches on the results of short-term load forecasting using neural networks, and to present two new improved scaling strategies that enhance forecast results. The cases that were studied for short-term load forecasting, as well as the structure of the neural networks used, are presented in detail in this chapter. The cases under consideration are the next day load forecast using the average value of daily loads, the hourly load prediction using historical data from previous days, and the next day's load forecast using historical data from previous days and previous hour. The latter case is what sets the current postgraduate thesis apart from others by proposing a more efficient approach of normalizing the input data

## 5.2 Implementation for short-term load forecasting

The data used in the above studies came from the Greek power system for the years 2013, 2014, 2015, 2016, and 2017. To make a more accurate forecast, weather data, such as temperature, must be used in addition to the historical data of the loads. As stated in previously, data is separated into training and test sets in a ratio of 80% - 20% of the total data. As a result, the training set is made up of data from 2013, 2014, 2015, and 2016, while the predictions are for 2017 (test set). Figure 5.1 illustrates the separation of data into training and test sets. In this master's thesis, neural networks were built using the Python programming language and specifically the library scikit-learn.

### 5.2.1 Case A: Forecast of the average daily load of the following day

As a first case, different techniques of normalizing the input data for the average daily load forecast using MLP neural networks are considered. The neural network used aims to predict the average value of the next day's load and is shown in Figure 5.2. The data in the input layer is as follows:

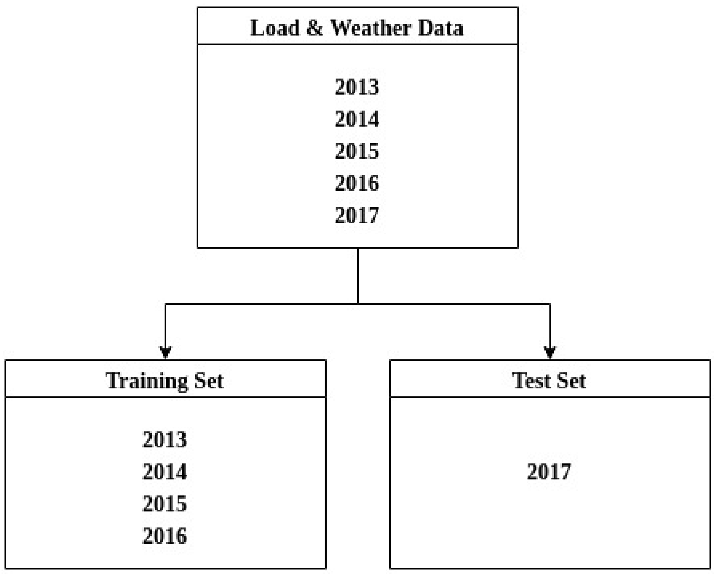


Figure 5.1: Data is divided into two sets: training and test. The data from the previous four years is used as a training set, while the data from the most recent year is used as a test set.

- Date: The day of the month as an integer.

- Week Day: It is a characteristic coding to determine the day of the week. The coding is achieved by using integers from 1-7, where 1 symbolizes Sunday, 2 symbolizes Monday, etc.

- Holiday: Binary coding of the days that are considered holidays for the Greek state (special days and holidays). Holidays and weekends are denoted by 1, while other days are denoted by 0.

- Temperature: The average value of the day temperature for which the forecast is made.

- D-1 Load: The average value of the previous day's load from the forecast.

The data is initially entered into the neural network without being normalized in either way. The results of this method are contrasted to the true average values of the loads for the particular day. Figure 5.3 shows a schematic representation of real load values and predicted values from the neural network, as well as a comparison of the two. The MSE, MAE, and MAPE results are then computed and reported in Table 5.1.

The effect of a simple scaling of the input data on the neural network's prediction outcome is then investigated. By dividing each input variable by the maximum value of the corresponding data set, simple scaling is achieved. This method is only used to get values in-

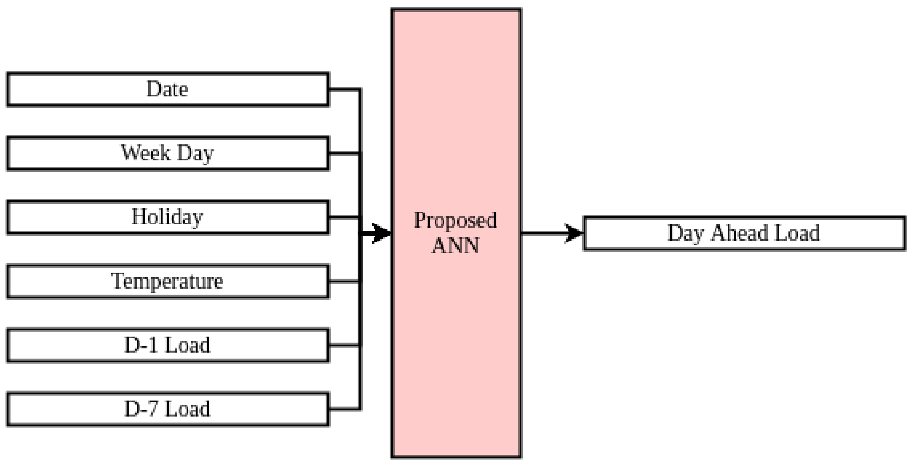


Figure 5.2: The structure of the proposed three-layer MLP for the average daily load forecast.

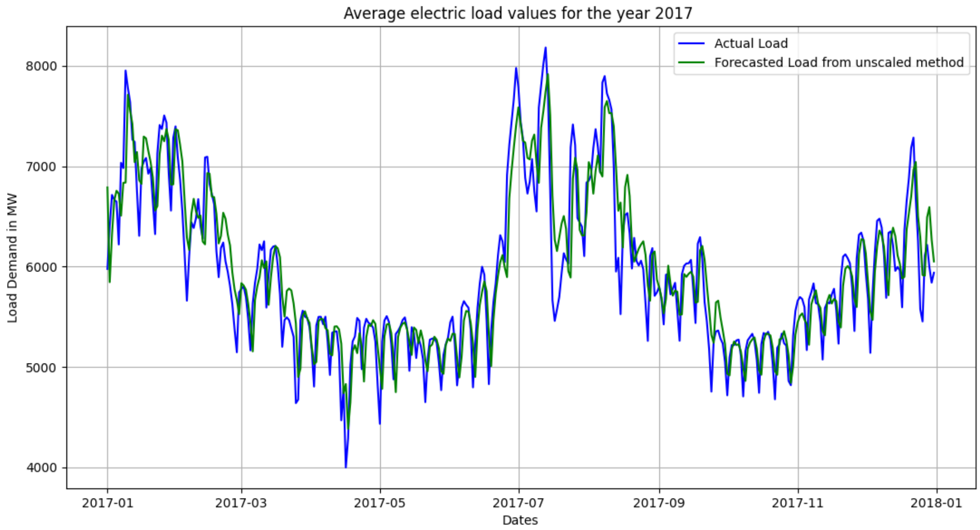


Figure 5.3: Comparison of the real and the predicted values of the electric load for the year 2017 without any kind of normalization.

Table 5.1: MSE, MAE and MAPE values (based on MW) for the average daily load forecast. Input data is not subject to any scaling technique.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 109120.64875 | 250.32491 | 4.24 % |

side the area [0,1] for the variables $Temperature$, $D-1Load$, and $D-7Load$. The following mathematical relation gives the simple scaling for temperature:

$$Temp_{\text{scaled}} = \frac{Temp_i}{Temp_{\text{max}}} \tag{5.1}$$

where, $Temp_i$ is the average temperature of day i and $Temp_{\text{max}}$ is the maximum temperature value of the data set. Exactly the same relationship applies to $D - 1Load$ and $D - 7Load$ data.

The now-normalized data is fed into the neural network in order to estimate the average value of the next day's load, much as in the previous example. The corresponding results are compared to the real load figures for 2017. The predicted outcomes are graphically depicted in Figure 5.4, and the MSE, MAE, and MAPE equations are listed in Table 5.2. As can be shown, relative to basic data entry without preprocessing, this form of preprocessing of temperature and historical load data yields a higher MAPE.

In most papers of the literature review the scaling of the input data is considered necessary

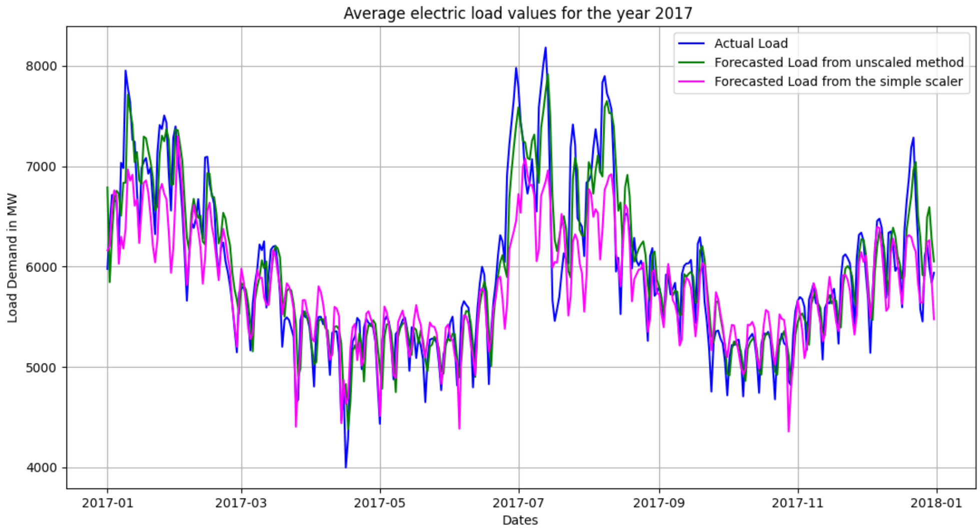


Figure 5.4: Comparison of the real and the predicted values of the electric load for the year 2017. Now the values predicted through the simple scaling method are added.

