ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Kolmogorov-Arnold Νευρωνικά Δίκτυα και Εφαρμογές τους σε Χρονοσειρές

Διπλωματική Εργασία

Μυρτώ Δραγανίγου

Επιβλέπων: Κατσαρός Δημήτριος

Volos, 09/2025

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

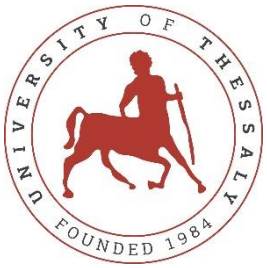
Kolmogorov-Arnold Νευρωνικά Δίκτυα και Εφαρμογές τους σε Χρονοσειρές

Διπλωματική Εργασία

Μυρτώ Δραγανίγου

Επιβλέπων: Κατσαρός Δημήτριος

Volos, 09/2025



UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Kolmogorov-Arnold Neural Networks and their Applications in Time Series

Diploma Thesis

Myrto Draganigou

Supervisor: Dimitrios Katsaros

Volos, 09/2025

Εγκρίνεται από την Επιτροπή Εξέτασης:

Επιβλέπων	**Δημήτριος Κατσαρός** Αναπληρωτής Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας
Μέλος	**Γεώργιος Θάνος** Μέλος Εργαστηριακού Διδακτικού Προσωπικού, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας
Μέλος	**Δημήτριος Ραφαηλίδης** Αναπληρωτής Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΑΚΑΔΗΜΑΪΚΗΣ ΔΕΟΝΤΟΛΟΓΙΑΣ ΚΑΙ ΠΝΕΥΜΑΤΙΚΩΝ ΔΙΚΑΙΩΜΑΤΩΝ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα διπλωματική εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας, αποτελούν αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλουν οποιασδήποτε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχουν έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Δηλώνω επίσης ότι τα αποτελέσματα της εργασίας δεν έχουν χρησιμοποιηθεί για την απόκτηση άλλου πτυχίου. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Η Δηλούσα

Μυρτώ Δραγανίγου

DISCLAIMER ON ACADEMIC ETHICS AND INTELLECTUAL PROPERTY RIGHTS

Being fully aware of the implications of copyright laws, I expressly state that this diploma 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

Myrto Draganigou

Διπλωματική Εργασία

Kolmogorov-Arnold Νευρωνικά Δίκτυα και Εφαρμογές τους σε Χρονοσειρές

Μυρτώ Δραγανίγου

Περίληψη

Η παρούσα διπλωματική εργασία εξετάζει την εφαρμογή των Kolmogorov–Arnold Νευρωνικών Δικτύων στην ανάλυση χρονοσειρών. Οι χρονοσειρές χαρακτηρίζονται από μη γραμμικότητα, εποχικότητα και θόρυβο, στοιχεία που δυσχεραίνουν τη μοντελοποίησή τους. Παρά την πρόοδο με στατιστικές μεθόδους και νευρωνικές αρχιτεκτονικές, εξακολουθούν να υπάρχουν προκλήσεις που αφορούν τη γενίκευση, την αποδοτικότητα και την ερμηνευσιμότητα.

Στην συγκεκριμένη εργασία μελετήθηκαν και αξιολογήθηκαν διαφορετικές παραλλαγές KAN, όπως τα SplineKAN, ChebyshevKAN και υβριδικές εκδοχές, σε σύνολα δεδομένων του UCR Time Series Archive αλλά και σε πραγματικά δεδομένα πρόγνωσης θερμοκρασίας. Η απόδοσή τους συγκρίθηκε με καθιερωμένα μοντέλα όπως τα MLP, LSTM, ARIMA και Prophet, με χρήση μετρικών ταξινόμησης (accuracy, F1-score) και πρόβλεψης (MAE, RMSE).

Τα αποτελέσματα αποδεικνύουν ότι οι εκδοχές των KAN παρουσιάζουν ανταγωνιστική, και σε αρκετές περιπτώσεις ανώτερη απόδοση, ενώ ταυτόχρονα προσφέρουν καλύτερη ερμηνευσιμότητα με μικρότερα μεγέθη δικτύου. Η εργασία συμβάλλει στην αποσαφήνιση του κατά πόσο τα KANs μπορούν να αποτελέσουν μια αξιόπιστη και επεκτάσιμη εναλλακτική τεχνική για την ανάλυση χρονοσειρών.

Λέξεις-κλειδιά: KAN, Χρονοσειρές, Νευρωνικά Δίκτυα, Μηχανική Μάθηση, Πρόβλεψη, Κατηγοριοποίηση, Grid, Chebyshev, Υβριδικές Τεχνικές, Στατιστική Ανάλυση.

Diploma Thesis

Kolmogorov-Arnold Neural Networks and their Applications in Time Series

Myrto Draganigou

Abstract

This thesis examines the application of Kolmogorov Arnold Neural Networks in time series analysis. Time series are characterized by nonlinearity, seasonality, and noise, which complicate their modeling. Despite the progress achieved with statistical methods and neural architecture, challenges remain in terms of generalization, efficiency, and interpretability.

In this work, different KAN variants, such as SplineKAN, ChebyshevKAN, and hybrid approaches, were studied and evaluated on datasets from the UCR Time Series Archive as well as on real world temperature forecasting data. Their performance was compared with established models such as MLP, LSTM, ARIMA and Prophet, using both classification metrics (accuracy, F1-score) and forecasting metrics (MAE, RMSE).

The results demonstrate that KAN variants achieve competitive, and in several cases superior, performance while also offering improved interpretability with smaller network sizes. This study contributes to clarifying the extent to which KAN can serve as a reliable and scalable alternative technique for time series analysis.

Keywords: KAN, Time Series, Neural Networks, Machine Learning, Forecasting, Classification, Grid, Chebyshev, Hybrid Techniques, Statistical Analysis.

Πίνακας περιεχομένων

Κατάλογος εικόνων

Κατάλογος σχημάτων

Κατάλογος πινάκων

Συντομογραφίες

et al.	και άλλοι
κλπ.	και λοιπά
πχ.	Παραδείγματος χάρη
ΑΕΙ	Ανώτατο Εκπαιδευτικό Ίδρυμα
KAN	Kolmogorov Arnold Networks
KAT	Kolmogorov Arnold Theorem
MLP	Multi Layer Perceptron
MoE	Mixture of Experts
MoK	Mixture of KAN
MMK	Multi layer Mixture of KAN
TSF	Time Series Forecasting
ARIMA	Autoregressive integrated moving average
SARIMA	Seasonal Autoregressive integrated moving average
SARIMAX	Seasonal Autoregressive integrated moving average with Exogenous Regressors
RNN	Recurrent Neural Network
LSTM	Long Short Term Memory
CNN	Convolutional Neural Network
μ	μέσος όρος
sd	standard deviation
ACF	Auto Correlation Function
PACF	Partial Auto Correlation Function
GRU	Gated Recurrent Unit
NN	Neural Network

ML	Machine Learning
XAI	Explainable artificial intelligence
SHAP	SHapley Additive exPlanations
GRAD	Gradient weighted
TSR	Temporal Saliency Rescaling
DTW	Dynamic time warping
TSV	Tab Separated Values
MNIST	*Modified National Institute of Standards and Technology*
Tanh	hyperbolic tangent
SiLU	Sigmoid Linear Unit
ReLU	Rectified Linear Unit
MSE	Mean Squared Error
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
CPU	Central Processing Unit
Argmax	Arguments of the Maxima

Κεφάλαιο 1 Εισαγωγή

Τα τελευταία χρόνια εκτελείται η ραγδαία εξέλιξη νέων τεχνικών και τεχνολογιών, γεγονός που έχει δώσει επιπλέον δυνατότητες στην οπτική των πάλαι πότε προβλημάτων. Έτσι, η κατανόηση και ανάλυση δεδομένων, η οποία μπορεί να δώσει πολύτιμες πληροφορίες, έχει αποκτήσει ολοένα και μεγαλύτερη σημασία, ειδικά σε απαιτητικά πεδία εφαρμογής όπως οι χρονοσειρές που θα μελετήσουμε. Πρόκειται για ακολουθίες μετρήσεων που συλλέγονται διαδοχικά στο χρόνο, οι οποίες εμφανίζουν συχνά μη γραμμικότητα, εποχικότητα, θορυβώδη συμπεριφορά και υψηλή διαστατικότητα [25]. Η ανάλυση και πρόβλεψη χρονοσειρών βρίσκει εφαρμογή σε μείζονα ζητήματα της καθημερινής ζωής όπως η πρόγνωση καιρού, η κατανομή πρώτων αναγκών, η ενέργεια και η υγεία.

Η μηχανική μάθηση έχει αποδειχτεί ένας κλάδος που μπορεί να συνεισφέρει καθοριστικά στην εκμάθηση τέτοιων δεδομένων και την βελτίωση της απόδοσης του. Από την σύλληψη μακροπρόθεσμων εξαρτήσεων και μη γραμμικών σχέσεων μέχρι την γενίκευση σε ετερογενείς πηγές δεδομένων αλλά και την αυτόματη εξαγωγή χαρακτηριστικών φαίνεται ότι μπορεί να συνεισφέρει καθοριστικά.

Παρά την πρόοδο που συντελείται με προσεγγίσεις όπως οι στατιστικές και οι νευρωνικές, οι προκλήσεις που σχετίζονται με τη γενίκευση, την υπολογιστική αποδοτικότητα και τη σταθερότητα των μοντέλων σε δεδομένα με θόρυβο ή μεγάλη ποικιλία χαρακτηριστικών παραμένουν [43]. Αυτό καθιστά αναγκαία την ανάπτυξη νέων αρχιτεκτονικών, οι οποίες θα μπορούν να συνδυάζουν εκφραστικότητα και υπολογιστική αποτελεσματικότητα, αξιοποιώντας θεωρητικά θεμέλια της μαθηματικής ανάλυσης.

1.1Αντικείμενο της διπλωματικής

Η παρούσα διπλωματική εργασία επικεντρώνεται στη μελέτη και εφαρμογή των Kolmogorov–Arnold Networks (KANs) σε προβλήματα ταξινόμησης και πρόβλεψης χρονοσειρών. Σκοπός μας είναι να μελετήσουμε εκτενώς αν τα KAN μπορούν να αποτελέσουν μία σύγχρονη τεχνική που θα ξεπεράσει τους υπολογιστικούς περιορισμούς των παραδοσιακών νευρωνικών δικτύων, όπως τα MLP.

Τα KAN βασίζονται στο θεώρημα Kolmogorov–Arnold, σύμφωνα με το οποίο οποιαδήποτε πολυδιάστατη συνεχής συνάρτηση μπορεί να αναπαρασταθεί μέσω πεπερασμένου αριθμού μονοδιάστατων συναρτήσεων. Το θεωρητικό αυτό υπόβαθρο τα καθιστά ιδιαίτερα κατάλληλα για εφαρμογές σε πεδία όπου η μη γραμμικότητα και οι πολύπλοκες εξαρτήσεις μεταξύ μεταβλητών αποτελούν πρόκληση. Βάση των ερευνών η αρχιτεκτονική τους, τους δίνει την δυνατότητα να έχουν πιο ευέλικτες και ακριβείς προσεγγίσεις μετριάζοντας προβλήματα, όπως η πολυπλοκότητα και η μη γραμμικότητα διατηρώντας παράλληλα υψηλή ικανότητα ακρίβειας και ερμηνευσιμότητας [44]. Για αυτό τον λόγω θεωρούνται και μία ανερχόμενη μέθοδος για ανάλυση δεδομένων όπως οι χρονοσειρές.

Παρά τις δυνατότητες τους, τα KAN αντιμετωπίζουν προκλήσεις σε σχέση με την υπολογιστική τους ικανότητα σε καταστάσεις υψηλής διάστασης αλλά και αυξημένου θορύβου. Το γεγονός αυτό μαζί με την πρόκληση του νεαρού χρόνου ζωής τους, συνιστούν σε συνεχείς έρευνες για την βελτιστοποίηση τους, με τεχνικές όπως η κανονικοποίηση και τα υβριδικά μοντέλα.

1.2 Συνεισφορά

Σε αυτή την διπλωματική συνεισφέρουμε ως προς την μελέτη τους σκοπού μας, την εξαγωγή δηλαδή συμπερασμάτων για το αν μπορούν τα KAN να αποτελέσουν μια επαρκή και αποδοτική εναλλακτική λύση για την ανάλυση χρονοσειρών. Η συμβολή μας εστιάζει τόσο στην εφαρμογή όσο και στη συγκριτική αξιολόγηση των KAN, με στόχο να αποσαφηνιστεί η πρακτική τους αξία έναντι καθιερωμένων μεθόδων. Σημειώνουμε συγκεκριμένα παρακάτω την συμβολή μας:

1. **Θεωρητική μελέτη** του Kolmogorov–Arnold θεωρήματος και των παραλλαγών των KANs (SplineKAN, ChebyshevKAN, HybridKAN, Mixture of KANs), ώστε να κατανοηθεί το μαθηματικό υπόβαθρο και οι ιδιαιτερότητές τους.
2. **Υλοποίηση σε δύο διαφορετικά παιδία μελέτης, αυτό της πρόβλεψης και αυτό της ταξινόμησης.** Εφαρμογή διαφορετικών εκδοχών KANs, με έμφαση σε τεχνικές βελτιστοποίησης της απόδοσης, την κανονικοποίηση και μεθόδους πρόωρης διακοπής για την αντιμετώπιση θορύβου και υπερεξειδίκευσης.

3. **Εφαρμογή σε σύνολα δεδομένων**, με αξιολόγηση σε τυπικά δεδομένα όπως του UCR Time Series Archive, καθώς και σε πραγματικά δεδομένα πρόγνωσης θερμοκρασίας, ώστε να δοκιμαστεί η γενίκευση των μοντέλων.
4. **Συγκριτική ανάλυση** με παραδοσιακές και σύγχρονες μεθόδους, όπως τα MLP, LSTM, ARIMA και Prophet, για να αποτιμηθεί η αποδοτικότητα των KAN.
5. **Στατιστική αξιολόγηση** των αποτελεσμάτων μέσω μετρικών απόδοσης (Accuracy, F1-score, MAE, RMSE) και στατιστικοί έλεγχοι σημαντικότητας (π.χ. Wilcoxon test), για την συγκριτική τεκμηρίωση των αποτελεσμάτων.
6. **Πειραματική διερεύνηση υπερπαραμέτρων**, όπως των μεγεθών window size, grid size και Chebyshev degree για να εντοπιστούν οι παράγοντες που επηρεάζουν κρίσιμα την απόδοση των KAN σε διαφορετικά δεδομένα.
7. **Πρόταση μελλοντικών επεκτάσεων** με σκοπό να επεκταθεί η ανάλυση στο συγκεκριμένο πεδίο με προτάσεις που μπορούν να αναδείξουν τα KAN σε μία σταθερή τεχνική προς μαζική ερευνητική αξιοποίηση.

1.3 Οργάνωση του τόμου

Η εργασία μας είναι οργανωμένη σε έξι κεφάλαια, παρέχοντας μία εκτενή μελέτη για την εφαρμογή των Kolmogorov Arnold Networks στο πεδίο της ανάλυσης χρονοσειρών. Τα κεφάλαια δομούνται με τον εξής τρόπο:

Στο κεφάλαιο 2 παρουσιάζεται το θεωρητικό υπόβαθρο των KAN, αναλύοντας το αντίστοιχο θεώρημα και την αρχιτεκτονική του, καθώς και οι βασικές παραλλαγές του.

Στο **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.** αναλύονται τα βασικά χαρακτηριστικά των χρονοσειρών, οι μέθοδοι που ευδοκιμούν στον χώρο της στατιστικής τους ανάλυσης καθώς και η συμβολή της μηχανικής μάθησης σε αυτό, μαζί με την παρουσίαση των μοντέλων που επιλέγονται στο πεδίο αυτό, εκ των οποίων γίνεται επιλογή για χρήση ως μοντέλα σύγκρισης. Τέλος παρουσιάζονται τα σετ δεδομένων που θα χρησιμοποιήσουμε στην εκπαίδευση.

Στο **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**περιγράφεται η πειραματική διαδικασία και συγκεκριμένα η δομή των δικτύων KAN που γίνεται χρήση, και τα στάδια της προ επεξεργασίας, της εκπαίδευσης αλλά και της αξιολόγησης για

καθένα από τα δύο πεδία εφαρμογής καθώς και οι διάφορες τεχνικές βελτιστοποίησης που χρησιμοποιήθηκαν.

Στο κεφάλαιο 5 παρουσιάζονται και σχολιάζονται τα αποτελέσματα των πειραμάτων, τα διαγράμματα ρύθμισης υπερπαραμέτρων και η επίδοση σε ομαδοποιήσεις των δεδομένων με συγκριτική αξιολόγηση των σε σχέση με τα αποτελέσματα των μοντέλων αναφοράς.

Τέλος, στο Κεφάλαιο 6 παρουσιάζονται συγκεντρωτικά τα κυριότερα συμπεράσματα της παρούσας εργασίας σε σχέση με όλα τα μοντέλα KAN, οι λόγοι της ερευνητικής τους αξίας, η συμβολή της στη διερεύνηση του θέματος, καθώς και συγκεκριμένες προτάσεις για περαιτέρω μελλοντική διερεύνηση του θέματος.

Κεφάλαιο 2 Kolmogorov – Arnold Νευρωνικά Δίκτυα

Τα Kolmogorov – Arnold Networks (KAN) είναι ένας τύπος αρχιτεκτονικής νευρωνικών δικτύων βασισμένα στο θεώρημα αναπαράστασης Kolmogorov – Arnold (KAT). Εφαρμόζονται στο πεδίο της μηχανικής μάθησης και αναφέρονται στις μελέτες ως μία πολλά υποσχόμενη εναλλακτική στα (Multi-Layer Perceptrons) MLPs. Η μεγάλη διάκριση τους είναι η εξής: Ενώ τα MLPs έχουν σταθερές συναρτήσεις ενεργοποίησης (activation functions) στους κόμβους (nodes), τα KAN έχουν εκπαιδεύσιμες συναρτήσεις ενεργοποίησης στις ακμές (edges). Τα KAN αντικαθιστούν τα γραμμικά βάρη στις ακμές από μια μονομεταβλητή εκπαιδεύσιμη συνάρτηση παραμετροποιημένη ως B - Spline. [2] Αυτή είναι μία βασική αλλαγή στον τρόπο εκμάθησης των παραμέτρων και παρότι φαινομενικά είναι απλή, κάνει τα KAN να ξεπερνούν πολλές φορές τα MLP όσον αφορά την ακρίβεια αλλά και την ερμηνευσιμότητα.

2.1 Το Θεώρημα Kolmogorov – Arnold

Το Θεώρημα Kolmogorov – Arnold διατυπώθηκε αρχικά από τον Andrey Kolmogorov το 1957 [4] και διαμορφώθηκε ως το τελικό Θεώρημα Kolmogorov – Arnold μετά την παρέμβαση του Vladimir Arnold το 1963 [5]. Σύμφωνα με αυτό, κάθε πολυμεταβλητή συνεχής συνάρτηση f(x) μπορεί να αναπαρασταθεί ως πεπερασμένος συνδυασμός συνεχών μονομεταβλητών συναρτήσεων $\Phi_q: \mathrm{R} \to R$ και $\varphi_{q,p}: [0,1]^n \to R$ [1]με την μορφή που πρότειναν στις πρωτότυπες εργασίες οι Kolmogorov και Arnold [4], [5]:

$$f(\mathbf{x}) = f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right).$$

where $\phi_{q,p}: [0,1] \to \mathbb{R}$ and $\Phi_q: \mathbb{R} \to \mathbb{R}$.

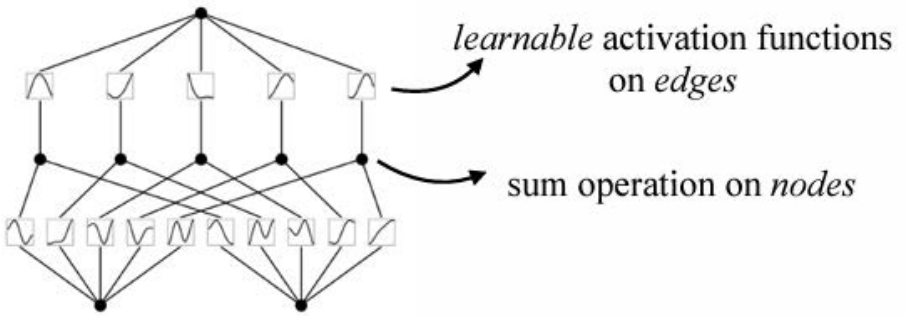
Σχήμα 2.1: Το Θεώρημα Kolmogorov – Arnold

Κατά μία έννοια, έδειξαν ότι η μόνη πολυμεταβλητή συνάρτηση είναι η πρόσθεση, καθώς κάθε άλλη συνάρτηση μπορεί να αναπαρασταθεί μόνο από μονομεταβλητές συναρτήσεις και άθροισμα.