Table 5.2: MSE, MAE and MAPE values (based on MW) for the average daily load forecast after a simple scaling of the input data.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 109120.64875 | 250.32491 | 4.24 % |
| Simple Scaling | 174437.87768 | 289.10109 | 4.60 % |

to achieve better predictions and reduce the error rate. For this reason, after an extensive study of the correlation of the data but also of the way in which the neural network manages them, this work proposes an innovative scaling method that will give the appropriate weight to the parameters $D-1Load$ and $D-7Load$ greatly improving the forecast results. We investigate whether or not the scaling techniques are affecting the accuracy of the predicted values. Based on the Greek power system data, it turns out that a weighting factor of 10 on the load data of existing scaling methods could greatly improve the accuracy of the forecasted values. For different type of data possibly a new investigation could lead to different weighting factors.

In the first stage, temperature data is scaled the same as it was in the previous scaling process. Therefore, the temperature input data are transformed according to the equation above. After extensive study and considerable experimentation, it was observed that the $D-1Load$ and $D-7Load$ data determine to a greater extent the effect of the neural network. Therefore, it is considered reasonable and necessary to give them due consideration. As a result, the following mathematical relationship was found to be a better scaling technique for these data:

$$D - 1Load_{\text{scaled}} = \frac{D - 1Load_i}{D - 1Load_{\text{max}}} \cdot 10 \tag{5.2}$$

where, $D - 1Load_i$ is the average value of the i$^{\text{th}}$ day's load for which the forecast is made and the maximum value of the total $D - 1Load$.

The data is then scaled and entered into the network. The neural network's prediction outcomes are related to the actual values. Figure 5.5 depicts this contrast. In Table 5.3, the MSE, MAE, and MAPE metrics are estimated and presented. As compared to the previous two techniques, the value of MAPE seems to drop significantly, highlighting the fact that the proposed scaling approach performs better in estimating the average value of the next day's load.

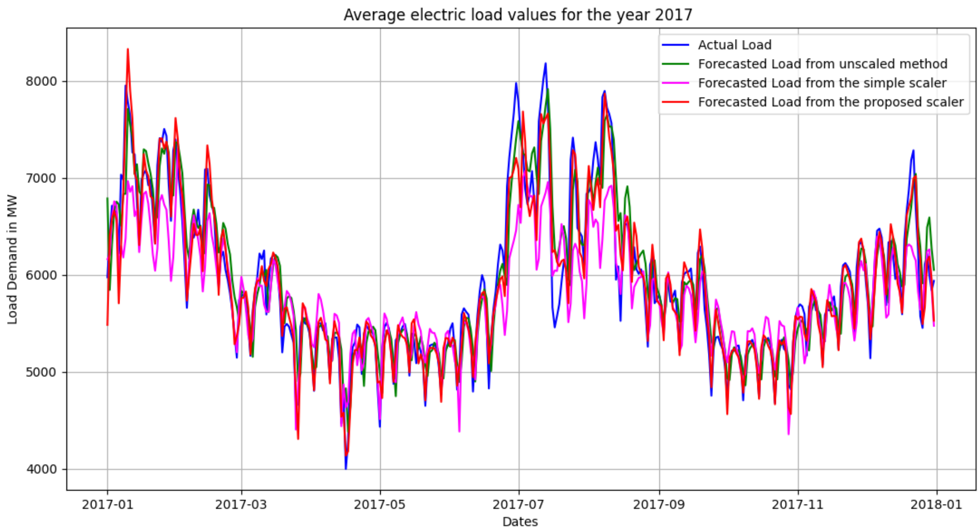


Figure 5.5: Comparison of the real and the predicted values of the electric load for the year 2017. Now the values predicted through the proposed scaling method are added.

In respect of the other papers of the literature review, the min-max scaling technique is applied to the input data separately. Therefore, a separate scaling is applied to the temperature data from the load data. This action is desirable as it better attributes the weight of the data to the neural network and separates the relationship between them. As mentioned in a previous chapter, the mathematical formula for this scaling technique is as follows:

$$y = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{5.3}$$

Table 5.3: MSE, MAE and MAPE values (based on MW) for the average daily load forecast after an enhanced scaling of the input data.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 109120.64875 | 250.32491 | 4.24 % |
| Simple Scaling | 174437.87768 | 289.10109 | 4.60 % |
| Enhanced Scaling (Proposed) | 48804.36457 | 159.42596 | 2.62 % |

where, $y$ is the new scaled value, $x$ is the initial value, $x_{\text{min}}$ and $x_{\text{max}}$ are the minimum and maximum values of the set respectively. The data is fed into the neural network after this preprocessing technique. The outcomes of the MLP neural network's mean value of daily load are compared to actual values for the year 2017. Figure 5.6 depicts the contrast, while Table 5.4 summarizes the MSE, MAE, and MAPE estimates from this prediction.

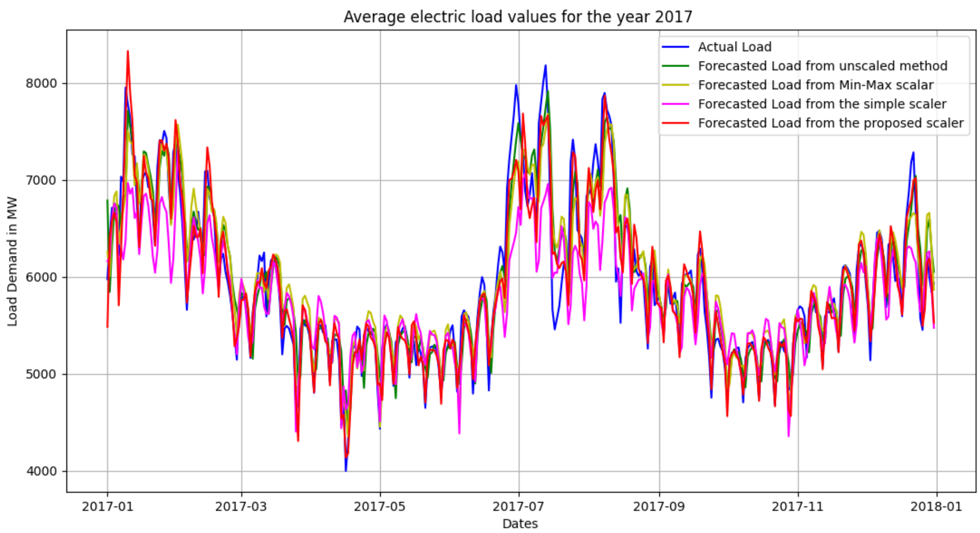


Figure 5.6: Comparison of the real and the predicted values of the electric load for the year 2017. Now the values predicted through the min-max scaling method are added.

As shown in Table 5.4, the load forecast from the proposed neural network using the Min - Max Scaling Method for the preprocessing of the input data shows a smaller MAPE compared to the simple scaling technique and the case where the input data without preprocessing. However, an important observation is the fact that our proposed enhanced scaling method has much better prediction results than the min-max scaling technique widely used in regression problems.

Table 5.4: MSE, MAE and MAPE values (based on MW) for the average daily load forecast after a min-max scaling of the input data.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 109120.64875 | 250.32491 | 4.24 % |
| Simple Scaling | 174437.87768 | 289.10109 | 4.60 % |
| Enhanced Scaling (Proposed) | 48804.36457 | 159.42596 | 2.62 % |
| Min – Max Scaling | 104367.85298 | 245.74374 | 4.11 % |

Despite the fact that min-max scaling does not seem to do as well as our proposed approach, the case of an improved Min-Max Scaling process should be considered, stressing the value and significance in the outcome of predicting the data that involve the inputs $D-1Load$, $D-7Load$. Initially, the temperature data undergoes simple min-max scaling, while the historical load data undergoes the following scaling:

$$y = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \cdot 10 \tag{5.4}$$

The data is then entered into the input layer of the MLP neural network. The results have values between [0, 10]. They reflect a scaled average value of the next day's load. Therefore, in order to be able to compare with the actual values of the daily load, they will have to return to the unit of measurement, in order to obtain their real value. The following mathematical formula is used to obtain the values in units:

$$x = \frac{y \cdot (x_{\text{max}} - x_{\text{min}})}{10} + x_{\text{min}} \tag{5.5}$$

Figure 5.7 illustrates the load behaviour of all scaling methods tested. Table 5.5 summarizes the results of the MSE, MAE and MAPE metrics calculated from this forecast as a measure of comparison of all scaling methods for the average n-day load forecast for the next day.

From Table 5.5 we conclude that our proposed method gives the best forecast of average daily load, as it displays the lowest MAPE. However, the enhanced min-max scaling technique using the result of our study on the gravity of historic load data dramatically reduces MAPE compared to the corresponding conventional technique widely used in the literature.