Το θεώρημα Kolmogorov-Arnold βάση των ιδιοτήτων του φάνηκε ιδιαίτερο ελπιδοφόρο και ανταγωνιστικό προς παρόμοια τρέχοντα μοντέλα όπως τα MLP. Παρά την θεωρητική του πληρότητα όμως, αρχικά δεν θεωρήθηκε ιδιαίτερα χρήσιμο. Η κύρια δυσκολία έγκειται στο γεγονός ότι οι μονομεταβλητές συναρτήσεις που προκύπτουν από την αναπαράσταση του θεωρήματος εμφανίζουν συχνά έλλειψη ομαλότητας, καθιστώντας την εκμάθησή τους ιδιαιτέρως απαιτητική στην πράξη. Όμως πληθώρα μελετών το τελευταίο διάστημα, επαναφέρουν το Θεώρημα σαν μια πολύ υποσχόμενη μέθοδο, μέσω της εφαρμογής των Kolmogorov – Arnold Νευρωνικών Δικτυών (KAN). Αυτό οφείλεται κυρίως στις νέες πηγές που προτείνουν τη γενίκευση του δικτύου πέρα από την αρχική αναπαράσταση βάθους 2 και πλάτους (2n+1) όρων στο κρυφό επίπεδο.

2.2 KAN Αρχιτεκτονική

Η αρχιτεκτονική των Kolmogorov–Arnold Networks διαφοροποιείται θεμελιωδώς από τα κλασικά νευρωνικά δίκτυα, καθώς αντικαθιστά τα συμβατικά βάρη των ακμών με εκπαιδεύσιμες μονομεταβλητές συναρτήσεις ενεργοποίησης.



Σχήμα 2.2: Η αρχιτεκτονική του KAN

Η αρχική αναπαράσταση που προτείνεται από τους Ziming Liu et al (2024), αποτελείται από ένα δίκτυο δύο επίπεδων με διαμόρφωση [n, 2n + 1, 1], όπου η είσοδος είναι n διάστασης και η έξοδος είναι μονοδιάστατη. Τα δεδομένα εισόδου θεωρούνται πολυδιάστατα διανύσματα της μορφής [2]:

$$x = [x1, x2, \ldots, xn]$$

Χαρακτηρίζονται από υψηλή διαστασιμότητα, δηλαδή μεγάλο πλήθος εισόδων, που αντιστοιχούν στις μεταβλητές της συνάρτησης στόχου f(x1,...,xn), η οποία και συνιστά την βασική πρόκληση στη μοντελοποίηση.

Πριν την είσοδο στο δίκτυο, τα δεδομένα υπόκεινται σε προ επεξεργασία και κανονικοποίηση ώστε να ευθυγραμμίζονται με το πεδίο των spline συναρτήσεων ή με τα όρια των μεγεθών Grid. Στην διαδικασία της εκπαίδευσης περνάνε δέσμες (batch) από πολυδιάστατα διανύσματα που τελικά αναπαρίσταται ως τένσορες του σχήματος :

Input Shape: (batch_size,num_features)

Κάθε επίπεδο KAN με δεδομένα εισόδου n διαστάσεων διαμορφώνεται ως ένας πίνακας μονομεταβλητών συναρτήσεων Φ = $\{\phi_{q,p}\}$ όπου κάθε συνάρτηση $\phi_{q,p}$ μπορεί να αναπαρασταθεί ως μία B – spline

$$s(x) = \sum_n c_n B_n(x)$$

Ως εκ τούτου η συνάρτηση ενεργοποίησης ϕ(x) ορίζεται ως το άθροισμα των βασικών συναρτήσεων b(x) και των συναρτήσεων spline που μπορεί να αναπαρασταθεί με την εξίσωση:

$$\phi(x) = \alpha(b(x) + s(x))$$

Για να γενικευτεί η αρχιτεκτονική πέρα από την αρχική αναπαράσταση των 2 επιπέδων, επεκτείνεται σε σύνθεση L τέτοιων επιπέδων δημιουργώντας ευρύτερα και βαθύτερα KAN. Αν το αρχικό διάνυσμα θεωρηθεί ως $x_0 \in \mathbb{R}^{n0}$, τότε η έξοδος του πολυεπίπεδου KAN διαμορφώνεται ως

$$KAN(x) = (\Phi L-1 \circ \Phi L-2 \circ \cdots \circ \Phi 1 \circ \Phi 0)\, x.$$

Η ενεργοποίηση ενός νευρώνα στο επίπεδο l + 1 προκύπτει ως το άθροισμα όλων των εξόδων από τις εισερχόμενες ακμές του επιπέδου l, με κάθε ακμή να εφαρμόζει μια ξεχωριστή, εκπαιδεύσιμη, μονομεταβλητή συνάρτηση στην τιμή ενεργοποίησης του προηγούμενου κόμβου. Όλες οι πράξεις είναι διαφορίσιμες, επιτρέποντας την εκπαίδευση του δικτύου μέσω της διαδικασίας οπισθοδιάδοσης σφάλματος (backpropagation) [2].

2.2.1To B - Spline KAN

Όπως αναφέρεται νωρίτερα οι B-splines είναι οι βάσεις με τις οποίες αναπαρίσταται μια μονομεταβλητή συνάρτηση $\phi_{q,p}$. Σύμφωνα με τον de Boor [6] κάθε spline ορίζεται με

βάση μια ακολουθία κόμβων (knots) και μπορεί να εκφραστεί ως γραμμικός συνδυασμός από B-spline βάσεις με την μορφή:

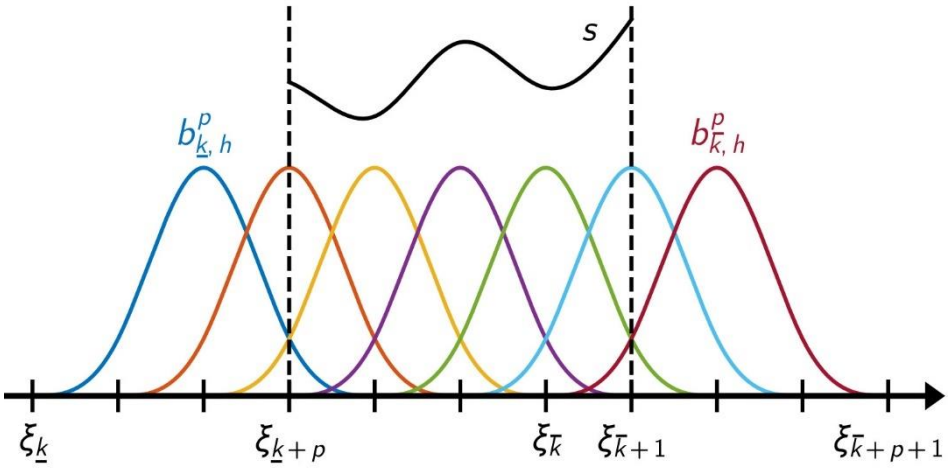
$$s(x) = \sum_{n} c_n B_n(x)$$

όπου $B_n(x)$ η n-οστή βάση και c_n οι αντίστοιχοι συντελεστές. Μαθηματικά ο ορισμός όπως τον περιέγραψαν οι Curry & Schoenberg [1947] είναι ο εξής:

Έστω t = (tj) μία μη φθίνουσα ακολουθία η οποία μπορεί να είναι πεπερασμένη, άπειρη ή δι-άπειρη). Η j-οστή (κανονικοποιημένη) B-spline τάξης k για την ακολουθία κόμβων t συμβολίζεται με $B_{j,k,t}$ και ορίζεται από τον κανόνα:

$$\mathrm{B_{j,k,t}(x) = (t_{j+k} - t_j)(t_j, \dots, t_{j+k})(-x)^{k-1}}, \qquad \mu\varepsilon \;\; x \,\epsilon\, \mathbb{R}$$

Οι B-splines αποτελούν λείες συναρτήσεις με ισχυρή τοπικότητα, προσέγγιση δηλαδή σε τοπικό επίπεδο, γεγονός που τις καθιστά κατάλληλες για χρήση ως βάσεις σε προβλήματα προσεγγιστικής μοντελοποίησης [6]. Μέσω αυτού, η τροποποίηση ενός συντελεστή c_i επηρεάζει τη συνάρτηση μόνο σε μια πεπερασμένη περιοχή γύρω από τον αντίστοιχο κόμβο και χωρίς να επηρεάζεται ολόκληρη η έξοδος από μικρές αλλαγές στα δεδομένα. Στο σχήμα αναπαρίσταται η δομή μίας B Spline συνάρτησης, όπου ξ θεωρούνται οι κόμβοι, κ ο βαθμός και b οι καμπύλες $B_n(x)$ που περιγράφουμε παραπάνω. Τέλος, το S είναι το τελικό άθροισμα $s(x)$ που αναφέρεται παραπάνω.



Σχήμα 2.2 Αναπαράσταση των B-Spline Συναρτήσεων

Η χρήση των B-splines στα KAN ευθυγραμμίζεται και θεωρητικά με το θεώρημα Kolmogorov–Arnold, ενώ πρακτικά παρέχει τη δυνατότητα κατασκευής εύκαμπτων και ερμηνεύσιμων μοντέλων. Σύμφωνα με τους Liu et al. [2], οι spline-based βάσεις συνδυάζουν την εκφραστικότητα με τη σταθερότητα, επιτρέποντας την εκμάθηση σύνθετων συναρτήσεων με μεγαλύτερη ακρίβεια σε σχέση με παραδοσιακές ενεργοποιήσεις. Η πρωτότυπη υλοποίηση του έργου των Liu et al. είναι η πρώτη πρακτική εφαρμογή των KAN και περιγράφεται στο [38] με την ονομασία pykan.

Παρά τον λίγο καιρό εφαρμογής των KAN στο επίπεδο της μηχανικής μάθησης έχουν προκύψει πληθώρα παραλλαγών του παραδοσιακού KAN με σκοπό να αυξήσουν την υπολογιστική αποδοτικότητα του και την εκφραστικότητα των μοντέλων. Θα αναφέρουμε τις εξής: Το FastKAN αντικαθιστά τις κυβικές B-splines με ακτινικές συναρτήσεις βάσης (Radial Basis Functions, RBFs) για ταχύτερη εξαγωγή συμπερασμάτων [11], τα rKAN και fKAN εισάγουν εκπαιδεύσιμες ορθολογικές και κλασματικά ορθογώνιες βάσεις Jacobi [12, 13], το FourierKAN αντικαθιστά τους συντελεστές spline με μονομεταβλητές συναρτήσεις Fourier [14] .

2.2.2To efficient KAN

Η υλοποίηση που χρησιμοποιήθηκε στη συγκεκριμένη εργασία βασίζεται στο efficient-KAN [3] και είναι μία απλοποιημένη παραλλαγή της πρωτότυπης υλοποίησης του KAN [38]. Αναδιατυπώνει τον υπολογισμό που υλοποιείται στην είσοδο με διαφορετικές B-Spline συναρτήσεις βάσης και τις συνδυάζει γραμμικά σε ένα πίνακα πολλαπλασιασμού, μειώνοντας έτσι δραστικά το υπολογιστικό κόστος. Αντικαθιστά επίσης την L1 κανονικοποίηση στα δεδομένα εισόδου, η οποία χρησιμοποιείται στο αρχικό δίκτυο KAN και απαιτεί μη-γραμμική διαχείριση με μία αντίστοιχη στα βάρη, η οποία είναι συμβατή με τις βάσεις που εισήχθησαν προηγουμένως.

2.2.3 Το Chebyshev KAN

Σε αυτή την εργασία θα ασχοληθούμε με την παραλλαγή Chebyshev KAN [7]. Σε αυτό οι μονομεταβλητές συναρτήσεις ενεργοποίησης στις ακμές παραμετροποιούνται όχι μέσω B-splines αλλά μέσω πολυωνύμων Chebyshev πρώτου είδους. Τα πολυώνυμα Chebyshev είναι μια ακολουθία ορθογωνίων πολυωνύμων και χαρακτηρίζονται ιδιαίτερα σημαντικά

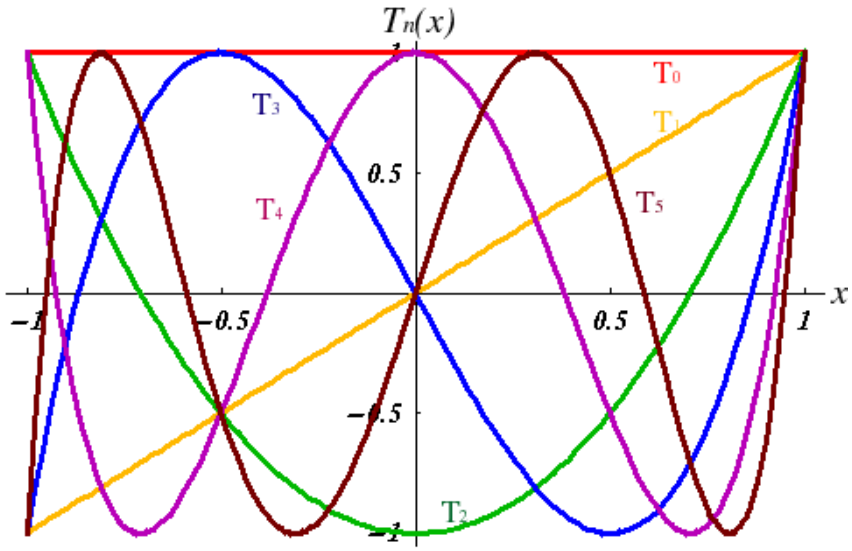
στην θεωρία της προσέγγισης συναρτήσεων λόγω της ελαχιστοποίησης του μέγιστου σφάλματος [8]. Χωρίζονται σε πολυώνυμα πρώτου και δεύτερου είδους με τα πρώτα $T_n(x)$ να έχουν την ιδιότητα να ελαχιστοποιούν τη μέγιστη απόκλιση από το μηδέν μεταξύ όλων των πολυωνύμων του ίδιου βαθμού με κύριο συντελεστή (coefficient) 1, γεγονός που τα καθιστά χρήσιμα στην προσέγγιση πολυωνύμων (approximation) [7].

Τα Chebyshev πολυώνυμα πρώτου είδους ορίζονται ως:

$$T_n(x) = \cos\left(n \arccos(x)\right)$$

Και τα Chebyshev πολυώνυμα δεύτερου είδους ορίζονται αντίστοιχα ως:

$$U_n(x) = \frac{\sin\left((n+1)\arccos(x)\right)}{\sqrt{1-x^2}}$$



Σχήμα 5.3: Αναπαράσταση των Chebyshev πολυωνύμων πρώτου είδους

Η χρήση των $T_n(x)$ προσφέρει αριθμητική σταθερότητα και αποτελεσματική αναπαράσταση λόγω της ορθογωνικότητας και της ταχείας σύγκλισης τους. Καθώς οι είσοδοι των KANs κανονικοποιούνται στο [−1,1], τα πολυώνυμα Chebyshev μπορούν να χρησιμοποιηθούν ως βάσεις για την παραμετροποίηση των ακμών του δικτύου. Η υλοποίηση τους ως Chebyshev KAN παίρνει σαν είσοδο τένσορες x με

$$x \in \mathbb{R}^{batch\ size \times input\ dim}$$

και για κάθε είσοδο υπολογίζει τις τιμές όλων των πολυωνύμων Chebyshev. Αυτό καταλήγει σε τένσορες

$$T \in \mathbb{R}^{batch\ size \times input\ dim \times (degree+1)}$$

όπου degree ο βαθμός του αντίστοιχου πολυωνύμου που έχει οριστεί. Κάθε συνάρτηση ενεργοποίησης του σχηματοποιείται τελικά ως:

$$\varphi(x) = \sum_{k=0}^{d} \theta_k T_k(\hat{x})$$

όπου $\hat{x}$ είναι η κανονικοποιημένη είσοδος στο διάστημα [−1,1], και θ_k είναι οι εκπαιδεύσιμοι συντελεστές.

Για δίκτυα με πολλαπλά επίπεδα, η υπολογιστική διαδικασία προχωρά αναδρομικά, με χρήση της συνάρτησης ενεργοποίησης τύπου tanh πριν την εφαρμογή των πολυωνύμων και με κανονικοποίηση στο τέλος κάθε επιπέδου για αποφυγή εξαφάνισης των παραγώγων (gradient vanishing). Η συνολική έξοδος ενός Chebyshev KAN με L επίπεδα δίνεται ως:

$$\text{ChebyKAN(x)} = (\Phi_{L-1} \circ \Phi_{L-2} \circ \cdots \circ \Phi_0)(\text{x})$$

όπου κάθε Φ_l είναι πίνακας από μονομεταβλητές συναρτήσεις Chebyshev για το αντίστοιχο επίπεδο l. Το κάθε επίπεδο υπολογίζεται ως:

$$x^{(l+1)} = \Phi_l\left(\tilde{x}^{(l)}\right) \quad με \quad \tilde{x}^{(l)} = \tanh\left(x^{(l)}\right)$$

Η χρήση των πολυωνύμων Chebyshev καθιστά το μοντέλο συμπαγές και ενισχύει την ερμηνευσιμότητα των συναρτήσεων ενεργοποίησης [10].

2.2.4 Το Hybrid KAN και Mixture of KANs (MoK)

Βάση της ανάλυσης που ακολούθησε για τα B – spline based KAN και τα Chebyshev based KAN προκύπτει από διάφορες εργασίες [15] αλλά και από την συγκεκριμένη, η ιδέα του συνδυασμού διαφορετικών αρχιτεκτονικών νευρωνικών δικτύων στο ίδιο μοντέλο. Ο σκοπός αυτού έγκειται στην δυνατότητα τα υβριδικά μοντέλα να αποδώσουν εκθετικά στα πεδία που οι διάφορες αρχιτεκτονικές μπορεί να υπερισχύουν. Οι Sainath Dey et al. αναφέρουν στο έργο τους Hyb-KAN ViT ότι η συνεργατική προσέγγιση αντισταθμίζει τους υπολογιστικούς περιορισμούς και επιδιώκει να ενισχύσει τη χώρο-συχνότητα μοντελοποίησης, ενώ μετριάζει τα υπολογιστικά bottlenecks. Προτείνουν την εφαρμογή του ενός δικτύου στον κωδικοποιητή (Encoder) για να αποσυνθέσει τις εισόδους σε συνιστώσες πολλαπλής κλίμακας και του άλλου στην κεφαλή (Head) για να αποστάζει αυτές τις αναπαραστάσεις σε βελτιωμένα χαρακτηριστικά για τελικές προβλέψεις.

Εμείς στην συγκεκριμένη εργασία υλοποιούμε αρχικά την σειριακή σύνδεση στρωμάτων Spline και Chebyshev KAN μέσω της δημιουργίας του μοντέλου Hybrid KAN και έπειτα την εφαρμογή της (MoK) προσέγγισης για την πρόβλεψη πολυμεταβλητών χρονοσειρών. Το Υβριδικό μας μοντέλο δημιουργεί ένα πολυεπίπεδο δίκτυο εφαρμόζοντας ποικίλους συνδυασμούς Spline και Chebyshev kan με εναλλαγές επιπέδων και υπερπαραμέτρων.

Το (MoK) που αναφέρθηκε παραπάνω είναι μία πρόταση την οποία προσεγγίζουμε θεωρητικά και υπολογιστικά και ορίστηκε ως Multi-layer Mixture-of-KAN Network (MMK) από τους X. Han et al [16]. Βάση αυτής της ιδέας δημιουργήθηκε το MoK layer, το οποίο είναι ένας συνδυασμός των Kolmogorov-Arnold Networks (KAN) και της αρχιτεκτονικής "mixture-of-experts" (MoE) και εφαρμόζεται για την πρόβλεψη των πολυμεταβλητών χρονοσειρών. Σχεδιάστηκε για να αντιμετωπίσει την μη-στασιμότητα και τις σημαντικές αποκλίσεις στην κατανομή μεταξύ των μεταβλητών που συχνά εμφανίζονται σε δεδομένα χρονοσειρών. Λόγω του ότι τα KAN έχουν πολλές παραλλαγές με διαφορετικές συναρτήσεις βάσης (όπως spline, wavelet, Taylor, Jacobi), το MoK προσπαθεί να συνδυάσει αυτά τα διαφορετικά KAN σε ένα ενιαίο επίπεδο και να τις προγραμματίσει προσαρμοστικά ανάλογα με τα δεδομένα εισόδου.
Ο συνδυασμός και η εναλλαγή KAN αρχιτεκτονικών είτε σε διαφορετικά επίπεδα, είτε σε ίδια επίπεδα, ή μέσω του κωδικοποιητή και της κεφαλής φαίνεται ότι είναι μια τεχνική που δοκιμάζεται και δίνει ελπιδοφόρα αποτελέσματα. Παρόλα αυτά η κλιμάκωση σε μεγαλύτερα μοντέλα αποκαλύπτει κρίσιμους περιορισμούς. Η αυξημένη πολυπλοκότητα των μοντέλων, η διαχείριση ετερογενών βάσεων και η έλλειψη σταθερότητας κατά την εκπαίδευση είναι από τους περιορισμούς που συναντάμε σε αυτά τα μοντέλα ειδικά καθώς τα μοντέλα μεγαλώνουν.

2.3 Εφαρμογές των KAN στις Χρονοσειρών

2.3.1 Χρήση KAN σε Πρόβλεψη Χρονοσειρών

Η πρόβλεψη χρονοσειρών είναι μία διαδικασία χρήσης δεδομένων με χρονική ακολουθία για την πρόβλεψη μελλοντικών τιμών ή καταστάσεων μεταβλητών. Είναι από

τα κρίσιμα πεδία που βρίσκουν εφαρμογή τα υπολογιστικά μοντέλα αφού η μελέτη τέτοιων καταστάσεων αποδεικνύεται να είναι επιτακτικής σημασίας σε διάφορους κλάδους της καθημερινής ζωής (ιατρική, καιρικά φαινόμενα, χρηματοοικονομικά κλπ). Η τεχνολογία της πρόβλεψης Χρονοσειρών (TSF) περιλαμβάνει τόσο παραδοσιακές στατιστικές τεχνικές όπως το ARIMA, το Facebook Prophet και η Holt-Winters όσο και τεχνικές από την κοινότητα της μηχανικής μάθησης με χαρακτηριστικές αυτής να είναι τα μοντέλα που βασίζονται σε Transformers, MLPs, RNN (πχ. LSTM), CNN κλπ. Τα μοντέλα μηχανικής μάθησης έχουν επηρεάσει αυτό το πεδίο λόγω της ικανότητας τους να χειρίζονται δεδομένα μεγάλου όγκου και να αποτυπώνουν πολύπλοκες μη-γραμμικές σχέσεις, στις οποίες τα στατιστικά μοντέλα υστερούν [17].

Οι προκλήσεις που εγκυμονεί παρόλα αυτά το συγκεκριμένο πεδίο σχετίζονται κυρίως με την εποχικότητα που παρουσιάζουν τα μοντέλα, την δυσκολία να εξασφαλίζεται η συνεκτικότητα στην ακρίβεια στο πέρας των δεδομένων αλλά και η έλλειψη ερμηνευσιμότητας που οφείλεται στην ασαφή σχέση μεταξύ του μεγέθους του δικτύου και της δυνατότητας προσαρμογής στα δεδομένα [18].

Τα KAN αποτελούν μία καινοτόμο εναλλακτική των παραπάνω μοντέλων, με σκοπό να αντιμετωπίσουν τις προκλήσεις λόγω της προσεγγιστικής τους ικανότητας. Η ιδιότητα τους να αναπαριστούν τις πολυμεταβλητές συναρτήσεις ως συνδυασμό μονομεταβλητών συνεχών συναρτήσεων μπορεί να επιδράσει καταλυτικά στην σχέση του δικτύου με τα δεδομένα εισόδου. Φαίνεται επίσης ότι μπορούν να προσφέρουν υψηλότερη ακρίβεια και ερμηνευσιμότητα σε σχέση με εναλλακτικά μοντέλα MLP, ακόμη και με μικρότερες αρχιτεκτονικές. Επιπλέον, η προσαρμοστικότητά τους στις περιοδικότητες και στις τάσεις που χαρακτηρίζουν τις χρονοσειρές οδηγεί σε πιο ομαλή και αποτελεσματική ενσωμάτωση στη δομή του δικτύου [17].

Με βάση αυτή την ανάλυση, έχουν αναπτυχθεί και οι διάφορες υλοποιήσεις των KAN που αναπτύξαμε, όπως το MoK [16], το RMoK [18] αλλά και το HIPPO-KAN [20], αλλά και περαιτέρω τεχνικές γύρω από το ζήτημα. Στην συγκεκριμένη εργασία θα μελετήσουμε την πρόβλεψη σε δεδομένα χρονοσειρών των μοντέλων Efficient Spline KAN, Chebyshev KAN και τον συνδυασμό τους σε υβριδικό και MoK μοντέλο.

2.3.2 Χρήση KAN σε Κατηγοριοποίηση Χρονοσειρών

Αντίστοιχα με την πρόβλεψη, η κατηγοριοποίηση χρονοσειρών είναι ένας άλλος τομέας που περιλαμβάνει μείζονα ζητήματα της καθημερινής ζωής όπως η διάγνωση στην υγεία, τα οικονομικά, η ανθρώπινη δραστηριότητα σε σχέση με τα ερεθίσματα που δίνει κλπ [20]. Σε αντίθεση όμως με την πρόβλεψη στοχεύει στην ακριβή κατανόηση των χρονικών ακολουθιών για να κατατάξει τα δεδομένα σε κατηγορίες δίνοντας τους μία ετικέτα που αντιστοιχεί σε μία κατηγορία. Φαίνεται ότι οι προκλήσεις που αντιμετωπίζονται για το συγκεκριμένο πεδίο βρίσκονται κυρίως στο μεγάλο υπολογιστικό χρόνο και κόστος της εκπαίδευσης που απαιτεί αλλά και του φαινομένου vanishing gradients που πλήττει τα MLPs [21].

Τα KAN παρά τις λιγοστές μελέτες στο αντικείμενο εισάγονται στην κατηγοριοποίηση κυρίως για να ανταγωνιστούν τα MLPs. Η εφαρμογή τους στις μελέτες [20,21] μας αναφέρουν τα εξής χαρακτηριστικά που παρουσιάζουν.

Παρουσιάζουν πιο σαφή ερμηνεία αφού το μοντέλο κατά την διάρκεια της εκπαίδευσης απλοποιείται και έτσι παρέχει καθαρή οπτικοποίηση αλλά και επιτυγχάνουν πιο ακριβή και σταθερά αποτελέσματα με λιγότερους παραμέτρους και σε μικρότερο χρόνο. Συγκεκριμένα το Efficient KAN [20] και το Chebyshev KAN [21] με τα οποία ασχολούμαστε φαίνεται να ανταγωνίζονται ισότιμα με τα MLPs, ειδικότερα στα μικρότερα μοντέλα και να δίνουν πολύτιμη τροφοδότηση στους ερευνητές.

Βάση των δυνατοτήτων που εγείρονται, παρακάτω θα εξετάσουμε την εφαρμογή των Efficient Spline KAN, Chebyshev KAN και τον συνδυασμό τους σε υβριδικό μοντέλο σε πληθώρα δεδομένων κατάλληλα για ταξινόμηση.

Κεφάλαιο 3 Χρονοσειρές και Στατιστική Ανάλυση

Οι χρονοσειρές ορίζονται ως μια σειρά παρατηρήσεων που λαμβάνονται διαδοχικά στον χρόνο με τα δεδομένα να συλλέγονται σειριακά στην διάρκεια του χρόνου και χωρίς να προκαθορίζεται κάποιο χρονικό όριο για να θεωρηθεί ως τέτοια. Χωρίζονται στην κατηγορία των μονομεταβλητών, δηλαδή σε παρατηρήσεις από ένα συγκεκριμένο φαινόμενο και των πολυμεταβλητών, οι οποίες περιέχουν πάνω από μία συνειστώσα, με σκοπό να εξετάζεται η κοινή συμπεριφορά τους. Είναι μία κλασσική μορφή συγκέντρωσης των δεδομένων, η οποία χρησιμοποιείται για στατιστική ανάλυση, πρόβλεψη, ανίχνευση αλλά και προσέγγιση.

Οι χρονοσειρές είναι συνήθως σύνθετες και δύσκολα ερμηνεύσιμες, γεγονός το οποίο έχει οδηγήσει στην ανάγκη να μετασχηματίζονται σε πιο συμπαγή και ερμηνεύσιμα σύνολα για να κατανοείται η συμπεριφορά τους [25]. Βάση αυτού έχουν οριστεί κατηγορίες, ή αλλιώς χαρακτηριστικά σε σχέση με τον χρόνο αυτοσυσχέτισης, την επιρροή από ακραίες τιμές, την ανίχνευση μη γραμμικών φαινομένων, την κατανομή, την συχνότητα κλπ. Η τεχνική που ορίστηκε γύρω από αυτά τα στοιχεία ονομάζεται εξαγωγή χαρακτηριστικών και στοχεύει στο εξής: Την συστηματική σύγκριση μεταξύ των χιλιάδων χαρακτηριστικών, με σκοπό την διάκριση τους, την κατανόηση των ιδιοτήτων τους και εν τέλει την ταξινόμηση τους [24].

Οι Lubba et al. αναφέρουν ότι έχουν οριστεί πάνω από 7500 χαρακτηριστικά τα οποία μπορεί να περιγράψουν μία χρονοσειρά ενώ έχουν δημιουργηθεί ειδικά εργαλεία όπως το hctsa [22] για matlab , το catch22 [24] για C, matlab και python και το tsfresh [23] για python με σκοπό να διευκολύνεται η έρευνα σε σχέση με τον τρόπο αναπαράστασης των χρονοσειρών. Αυτά σχετίζονται κυρίαρχα με τις στατιστιστικές ιδιότητες, την αυτοσυσχέτιση, την στασιμότητα και το σφάλμα, την εντροπία και την γραμμικότητα [22]. Αν προσπαθούσαμε να περιγράψουμε τις θεμελιώδεις ιδιότητες που σκοπεύουν να αποτυπώσουν αυτά τα χαρακτηριστικά, τότε θα καταγράφαμε τις έννοιες της τάσης, της εποχικότητας και της ακανόνιστης συμπεριφοράς (θόρυβος) ως τις τρείς ισχύουσες .

Πιο αναλυτικά πρόκειται για τις:

Τάση (Trend): Περιγράφει τα μοτίβα που δημιουργούν οι χρονοσειρές και μπορεί να είναι θετική ή αρνητική ανάλογα με το πως εξελίσσονται τα δεδομένα στην πάροδο του χρόνου.

Εποχικότητα (Seasonality): Δημιουργείται στις περιπτώσεις που παρουσιάζεται μια περιοδική εναλλαγή των δεδομένων με σταθερό τρόπο.

Υπόλοιπο ή ακανόνιστη συνιστώσα (Remainder(R)/Irregularity(I)): Είναι οι ακανόνιστες διακυμάνσεις που μπορεί να συμβαίνουν λόγω απρόβλεπτων παραγόντων και η έλλειψη δημιουργίας μοτίβων. Λέγεται και λευκός θόρυβος ή σφάλμα.

Αυτές οι τρεις ιδιότητες έχουν την ικανότητα να αποσυνθέσουν τα δεδομένα μίας χρονοσειράς και να τα περιγράψουν ως:

$$y_t = S_t + T_t + R_t$$

Όπου y_t τα δεδομένα, S_t η εποχικότητα, T_t η τάση και R_t το σφάλμα, όλα σε σχέση με τον χρόνο t [29].

3.1 Στατιστική Ανάλυση Δεδομένων

Αυτή η υπό ενότητα στοχεύει στο να κατανοήσουμε τον τρόπο που αναλύονται στατιστικά οι χρονοσειρές σε σχέση με την εφαρμογή που εξετάζονται, για την οποία και αξιολογούνται. Μέσα από αυτό θα επιδιώξουμε να αξιολογήσουμε και τα αποτελέσματα από τα μοντέλα μας, για να παρέχουμε αξιόπιστα αποτελέσματα και δίκαιες συγκρίσεις με άλλες μεθόδους που επικρατούν. Η ανάλυση χρονοσειρών είναι η στατιστική τεχνική που αναλύει τα δεδομένα που συλλέγονται στην πάροδο του χρόνου, με σκοπό να αναγνωριστούν τα μοτίβα σε σχέση με τις ιδιότητες που αναφέραμε. Βοηθάει να κατανοήσουμε τον τρόπο με τον οποίο οι μεταβλητές μετασχηματίζονται, ώστε να εξάγουμε πληροφορίες και να προβλέψουμε την μελλοντική τους συμπεριφορά.

Μια βασική ιδιότητα των χρονοσειρών είναι η στασιμότητα, η οποία μπορεί να αναλυθεί μέσω των ιδιοτήτων του μέσου όρου (μ), της διακύμανσης (sd) και την συνάρτησης αυτοσυσχέτισης (ACF) [26]. Η συνάρτηση αυτοσυσχέτισης μετρά την συν διακύμανση μεταξύ δύο τιμών zt και zt+k με χρονική διαφορά k μεταξύ τους και παράγει το διάγραμμα autocorrelation plot, που είναι βασικό στάδιο για την ανάλυση των χρονοσειρών. Το συγκεκριμένο χρησιμοποιείται για τον εντοπισμό των τάσεων της

χρονοσειράς και συγκεκριμένα της εποχικότητας καθώς και για την εύρεση της καταλληλότητας μοντέλων όπως το ARIMA.

Η οπτικοποίηση είναι θεμελιώδες βήμα στην ανάλυση χρονοσειρών και μαζί με το ACF Plot θα ξεχωρίσουμε τα διαγράμματα απλής αναπαράστασης (time-series plot) που εμφανίζουν τις τάσεις μίας σειράς σε σχέση με τον χρόνο, της μερικής αυτοσυσχέτισης (PACF), της σκέδασης (Scatter) που τονίζει την σχέση μεταξύ των μεταβλητών, τα εποχικά που εξετάσουν περιπτώσεις κόντρα στην τάση και αυτό της φασματικής πυκνότητας (Spectral Density) που εντοπίζει την εποχική συχνότητα μετά από μετασχηματισμούς Fourier [26],[27],[29]. Μετά την κατανόηση της συμπεριφοράς των χρονοσειρών μέσα από την ερμηνεία των ιδιοτήτων της και την οπτικοποίηση μέσα από τα διάφορα διαγράμματα η χρονοσειρά αναλύεται από τα αντίστοιχα μοντέλα.

Τα μοντέλα που παρέχουν ανάλυση χρονοσειρών χωρίζονται στα στατιστικά μοντέλα, που περιέχουν το ARIMA, το Holt-Winters αλλά και μοντέλα παλινδρόμησης όπως το Prophet και αυτά που βασίζονται στην μηχανική μάθηση μεταξύ των οποίων τα LSTM, RNN/GRU, k-NN, MLP και οι Transformers και θα εξεταστούν στην επόμενη υπό ενότητα. Οι δύο βασικές κατηγορίες που επικεντρωνόμαστε είναι η πρόβλεψη και η κατηγοριοποίηση, όπου εξετάζουμε την στατιστική σημαντικότητα των δεδομένων με αντίστοιχα στατιστικά τεστ, παρουσιάζουμε μοντέλα που επικρατούν στο κάθε κομμάτι και παρέχουμε κατάλληλες μετρικές αξιολόγησης.

Για την πρόβλεψη [27] θα εξετάσουμε κάποιες από τις κλασσικές μεθόδους στατιστικής ανάλυσης:

ADF Test (Augmented Dickey-Fuller): Είναι ένα στατιστικό τεστ που αποκαλείται και τεστ μοναδιαίας ρίζας (unit root test) και καθορίζει το αν μία χρονοσειρά είναι στάσιμη, αν δηλαδή εξαρτάται από τον χρόνο στον οποίο παρατηρείται ή όχι [29]. Σε περίπτωση που βρεθεί μη-στάσιμη, πρέπει να διαφοροποιηθεί πριν από περαιτέρω ανάλυση με στατιστικά μοντέλα ή εκπαίδευση με μοντέλα μηχανικής μάθησης. Το όριο που καθορίζει το κριτήριο στασιμότητας συνήθως είναι το $p < 0.05$ και ονομάζεται significance level.

AutoRegressive Integrated Moving Average (ARIMA): Είναι ενδεχομένως η πιο γνωστή στατιστική τεχνική και χρησιμοποιείται ευρέως για την πρόβλεψη. Είναι γραμμικό μη

στατικό μοντέλο [26] όπως σημειώνουν οι George EP Box el al. και συνδυάζει την αυτόματη παλινδρόμηση (auto-regression), την ολοκλήρωση (integration) και τους κινητούς μέσους όρους (moving average). Η επέκταση του είναι το SARIMA και SARIMAX όπου προστίθεται η παράμετρος της εποχικότητας (seasonality) και στο SARIMAX των εξωγενών συντελεστών με σκοπό να μοντελοποιούνται πιο εύκολα τέτοια μοτίβα.

Prophet: Είναι μια open source βιβλιοθήκη για μονομεταβλητές χρονοσειρές και ορίζεται ως μία μη παραμετρική μέθοδος παλινδρόμησης στον χώρο της μηχανικής μάθησης. Μπορεί να αναλυθεί και ως

$$\mathrm{y(t)} = \mathrm{g(t)} + \mathrm{s(t)} + \mathrm{h(t)} + e_t$$

με g την τάση, s την εποχικότητα, h την ακανόνιστη συμπεριφορά και e το σφάλμα.
Παρόλο που τα παραπάνω είναι στατιστικά μοντέλα μερικές φορές κατατάσσονται στην κοινότητα της μηχανικής μάθησης. Αυτό συμβαίνει χάρη στο γεγονός ότι είναι μοντέλο παλινδρόμησης, παρόλο που δεν λειτουργούν μέσω εκμάθησης δεδομένων όπως τα ML μοντέλα, και για αυτό ανήκουν στα μοντέλα στατιστικής ανάλυσης.

Για την κατηγοριοποίηση μελετάμε τα προτεινόμενα στατιστικά τεστ [28] και μοντέλα που παρέχουν αξιόπιστες αξιολογήσεις:

Τα τεστ σημαντικότητας (significance tests) είναι εργαλεία για την σύγκριση των δεδομένων μας με την μηδενική υπόθεση, η οποία υποστηρίζει ότι δεν υπάρχει συσχέτιση μεταξύ δύο ομάδων ή μεταβλητών και εξετάζεται αν η συγκεκριμένη συνθήκη αληθεύει [30]. Το αποτέλεσμα είναι μία τιμή p, η οποία εκφράζεται σε ποσοστό και δηλώνει το κατά πόσο τα δεδομένα συμμορφώνονται με την υπόθεση της συσχέτισης. Αποτελούν την βάση της στατιστικής ανάλυσης και προσφέρουν ερμηνεία για το σύνολο των δεδομένων με σκοπό να οριστούν αντίστοιχα τα κατάλληλα μοντέλα για την ανάλυση τους. Ας παρουσιάσουμε μερικά από αυτά:

Paired T-test (p<0.05) και Wilcoxon ranged test (α<0.05): Είναι στατιστικά τεστ σημαντικότητας και συνήθως χρησιμοποιούνται για την σύγκριση δύο κατηγοροποιητών και άρα την ανάλυση των δεδομένων που προκύπτουν σε σχέση με την τιμή (p/a<0.05) που κρίνει αν διαπιστώνεται στατιστικά σημαντική διαφορά. Το paired t-test

χρησιμοποιείται όταν τα δεδομένα έχουν κανονική κατανομή ενώ το Wilcoxon είναι μία μη-παραμετρική τεχνική και προτιμάται στα δεδομένα χωρίς κανονική κατανομή. Το μέτρο σύγκρισης τους θεωρείται η διάμεσος ενός συνόλου μετρήσεων ακρίβειας [28].

Friedmann test και Nemenyi post-hoc test: Το πρώτο είναι το εναλλακτικό τεστ που χρησιμοποιείται όταν απαιτείται σύγκριση περισσότερων από δύο μεθόδων κατηγοριοποίησης [28] και διακρίνει αν εμφανίζεται στατιστικά άξια διαφοροποίηση μεταξύ των πολλαπλών κατηγοριοποιητών. Το δεύτερο εφαρμόζεται στην περίπτωση που το Friedmann εντοπίσει γενική διαφορά και έχει σκοπό να παρουσιάσει τα συγκεκριμένα ζεύγη που το προκαλούν αυτό [9].