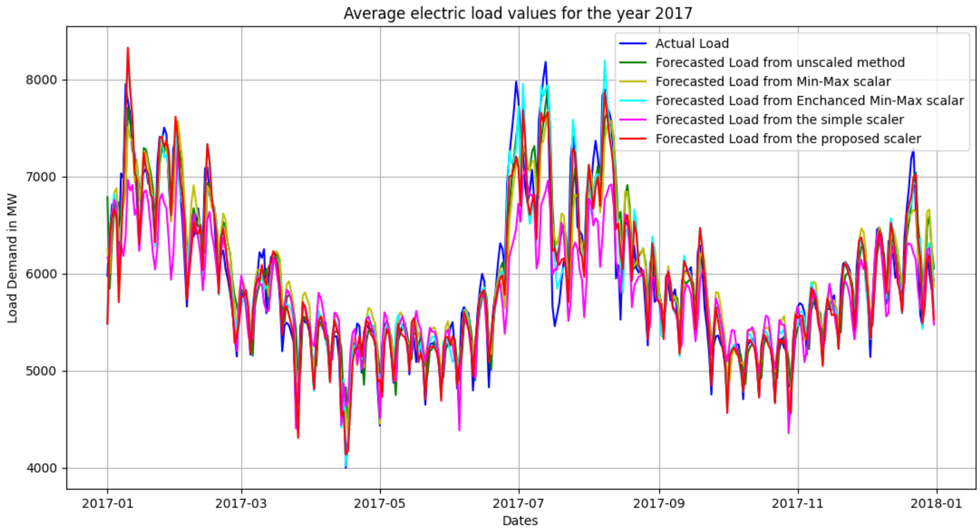


Figure 5.7: Graphical representation of the results obtained from the various methods of pre-processing the input data.

Table 5.5: MSE, MAE and MAPE values (based on MW) for the average daily load forecast.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 109120.64875 | 250.32491 | 4.24 % |
| Simple Scaling | 174437.87768 | 289.10109 | 4.60 % |
| Enhanced Scaling (Proposed) | 48804.36457 | 159.42596 | 2.62 % |
| Min – Max Scaling | 104367.85298 | 245.74374 | 4.11 % |
| Enhanced Min-Max Scaling (Proposed) | 46757.39119 | 162.70983 | 2.71 % |

### 5.2.2 Case B: Hourly load forecasting

Case B is based on the hourly load data of the Greek power system for the years 2013, 2014, 2015, 2016 in order to predict the hourly load demand of the next day for the year 2017. For this reason, a three-layer MLP neural network (Figure 5.8) is constructed as they also report a plethora of literature studies presented in Chapter 3. More specifically, the following variables are introduced at the input level:

- Hour: The time of day for which the load forecast will be made. The time is entered as an integer taking variables from 0 - 23.
- Week Day: As in case A, it is a characteristic coding to determine the day of the week.

The coding is achieved by the use of integers from 1-7, where 1 symbolizes Sunday, 2 symbolizes Monday, etc.

- Holiday: Binary coding to characterize a day as a holiday or a working day. The holidays of the Greek state, i.e., the days concerning national anniversaries and major religious holidays, as well as the weekends are marked with the number 1. On the contrary, the other days are coded with the number 0.
- Temperature: The hourly value of the temperature of the day for which the load is forecast.
- D-1 Load: The value of the previous day's load from that predicted at the corresponding time. For example, if the MLP neural network predicts Wednesday at 17:00 then the price of Tuesday load at 17:00 is entered as entry to the neural network.
- D-1 Load: The value of the load corresponding to the same day of the previous week at the corresponding time. For example, if the output of the neural network is the load forecast for Wednesday 12/05 at 17:00, then until then, as the input to the neural network, the value of the load of Wednesday 5/05 at 17:00 is entered.

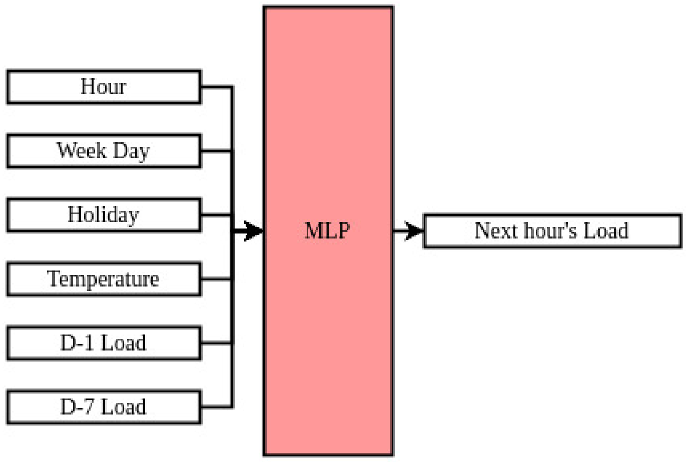


Figure 5.8: Typical structure of the proposed three-layer MLP for hourly load forecasting. Particular attention is paid to input variables.

The coding of the neuronal input data described above differentiates the present master's thesis from the coding of papers [24], [28] as it has been observed, following experiments, that it describes the seasonality of the charge more efficiently. The season cycle of papers [24], [28] was experimentally tested as neuronal input yielding higher MAPE values.

As in the previous case, the effect on the prediction result of the hourly value of the load of some scaling methods on the input data is examined by calculating the metric MSE, MAE and MAPE. Initially, the input data is not subject to any kind of configuration and is entered as is in the three-layer perceptron. The results produced through this process are compared with the actual hourly values of the loads of the respective day for the whole of 2017. Figure 5.9 graphically shows the real and forecasted values, while Table 5.6 summarizes the metric calculations. The input data is then pre-processed separately using a simple scaling method.

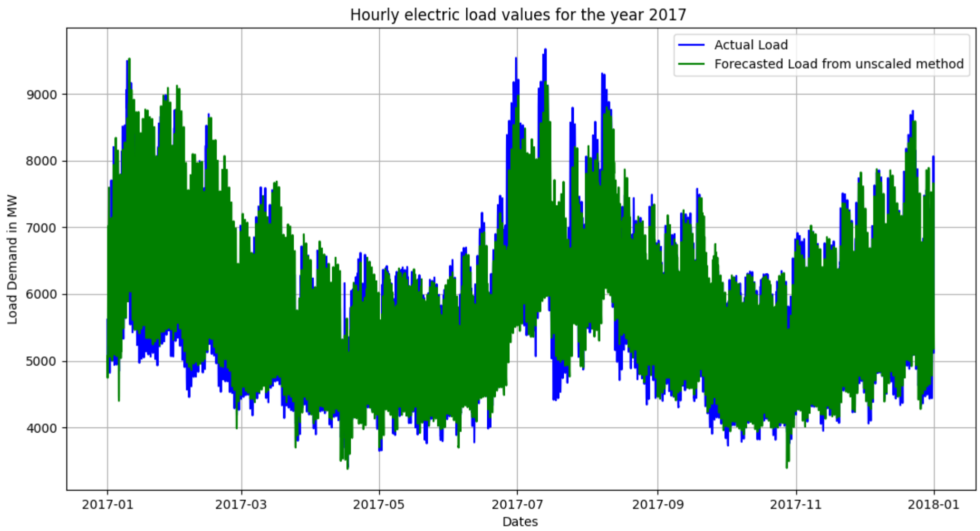


Figure 5.9: Graphic comparison of actual and predicted hourly load values. In the first case, the input data is entered as is.

Table 5.6: MSE, MAE and MAPE values (based on MW) for the hourly load forecast. Input data is not subjected to any scaling technique.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 111390.38535 | 239.36012 | 3.99 % |

Simple scaling is performed by dividing each input variable by the maximum value of the corresponding data set. This procedure is performed only for the variables $Temperature$, $D-1Load$, $D-7Load$ in order to get values within the field [0,1]. Equation 5.1 gives the simple scaling for Temperature.The corresponding relationship also applies to $D-1Load$ and $D-7Load$ data.

The scaled data is then entered into the neuron in order to predict the hourly value of the load. The values extracted from the proposed neural network are compared with the real hourly values of 2017. Its graphic representation is shown in Figure 5.10. Table 5.7 summarizes the results of the MSE, MAE and MAPE metrics calculated from this prediction with the specific input data preprocessing technique.

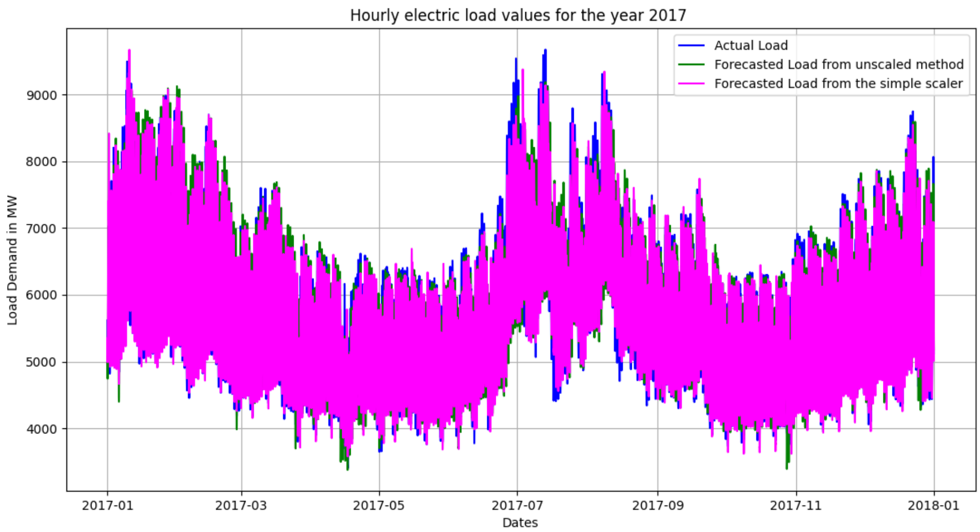


Figure 5.10: Graphic comparison of actual and predicted hourly load values. The values predicted through the simple scaling method are added.