3.2 Οι χρονοσειρές στην Μηχανική Μάθηση

Τα μοντέλα της βαθιάς και ευρύτερα της μηχανικής μάθησης χαρακτηρίζονται από πολλούς ως μοντέλα μαύρων κουτιών (black-box models) αφού αρχικά εισήγαγαν τεχνικές οι οποίες φαινομενικά ήταν πιο δύσκολο να ερμηνευτούν. Θεωρούνται πολύπλοκα και συχνά πάσχουν από την κατανόηση των πλευρών και των τεχνικών που φέρνουν τα θετικά αποτελέσματα στο μοντέλο [1]. Η ανάγκη όμως για πιο ερμηνεύσιμα μοντέλα έχει αυξηθεί, ιδίως σε κλάδους όπως η ιατρική, όπου ο παράγοντας λάθους είναι κρίσιμος και οι αποφάσεις που επιλέγονται πρέπει να συνοδεύονται από επαρκή εξήγηση, ώστε να διασφαλίζεται η αξιοπιστία σε αυτά. Σε αυτή την κατεύθυνση έχουν δημιουργηθεί τεχνικές οι οποίες συμβάλουν καθοριστικά στην ερμηνευσιμότητα των μοντέλων. Τέτοιες, είναι η χρήση των μηχανισμών προσοχής, που εστιάζουν σε συγκεκριμένα στοιχεία στην πάροδο του χρόνου και παρέχουν βελτιωμένη εικόνα των τάσεων του μοντέλου ως προς την μεροληψία [31]. Ακόμα, χρήσιμες είναι και οι μέθοδοι σημαντικότητας (saliency methods) και Εξηγήσιμων AI Τεχνικών (XAI) όπως το SHAP και το GRAD [32] και η εξέλιξη τους σε Temporal Saliency Rescaling (TSR), [33] οι οποίες προσφέρουν εξήγηση για τα αποτελέσματα των μοντέλων και τις αποφάσεις που ακολουθούν. Η TSR μπορεί να προσφέρει βελτίωση στην σχέση μεταξύ του χρόνου και των χαρακτηριστικών, με ιδιαίτερη συμβολή και στην κατηγοριοποίηση των πολυμεταβλητών χρονοσειρών.

Τα μοντέλα όπως το LSTM ή το GRU χαρακτηρίζονται ως αναδρομικές αρχιτεκτονικές, είναι σχεδιασμένα να κρατάνε μνήμη και έτσι να υπολογίζουν βάση των μακροχρόνιων μοτίβων και εξαρτήσεων που διακρίνουν, αποφεύγοντας προκλήσεις όπως η εξαφάνιση των κλίσεων (vanishing gradients). [25] Τα CNN και τα MLPs έχουν την ικανότητα να αναγνωρίζουν χρονικά και χωρικά μοτίβα και έτσι να μοντελοποιούν τις διάφορες ιδιότητες που παρουσιάζουν οι χρονοσειρές αλλά εμφανίζουν περιορισμούς, όπως η χαμηλή αποδοτικότητα σε σχέση με την αύξηση των παραμέτρων τους.
Το KAN συγκεκριμένα με το οποίο ασχολούμαστε, προσεγγίζει και μαθαίνει παραμετρικά αυτές τις ιδιότητες μέσω των μη γραμμικών συναρτήσεων βάσης (Spline, Chebyshev κλπ). Για αυτό τον λόγο δεν χρειάζεται να εξάγει συγκεκριμένα χαρακτηριστικά και έτσι θεωρείται μαζί και με τα υπόλοιπα μοντέλα βαθιάς μάθησης πιο σύγχρονη εναλλακτική για την ανάλυση των χρονοσειρών. Θα μελετήσουμε πιο συγκεκριμένα κάποια από τα παραπάνω μοντέλα. Σε αντίθεση με την στατική ανάλυση, τα μοντέλα αυτά δεν διακρίνονται προς πρόβλεψη ή κατηγοριοποίηση, αλλά παρέχουν την δυνατότητα να διαφοροποιούνται σε μετρικές και υπερπαραμέτρους και έτσι να παρέχουν ανάλυση στο εκάστοτε πεδίο.

MLP: Είναι είδος τεχνικού νευρωνικού δικτύου και η βασική εναλλακτική στα KAN. Ενεργοποιείται μέσω ορισμένων συναρτήσεων ενεργοποίησης στους κόμβους και εκπαιδεύσιμων βαρών στις ακμές, παρέχοντας την δυνατότητα να δημιουργούν βαθιά μοντέλα μέσω της διαδοχικής τοποθέτησης επιπέδων.

LSTM: Θεωρείται επαναληπτικό νευρωνικό δίκτυο (RNN) και υπερέχει στο να χειρίζεται διαδοχικά δεδομένα και να μαθαίνει μακροχρόνιες εξαρτήσεις, το οποίο το καθιστά και ιδανικό για την ανάλυση χρονοσειρών. Υπερέχει σε σχέση με τα παραδοσιακά RNN λόγω της μακροχρόνιας μνήμης του και θεωρείται μία από τις πιο αξιόπιστες εναλλακτικές για την πρόβλεψη.

Support Vector Machines (SVM): Είναι αλγόριθμος μηχανικής μάθησης ειδικά σχεδιασμένος για να διαχωρίζει τα δεδομένα μέσω μίας βέλτιστης γραμμής, η οποία προκύπτει από την μέγιστη απόσταση μεταξύ των επιθυμητών κλάσεων σε έναν

πολυδιάστατο χώρο. Θεωρείται ισχυρή επιλογή για εφαρμογή σε κατηγοριοποίηση και πρόβλεψη και μπορεί να χειρίζεται γραμμικά και μη γραμμικά προβλήματα [34].

Naive Bayes: Είναι κατηγορία μοντέλων που βασίζονται στην προσέγγιση βάση πιθανοτήτων με οδηγό το θεώρημα Bayes που ορίζει την σχέση μεταξύ της υπό περίπτωση και της γινομένης πιθανότητας. Χρησιμοποιούνται κατά βάση για κατηγοριοποίηση αλλά προσφέρουν εναλλακτική και για την ανάλυση παλινδρόμησης [35].

1-NN DTW: Το Dynamic Time Warping (DTW) είναι μία μη-παραμετρική προσέγγιση βασισμένη στην απόσταση, συχνά αναφερόμενη και ως η εναλλακτική στην ευκλείδεια απόσταση. Υπολογίζει την απόσταση δύο χρονοσειρών μέσω της ομοιότητας που παρουσιάζουν και μετράει το βέλτιστο υπολογιστικό κόστος. Συχνά, όπως και στην περίπτωση που μελετάμε, χρησιμοποιείται συνεργατικά με τον k-nearest neighbor (k-nn) αλγόριθμο, προσφέροντας έναν ισχυρό αλγόριθμο για ταξινόμηση που χρησιμοποιείται και ως βάση σύγκρισης με πιο πολύπλοκες προσεγγίσεις [36].

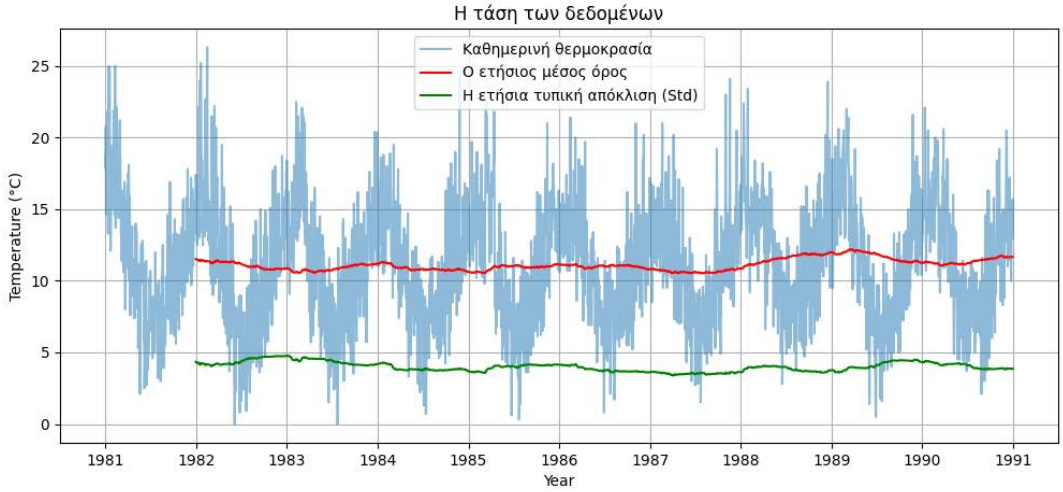
3.3 Περιγραφή Δεδομένων για Πρόβλεψη

Για την πρόβλεψη χρησιμοποιήσαμε ένα σετ δεδομένων μονοδιάστατων χρονοσειρών που αναπαριστά την καθημερινή ελάχιστη θερμοκρασία στην Μελβούρνη σε βαθμούς κελσίου στην πάροδο δέκα χρόνων μετρημένες από το 1981 έως το 1990 [37]. Η μία στήλη περιέχει την ημερομηνία στην μορφή YYYY-MM-DD και η άλλη την ελάχιστη θερμοκρασία.

Date	Daily minimum temperatures
1981-01-01	20.7
1981-01-02	17.9
1981-01-03	18.8
1981-01-04	14.6
1981-01-05	15.8
1981-01-06	15.8
1981-01-07	15.8
1981-01-08	17.4
1981-01-09	21.8
1981-01-10	20.0

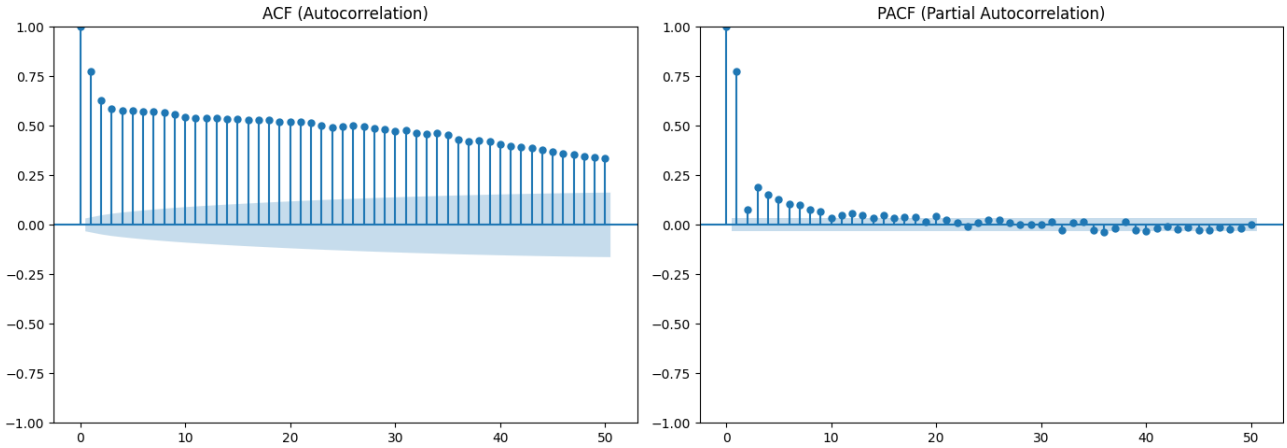
Πίνακας 3.1: Τα πρώτα 10 δείγματα των δεδομένων πρόβλεψης

Τα δεδομένα παρουσιάζουν ισχυρή εποχικότητα λόγω της μεγάλης μεταβολής της θερμοκρασίας στην πάροδο των εποχών και δεν παρουσιάζει ιδιαίτερο θόρυβο που θα προσέδιδε δυσκολία στην ερμηνεία. Η τάση και η εποχικότητα αναπαρίστανται παρακάτω στο σχήμα 3.1 όπου παρουσιάζονται οι κύκλοι εποχών σε σταθερά επαναλαμβανόμενα μοτίβα. Ο σταθερός μέσος και η σταθερή τυπική απόκλιση επιβεβαιώνουν την περιοδικότητα.



Σχήμα 3.1: Καθημερινή ελάχιστη θερμοκρασία (1981-1990) με αναπαράσταση της τάσης και της εποχικότητας

Στα δεδομένα εφαρμόστηκαν επίσης και η συνάρτηση αυτοσυσχέτισης με αποτέλεσμα τα διαγράμματα ACF και PACF plot που εμφανίζονται στο σχήμα 4.3. Σε αυτά παρατηρείται ισχυρή θετική αυτοσυσχέτιση στις πρώτες καθυστερήσεις καθώς και επαναλαμβανόμενες κορυφές που επιβεβαιώνουν την εποχικότητα. Ο συνδυασμός των δύο διαγραμμάτων μας προτρέπει στην χρήση μοντέλων ARIMA/SARIMA το οποίο θα μελετηθεί στο επόμενο κεφάλαιο μετά την εφαρμογή του ADF τεστ.



Σχήμα 3.2: Διαγράμματα αυτοσυσχέτισης ACF και PACF

Τα σύνολο δεδομένων έχει χρησιμοποιηθεί πολλαπλώς από ερευνητές για την πρόβλεψη, γεγονός που συνιστά και έναν από τους λόγους που επιλέχθηκε για την παρούσα εργασία. Τα δεδομένα ακολούθησαν προ επεξεργασία με κανονικοποίηση και δημιουργία ακολουθιών που θα περιγραφούν αναλυτικά στο επόμενο κεφάλαιο.

3.4 Περιγραφή Δεδομένων για Κατηγοριοποίηση

Για την ταξινόμηση χρησιμοποιήσαμε το πιο ευρέως διαδεδομένο σύνολο δεδομένων, το UCR Time Series Archive. Ξεκίνησε την χρήση του το 2002 από τον Eamonn Keogh και την ομάδα του, με 16 σύνολα χρονοσειρών και με την τελευταία επέκταση του 2018 μετράει 128 σύνολα από προβλήματα όπως την βιολογία, την ανθρώπινη δραστηριότητα, τα χρηματοοικονομικά και άλλα. Αυτά αποτελούνται από μονοδιάστατες χρονοσειρές σταθερού μήκους, διαχωρισμένες σε αρχεία TSV, έτοιμα για εκπαίδευση και δοκιμή/επαλήθευση [28]. Ο χρονοσειρές μέσα στο αρχείο διαφέρουν στο πλήθος μεταξύ των dataset, με την πιο μικρή να απαριθμεί 60 δείγματα και την πιο μεγάλη πάνω από 9000. Τα δεδομένα όπως παρέχονται στην έκδοση του 2018 που χρησιμοποιούμε, έχουν υποστεί z-κανονικοποίηση για να εξαλειφθούν οι τάσεις της κλιμάκωσης και να καταστήσουν τις χρονοσειρές συγκρίσιμες.

Μαζί με το σύνολο των μονοδιάστατων χρονοσειρών παρέχεται και μία αναφορά σχετικά με όλα τα dataset. Αυτή αναφέρει τον τύπο των δεδομένων, το σύνολο εκπαίδευσης και δοκιμής, το μήκος των χρονοσειρών τον αριθμό των κλάσεων που καλείται να ταξινομηθεί, τον πάροχο των δεδομένων και τρία αποτελέσματα σφάλματος από δοκιμασμένες τεχνικές που παρέχονται για σύγκριση με τα διάφορα μοντέλα εκπαίδευσης. Αυτά είναι το σφάλμα από την ευκλείδεια απόσταση, την απόσταση DTW με πλάτος που υπολογίζεται μέσω διασταυρούμενης επικύρωσης, την απόσταση DTW χωρίς περιορισμό στο πλάτος (w=100) και ένα τελικό σκορ (default rate) που βασίζεται στην επιλογή του να ταξινομούνται όλα στην πιο συχνή κατηγορία και παρέχεται για την τελική σύγκριση. Παρακάτω παρέχεται ο πίνακας με τα 15 πρώτα στοιχεία της αναφοράς των συνόλων δεδομένων.

Type	Name	Train	Test	Class	Length	ED (w=0)	DTW (learned_w)	DTW (w=100)	Default rate	Data donor/editor
Image	Adiac	390	391	37	176	0.3887	0.3913	0.3964	0.9591	A. Jalba
Image	ArrowHead	36	175	3	251	0.2000	0.2000	0.2971	0.6057	L. Ye & E. Keogh
Spectro	Beef	30	30	5	470	0.3333	0.3333	0.3667	0.8000	K. Kemsley & A. Bagnall
Image	BeetleFly	20	20	2	512	0.2500	0.3000	0.3000	0.5000	J. Hills & A. Bagnall
Image	BirdChicken	20	20	2	512	0.4500	0.3000	0.2500	0.5000	J. Hills & A. Bagnall
Sensor	Car	60	60	4	577	0.2667	0.2333	0.2667	0.6833	J. Gao
Simulated	CBF	30	900	3	128	0.1478	0.0044	0.0033	0.6644	N. Saito
Sensor	Chlorine Concentra	467	3840	3	166	0.3500	0.3500	0.3516	0.4674	L. Li & C. Faloutsos
Sensor	CinCECGTorso	40	1380	4	1639	0.1029	0.0696	0.3493	0.7464	physionet.org
Spectro	Coffee	28	28	2	286	0.0000	0.0000	0.0000	0.4643	K. Kemsley & A. Bagnall
Device	Computers	250	250	2	720	0.4240	0.3800	0.3000	0.5000	J. Lines & A. Bagnall
Motion	CricketX	390	390	12	300	0.4231	0.2282	0.2462	0.8974	A. Mueen & E. Keogh
Motion	CricketY	390	390	12	300	0.4333	0.2410	0.2564	0.9051	A. Mueen & E. Keogh
Motion	CricketZ	390	390	12	300	0.4128	0.2538	0.2462	0.8974	A. Mueen & E. Keogh
Image	DiatomSize Reduce	16	306	4	345	0.0654	0.0654	0.0327	0.6928	ADIAC project

Πίνακας 3.2: Τα πρώτα 10 δείγματα της σύνοψης δεδομένων UCR

Εμείς εκπαιδεύσαμε τα μοντέλα μας σε 40 σύνολα δεδομένων από το UCR αρχείο για να παρέχεται ένα επαρκές σύνολο δεδομένων που να μπορεί να θεωρηθεί πλήρες με βάση τον σκοπό της εργασίας. Ο λόγος που δεν κάναμε χρήση όλων των συνόλων δεδομένων ήταν οι υπολογιστικοί περιορισμοί που αντιμετωπίσαμε καθώς και η κοστοβόρα σε χρόνο και υπολογιστικούς πόρους εκμάθηση της κατηγοριοποίησης.

Κεφάλαιο 4 Πειραματική διαδικασία

4.1 Τα Kolmogorov Arnold Νευρωνικά Δίκτυα

4.1.1 Το Efficient KAN

Στο πλαίσιο των πειραμάτων μας αξιοποιήθηκε το spline μοντέλο efficient KAN που παρουσιάζει αρκετά καλύτερα αποτελέσματα στις πειραματικές μετρήσεις δεκάδων ερευνητών σε σχέση με την αρχική υλοποίηση. Το δίκτυο δομείται με στρώματα KANLinear. Αρχικοποιεί τις παραμέτρους του δικτύου με kaiming uniform και έπειτα ορίζει τις B-spline βάσεις για τους δοσμένους τένσορες εισόδου με χρήση της βιβλιοθήκης pytorch. Έτσι προσαρμόζει τις εισόδους του μοντέλου στις συγκεκριμένες βάσεις με σκοπό να αναπαριστά πολύπλοκα μοτίβα. Αυτές οι spline συναρτήσεις μετατρέπονται σε συντελεστές συσχέτισης (coefficients) και στην έξοδο παράγουν τένσορα σχήματος:

(*out_features, in_features, grid_size + spline_order)*

Κανονικοποιεί τα βάρη των spline και ανανεώνει δυναμικά το πλέγμα grid στην διαδικασία της εκπαίδευσης, με σκοπό να προσφέρει ευελιξία για να προσαρμόζεται στις διάφορες κατανομές δεδομένων. Τέλος υλοποιεί την προαναφερόμενη L1 κανονικοποίηση στα βάρη, η οποία υπολογίζεται ως ο μέσος όρος της απόλυτης τιμής των βαρών spline για να αποφευχθεί ο υπολογισμός απόλυτων τιμών και εντροπίας στο ενδιάμεσο τένσορα που θα αύξανε σημαντικά το κόστος στην μνήμη.