Table 5.7: MSE, MAE and MAPE values (based on MW) for the average daily load forecast. Input data is subjected to a simple scaling technique.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 111390.38535 | 239.36012 | 3.99 % |
| Simple Scaling | 71917.69612 | 189.87314 | 3.17 % |

From the above results it is understood that a proper preprocessing of input data can yield better results in forecasting compared to unscaled data entry. For this reason, after an extensive study of the correlation of the data but also of the way in which the neural network manages them from this work, an innovative scaling method is proposed that will give the appropriate weight to the parameters $D-1Load$ and $D-7Load$ greatly improving the results of the forecast.

In the first step the temperature data undergoes the previous preprocessing described in the above equation. After several experiments in the data, it was observed that the variables $D-1Load$ and$D-7Load$ are the ones that greatly influence the result of the prediction of the neural network. Therefore, more weight should be given to these variables in order to obtain a more accurate forecast. In this work, we investigate whether or not the scaling techniques are affecting the accuracy of the predicted values. Based on the Greek power system data, it turns out that a weighting factor of 10 on the load data of existing scaling methods could greatly improve the accuracy of the forecasted values. For different type of data possibly a new investigation could lead to different weighting factors. The proposed scaling technique is described by Equation 5.2.

The scaled data is entered into the neural network. The resulting results are shown graphically in Figure 5.11 in correspondence with the real hourly values of the year 2017. Table 5.8 compares the values of MSE, MAE and MAPE calculated from this forecast. MAPE values fail because our proposed data preprocessing technique significantly improves forecasting.

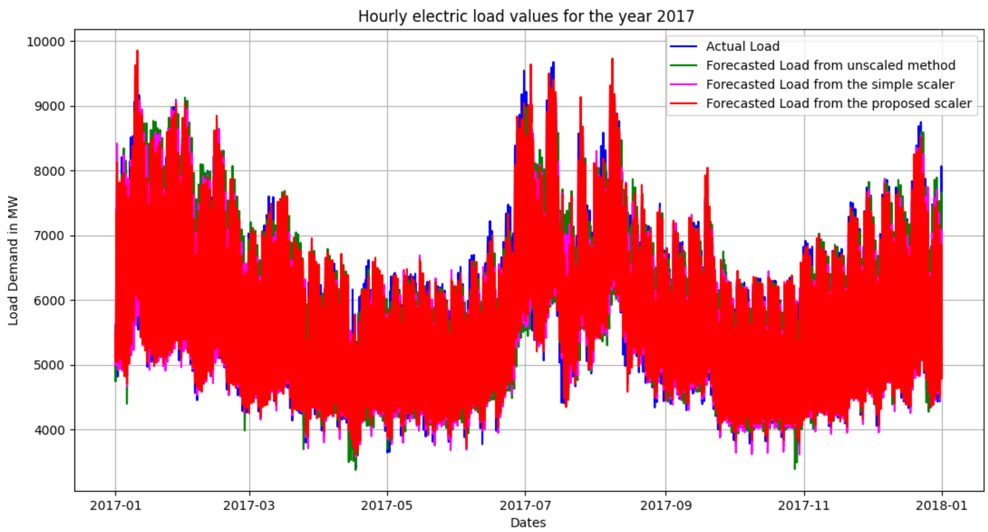


Figure 5.11: Graphic comparison of actual and predicted hourly load values. The values predicted through the enhanced scaling method are added.

A common method of scaling neural network input data related to regression problems is the Min - Max method. Due to the extensive references of this method in the literature, its impact on the issue of this thesis should also be considered.

Table 5.8: MSE, MAE and MAPE values (based on MW) for the hourly load forecast. Input data is subjected to an enhanced scaling technique.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 111390.38535 | 239.36012 | 3.99 % |
| Simple Scaling | 71917.69612 | 189.87314 | 3.17 % |
| Enhanced Scaling (Proposed) | 62238.40502 | 174.52799 | 2.92 % |

In the first stage, the temperature data and the historical load data are subjected to a separate Min-Max scaling through the Equation 5.3. In the next step, the scaled data are now entered in the input layer for the forecast of the hourly load values of the year 2017. The values of the forecast results with the help of the proposed MLP neural network are compared graphically with the real values, as shown in the Figure 5.12. Table 5.9 summarizes the calculated MSE, MAE and MAPE metric values.

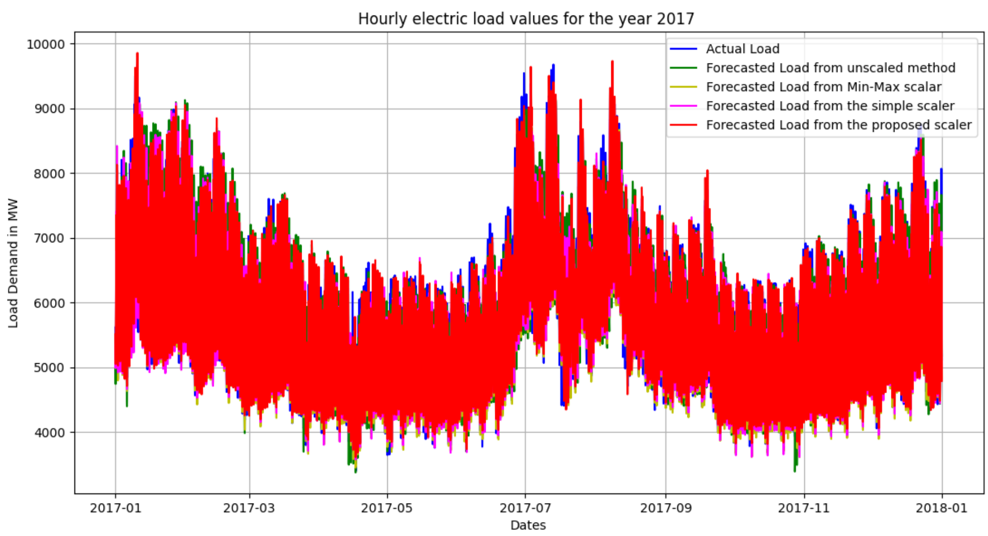


Figure 5.12: Graphic comparison of actual and predicted hourly load values. The values predicted through the min-max scaling method are added.

Particular attention should be paid to the results in Table 5.9. Initially, it became apparent that the Min-Max scaling method only improved the prediction compared to the case where data is entered into the neural network without any preprocessing confirming what is stated in the neural network theory. It also brings higher MAPE values than our proposed scaling technique. Therefore, the conclusion that emerges is that our proposed method highlights the

Table 5.9: MSE, MAE and MAPE values (based on MW) for the hourly load forecast. Input data is subject to a min-max scaling technique.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 111390.38535 | 239.36012 | 3.99 % |
| Simple Scaling | 71917.69612 | 189.87314 | 3.17 % |
| Enhanced Scaling (Proposed) | 62238.40502 | 174.52799 | 2.92 % |
| Min – Max Scaling | 75953.72593 | 204.73686 | 3.40 % |

direct correlation of historical load data in the input layer with the output of the MLP neural network and better approaches the issue of STLF.

Although Min-Max scaling does not seem to perform well compared to the other two data preprocessing methods, the case of an enhanced Min-Max Scaling method that emphasizes the importance of input variables such as $D-1Load$ and $D-7Load$ should also be considered. Initially, the temperature data undergoes the simple min-max scaling of Equation 5.3, while the historical load data undergoes the enhanced min-max scaling of Equation 5.4. The variables D-1 Load and D-7 Load get values in the region [0,10] using this mathematical relationship.

The data has now been scaled and entered into the proposed neural network. The result of the forecast is a value within the field [0,10] and refers to the hourly load of the day for which the forecast is executed. This value should be converted to MW in order to compare with the corresponding actual hourly load value and to be able to correctly calculate the metrics of MSE, MAE and MAPE. The conversion of this value in MW is done by solving the equation Equation 5.5.

Figure 5.13 illustrates the load behaviour of all scaling methods tested. Table 5.10 summarizes the results of the MSE, MAE and MAPE metrics calculated from this forecast as a measure of comparison of all scaling methods for the hourly load forecast for 2017.

Table 5.10 is a measure of comparison of all scaling techniques presented and related to case B. Apparently, a smaller MAPE error in the forecast shows our enhanced proposed technique. Despite its relative simplicity, this method appropriately emphasizes the weight and importance of the input variables $D-1Load$ and $D-7Load$ in terms of MLP output. Both Min - Max Scaling and Enhanced Min - Max Scaling improve MAPE compared to unscaled input data, but do not produce better predictive results for the proposed enhanced technique.