4.1.2 Το Chebyshev KAN

Η υλοποίηση που επιλέχτηκε για την δοκιμή των Chebyshev πολυωνύμων είναι το Chebyshev_kan [39], ως το πειραματικό κομμάτι της εργασίας [7]. Το μοντέλο βρίσκεται σε πλήρη παραλληλία με το spline efficient kan και τις συναρτήσεις που εφαρμόζει και αρχικά σχεδιάστηκε για να εφαρμοστεί στο σύνολο δεδομένων MNIST. Το δίκτυο αποτελείται από στρώματα ChebyshevKANLinear όπου κάθε στρώμα ορίζεται ως ένα άθροισμα ενός γραμμικού μέρους και ενός πολυωνυμικού μέρους. Αρχικά εφαρμόζεται μη γραμμική ενεργοποίηση SiLU στον τένσορα εισόδου, ενώ εισάγεται θόρυβος κατά την αρχικοποίηση των βαρών για μεγαλύτερη εκφραστικότητα του μοντέλου. Στη συνέχεια,

οι είσοδοι κανονικοποιούνται με την υπερβολική εφαπτομένη (tanh) ώστε να περιοριστούν στο διάστημα [−1,1], στο οποίο είναι ορισμένα τα πολυώνυμα Chebyshev και υπολογίζονται αναδρομικά μέσα από την σχέση:

$$Tk(x)=2xTk−1(x)−Tk−2(x)$$

στην συνάρτηση Chebyshev polynomials. Οι βάσεις συνδυάζονται με τα βάρη μέσω του αθροίσματος einsum ώστε να παραχθεί η πολυωνυμική συνεισφορά και τελικά παράγεται η έξοδος y = ybase+ycheb ως το γραμμικό και το πολυωνυμικό άθροισμα. Τέλος, εφαρμόζεται L1 κανονικοποίηση στα πολυωνυμικά βάρη καθώς και ο αντίστοιχος όρος εντροπίας ώστε να διατηρείται η γενίκευση και να αποφεύγεται η υπερπροσαρμογή.

4.1.3 Το Mixture of KAN (MoK)

Ακολουθώντας την ροή του κεφαλαίου 2 θα παρουσιάσουμε το επίπεδο MoK Layer [40] που χρησιμοποιήθηκε για να αναπτύξουμε τα πειράματα μας για στον συνδυασμό των KAN, το οποίο αποτελεί κομμάτι της συνολικής εργασίας “Easy TSF” [16]. Το μοντέλο κάνει χρήση πολλαπλών ειδικών (experts) διαφορετικού τύπου, παράγοντας έξοδο η οποία δεν καθορίζεται από έναν μόνο ειδικό αλλά από τον συνδυασμό τους. Ερευνώνται πέντε διαφορετικοί τύποι ειδικών, ένας γραμμικού, ένας Spline KAN, ενός Chebyshev KAN και δύο συνδυαστικοί με εναλλαγές Spline και Chebyshev υπολογισμών. Ο συνδυασμός των διάφορων τύπων καθορίζεται μέσω μίας πύλης η οποία μαθαίνει να κατανέμει πιθανότητες στους ειδικούς για κάθε είσοδο, έτσι ώστε το δίκτυο να προσαρμόζεται δυναμικά στα δεδομένα ανάλογα με την αποτελεσματικότητα των ειδικών. Η πύλη υλοποιείται ως ένα υπολογιστικό στρώμα, που στην προκειμένη επιλέγεται μεταξύ γραμμικού ή KANLinear. Ακολουθεί softmax κανονικοποίηση των πιθανοτήτων για την στάθμιση της εντροπίας και για να σχηματιστεί η κατανομή τους. Τέλος, παράγεται έξοδος σχήματος [B, out_features] ως το σταθμισμένο einsum άθροισμα των επιμέρους εξόδων.

4.1.4 Το Hybrid KAN

Το συγκεκριμένο μοντέλο δημιουργήθηκε με σκοπό να μελετηθεί η συμβολή του συνδυασμού Spline και Chebyshev KAN βάσεων στην ίδια αρχιτεκτονική αλλά σε διαφορετικά στρώματα. Το δίκτυο δομείται από στρώματα KANLinear, Chebyshev

KANLinear και Linear σύμφωνα με την αρχιτεκτονική που ορίζεται εξαρχής. Τα Spline στρώματα προσφέρουν στην τοπική εκφραστικότητα του μοντέλου ενώ τα Chebyshev στρώματα προσδίδουν στην παγκόσμια προσεγγιστική ικανότητα των Chebyshev με σκοπό η συμβολή αυτών των δύο ιδιοτήτων να προσδίδουν βελτίωση στην συνολική αποτελεσματικότητα. Η παρεμβολή των γραμμικών στρωμάτων ενδιάμεσα των KAN μειώνει την πολυπλοκότητα και λειτουργεί ως ουδέτερο στρώμα για να αποφευχθεί η διαστατικότητα του μοντέλου. Μετά από κάθε στρώμα ακολουθεί ένα αντίστοιχο κανονικοποίησης (Layer Normalization) για να σταθεροποιήσει την εκπαίδευση καθώς και ενεργοποίηση SiLU η οποία τείνει να αποδίδει καλύτερα στις χρονοσειρές.

Τέλος ορίζεται η συνάρτηση σφάλματος κανονικοποίησης, στην οποία συλλέγονται οι αντίστοιχοι όροι από το κάθε στρώμα όπως αυτοί περιγράφηκαν νωρίτερα. Η επιλογή αυτή έγινε διότι κάθε επιμέρους KAN στρώμα εφαρμόζει τον όρο κανονικοποίησης με τρόπο προσαρμοσμένο στη δική του λειτουργία, εξασφαλίζοντας ότι το υβριδικό μοντέλο παραμένει πλήρως συμβατό με όλα τα υπόλοιπα μοντέλα που ερευνούμε.

4.2 Pipeline Πρόβλεψης Χρονοσειρών

Στην συγκεκριμένη ενότητα θα παρουσιάσουμε την διαδικασία που ακολουθήθηκε για την πρόβλεψη της θερμοκρασίας με την χρήση των KAN μοντέλων μας και των μοντέλων σύγκρισης στα δεδομένα που παρουσιάστηκαν παραπάνω. Τα μοντέλα μελέτης KAN και τα μοντέλα αναφοράς, εκπαιδεύτηκαν και αξιολογήθηκαν με την ίδια επεξεργασία δεδομένων ώστε να παραχθούν αξιόπιστες συγκρίσεις.

4.2.1 Προ επεξεργασία δεδομένων

Τα ακατέργαστα δεδομένα χρειάζεται να περάσουν μετασχηματισμούς πριν την χρήση τους στα δίκτυα της βαθιάς μάθησης, καθώς έτσι βελτιώνεται σημαντικά η σύγκλιση και η επίδοση του μοντέλου. Θα εναρμονιστούμε με το πείραμα του S. Willson [41], το οποίο προβλέπει θερμοκρασίες με το πρωτότυπο μοντέλο KAN [38] και με χρήση της ίδιας χρονοσειράς. Αρχικά τα δεδομένα χωρίζονται στις εισόδους x και τους στόχους y με την τεχνική του sliding window, την ομαδοποίηση δηλαδή δεδομένων κατά την είσοδο και την σταδιακή χρήση τους (γλίστρημα) ανάλογα με τα μεγέθη που ορίζονται. Δημιουργείται το πλαίσιο δεδομένων με αριθμητικούς όρους και αφαιρούνται οι μηδενικές τιμές. Ακολουθεί η μηχανική των χαρακτηριστικών, για την μετατροπή των

δεδομένων σε μεταβλητές εισόδου με τρόπο που το μοντέλο μπορεί να αξιοποιήσει για την αύξηση της αποδοτικότητας.

Συγκεκριμένα κάνουμε την εξής διάκριση. Η χρονοσειρά είναι μονομεταβλητή, κάτι το οποίο είναι λειτουργικό για τα στατιστικά μοντέλα, αλλά δίνει περιορισμένες δυνατότητες στα μοντέλα μηχανικής μάθησης να αξιοποιήσουν τα μοτίβα και τις εξαρτήσεις που υπάρχουν. Δημιουργούμε δύο πλαίσια δεδομένων με την παραπάνω διάκριση με το μονομεταβλητό να κάνει χρήση των έτοιμων στηλών της θερμοκρασίας και της ημερομηνίας ως όρος datetime. Το δεύτερο περιέχει πέντε στήλες εισόδου, οι τέσσερις με στοιχεία από την ημερομηνία σε σχέση με την ημέρα, τον μήνα και τον χρόνο και η πέμπτη την στήλη εξόδου με την θερμοκρασία. Όλες οι τιμές έχουν κανονικοποιηθεί στο διάστημα [-1,1].

Το σετ χωρίζεται σε τρία υποσύνολα της εκπαίδευσης, επαλήθευσης και εξέτασης με αντίστοιχα μεγέθη 0.85, 0.15, 0.15 του συνολικού, με την μέθοδο του train test split. Το μονοδιάστατο σετ χωρίζεται σε 0.85 για εκπαίδευση και 0.15 για εξέταση. Μετατρέπουμε τις μεταβλητές εισόδου και εξόδου που θα εισέλθουν στα μοντέλα μηχανικής μάθησης σε τένσορες, με την βιβλιοθήκη torch. Αυτό είναι απαραίτητο αφού τόσο το KAN όσο και το MLP και το LSTM δέχονται σαν είσοδο τένσορες δύο ή τριών διαστάσεων, οι οποίες οργανώνουν τα δεδομένα τους που τελικά διαμορφώνονται στα:

Multivariate Train Loader = (Train df, batch_size, shuffle=True)

Multivariate Val Loader = (Validation df, batch_size)

Multivariate Test Loader = (Test df, batch_size)

4.2.2 Ανάπτυξη των μοντέλων KAN

Η ανάπτυξη ενός νευρωνικού μοντέλου είναι η διαδικασία με την οποία το δίκτυο διαμορφώνεται, βελτιστοποιείται, αξιολογείται και περιέχει σαν στάδια τον ορισμό του, την εκπαίδευση, την αξιολόγηση και την πρόβλεψη.

- **Ορισμός**

Ο ορισμός είναι το στάδιο στο οποίο αποσαφηνίζονται οι παράμετροι του KAN δικτύου και δημιουργείται το αντίστοιχο μοντέλο. Ως είσοδο παίρνουν το μέγεθος του sliding

παραθύρου επί τα χαρακτηριστικά του συνόλου δεδομένων και ως έξοδο τον αριθμό 1 για την επόμενη μέρα.

Spline KAN / Chebyshev KAN
Layers hidden = [Window Size*5 , hidden 1 , hidden n , 1]
Grid Size/Chebyshev Degree
Spline order (only for Spline KAN)
Seed
Base activation = SiLU activation

Πίνακας 4.1: Η διαμόρφωση του Spline και Chebyshev KAN

Hybrid KAN
Input dim= Window Size*5
Output dim = 1
Grid Size
Chebyshev Degree
Spline order
Seed
Architecture
SiLU activation

Πίνακας 4.2: Η διαμόρφωση του Hybrid KAN

Mixture of Experts KAN (MoK)
In features=Window Size*5
Out features = 1
Grid Size
Chebyshev Degree
Spline order
Seed
Experts type
Gate type

Πίνακας 4.3: Η διαμόρφωση του Mixture of KAN

- **Εκπαίδευση**

Η εκπαίδευση του μοντέλου βασίζεται σε έναν επαναληπτικό βρόχο βελτιστοποίησης, όπως αυτός ορίζεται στη συνάρτηση evaluate_kan και ακολουθεί την ροή της κλασσικής εκπαίδευσης νευρωνικών δικτύων. Εφαρμόζουμε προώθηση εμπρός (forward pass) για κάθε batch αλλά και οπισθοδιάδοση (backpopagation) και υπολογίζουμε το σφάλμα

μέσω της συνάρτησης κόστους MSE. Οι παράμετροι με τις οποίες μαθαίνει το δίκτυο αντιστοιχούν στα γραμμικά βάρη αλλά και στους συντελεστές των εκπαιδεύσιμων συναρτήσεων ενεργοποίησης, δηλαδή τα spline weights και τις Chebyshev coefficients που επιλέγονται χωριστά ή ταυτόχρονα βάση του μοντέλου. Για την ενημέρωση αυτών χρησιμοποιήθηκε ο Adam optimizer με Learning Rate που ξεκινάει από 0.001.

- **Αξιολόγηση**

Στο στάδιο αυτό υπολογίζεται η τιμή του σφάλματος στα δεδομένα επαλήθευσης, έτσι ώστε να βελτιωθεί η ικανότητα γενίκευσης. Ένα από τα βασικά προβλήματα τα οποία εντοπίστηκαν ήταν η υπερπροσαρμογή του μοντέλου λόγω της ισχυρής εκφραστικότητας των KAN. Για αυτό τον λόγο εφαρμόστηκε ο μηχανισμός μείωσης του ρυθμού μάθησης ReduceLROnPlateau, ο οποίος υποχρεούται να μειώνει τον ρυθμό μάθησης όταν το μέσο σφάλμα επαλήθευσης δεν βελτιώνεται περαιτέρω. Τότε γίνεται αποθήκευση του καλύτερου μοντέλου για την χρήση του έπειτα με σκοπό την εξαγωγή μετρικών. Σαν επιπλέον εργαλείο βελτιστοποίησης και μείωσης της υπερπροσαρμογής χρησιμοποιήθηκε το early stopping, το οποίο διακόπτει την εκπαίδευση αν το σφάλμα επικύρωσης παραμείνει σταθερό για αριθμό ίσο με το προκαθορισμένο βαθμό υπομονής (patience).

- **Πρόβλεψη**

Για την πρόβλεψη φορτώνεται το καλύτερο μοντέλο που παράχθηκε από το στάδιο αξιολόγησης και προβλέπει στα άγνωστα δεδομένα του σετ εξέτασης. Από αυτό παράγονται δύο πίνακες, ένας με τις πραγματικές τιμές στόχου και ένας με τις προβλέψεις του μοντέλου KAN. Οι τιμές των πινάκων υφίστανται αντίστροφη κανονικοποίηση έτσι ώστε να εκφράζονται σε βαθμούς θερμοκρασίας. Οι τελικές μετρικές που χρησιμοποιούνται για την πρόβλεψη είναι το μέσο απόλυτο σφάλμα (MAE) και η ρίζα του μέσου τετραγωνικού σφάλματος (RMSE) που ορίζονται ως εξής:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|E_t| \qquad \text{και} \qquad RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}E_t^2}$$

4.2.3 Μοντέλα Αναφοράς : ARIMA, SARIMA, Prophet, MLP, LSTM

Τα στατιστικά μοντέλα ARIMA, SARIMA, Prophet που παρουσιάστηκαν στο Κεφάλαιο 3 έχουν προκαθορισμένες αρχιτεκτονικές και υλοποιούνται στα πειράματα μας μέσω αντίστοιχων βιβλιοθηκών. Εκπαιδεύονται στο μονοδιάστατο σύνολο δεδομένων ημερομηνίας και θερμοκρασίας και τελικά παράγονται οι τελικές μετρικές MAE, RMSE σε αντιστοιχία με των KAN.

Τα νευρωνικά μοντέλα αναφοράς MLP και LSTM ορίστηκαν με βάση τις παραμέτρους που προκύπτουν στην προ επεξεργασία δεδομένων και ακολούθησαν εκπαίδευση, αξιολόγηση και πρόβλεψη σε αναλογία με αυτή των KAN έτσι ώστε να παραχθούν ισότιμες συγκρίσεις. Οι βασικές παράμετροι που χρησιμοποιήθηκαν για τον ορισμό των μοντέλων φαίνονται παρακάτω:

MLP
Input size = Window Size*5
Hidden1, Hidden2, Hidden n
Seed

LSTM
Input size = 5
Hidden dim
Number of layers
Seed

Πίνακας 4.4: Η διαμόρφωση των MLP και LSTM

Το LSTM σε αντίθεση με τα υπόλοιπα μοντέλα βαθιάς μάθησης δέχεται ως είσοδο μόνο τον αριθμό χαρακτηριστικών (features), ενώ το μήκος του παραθύρου (Window Size) ενσωματώνεται αυτόματα στη διάσταση seq_len κατά τη δημιουργία των δεδομένων. Έτσι το τελικό σχήμα εισόδου προκύπτει ως:

[batch size, Window Size, Features]

4.2.4 Διερεύνηση υπερπαραμέτρων

Για την αξιολόγηση των μοντέλων πραγματοποιήθηκαν πολλαπλές δοκιμές με εναλλαγές στις τιμές των υπερπαραμέτρων, με στόχο τον ορισμό αυτών που μεγιστοποιούν την απόδοση. Η επιλογή τους ήταν κυρίως πειραματική, με την διερεύνηση και των επιλογών

που φαίνεται να υπερτερούν σε αντίστοιχες μελέτες αλλά και σε θεωρητικές εργασίες. Ως κρίσιμες παράμετροι κρίθηκαν το μέγεθος παραθύρου(window size), τα κρυφά επίπεδα και το μέγεθος τους (hidden layers), καθώς και το μέγεθος grid και ο βαθμός των Chebyshev πολυωνύμων. Τα βέλτιστα αποτελέσματα, τα οποία κρίθηκαν με βάση το σφάλμα MAE θα παρουσιαστούν στο κεφάλαιο 5 μαζί με τα διαγράμματα πρόβλεψης και τα διαγράμματα tuning.

4.3 Pipeline Κατηγοριοποίησης Χρονοσειρών

Αυτή η υποενότητα αφορά τα μοντέλα και τα πειράματα που δημιουργήθηκαν γύρω από την κατηγοριοποίηση των μονομεταβλητών χρονοσειρών που περιέχονται στο UCR Αρχείο. Το πείραμα μας υιοθέτησε την μεθοδολογία που προτείνουν οι I. Barašin et al. στην μελέτη [42].

4.3.1 Προ επεξεργασία δεδομένων

Τα δεδομένα προέρχονται από το UCR Univariate Time Series Archive και παρέχουν χρονοσειρές με εφαρμοσμένη z-κανονικοποίηση καθώς και διαχωρισμό σε σύνολο εκπαίδευσης και εξέτασης. Σε αυτό το βήμα θα εφαρμόσουμε επιπλέον στάδια προ επεξεργασίας, ώστε τα δεδομένα να είναι συμβατά με τις απαιτήσεις των μοντέλων που αναπτύχθηκαν.

Αρχικά αφαιρέθηκαν τα σύνολα δεδομένων που περιείχαν μηδενικές τιμές, για να αποφευχθούν σφάλματα κατά την εκπαίδευση, καταλήγοντας έτσι σε 117 dataset σε σχέση με τα αρχικά 125. Ακολουθεί αναπροσαρμογή των ετικετών (labels) ώστε να ξεκινάνε από το 0, όπως απαιτεί το κριτήριο κόστους CrossEntropyLoss που χρησιμοποιήθηκε κατά την εκπαίδευση.

Δημιουργούνται τρία σύνολα δεδομένων (train, validation, test) με μεγέθη 0.6, 0.2, 0.2 και διαχωρισμό επί του δοσμένου από τους δημιουργούς συνόλου εκπαίδευσης. Ένα από τα πιο κρίσιμα βήματα ήταν η εφαρμογή κανονικοποίησης σε κάθε χρονοσειρά με τον όρο StandardScaler, που εφάρμοσε την μετατροπή:

$$x_{t,scaled} = \frac{x_t - \mu_t}{\sigma_t}$$

ώστε να αποκτήσει κάθε χρονοσειρά στο συνολικό σετ, μηδενικό μέσο όρο και μοναδιαία τυπική απόκλιση. Τα δεδομένα μετατράπηκαν σε τένσορες και οργανώθηκαν σε DataLoaders, όπως κάναμε και στην πρόβλεψη, με χρήση του παράγοντα τυχαιότητας shuffle στο σύνολο εκπαίδευσης, ώστε τα δείγματα να εισέρχονται με τυχαία σειρά για να μειώνεται ο κίνδυνος υπερπροσαρμογής σε συγκεκριμένα μοτίβα.

Τέλος έχουμε εφαρμόσει ένα επιπλέον βήμα για να αντιμετωπιστεί η ανισορροπία των κλάσεων. Οι χρονοσειρές δεν περιέχουν σταθερό αριθμό δειγμάτων ανά κλάση, με αποτέλεσμα να προκύπτουν κλάσεις με πληθώρα δειγμάτων, σε αντίθεση με άλλες που έχουν έλλειψη. Για να προκύπτει αμερόληπτη μάθηση ως προς των αριθμό των δειγμάτων εφαρμόζουμε την τεχνική βαρών μέσω της συνάρτησης υπολογισμού βαρών, ώστε οι πιο συχνές κλάσεις να σταθμίζονται με μικρότερο βάρος από τις πιο σπάνιες.