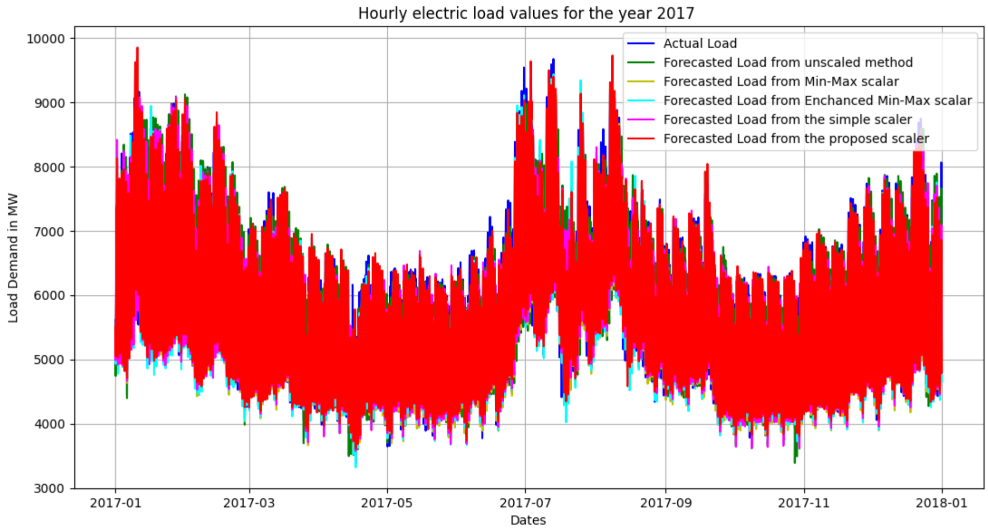


Figure 5.13: Graphical representation of the results obtained from the various methods of preprocessing the input data.

Table 5.10: MSE, MAE and MAPE values (based on MW) for for the hourly load forecast.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 111390.38535 | 239.36012 | 3.99 % |
| Simple Scaling | 71917.69612 | 189.87314 | 3.17 % |
| Enhanced Scaling (Proposed) | 62238.40502 | 174.52799 | 2.92 % |
| Min – Max Scaling | 75953.72593 | 204.73686 | 3.40 % |
| Enhanced Min-Max Scaling (Proposed) | 68369.70101 | 190.54487 | 3.17 % |

Figure 5.14 is a more detailed graphical representation of all the methods examined in case B, emphasizing the efficiency of the proposed technique.

### 5.2.3 Case C: Improved hourly load forecasting

In case C, a modified MLP neural network is used to predict the hourly value of the load. The only distinction between the current neural network and the present literature is that a new input variable called $H-1Load$ is applied that refers to the value of the load in the previous hour from that for which the prediction is made. Since the behaviour of the hourly value of the load is represented with greater precision knowing the load of the Greek interconnected system in the previous hour, the inclusion of this component improves the model prediction's

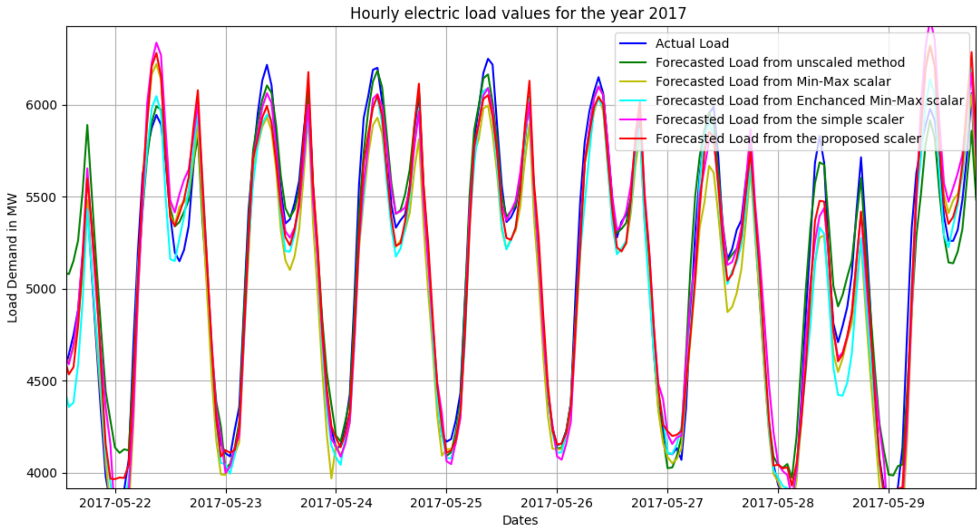


Figure 5.14: For the week of May, the hourly load demand is evaluated. The proposed approach for pre-processing input data more closely resembles the curve of real values.

success. The improved neural network model proposed in this thesis is illustrated in Figure 5.15 and consists of the following output variables:

- Hour: The time of day for which the load forecast will be made. The time is expressed as an integer with values ranging from 0 to 23.

- Week Day: It's a characteristic coding to decide the day of the week, much as in case A. The coding is done with integers ranging from 1 to 7, with 1 denoting Sunday, 2 denoting Monday, and so on.

- Holiday: Binary coding is used to indicate whether a day is a holiday or a working day. The number 1 is used to designate Greek state holidays, such as national anniversaries and major religious holidays, as well as weekends. The other days, on the other hand, are coded with number 0.

- Temperature: The hourly value of the temperature of the day for which the load is forecast.

- D-1 Load: The value of the previous day's load from that predicted at the corresponding time. For example, if the MLP neural network predicts Wednesday at 17:00 then the value of Tuesday load at 17:00 is entered as entry to the neural network.

- D-7 Load: The value of the load at the corresponding time on the same day of the previous week. If the load prediction for Wednesday 12/05 at 17:00 is the output of the neural network, then the value of the load for Wednesday 5/05 at 17:00 is entered as the neural network's input.

- H-1 Load: The value of the previous hour's load on which the forecast is based. If the load prediction for Wednesday 12/05 at 17:00 is the neural network's output, then the value of the load on Wednesday 12/05 at 16:00 is entered as the neural network's input.

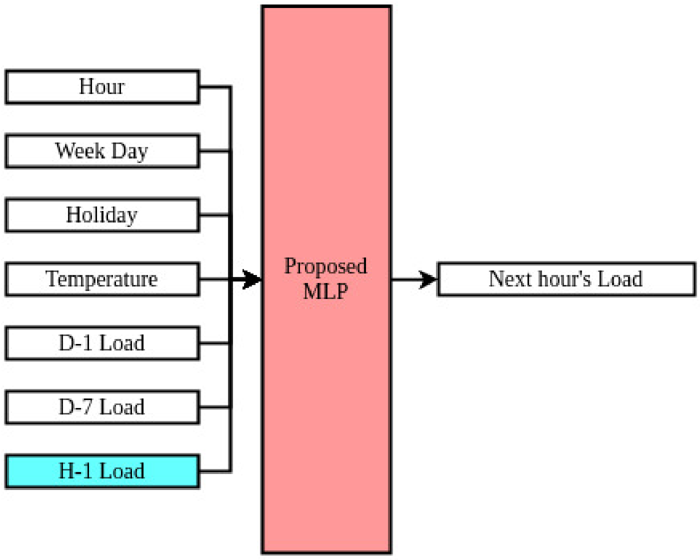


Figure 5.15: The form of MLP remains unchanged. However, the $H - 1Load$ variable is added to the input data to improve the forecast.

The MLP neural network architecture used to predict the hourly value of the load is shown in Figure 5.16. An input layer, a hidden layer, and an output layer represent the three layers of a neural network. Seven neurons make up the input level. Each neuron is associated with one of the variables listed above. There are 100 neurons in the hidden layer. The value 100 was chosen experimentally as it was found to produce better predictive values by dramatically reducing error. As can be seen from the papers in Chapter 4, neural networks with a single hidden layer approach the STLF problem quite accurately. The output layer is composed of a single neuron and refers to the hourly load value for which the prediction is developed.

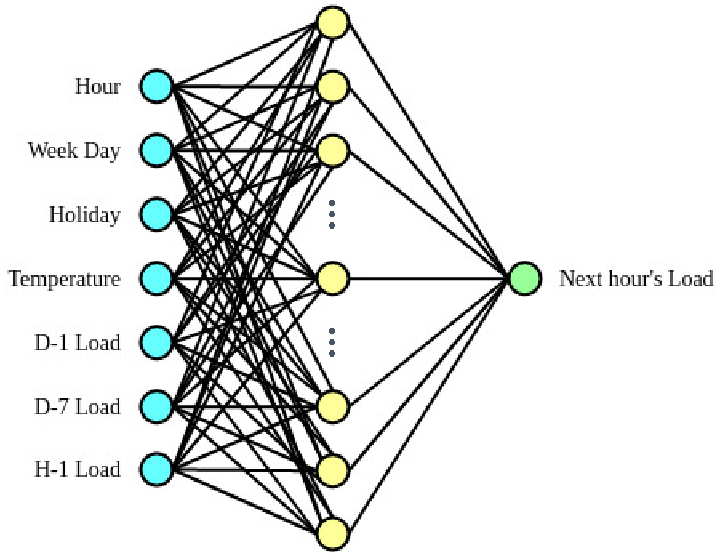


Figure 5.16: The structure of the proposed MLP neural network.

The aim of this study is to try to forecast hourly electricity prices for a year using data from the Greek power system. The forecast for 2017 (test set) is based on data from the previous four years, namely 2013, 2014, 2015, and 2016. (training set). The preprocessing techniques for the data entered into the neural network are of particular interest in order to develop the current prediction model for the Greek interconnected power system. As a result, the MSE, MAE, and MAPE metrics are used to compare the various scaling methods for the input data.

As previously mentioned, the impact of some scaling methods in the input data on the outcome of the prediction of the hourly value of the load by measuring the metrics MSE, MAE, and MAPE is investigated in this MSc Thesis. The input data is not subjected to some kind of modification at first, and it is entered into the three-layer perceptron as is. The results of this method are compared to the real hourly load values for the respective day for the entire year of 2017. The real and forecasted values are graphically depicted in Figure 5.17, while the metric equations are summarized in Table 5.11.

Next, the effect of a simple scaling of the input data on the prediction result of the neural network is studied. Simple scaling is performed by dividing each input variable by the maximum value of the corresponding data set. This procedure is performed only for the $Temperature$, $D-1Load$, $D-7Load$ and $H-1Load$ variables in order to obtain values within the field [0,1]. Equation 5.1 gives the simple scaling for temperature.

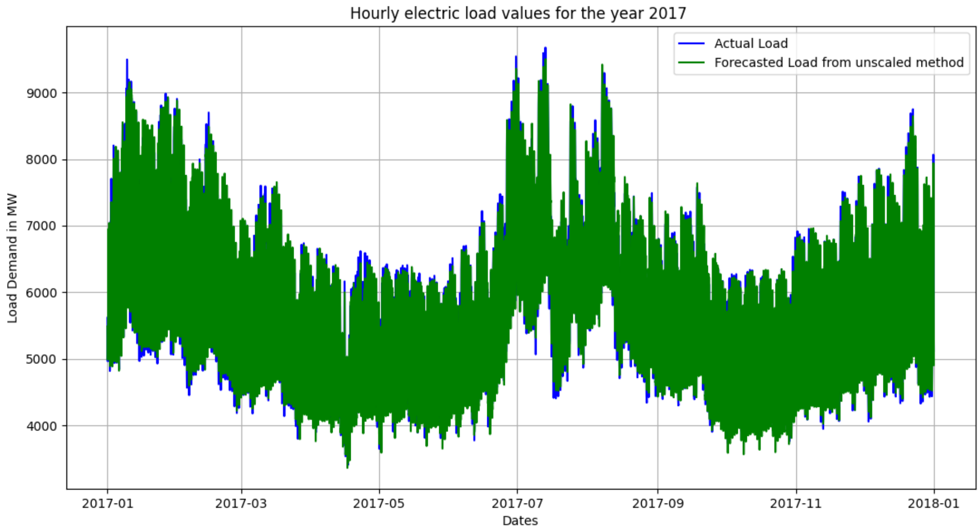


Figure 5.17: Graphic comparison of actual and predicted hourly load values for the year 2017. The input data is entered as is.

Table 5.11: MSE, MAE and MAPE values (based on MW) for the hourly load forecast. Input data is not subjected to any scaling technique.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 44111.39451 | 160.99752 | 2.72 % |

The scaled data is then fed into a neural network, which predicts the load's hourly value. The results of the proposed neural network are compared to real-time hourly values from 2017. Figure 5.18 depicts its graphic display. The MSE, MAE, and MAPE metrics calculated from this forecast using the same input data preprocessing technique are summarized in Table 5.12.

Table 5.12: MSE, MAE and MAPE values (based on MW) for the hourly load forecast. Input data is is subjected to a simple scaling technique.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 44111.39451 | 160.99752 | 2.72 % |
| Simple Scaling | 8208.88762 | 131.14385 | 2.24 % |

As a result of the above findings, it is clear that proper preprocessing of input data can produce better forecasting results than unoptimized data entry. As a result of this work, an

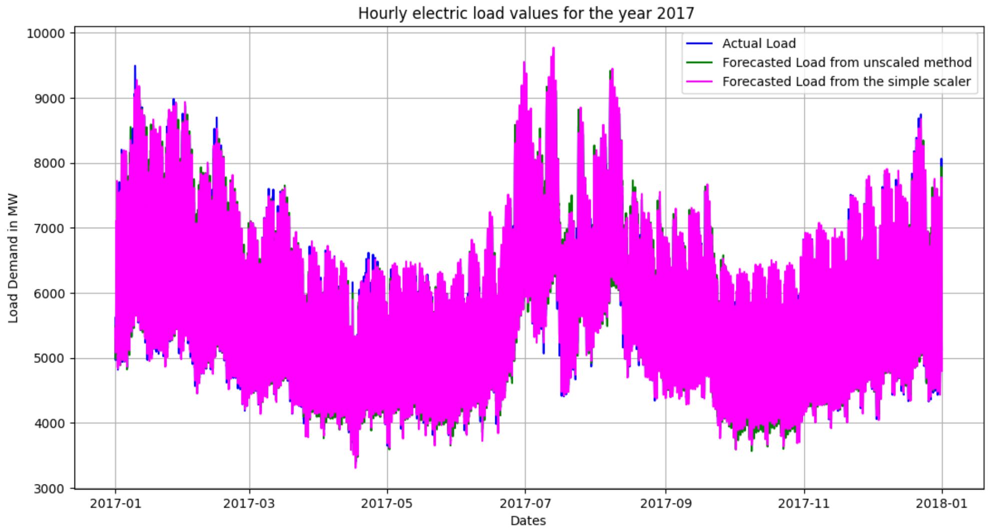


Figure 5.18: Graphic comparison of actual and predicted hourly load values for the year 2017. The values predicted through the simple scaling method are added.

innovative scaling method is proposed that will give the appropriate weight to the parameters $D-1Load$, $D-7Load$, and $H-1Load$, greatly improving the forecast results, following an extensive study of the correlation of the data as well as the way in which the neural network manages them.

In the first step, the temperature data is preprocessed as described in the previous equation. Following several data experiments, it was discovered that the variables $D-1Load$, $D-7Load$, and $H-1Load$ have a significant impact on the neural network's prediction result. As a result, these variables should be valued more highly in order to obtain a more accurate forecast. According to the findings of this study, similar to the previous cases, the coefficient 10 fully attributes the weight of these variables. Equation 5.2 describes the proposed scaling technique.

Afterwards, the scaled data is fed into the neural network. Figure 5.19 graphically depicts the resulting outcomes, which correspond to real and predicted hourly values for the year 2017. The MSE, MAE, and MAPE values calculated from this forecast are compared in Table 5.13. Our proposed data preprocessing technique improves forecasting significantly, so MAPE values fail.

The Min - Max method is a popular way of scaling neural network input data for regression problems. Because of the method's numerous references in the literature, its impact on the topic of this thesis should be considered as well. The temperature data and historical load

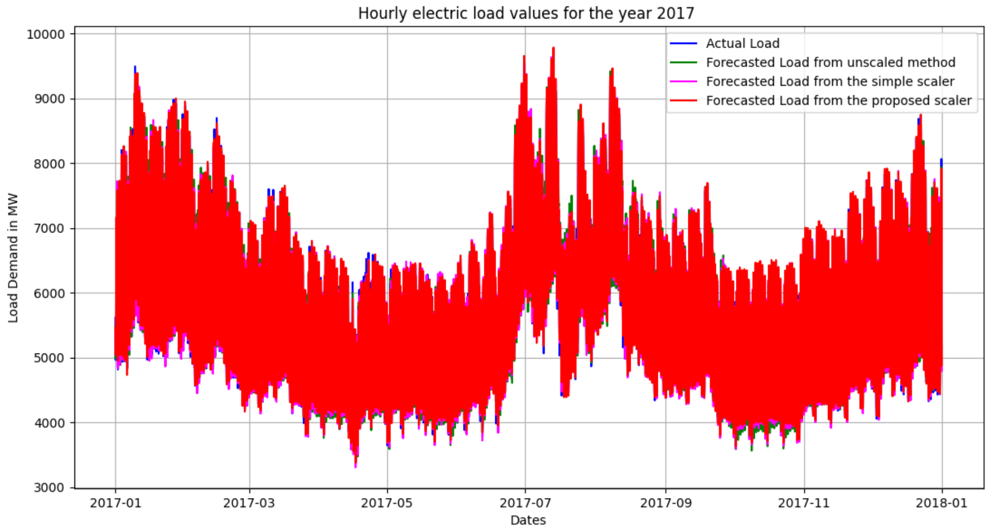


Figure 5.19: Graphic comparison of actual and predicted hourly load values for the year 2017. The values predicted through the proposed scaling method are added.

Table 5.13: MSE, MAE and MAPE values (based on MW) for the hourly load forecast. Input data is subjected to an enhanced scaling technique.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 44111.39451 | 160.99752 | 2.72 % |
| Simple Scaling | 8208.88762 | 131.14385 | 2.24 % |
| Enhanced Scaling (Proposed) | 22111.66683 | 112.91976 | 1.92 % |

data are subjected to a separate Min-Max scaling through the mathematical relation 5.3. The scaled data will now be inserted into the input layer to predict hourly load values for the year 2017. As shown in Figure 5.20, the values of the forecast results using the proposed MLP neural network are graphically compared to the real values. The values of the estimated MSE, MAE, and MAPE metrics are summarized in Table 5.14.

While Min-Max scaling does not seem to do as well as the other two data preprocessing methods, an improved Min-Max Scaling approach that stresses the value and weight of input variables $D-1Load$ and $D-7Load$ in the forecast outcome should also be considered. The temperature data is first scaled using the basic min-max equation, while the historical load data is scaled using the following formula 5.4.The variables $D-1Load$ and $D-7Load$ get values in the field [0,10] using this mathematical relationship.