4.3.2 Ανάπτυξη των μοντέλων KAN

- **Ορισμός**

Τα μοντέλα ορίζονται αφού εφαρμοστεί η προ επεξεργασία των δεδομένων και ακολουθούν την λογική που αναφέρθηκε στην πρόβλεψη. Τα KAN (Spline, Chebyshev, Hybrid) δέχονται ως είσοδο μία δομή αρχιτεκτονικής, η οποία είναι αναπαρίσταται ως μία λίστα της μορφής:

[input_size, hidden_layer_1_size, ..., output_size]

Ως είσοδο δέχεται ολόκληρη την κάθε χρονοσειρά, με το μέγεθος εισόδου (input size) να αντιστοιχεί στο μήκος της κάθε χρονοσειράς, όπως αυτό παρέχεται στο εκάστοτε σύνολο δεδομένων. Στις περιπτώσεις όπου το πεδίο του μήκους φέρει την τιμή “vary”, τα αντίστοιχα δεδομένα απορρίπτονται ως μη έγκυρα, ώστε να διασφαλιστεί η ομοιομορφία του χώρου εισόδου. Τα κρυφά στρώματα δηλώνονται μέσα από την λίστα αρχιτεκτονικής, έτσι ώστε να παρέχουν ευελιξία στο μοντέλο. Τέλος ως έξοδο δέχεται την κλάση στην οποία πρέπει να κατανεμηθεί η χρονοσειρά και ως μέγεθος εξόδου (output) τον αριθμό των κλάσεων που περιέχει το κάθε σύνολο δεδομένων.

- **Εκπαίδευση**

Η διαδικασία της εκπαίδευσης πραγματοποιήθηκε με την ροή που συνηθίζουν να ακολουθούν τα νευρωνικά δίκτυα. Σαν κριτήριο χρησιμοποιήθηκε η συνάρτηση κόστους CrossEntropyLoss, η οποία είναι κατάλληλη για προβλήματα ταξινόμησης πολλαπλών κλάσεων και κάνει χρήση της συνάρτησης υπολογισμού βαρών για ισορροπημένες κλάσεις που αναφέραμε παραπάνω. Για την βελτιστοποίηση του μοντέλου χρησιμοποιήθηκε ο AdamW με αρχικό ρυθμό μάθησης 0.001, ο οποίος σε αντίθεση με τον Adam εφαρμόζει την L2 κανονικοποίηση (weight decay) ξεχωριστά από το βήμα ενημέρωσης των παραμέτρων με αποτέλεσμα να προκύπτει πιο σταθερή σύγκλιση και μικρότερος κίνδυνος υπερπροσαρμογής. Η προσαρμογή του ρυθμού μάθησης εφαρμόζεται μέσω του ExponentialLR, ο οποίος μειώνει τον ρυθμό εκθετικά σε κάθε επανάληψη με συντελεστή μείωσης 0.98 επί του αρχικού.

Η εκπαίδευση πραγματοποιήθηκε σε CPU, χρησιμοποιώντας PyTorch. Τα πειράματα έτρεξαν με batch size = 32 και αριθμό εποχών 50. Για την αξιοπιστία των πειραμάτων εφαρμόστηκε εκπαίδευση με πολλαπλά seeds. Συγκεκριμένα χρησιμοποιήθηκαν οι τιμές 1, 42 και 123, ώστε να περιοριστεί η πιθανότητα τα αποτελέσματα να οφείλονται σε τυχαία αρχικοποίηση. Ο μέσος χρόνος εκπαίδευσης για τις χρονοσειρές μικρού μεγέθους ήταν λίγα δευτερόλεπτα, ενώ για τις μεγαλύτερες κυμαινόταν σε αρκετά λεπτά ανά seed. Η καταγραφή του συνολικού χρόνου εκπαίδευσης ανά μοντέλο γίνεται στο Κεφάλαιο 5.

- **Αξιολόγηση**

Σε αυτή την υποενότητα υπολογίζεται το σφάλμα ταξινόμησης στα δεδομένα επαλήθευσης και ελέγχου, με βάση της πιθανότητες αντιστοίχισης σε κάθε κλάση. Μετά το στάδιο επαλήθευσης εφαρμόζεται η διαδικασία μείωσης του ρυθμού μάθησης μέσω του ExponentialLR, ενώ το σφάλμα επαλήθευσης προκύπτει από τη σύγκριση των εξόδων του μοντέλου με τις πραγματικές ετικέτες, βάσει της τιμής της συνάρτησης κόστους. Ο συνολικός χρόνος εκπαίδευσης και επαλήθευσης μετρήθηκε σαν ξεχωριστή μεταβλητή και θα παρουσιαστεί στο κεφάλαιο πέντε. Ως μετρικές χρησιμοποιούνται οι εξής:

Ακρίβεια (Accuracy): Εκφράζει το ποσοστό των σωστών ταξινομήσεων επί του συνολικού αριθμού δειγμάτων.

$$\frac{Αριθμός\ σωστών\ προβλέψεων}{Συνολικών\ αριθμός\ προβλέψεων}$$

Ακρίβεια θετικών προβλέψεων (Precision): Μετρά το ποσοστό των σωστών θετικών προβλέψεων σε σχέση με όλες τις θετικές προβλέψεις που παρήγαγε το μοντέλο Χρησιμοποιείται για έμφαση στη μείωση των ψευδώς θετικών.

$$\frac{Σωστά\ θετικά}{Σωστά\ θετικά + Ψευδά\ θετικά}$$

Ανάκληση (Recall): Μετρά το ποσοστό των σωστών θετικών προβλέψεων σε σχέση με όλα τα πραγματικά θετικά δείγματα, εστιάζοντας στη μείωση των ψευδώς αρνητικών.

$$\frac{Σωστά\ θετικά}{Σωστά\ θετικά + Ψευδά\ αρνητικά}$$

Δείκτης F1: Αποτελεί τον μέσο των precision και recall, παρέχοντας έναν ενιαίο δείκτη που ισορροπεί ανάμεσα στα ψευδώς θετικά και τα ψευδώς αρνητικά και θεωρείται ιδιαίτερα χρήσιμος σε ανισόρροπα σύνολα δεδομένων.

$$2 \times \frac{Precision \times Recall}{Precision + Recall}$$

- **Ταξινόμηση**

Η τελική απόφαση ταξινόμησης σε κατηγορίες προκύπτει με βάση τον όρο argmax της μέγιστης πιθανότητας. Αυτός εφαρμόζεται στο τελευταίο επίπεδο εξόδου του μοντέλου, το οποίο περιέχει αριθμό νευρώνων ίσο με τον αριθμό κλάσεων κάθε χρονοσειράς, όπως αναφέρθηκε και στο στάδιο του ορισμού.

4.3.3 Μοντέλα Αναφοράς: MLP, 1-NN DTW

Ως μοντέλα αναφοράς χρησιμοποιήθηκαν το MLP που θεωρείται ο κύριος ανταγωνιστής του KAN και ο αλγόριθμος 1-NN DTW που προτείνεται για χρήση στα συγκεκριμένα δεδομένα. Επιπλέον χρησιμοποιούνται τρία μεγέθη σφάλματος τα οποία παρέχονται στην σύνοψη των δεδομένων και αναφέρθηκαν παραπάνω. Το LSTM, όπως και τα ARIMA και Prophet που εξετάστηκαν παραπάνω δεν θα συμπεριληφθούν αφού δεν κρίνονται κατάλληλα για τα προβλήματα ταξινόμησης.

Το MLP υλοποιήθηκε μέσω της βιβλιοθήκης scikit-learn και ορίστηκε με είσοδο, έξοδο και κρυφά στρώματα ίδια με αυτά των KAN. Ως συνάρτηση ενεργοποίησης επιλέχθηκε η ReLU και ως βελτιστοποιητής ο Adam. Τέλος εφαρμόστηκε L2 κανονικοποίηση με σταθερό ρυθμό μάθησης.

Ο 1-NN DTW υλοποιήθηκε μέσω της βιβλιοθήκης tslearn και με προκαθορισμένη αρχιτεκτονική.

4.3.4 Διερεύνηση Υπερπαραμέτρων

Ένα σημαντικό κομμάτι του πειράματος ήταν και η μελέτη των υπερπαραμέτρων, με διερεύνηση πολλαπλών επιλογών, έτσι ώστε να προκύπτει η μεγαλύτερη δυνατή απόδοση στην πλειοψηφία των σετ δεδομένων. Η δυσκολία αυτού κρίθηκε στο γεγονός ότι ένας συνδυασμός υπερπαραμέτρων μπορεί να ευνοούσε μερικές χρονοσειρές και να αδικούσε άλλες. Για τον σκοπό αυτό το πείραμα διαθέτει επιλογή για επαναληπτικές λούπες ρύθμισης του ρυθμού μάθησης, της τυχαιότητας (seed) καθώς και των βασικών υπερπαραμέτρων των KAN: το μέγεθος του grid και τον βαθμό Chebyshev.
Η σημασία αυτών των υπερπαραμέτρων έγκειται στα εξής:
Το μέγεθος του grid καθορίζει την λεπτότητα με την οποία χωρίζεται ο χώρος εισόδου στις spline βάσεις, με τα μικρά μεγέθη να οδηγούν σε πιο απλή αναπαράσταση και τα μεγαλύτερα να προσφέρουν περισσότερη εκφραστικότητα με κίνδυνο υπερπροσαρμογής.
Αντίστοιχα, ο βαθμός Chebyshev επηρεάζει άμεσα την ικανότητα του μοντέλου να προσεγγίζει μη γραμμικές σχέσεις, με τους χαμηλούς βαθμούς να ευνοούν την ομαλότητα και τους υψηλούς βαθμούς να ενισχύουν την πολυπλοκότητα.
Τέλος, το πλήθος και το μέγεθος των κρυφών στρωμάτων καθορίζουν τη συνολική χωρητικότητα του δικτύου, επηρεάζοντας τόσο την ικανότητα εκμάθησης λεπτομερών χαρακτηριστικών όσο και τον κίνδυνο υπερπροσαρμογής σε μικρότερα σετ δεδομένων.

Στο επόμενο κεφάλαιο θα παρουσιαστούν τα συγκεντρωτικά αποτελέσματα, οι πίνακες και τα διαγράμματα που αναδεικνύουν τις επιδόσεις κάθε μοντέλου και τις επιδράσεις των παραπάνω υπερπαραμέτρων.

Κεφάλαιο 5 Αποτελέσματα και Συγκρίσεις

5.1 Αποτελέσματα Πρόβλεψης

Στο κεφάλαιο αυτό θα αναπτύξουμε τα αποτελέσματα των μοντέλων KAN που παράχθηκαν για την πρόβλεψη μονομεταβλητών χρονοσειρών στα δεδομένα θερμοκρασίας και τις συγκρίσεις με σύγχρονες τεχνικές και μοντέλα αναφοράς. Έπειτα θα παρουσιάσουμε τα διαγράμματα από την απόδοση των KAN στις αλλαγές κρίσιμων υπερπαραμέτρων.

5.1.1 Μετρικές Αξιολόγησης Απόδοσης

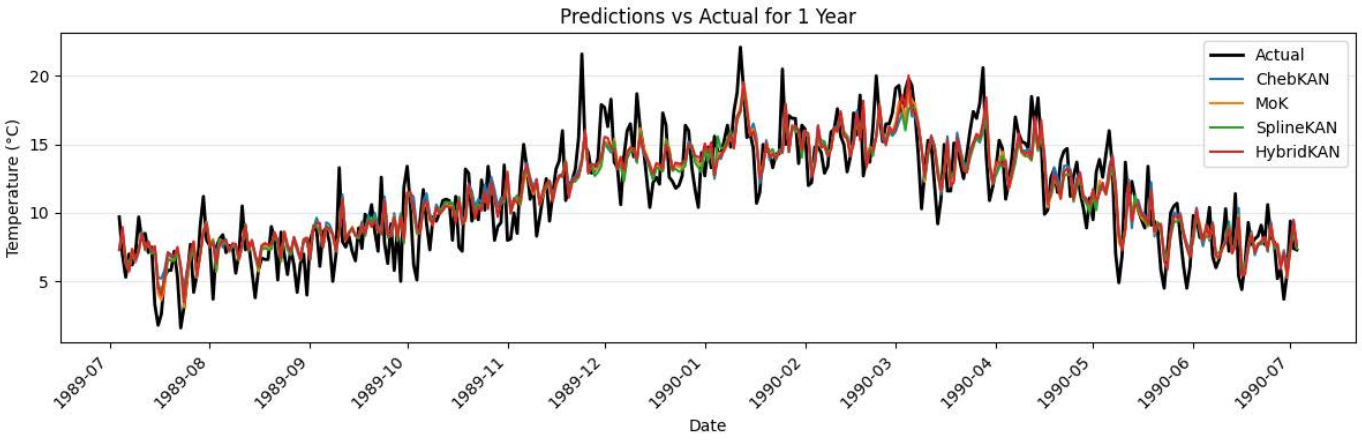
Παρουσιάζεται ο πίνακας μετρικών αξιολόγησης με τα καλύτερα αποτελέσματα από τα μοντέλα KAN, τα μοντέλα αναφοράς μαζί με τα αποτελέσματα των στατιστικών τεχνικών που δοκιμάσαμε. Τα σφάλματα MAE και RMSE είναι κανονικοποιημένα σε βαθμούς θερμοκρασίας κελσίου. Ως (g) θεωρούμε το μέγεθος του πλέγματος grid, ως (c) τον βαθμού του Chebyshev πολυωνύμου, ως (w) το μέγεθος του παραθύρου sliding window και ως (m) την περίοδο εποχικότητας του ARIMA.

Μοντέλο	Αρχιτεκτονική	MAE	RMSE
Auto ARIMA	(3,0,1)	3.0746	3.834
Auto SARIMA	(3,0,1), m=30	3.0746	3.834
Prophet		2.1605	2.687
MLP	(256,128,128)	1.7183	2.203
LSTM	(40,40)	1.7052	2.179
Spline KAN	(128,64,64), g=7, w=3	1.6938	2.170
Chebyshev KAN	(256,256), c=5, w4	1.6989	2.181
Hybrid KAN	(128,64,64,32,32,1), c=5, g=5, w=4	1.6835	**2.159**
Mix of KAN experts	(256,128,1), c=5, g=7, w=4	**1.6782**	2.160

Πίνακας 5.1 Μετρικές Αξιολόγησης Απόδοσης

Όπως παρατηρούμε από τον πίνακα 5.1 τα κλασσικά στατιστικά μοντέλα (Auto ARIMA, Auto SARIMA) παρουσιάζουν το υψηλότερο σφάλμα, με αμέσως επόμενο το Prophet, επιβεβαιώνοντας ότι τα γραμμικά μοντέλα δεν επαρκούν για να αποτυπώσουν τη μη γραμμικότητα της χρονοσειράς. Τα νευρωνικά μοντέλα (MLP, LSTM) παρουσιάζουν βελτιωμένη απόδοση, με το LSTM να υπερέχει ελαφρά λόγω της αξιοποίησης της χρονικής του ακολουθίας. Όλα τα μοντέλα KAN που μελετάμε ξεπερνούν τις αποδόσεις των μοντέλων αναφοράς και στις δύο μετρικές σφάλματος. Τα δύο υβριδικά (Hybrid και MoK) καταγράφουν τα χαμηλότερα σφάλματα, δείχνοντας έτσι τις δυνατότητες που μπορεί να παρέχει ο συνδυασμός Spline και Chebyshev βάσεων.

Παρουσιάζεται το διάγραμμα των προβλέψεων για τα καλύτερα KAN μοντέλα σε σχέση με τις πραγματικές τιμές στο βάθος ενός χρόνου.



Σχήμα 5.1: Προβλέψεις ενός έτους KAN Μοντέλων σε σχέση με την πραγματική τιμή

Τα υβριδικά μοντέλα δημιουργήθηκαν από την στοίβαξη στρωμάτων αξιοποιώντας από κοινού grid, πολυωνυμικές Chebyshev αλλά και γραμμικές βάσεις . Αναπτύσσουμε αναλυτικά τις αρχιτεκτονικές των υβριδικών μοντέλων που κατέγραψαν τα παραπάνω αποτελέσματα.

Το καλύτερο Hybrid model αποτελείται από τα παρακάτω στρώματα:

Spline KAN (input size, 128) → Spline KAN (128, 64) → Linear(64, 64) → ChebyshevKAN (64, 32) → ChebyshevKAN (32, 32) → Linear(32, 1)

Είναι δίκτυο 5 κρυφών επιπέδων και ενός επιπέδου εξόδου, όπου μετά από κάθε στρώμα ακολουθεί ένα στρώμα κανονικοποίησης Layer Normalization και ενεργοποίηση της συνάρτησης SiLU.

Το καλύτερο MoK model αποτελείται από τα παρακάτω στρώματα:

1 στρώμα (Expert: E, Gate: KAN)(input size, 256) → 1 στρώμα (Expert: D, Gate: Linear) (256, 128) → 1 στρώμα (Expert: E, Gate: KAN)(128,1)

Το συγκεκριμένο δίκτυο θεωρείται ένα πολυεπίπεδο Mixture of KAN experts (MMK) 2 κρυφών στρωμάτων και ενός στρώματος εξόδου. Ο τύπος των ειδικών που αναγράφεται ως 'E' και 'D' εξηγείται παρακάτω αναλυτικά.

Θα παρουσιάσουμε όλους τους τύπους MoK στρωμάτων που μελετήσαμε και τις αποδόσεις για κάθε τύπο. Η εκφραστικότητα του MoK δεν προκύπτει από το βάθος (hidden layers), αλλά από τις μη-γραμμικές βάσεις (KAN, Chebyshev) και από το μηχανισμό μίξης, μέσω της πύλης (gate), η οποία συνδυάζει τα διάφορα στρώματα KAN και αποδίδει πιθανότητες (softmax). Για αυτό τον λόγο ένα στρώμα MoK είναι πολλές φορές ικανό να αποτυπώσει την πολυπλοκότητα του μοντέλου με βάση την επιλογή του τύπου των experts και της αντίστοιχης πύλης, χωρίς την ανάγκη επιπλέον βάθους κρυφών στρωμάτων.

Gate	Experts	Architecture
KAN	**A**	KANLinear, KANLinear, KANLinear, KANLinear
	B	ChebyshevKAN, ChebyshevKAN, ChebyshevKAN, ChebyshevKAN
	C	ChebyshevKAN, ChebyshevKAN, KANLinear, KANLinear
Linear	**D**	KANLinear, ChebyshevKAN, KANLinear, ChebyshevKAN
	E	KANLinear, KANLinear, ChebyshevKAN, ChebyshevKAN
	L	Linear, Linear, Linear, Linear

Πίνακας 5.2 Αρχιτεκτονικές στρωμάτων MoK Layer

	Gate = KAN		Gate = Linear	
Expert	**MAE**	**RMSE**	**MAE**	**RMSE**
A	1.7141	2.201	1.7037	2.186
B	1.7206	2.211	1.7146	2.206
C	1.7036	2.176	1.7310	2.206
D	1.7021	2.182	**1.6932**	**2.174**
E	**1.6835**	**2.159**	1.6959	2.175
L	1.7247	2.214	1.7644	2.261

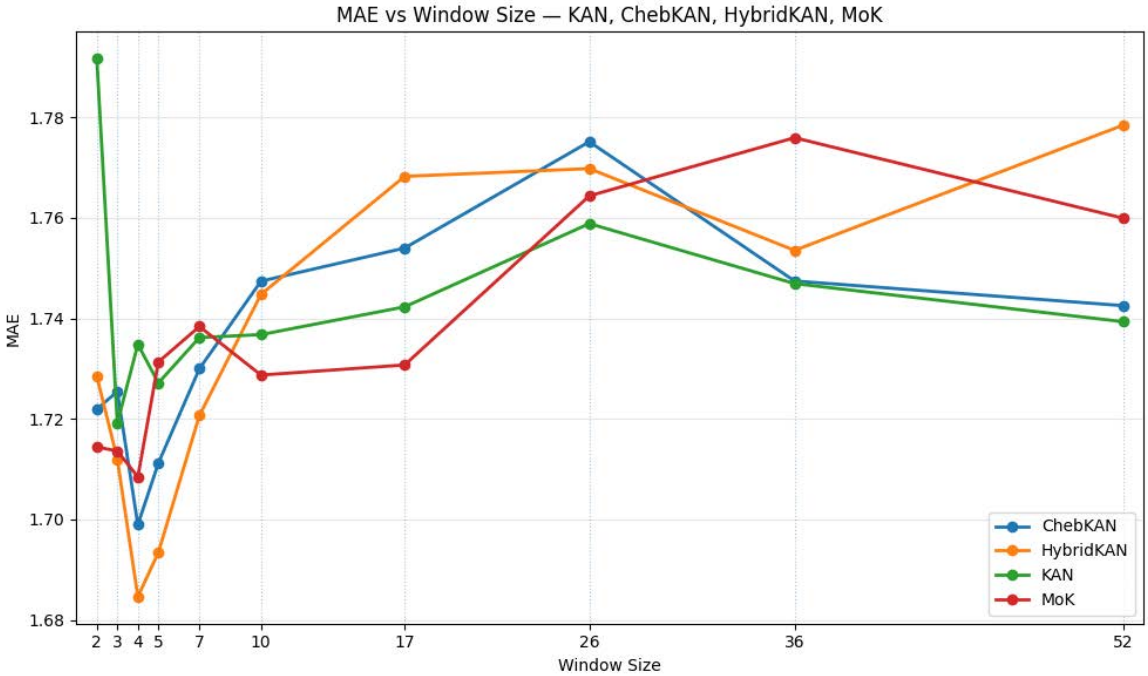
Πίνακας 5.3 Αποτελέσματα στρωμάτων MoK Layer

Τα αποτελέσματα παράχθηκαν με κοινό μέγεθος grid ίσο με 7, βαθμό πολυωνύμου ίσο με 5 και μέγεθος παραθύρου ίσο με 4. Η γραμμική αρχιτεκτονική L εμφανίζει τα πιο αδύναμα αποτελέσματα, δείχνοντας τον περιορισμό των απλών γραμμικών μετασχηματισμών. Οι διαμορφώσεις A και B εμφανίζουν τα μεγαλύτερα σφάλματα γεγονός που δείχνει ότι η εξειδίκευση σε έναν τύπο πυρήνα περιορίζει την εκφραστική ικανότητα του μοντέλου. Οι υβριδικοί συνδυασμοί C, D, E που εναλλάσσουν spline και Chebyshev βάσεις μειώνουν το σφάλμα εμφανίζοντας τα καλύτερα αποτελέσματα, επιβεβαιώνοντας ότι η ετερογένεια αυτή μπορεί να προσφέρει βελτίωση στην απόδοση. Η μέγιστη απόδοση επιτυγχάνεται από τον expert E και έπειτα D, οι οποίοι έχουν στην αρχή της διαμόρφωσης τους Spline KAN και ακολουθούν τα Chebyshev KAN. Αυτό μπορεί να ερμηνευθεί από τις ιδιότητες της κάθε βάσης. Τα spline προσφέρουν ικανότητα τοπικής προσαρμογής προσδίδοντας αρχικά λεπτομέρεια στο μοντέλο ενώ τα Chebyshev βοηθάνε στην γενίκευση σε όλο το πεδίο, εξομαλύνοντας και προστατεύοντας από υπερεξειδίκευση.

Όσον αφορά τον μηχανισμό πύλης, η KAN προσφέρει πλεονέκτημα καθώς επιτρέπει πιο εκφραστική μη γραμμική κατανομή βαρών στους experts, ενώ η γραμμική μπορεί να αποδώσει καλύτερα μόνο σε περιπτώσεις όπου οι ίδιοι οι experts είναι ιδιαίτερα ισχυροί.

5.1.2 Ρύθμιση Μεγέθους Παραθύρου

Σε αυτή την υποενότητα θα παρουσιάσουμε τις διαφοροποιήσεις στην απόδοση των KAN ανάλογα με τις εναλλαγές του μεγέθους παραθύρου που καθορίζει το μήκος της ιστορικής πληροφορίας χρήσης του μοντέλου για την πρόβλεψη.
Το KAN αναπαριστά το Spline KAN [128,64] με g=3, το Chebyshev KAN [256,256] με c=5, το Hybrid είναι της αρχιτεκτονικής [128,64,64,32,32] με g=3, c=5 και το MoK περιέχει τον ειδικό 'A' με πύλη KAN.



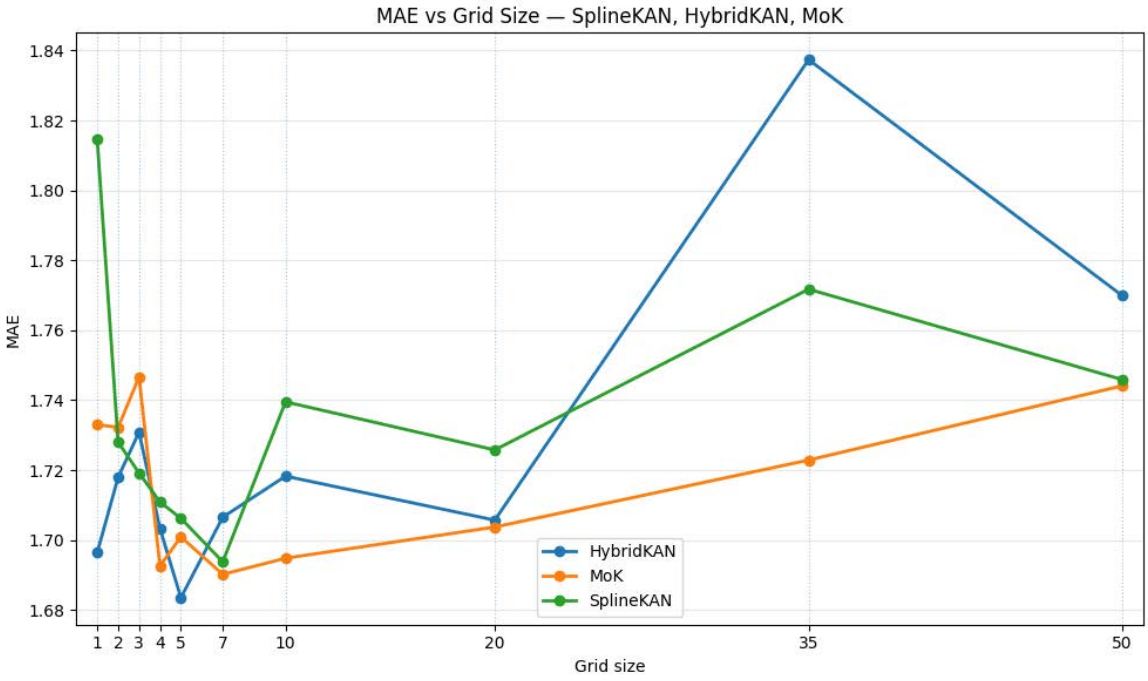
Σχήμα 5.2: Διαβάθμιση σφάλματος στις εναλλαγές του μεγέθους παραθύρου

Όλα τα μοντέλα φαίνεται να αποδίδουν καλύτερα στα μεγέθη 3-5, με ελάχιστο MAE στο 4-5, υποδηλώνοντας ισχυρή βραχυπρόθεσμη εξάρτηση μεταξύ των πρόσφατων δεδομένων. Το σφάλμα αυξάνεται σταδιακά μέχρι τον αριθμό 26, καθώς η εισαγωγή πρόσθετης πληροφορίας δεν φαίνεται να προσθέτει κάτι στο μοντέλο παρά θόρυβο. Παρόλα αυτά το MoK και το KAN κρατάνε τις πιο σταθερές επιδόσεις σε αυτά τα μεγέθη, ενδεχομένως λόγω της τοπικής προσαρμοστικότητας των spline και της ετερογένειας των expert. Στα μεγαλύτερα μεγέθη τα υβριδικά μοντέλα είναι λιγότερο σταθερά, καθώς η πολυπλοκότητα τους μαζί με την μακροχρόνια πληροφορία αποπροσανατολίζουν και οδηγούν σε υπερεξειδίκευση. Συνολικά φαίνεται ότι όλα τα KAN ακολουθούν την ίδια

τάση με διαφοροποιήσεις, οι οποίες είναι κρίσιμες για την ελαχιστοποίηση του σφάλματος.

5.1.3 Ρύθμιση Grid Υπερπαραμέτρου (Grid Tuning)

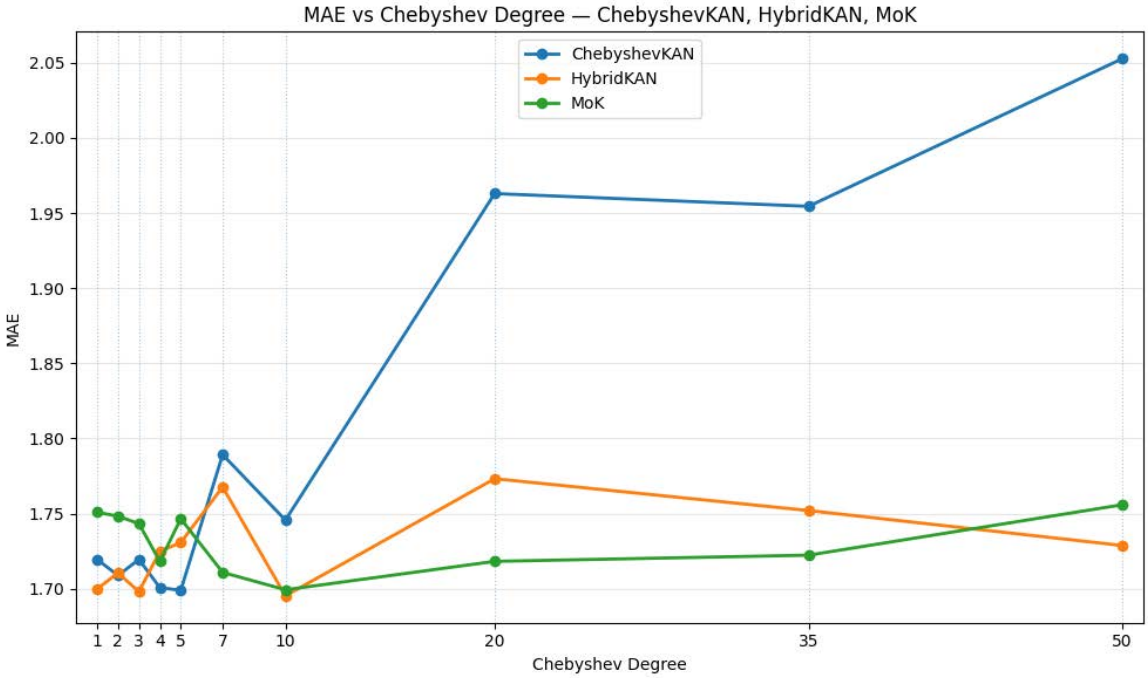
Τα μοντέλα KAN εξετάστηκαν σε σχέση με την συμπεριφορά τους απέναντι σε αλλαγές στις υπερπαραμέτρους με σκοπό να μελετήσουμε τα μοτίβα και τις τάσεις που εμφανίζονται. Παρουσιάζονται τα αντίστοιχα μοντέλα με αυτά που δοκιμάστηκαν στην ρύθμιση του window size.



Σχήμα 5.3: Διαβάθμιση σφάλματος στις εναλλαγές του μεγέθους grid

Από το διάγραμμα παρατηρούμε ότι τα μεγέθη 1-3 και τα μεγέθη 10-50 προκαλούν τα μεγαλύτερα σφάλματα και στα τρία μοντέλα. Τα μικρά grid δεν είναι ικανά να παράξουν τοπική προσαρμοστικότητα με το Spline KAN να εμφανίζει την μεγαλύτερη ευαισθησία λόγω της απλούστερης αρχιτεκτονικής του. Τα μεγάλα grid παρέχουν υπερβολικά λεπτομερή παραμετροποίηση, οδηγώντας σε σφάλματα, με το Hybrid KAN να επιδεινώνεται περισσότερο λόγω της πρόσθετης πληροφορίας από την αρχιτεκτονική του. Τα μεσαία μεγέθη οδηγούν στα βέλτιστα αποτελέσματα, αφού βρίσκουν ισορροπία μεταξύ προσαρμογής και γενίκευσης. Το MoK παρουσιάζεται συνολικά ως η πιο συνεπής αρχιτεκτονική πετυχαίνοντας σταθερά καλές επιδόσεις.

5.1.4 Ρύθμιση Chebyshev Υπερπαραμέτρου (Chebyshev degree Tuning)



Σχήμα 5.4: Διαβάθμιση σφάλματος στις εναλλαγές του μεγέθους Chebyshev

Όπως φαίνεται στο σχήμα, οι βαθμοί 3-5 προσδίδουν το μικρότερο σφάλμα με μικρές διακυμάνσεις ανάμεσα στα μοντέλα, όπως το MoK που εμφανίζει καλύτερη επίδοση για βαθμό 10. Στα μεσαία μεγέθη 5-10 παρατηρείται μια άνοδος και έπειτα σταθεροποίηση εκ νέου στο 10, ενδεχομένως λόγω της απότομης εισαγωγής θορύβου. Στους βαθμούς μεγαλύτερους από το 10, το Chebyshev εμφανίζει πολύ αδύναμα αποτελέσματα λόγω της έντονης εξάρτησης του με τους πολυωνυμικούς βαθμούς. Σε αντίθεση, τα υβριδικά μοντέλα Hybrid και MoK εμφανίζονται πιο σταθερά με σφάλμα 1.72-1.77 δείχνοντας ότι ο υπολογισμός μέσω και των spline συναρτήσεων μπορεί να εμφανίσει σταθερότητα στις εναλλαγές των βαθμών Chebyshev. Συνολικά, τα αποτελέσματα δείχνουν ότι το βέλτιστο εύρος για την ισορροπία ανάμεσα στην εκφραστικότητα και στην αποφυγή υπερπροσαρμογής βρίσκεται σε χαμηλές τιμές (3–7).

5.2 Αποτελέσματα κατηγοριοποίησης

Στο κεφάλαιο 5.2 θα αναπτύξουμε τα αποτελέσματα των μοντέλων KAN που παράχθηκαν για την κατηγοριοποίηση μονομεταβλητών χρονοσειρών στο αρχείο UCR και τις συγκρίσεις με σύγχρονες τεχνικές και μοντέλα αναφοράς. Παρακάτω θα παρουσιάσουμε πίνακες από την απόδοση των KAN σε διάφορες κατηγορίες των δεδομένων και από την απόδοση στις αλλαγές κρίσιμων υπερπαραμέτρων.

5.2.1 Μετρικές Αξιολόγησης Απόδοσης

Σε αυτή την ενότητα παρουσιάζουμε και αξιολογούμε την καλύτερη επίδοση των μοντέλων KAN σε σχέση με τα μοντέλα αναφοράς που έχουν επιλεχθεί για το συγκεκριμένο πρόβλημα. Παρακάτω διεξάγονται στατιστικά τεστ μεταξύ των κατανομητών για την αξιολόγηση των επιδόσεων.

Μοντέλο	**Αρχιτεκτονική**	**Accuracy**		**Precision**		**Recall**		**F1**		**Training Time (s)**
		Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	
Spline KAN	(128,64), g=5	0.709	0.15	0.702	0.15	0.684	0.16	0.681	0.16	40.60
Chebyshev KAN	(256,128,64), c=4	0.710	0.17	0.700	0.17	0.687	0.17	0.682	0.17	38.13
Hybrid KAN	(128,128,64,64,32) g=5, c=5	0.717	0.16	0.704	0.17	0.691	0.17	0.685	0.17	45.48
MLP	(128,64,64)	0.738	0.16	0.725	0.16	0.712	0.17	0.708	0.17	2.45
1-NN DTW	--	0.723	0.16	0.716	0.16	0.699	0.17	0.696	0.17	1787.28

Πίνακας 5.4: Μετρικές Αξιολόγησης Απόδοσης

Όπως παρατηρούμε στον πίνακα παραπάνω, τα αποτελέσματα από όλα τα μοντέλα είναι αρκετά κοντινά με μέσο όρο το 0.72. Τα KAN μοντέλα πετυχαίνουν απόδοση κοντινή μεταξύ τους με το υβριδικό KAN να έχει την πιο υψηλή από τα υπόλοιπα σε όλες τις μετρικές. Αυτό μπορούμε να το αποδώσουμε στην πολυπλοκότητα που του αποδίδει η αρχιτεκτονική του και άρα την ικανότητα να εντοπίζει καλύτερα τις διακυμάνσεις στα

δεδομένα. Συνολικά το MLP υπερτερεί σε όλα τα μεγέθη και ακολουθεί το DTW με πολύ μικρή διαφορά από το Υβριδικό KAN. Ακολουθεί ο πίνακας με τα μέσα σφάλματα απόδοσης των KAN σε συνάρτηση με τα μοντέλα βάσεις.

Model	SplineKAN	Chebyshev KAN	HybridKAN	1-NN DTW	MLP	ED (w=0)	DTW (w=100)	DTW (learned)	Default Rate
Error Rate	0.291	0.290	0.283	0.277	0.262	0.302	0.276	**0.250**	0.692

Πίνακας 5.5: Μέσο Σφάλμα Απόδοσης

Τα μοντέλα KAN εμφάνισαν μέσο σφάλμα 0.28-0.29 ξεπερνώντας τον απλό ταξινομητή της ευκλείδειας απόστασης και κινούμενα σε επίπεδα αντίστοιχα με του 1-NN DTW, DTW (w=100). Το DTW (learned) συμπλήρωσε το πιο χαμηλό σφάλμα, αφού θεωρείται ο πιο κλασσικός ταξινομητής για τα δεδομένα του UCR αρχείου και το MLP ακολούθησε με το δεύτερο μικρότερο σφάλμα.

Τα αποτελέσματα ελέγχθηκαν με το Wilcoxon signed-rank test ανά δυάδες dataset, έτσι ώστε να εντοπιστούν οι διαφορές που θεωρούνται στατιστικά σημαντικές. Το τεστ αυτό δεν ελέγχει το μέσο σφάλμα αλλά κάνει έλεγχο με βάση το κάθε σύνολο και εντοπίζει το πόσο σταθερά εμφανίζονται τα μοντέλα. Για παράδειγμα αν ένα μοντέλο συμπληρώνει το μικρότερο μέσο σφάλμα αλλά δεν υπερτερεί έναντι ενός άλλου σταθερά στα σύνολα δεδομένων, μπορεί να θεωρηθεί στατιστικά μη σημαντική η διαφορά μεταξύ τους.

Και τα τρία KAN βρέθηκαν στατιστικά ισάξια με τους ταξινομητές ED, DTW (w=100), DTW (learned) με μόνη εξαίρεση το Spline KAN που θεωρήθηκε στατιστικά κατώτερο από το DTW-learned. Η καλύτερη επίδοση επιτεύχθηκε από το MLP που βρέθηκε στατιστικά ισχυρότερο από τα τρία μοντέλα KAN και ισάξιο με τα DTW. Από τα αποτελέσματα παράχθηκε το γράφημα heatmap με όλες τις τιμές p των μοντέλων από το Wilcoxon test για οπτική αναπαράσταση, το οποίο βρίσκεται στα παραρτήματα.

5.2.2 Απόδοση των KAN ανά κατηγορίες δεδομένων

Εδώ θα παρουσιάσουμε τα αποτελέσματα που αντιστοιχούν στην επίδοση των μοντέλων μας σε σχέση με τα βασικά χαρακτηριστικά των συνόλων δεδομένων. Η ομαδοποίηση

αυτή μας επιτρέπει να εντοπίσουμε το πως η συμπεριφορά των αρχιτεκτονικών επηρεάζεται από την πολυπλοκότητα του προβλήματος, όπως το πλήθος κλάσεων ή από την αυξανόμενη διάσταση εισόδου, ώστε να αναδειχθούν τυχόν επαναλαμβανόμενα μοτίβα στις επιδόσεις τους.

Τα παρακάτω μοτίβα που εξετάζονται αποτελούν γενικές τάσεις των μοντέλων και όχι αποτέλεσμα μιας αρχιτεκτονικής. Το συμπέρασμα αυτό προέκυψε αφού εξετάσαμε δύο διαφορετικές αρχιτεκτονικές από κάθε μοντέλο, ως προς τις παρακάτω συσχετίσεις και η αντίδραση ήταν κοινή. Τα παρακάτω αποτελέσματα αφορούν τις πιο αποδοτικές αρχιτεκτονικές και αναπαριστούν την μέση ακρίβεια από πολλαπλές δοκιμές μεγέθους grid και βαθμού Chebyshev.

	Ακρίβεια			
Κλάσεις	**HybridKAN** (128,128,64,64,32)	**SplineKAN** (128,64)	**ChebyshevKAN** (256,128,64)	**N datasets**
Δυαδική (≤2)	**0.763**	0.748	0.762	16
3–10	0.699	0.699	0.689	16
>10	**0.620**	0.617	0.607	8

Πίνακας 5.6: Μέση ακρίβεια ανά κατηγορία πλήθους κλάσεων

Οι κλάσεις χωρίζονται στις κατηγορίες της δυαδικής, των μεσαίων μεγεθών και των πολλών. Και τα τρία μοντέλα φαίνεται να αποδίδουν με υψηλή ακρίβεια στις δυαδικές κλάσεις ενώ η απόδοση μειώνεται όσο αυξάνει το πλήθος. Αυτό συμβαίνει διότι όσο αυξάνονται οι κλάσεις το πρόβλημα γίνεται πιο περίπλοκο και χρειάζεται μεγαλύτερη εκφραστικότητα για να διαχωριστούν σωστά οι κλάσεις. Το Chebyshev KAN φαίνεται να είναι πιο ευαίσθητο στις μεσαίες και τις πολλές κλάσεις αφού σημειώνει και την πιο χαμηλή απόδοση. Στις πολλές κλάσεις το Hybrid KAN κρατά σχετικά πιο σταθερή απόδοση, ενώ το SplineKAN βρίσκεται πολύ κοντά, με ελάχιστη διαφορά.