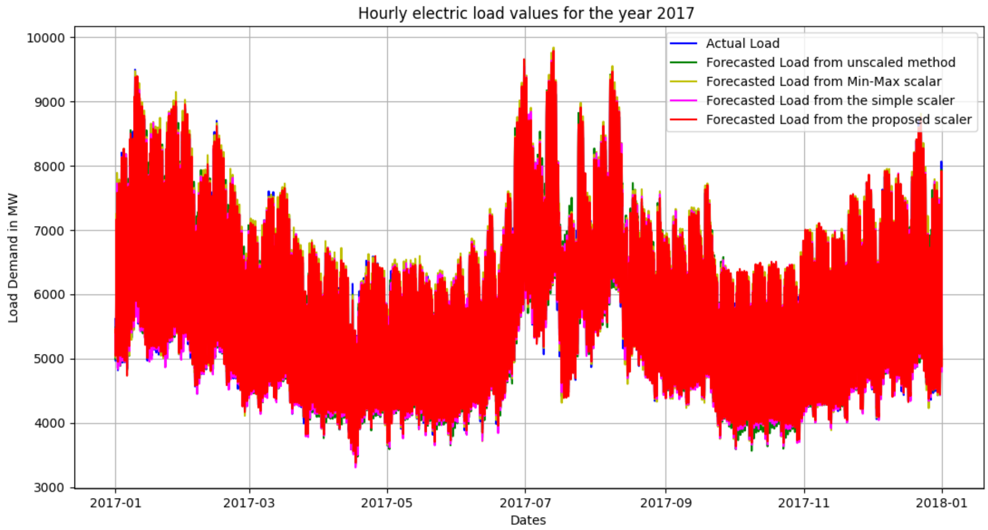


Figure 5.20: Graphic comparison of actual and predicted hourly load values for the year 2017. The values predicted through the min-max scaling method are added.

Table 5.14: MSE, MAE and MAPE values (based on MW) for the hourly load forecast. Input data is subjected to a min-max scaling technique.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 44111.39451 | 160.99752 | 2.72 % |
| Simple Scaling | 8208.88762 | 131.14385 | 2.24 % |
| Enhanced Scaling (Proposed) | 22111.66683 | 112.91976 | 1.92 % |
| Min – Max Scaling | 40720.48741 | 159.61260 | 2.73 % |

The data has now been scaled and entered into the proposed neural network. The forecast's outcome is a value in the region [0,10] that corresponds to the day's hourly load on which the forecast is conducted. This value can be translated to MW so that it can be compared to the corresponding real hourly load value and the MSE, MAE, and MAPE metrics can be calculated correctly. Equation 5.5 is used to convert this value to MW.

The load behaviour of all scaling methods evaluated is depicted in Figure 5.21. The results of the metrics MSE, MAE, and MAPE calculated are summarized in 5.15, which serve as a measure of reference with all scaling approaches for the hourly load forecast for 2017. Figure 5.22 and Figure 5.23 show the value of the weight attached to the input data in greater detail.

Our improved Min-Max Scaling strategy, it turns out, produces a lower MAPE value in the prediction. In terms of MLP performance, this approach sufficiently stresses the weight and value of the input variables $D-1Load$, $D-7Load$, and $H-1Load$, despite its relative simplicity. It's worth noting that when this improved MLP neural network is combined with our proposed enhanced scaling approach, the MAPE value drops below 2%, resulting in the lowest prediction value in the literature, based on data from the Greek interconnected power system.

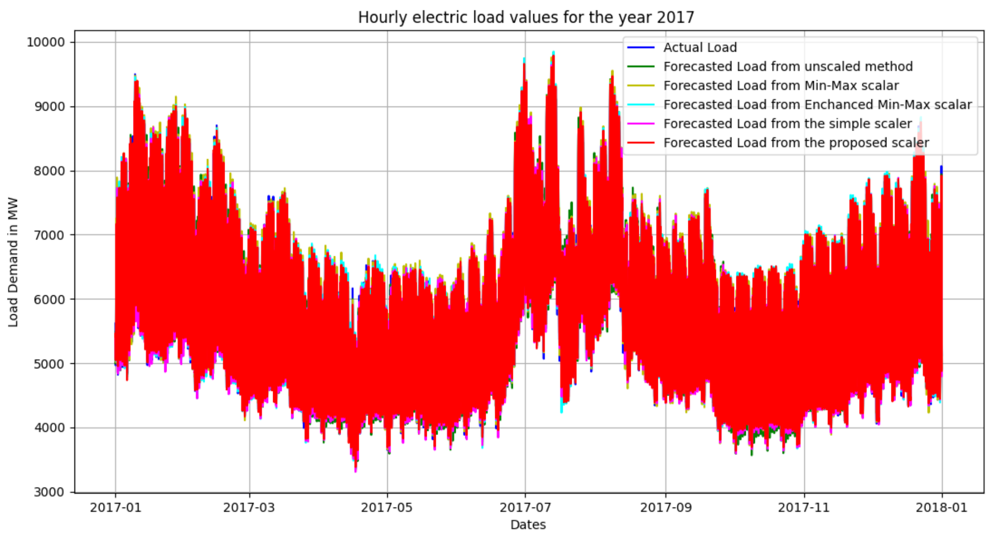


Figure 5.21: Graphic comparison of actual and predicted hourly load values for the year 2017. All preprocessing methods of the examined data are illustrated graphically.

Table 5.15: MSE, MAE and MAPE values (based on MW) for the hourly load forecast.The metrics for each scaling approach investigated are aggregated.

| Scaling Method | MSE | MAE | MAPE |
|---|---|---|---|
| Unscaled | 44111.39451 | 160.99752 | 2.72 % |
| Simple Scaling | 8208.88762 | 131.14385 | 2.24 % |
| Enhanced Scaling (Proposed) | 22111.66683 | 112.91976 | 1.92 % |
| Min – Max Scaling | 40720.48741 | 159.61260 | 2.73 % |
| Enhanced Min-Max Scaling (Proposed) | 18985.37885 | 103.60419 | 1.76 % |

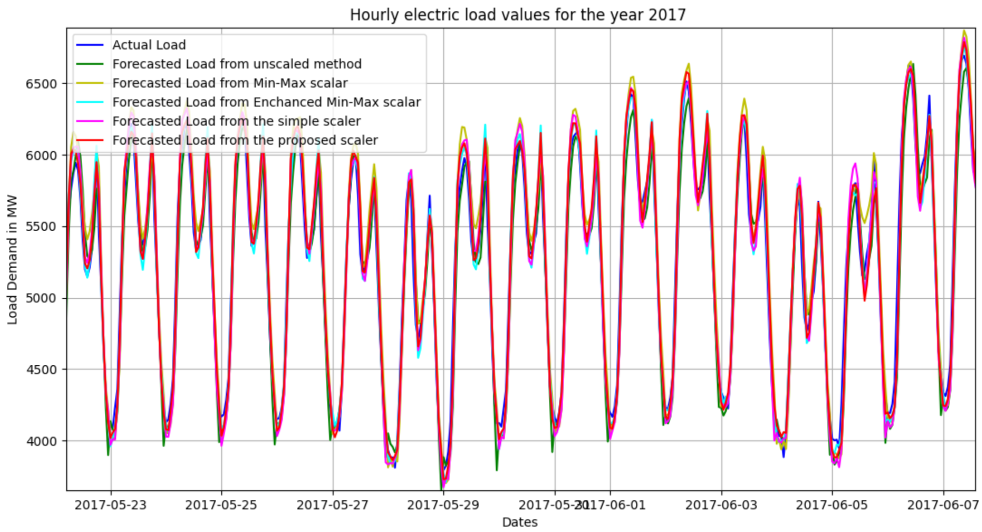


Figure 5.22: The various load curves for 2 weeks of 2017. The proposed method effectively follows the actual load curve.

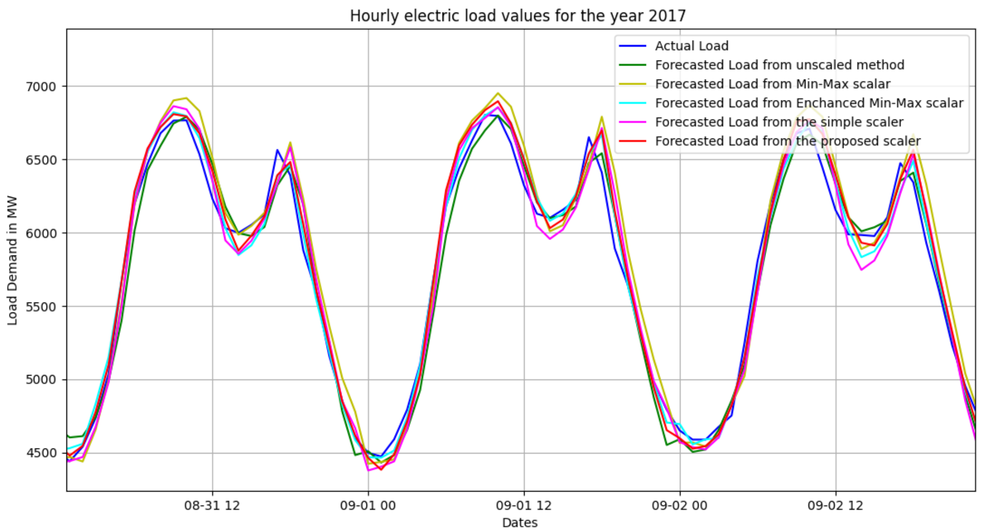


Figure 5.23: A more detailed depiction of the various load curves.

# Chapter 6

# Conclusion

Modelling, identification, and performance analysis are all used in developing statistical load forecasting models. For all stages, the system's active power generation matches the active power load. On an hourly basis, whole units must be brought online or rendered inaccessible, and load estimation over those time periods is critical. STLF based on 24-hour forecasts is needed for unit commitment and spinning reserve allocation. To predict sensitive conditions and take corrective action, security assessments depend on a priori awareness of planned bus load values ranging from 15 minutes to a few hours. Short-term load forecasting is a growing area of study that is expected to increase in the coming years. Because of their ability to generalize and simulate nonlinear dynamics, computational intelligence approaches are expected to be the driving force behind this field's study.

Because not every user is affected in the same way by time and weather effects, the behaviour of an electric power system load is affected by factors such as time, weather, and small random disturbances reflecting the inherent statistical nature of the load. Weekly periodicity and seasonal variations are examples of time factors. Temperature, humidity, light intensity, wind speed, precipitation, and cloud cover are all weather-related variables that have been shown to influence power consumption. The non-triviality of the forecasting problem has resulted in a plethora of methods for predicting load demand. These approaches can be divided into two categories: traditional intelligence techniques and computational intelligence techniques. Even today, it is unclear which method is the most effective. Those strategies perform well in some situations but are ineffective in others. Hybrid methods seem to have rekindled the interest of researchers in recent years, as they address the topic of STLF with high precision. The level of detail used in modelling, the collection of relevant influencing factors, such as social patterns and weather variables, and the level of testing that the methods undergo all influence the precision and pace of the forecast.

The results of various data preprocessing techniques on the outcome of the daily and hourly load prediction of the Greek electricity power system are presented in this master's thesis. The proposed neural network is based on historical load data from the day before, the week before, and the hour before. The architecture based on the suggested preprocessing techniques, with similar MLP architecture as in literature, provides even better prediction performance. In comparison to existing works, the MLP model, in conjunction with an improved method of preprocessing input data, greatly improves the forecast with the predicted loads being more accurate and closer to the real ones, effectively reducing the MAPE below 2%.

# Bibliography


[1] Weerakorn Ongsakul and Vo Ngoc Dieu. *Artificial Intelligence in Power System Optimization*. CRC Press, Inc., USA, 2013.

[2] R. Bayindir, Ilhami Colak, Gianluca Fulli, and Kenan Demirtas. Smart grid technologies and applications. *Renewable and Sustainable Energy Reviews*, 66:499–516, 12 2016.

[3] Elias Kyriakides and Marios Polycarpou. *Short Term Electric Load Forecasting: A Tutorial*, pages 391–418. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[4] James W. Taylor and Patrick E. McSharry. *Univariate Methods for Short-Term Load Forecasting*, chapter 2, pages 17–40. John Wiley and Sons, Ltd, 2017.

[5] K. Metaxiotis, A. Kagiannas, D. Askounis, and J. Psarras. Artificial intelligence in short term electric load forecasting: a state-of-the-art survey for the researcher. *Energy Conversion and Management*, 44(9):1525–1534, 2003.

[6] H.L. Willis and J.E.D. Northcote-Green. Spatial electric load forecasting: A tutorial review. *Proceedings of the IEEE*, 71(2):232–253, 1983.

[7] Deepa S. Kumar and J. S. Savier. Impact analysis of distributed generation integration on distribution network considering smart grid scenario. In *2017 IEEE Region 10 Symposium (TENSYMP)*, pages 1–5, 2017.

[8] Jizhong Zhu. *Optimization of Power System Operation*. 01 2015.

[9] Michael A. Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2018.

[10] 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.

[11] Himanshu Singh and Yunis Ahmad Lone. *Artificial Neural Networks*, pages 157–198. Apress, Berkeley, CA, 2020.

[12] J.M. Twomey and A.E. Smith. Performance measures, consistency, and power for artificial neural network models. *Mathematical and Computer Modelling*, 21(1):243–258, 1995.

[13] Miltiadis Alamaniotis, Dimitrios Bargiotas, and Lefteri Tsoukalas. Towards smart energy systems: Application of kernel machine regression for medium term electricity load forecasting. *SpringerPlus*, 5, 12 2016.

[14] Dimitrios Kontogiannis, Dimitrios Bargiotas, and Aspassia Daskalopulu. Fuzzy control system for smart energy management in residential buildings based on environmental data. *Energies*, 14(3), 2021.

[15] Miltiadis Alamaniotis, Andreas Ikonomopoulos, Aggelos Alamaniotis, Dimitrios Bargiotas, and Lefteri H. Tsoukalas. Day-ahead electricity price forecasting using optimized multiple-regression of relevance vector machines. In *8th Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MEDPOWER 2012)*, pages 1–5, 2012.

[16] Miltiadis Alamaniotis, Dimitrios Bargiotas, Nikolaos G. Bourbakis, and Lefteri H. Tsoukalas. Genetic optimal regression of relevance vector machines for electricity pricing signal forecasting in smart grids. *IEEE Transactions on Smart Grid*, 6(6):2997–3005, 2015.

[17] D.C. Park, M.A. El-Sharkawi, R.J. Marks, L.E. Atlas, and M.J. Damborg. Electric load forecasting using an artificial neural network. *IEEE Transactions on Power Systems*, 6(2):442–449, 1991.

[18] K.L. Ho, Y.-Y. Hsu, and C.-C. Yang. Short term load forecasting using a multilayer neural network with an adaptive learning algorithm. *IEEE Transactions on Power Systems*, 7(1):141–149, 1992.

[19] T.M. Peng, N.F. Hubele, and G.G. Karady. Advancement in the application of neural networks for short-term load forecasting. *IEEE Transactions on Power Systems*, 7(1):250–257, 1992.

[20] A.G. Bakirtzis, V. Petridis, S.J. Kiartzis, M.C. Alexiadis, and A.H. Maissis. A neural network short term load forecasting model for the greek power system. *IEEE Transactions on Power Systems*, 11(2):858–863, 1996.

[21] Paras Mandal, Tomonobu Senjyu, and Toshihisa Funabashi. Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market. *Energy Conversion and Management*, 47(15):2128–2142, 2006.

[22] Nahi Kandil, René Wamkeue, Maarouf Saad, and Semaan Georges. An efficient approach for short term load forecasting using artificial neural networks. *International Journal of Electrical Power and Energy Systems*, 28(8):525–530, 2006.

[23] G.J. Tsekouras, Fotis Kanellos, Vassiliki Kontargyri, Irene Karanasiou, Charalambos Elias, Anastasios Salis, and Nikos Mastorakis. A comparison of artificial neural networks algorithms for short term load forecasting in greek intercontinental power system. 01 2008.

[24] Miltiadis Alamaniotis and Lefteri H. Tsoukalas. Anticipation of minutes-ahead household active power consumption using gaussian processes. In *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pages 1–6, 2015.

[25] Dimitrios Kontogiannis, Dimitrios Bargiotas, and Aspassia Daskalopulu. Minutely active power forecasting models using neural networks. *Sustainability*, 12(8), 2020.

[26] Zhang Yun, Zhou Quan, Sun Caixin, Lei Shaolan, Liu Yuming, and Song Yang. Rbf neural network and anfis-based short-term load forecasting approach in real-time price environment. *IEEE Transactions on Power Systems*, 23(3):853–858, 2008.

[27] Wen-Yeau Chang. Short-term load forecasting using radial basis function neural network. *Journal of Computer and Communications*, 03:40–45, 01 2015.

[28] Carlo Cecati, Janusz Kolbusz, Paweł Różycki, Pierluigi Siano, and Bogdan M. Wilamowski. A novel rbf training algorithm for short-term electric load forecasting and comparative studies. *IEEE Transactions on Industrial Electronics*, 62(10):6519–6529, 2015.

[29] Madasu Hanmandlu and Bhavesh Kumar Chauhan. Load forecasting using hybrid models. *IEEE Transactions on Power Systems*, 26(1):20–29, 2011.

[30] Ioannis P. Panapakidis. Clustering based day-ahead and hour-ahead bus load forecasting models. *International Journal of Electrical Power and Energy Systems*, 80:171–178, 2016.

[31] Xishuang Dong, Lijun Qian, and Lei Huang. Short-term load forecasting in smart grid: A combined cnn and k-means clustering approach. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 119–125, 2017.

[32] Nima Amjady and Farshid Keynia. A new neural network approach to short term load forecasting of electrical power systems. *Energies*, 4(3):488–503, 2011.

[33] L. Ekonomou, C.A. Christodoulou, and Valeri Mladenov. A short-term load forecasting method using artificial neural networks and wavelet analysis. *International Journal of Power Systems*, 1:64–68, 07 2016.

[34] Fateme Fahiman, Sarah M. Erfani, Sutharshan Rajasegarar, Marimuthu Palaniswami, and Christopher Leckie. Improving load forecasting based on deep learning and k-shape clustering. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 4134–4141, 2017.

[35] Nian Liu, Qingfeng Tang, Jianhua Zhang, Wei Fan, and Jie Liu. A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids. *Applied Energy*, 129:336–345, 2014.

[36] Hamed H.H. Aly. A proposed intelligent short-term load forecasting hybrid models of ann, wnn and kf based on clustering techniques for smart grid. *Electric Power Systems Research*, 182:106191, 2020.

[37] Tareq Hossen, Siby Jose Plathottam, Radha Krishnan Angamuthu, Prakash Ranganathan, and Hossein Salehfar. Short-term load forecasting using deep neural networks (dnn). In *2017 North American Power Symposium (NAPS)*, pages 1–6, 2017.

[38] Muhammad Qamar Raza and Abbas Khosravi. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews*, 50:1352–1372, 2015.

[39] Seunghyoung Ryu, Jaekoo Noh, and Hongseok Kim. Deep neural network based demand side short term load forecasting. In *2016 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 308–313, 2016.