	Ακρίβεια			
Μήκος	**HybridKAN** (128,128,64,64,32)	**SplineKAN** (128,64)	**ChebyshevKAN** (256,128,64)	**N datasets**
<200	0.789	0.785	**0.799**	14
200–700	**0.698**	0.687	0.682	20
>700	0.559	**0.560**	0.553	6

Πίνακας 5.7: Μέση ακρίβεια ανά κατηγορία μήκους χρονοσειρών

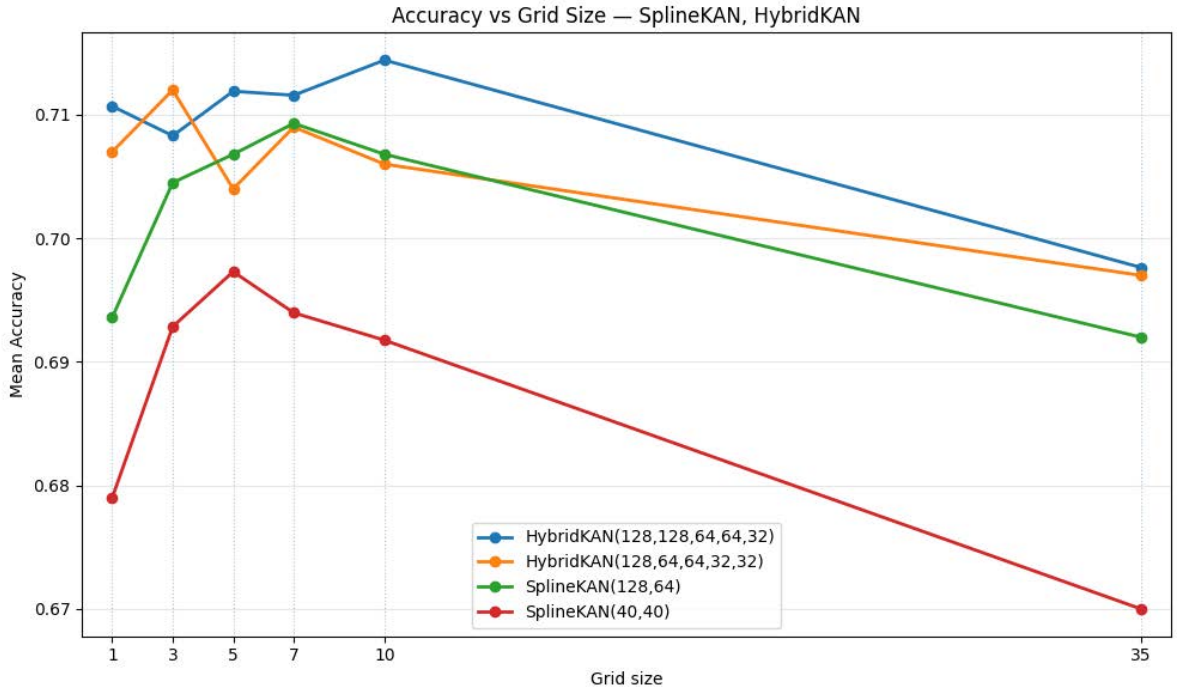
Σε σχέση με το μήκος των χρονοσειρών, η απόδοση όλων των μοντέλων εμφανίζεται υψηλότερη στα μικρότερα μήκη ενώ μειώνεται σημαντικά καθώς αυξάνεται το μέγεθος τους. Αυτό βρίσκει παρόμοια εξήγηση με τις αυξημένες κλάσεις, αφού αυξάνεται η πολυπλοκότητα και τα μοντέλα καλούνται να μάθουν σε πολύ μεγαλύτερο χώρο. Παρόλα αυτά το Chebyshev μοντέλο αποδίδει αρκετά καλύτερα στις μικρές χρονοσειρές ενώ υστερεί σημαντικά στα υπόλοιπα μεγέθη.

Τα μοντέλα συγκρίθηκαν επίσης με το μέγεθος από τα συνολικά δείγματα και τα αποτελέσματα βρίσκονται στα παραρτήματα. Το μέγεθος αυτό κρίθηκε ότι δεν παίζει τόσο σημαντικό ρόλο, αφού τα KAN χειρίζονται σχετικά λίγες παραμέτρους, και άρα δεν απαιτούν μεγάλα πλήθη για να γενικεύσουν. Αυτό επιβεβαιώθηκε και με τα αποτελέσματα, τα οποία εκφράζουν τα ίδια μοτίβα και στα τρία μοντέλα. Στα πολύ λίγα δείγματα (<200) έχουν επαρκή απόδοση με το Chebyshev να υπερτερεί, ενώ στα μεσαία μειώνεται με το spline να αποδίδει καλύτερα και στα πολλά να αυξάνονται και τα τρία.

5.2.3 Ρύθμιση Grid Υπερπαραμέτρου (Grid Tuning)

Παρουσιάζονται οι τιμές από δύο διαφορετικές ισχυρές αρχιτεκτονικές για κάθε μοντέλο Spline και Hybrid KAN σε σχέση με τα διάφορα μεγέθη grid που δοκιμάστηκαν. Αυτό το πείραμα συντελείται για να μελετήσουμε την τάση των μοντέλων στις τιμές του πλέγματος grid και να προκύψουν συμπεράσματα για το αν αυτές είναι γενικές τάσεις

των μοντέλων Spline και Hybrid KAN ή χαρακτηριστικό των συγκεκριμένων αρχιτεκτονικών.



Σχήμα 5.5: Μέση ακρίβεια για τα μεγέθη Grid στα μοντέλα Spline και Hybrid KAN

Η παρατήρηση της τιμής του grid δείχνει σαφή και επαναλαμβανόμενα μοτίβα για τα δύο μοντέλα.

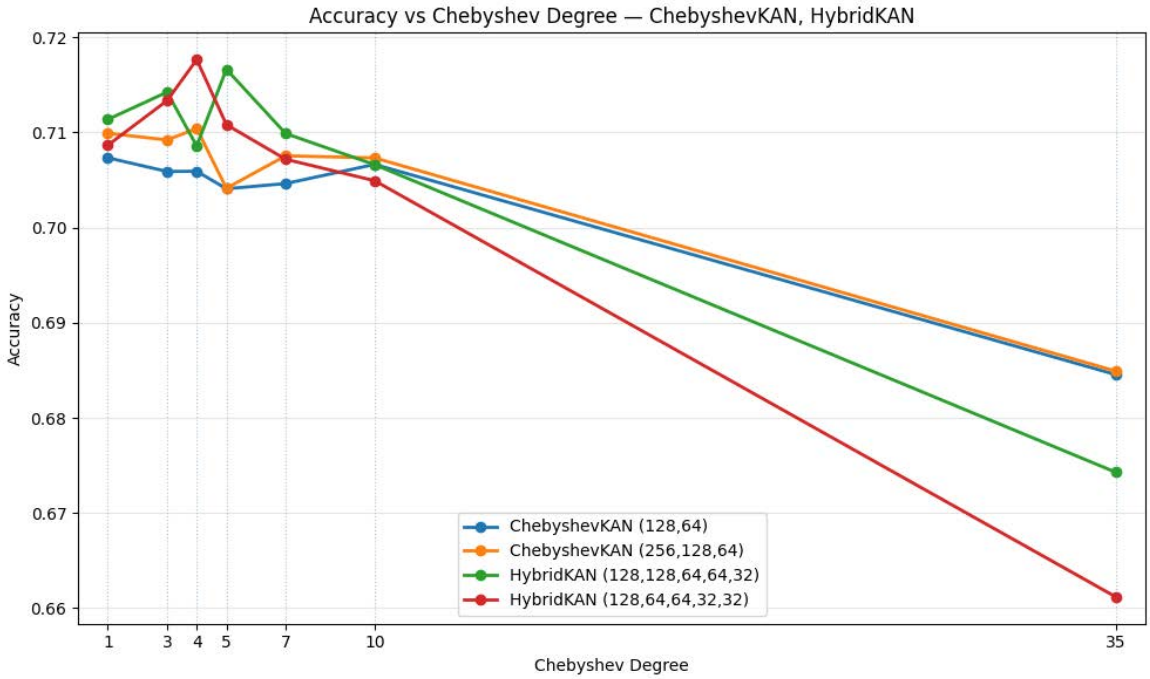
- Το υβριδικό μοντέλο αποδίδει καλύτερα σε σχέση με το spline στα μικρά grid (1–3), όπου το μοντέλο αξιοποιεί την υβριδική του αρχιτεκτονική, που του προσφέρει ήδη πολυπλοκότητα, χωρίς να χρειάζεται αυξημένο αριθμό grid σημείων. Η απόδοση σταθεροποιείται σε μεσαία grid (5–10) και παραμένει σχεδόν αμετάβλητη. Στις πολύ μεγάλες τιμές (35) η απόδοση πέφτει λόγω της εισαγωγής περαιτέρω παραμέτρων, παραμένοντας όμως πιο σταθερό από το spline. Η μεγαλύτερη αρχιτεκτονική φαίνεται να αποδίδει συνολικά καλύτερα στις μεταβολές του grid.

- Αντίθετα, το SplineKAN είναι πιο αποτελεσματικό στα μεσαία grid (5–10), καθώς οι spline βάσεις αποτελούν τον βασικό μηχανισμό εκφραστικότητας του μοντέλου. Στα πολύ μικρά grid (1–3) η ακρίβεια είναι χαμηλότερη, συμπεραίνοντας ότι δεν αρκούν

για να αποτυπώσει τα μοτίβα, ενώ στα πολύ μεγάλα η απόδοση μειώνεται ξανά, πιθανώς λόγω υπερπροσαρμογής όπως και στο υβριδικό. Αυτές φαίνεται να είναι γενικές τάσεις του μοντέλου καθώς έχουν κοινή συμπεριφορά και στις δύο αρχιτεκτονικές.

Τα αναλυτικά αριθμητικά αποτελέσματα βρίσκονται στα παραρτήματα.

5.2.4 Ρύθμιση Chebyshev Υπερπαραμέτρου (Chebyshev degree Tuning)



Σχήμα 5.6: Μέση ακρίβεια για τα μεγέθη Chebyshev στα μοντέλα Spline και Hybrid KAN

Από το διάγραμμα μπορούμε να βγάλουμε συγκεκριμένα συμπεράσματα:

Η παράμετρος του βαθμού Chebyshev, αν και θεωρητικά είναι κρίσιμη, στην πράξη φαίνεται να έχει περιορισμένη επίδραση στα συνολικά αποτελέσματα, αφού οι μεταβολές της ακρίβειας από τον βαθμό 1 έως 10 είναι της τάξης του 1%.

Το πολυωνυμικό μοντέλο φαίνεται να αποδίδει σταθερά στα μεγέθη 1-10 με μέγιστο στο 4. Παρόμοια συμπεριφορά εμφανίζει και το υβριδικό μοντέλο με την καλύτερη απόδοση να εμφανίζεται στα μεσαία μεγέθη (3-5). Η εκφραστικότητα που προσφέρουν οι βαθμοί Chebyshev φαίνονται ικανοί ακόμα και στους μικρούς βαθμούς για να περιγράψουν τις μη γραμμικότητες και να διαχωρίσουν τις κλάσεις.

Συνολικά το Υβριδικό μοντέλο φαίνεται να είναι πιο εκφραστικό και σταθερό στα μεσαία μεγέθη αλλά υστερεί σημαντικά έναντι του πολυωνυμικού Chebyshev στα μεγάλα μεγέθη. Τα αναλυτικά αριθμητικά αποτελέσματα βρίσκονται στα παραρτήματα.

Κεφάλαιο 6 - Σύνοψη και Συμπεράσματα

Σε αυτή την εργασία μελετήσαμε την επίδοση των KAN σε δύο κρίσιμους τομείς της καθημερινότητας, αυτόν της πρόβλεψης και της κατηγοριοποίησης, με την χρήση δεδομένων χρονοσειρών. Ο κύριος στόχος μας ήταν να ερευνήσουμε αν τα KAN μπορούν να σταθούν σαν μία αξιόλογη και ανταγωνιστική εναλλακτική σε σχέση με τα παραδοσιακά MLPs αλλά και τα υπόλοιπα νευρωνικά δίκτυα και στατιστικά μοντέλα. Αναλύσαμε πλήρως την συνεισφορά των δεδομένων χρονοσειρών στον χώρο της μηχανικής μάθησης και διαχειριστήκαμε τα δεδομένα μας με βήματα προ επεξεργασίας και στατιστικής ανάλυσης για την μέγιστη αξιοποίηση των ιδιοτήτων τους.

6.1 Συμπεράσματα

Η αξιολόγηση μας λαμβάνει υπόψη τις δύο κατηγορίες που μελετήσαμε και τους πολλαπλούς δείκτες απόδοσης που προέκυψαν. Θα αναπτύξουμε τα συμπεράσματα που προέκυψαν για κάθε μοντέλο KAN που ερευνήσαμε:

1. **Spline KAN**: Εμφανίζεται σαν μία σταθερή εναλλακτική κυρίως στην πρόβλεψη αφού ξεπερνά όλα τα μοντέλα αναφοράς και το αντίστοιχο Chebyshev KAN. Στην κατηγοριοποίηση υστερεί σχετικά, κυρίως απέναντι στο DTW-learned, με το οποίο βρέθηκε στατιστικά κατώτερο. Παρόλα αυτά μπορεί και προσφέρει σταθερή απόδοση στις ακραίες τιμές του UCR αρχείου, όπως τα μεσαία, μεγάλα μήκη και τις μεσαίες, πολλές κατηγορίες, αφού αξιοποιεί το grid για να προσαρμόσει τοπικά τη δομή του.

2. **Chebyshev KAN**: Φαίνεται μία ισχυρή εναλλακτική με σημαντική συνεισφορά κυρίως στην κατηγοριοποίηση που ξεπερνά το Spline μοντέλο. Παρόλα αυτά εμφανίζει ισχυρή αστάθεια στις διάφορες κατηγορίες δεδομένων. Ενώ στις μικρές κατηγορίες αποδίδει καλά στις μεσαίες και μεγάλες, με εξαίρεση τον αριθμό δειγμάτων, εμφανίζει διακυμάνσεις στην επίδοση. Αυτό μπορεί να αποδοθεί στις συναρτήσεις Chebyshev οι οποίες δυσκολεύονται να αναπαραστήσουν επαρκώς πολύ μεγάλες ακολουθίες: αφενός λόγω περιορισμένης χωρητικότητας, αφετέρου λόγω αριθμητικής αστάθειας που εμφανίζεται σε υψηλότερα μήκη.

3. Hybrid KAN: Μαζί με το MoK αποδείχτηκε το πιο ισχυρό μοντέλο αφού ξεπέρασε σε επιδόσεις τα Spline και Chebyshev KAN και φάνηκε ανώτερο κυρίως στην πρόβλεψη, σε σχέση με κλασσικά νευρωνικά δίκτυα, όπως το MLP και το LSTM. Παρουσίασε την πιο σταθερή συμπεριφορά στις διάφορες κατηγορίες δεδομένων, λόγω της αξιοποίησης της εσωτερικής του πολυπλοκότητας για να παραμείνει πιο σταθερό όταν αυξάνεται η δυσκολία.

4. MoK: Αποδείχτηκε η πιο αξιόπιστη αρχιτεκτονική αφού εμφάνισε τα πιο ισχυρά αποτελέσματα στην πρόβλεψη στην οποία και μελετήθηκε, αναδεικνύοντας την αξία του μηχανισμού experts–gate στη χρονοσειριακή πρόβλεψη. Αυτό δείχνει ότι, πρώτον η στρατηγική λεπτομέρεια και δεύτερον η εξομάλυνση, μπορούν να οδηγήσουν σε πιο ακριβείς προβλέψεις. Επιπλέον, διατήρησε πιο σταθερή συμπεριφορά σε διαφορετικές ρυθμίσεις υπερπαραμέτρων, απορροφώντας τις αδυναμίες των επιμέρους παραλλαγών.

Παρά το γεγονός ότι τα KAN εμφάνισαν κατώτερη απόδοση από τα MLP και DTW , τα αποτελέσματα του Wilcoxon κατέδειξαν ότι τα μοντέλα KAN παραμένουν ιδιαίτερα ανταγωνιστικά. Με εξαίρεση το SplineKAN έναντι του DTW-learned, δεν παρατηρήθηκε στατιστικά σημαντική υστέρηση σε σχέση με τον κλασικό ταξινομητή DTW ενώ ξεπέρασαν σε απόδοση τον ED, γεγονός που αποδεικνύει τη σταθερότητα των KAN. Η δυνατότητά τους να στέκονται ισάξια απέναντι στο DTW, έναν από τους πιο ισχυρούς παραδοσιακούς ταξινομητές, είναι αξιοσημείωτη, καθώς συνδυάζουν ανταγωνιστική ακρίβεια με το πλεονέκτημα της ερμηνευσιμότητας και του χρόνου εκπαίδευσης. Η καινοτομία τους έγκειται ακριβώς σε αυτόν τον συνδυασμό: αξιοποιούν πολυωνυμικές και spline συναρτήσεις για να αποτυπώνουν μη γραμμικές σχέσεις, προσφέροντας ένα πλαίσιο που μπορεί να επεκταθεί και να προσαρμοστεί σε πιο σύνθετα σενάρια χρονοσειρών.

Τέλος οι υβριδικές αρχιτεκτονικές κρίνονται ως οι πιο ανταγωνιστικές παραλλαγές αφού καταφέρνουν να αξιοποιήσουν από κοινού την τοπική προσαρμοστικότητα των spline και τη γενικευτική ικανότητα των Chebyshev βάσεων.

6.2 Μελλοντικές επεκτάσεις

Παρόλο που η εργασία μας συνεισέφερε στον τομέα της ανάλυσης των KAN θεωρούμε ότι είναι μία τεχνική που αξίζει να μελετηθεί περαιτέρω. Θα προτείνουμε κάποια σημεία για μελλοντική έρευνα.

Αρχικά τα ελπιδοφόρα αποτελέσματα του MoK δείχνουν ότι ακόμα και στο βάθος του ενός επιπέδου μπορεί να είναι ισχυρά ανταγωνιστικό. Η προσπάθεια μας να συνδυάσουμε τους διάφορους experts και να δημιουργήσουμε ένα βαθύτερο δίκτυο (MMK) φάνηκε να έχει ερευνητική ουσία αφού απέδωσε ισχυρά αποτελέσματα. Παρόλα αυτά η μελέτη των υβριδικών μοντέλων χρίζει περαιτέρω ενασχόλησης ως προς την επιλογή διαφορετικών KAN βάσεων πέρα των δύο που επιλέξαμε αλλά και διαφορετικών υπερπαραμέτρων.

Επιπλέον η αρχιτεκτονική των MoK έχει βρει εφαρμογή μόνο στην πρόβλεψη, κάτι που μπορεί να καταλήξει περιοριστικό στην εξέλιξη των KAN. Θεωρούμε ότι η δυνατότητα διεύρυνσης του σε άλλα πεδία όπως η ταξινόμηση, ανίχνευση κλπ. μπορεί να επεκτείνει τις υπολογιστικές δυνατότητες και να οδηγήσει τα KAN σε μία σταθερή επιλογή έναντι άλλων κλασσικών μοντέλων.

Επίσης, στην παρούσα εργασία αξιοποιήσαμε τις διάφορες ομαδοποιήσεις σχετικά με τα δεδομένα του UCR για να μπορέσουμε να κρίνουμε την απόδοση των μοντέλων. Η εξαγωγή συμπερασμάτων από το συγκεκριμένο μπορεί να προσφέρει γνώσεις, οι οποίες θα συμβάλουν για να εκπαιδευτούν ανάλογα τα μοντέλα μας καθώς και για να χρησιμοποιηθούν οι αντίστοιχες τεχνικές για να εξομαλύνουν τους τομείς υστέρησης τους.

Τέλος, η εφαρμογή των KAN νευρωνικών δικτύων σε αλγορίθμους και εφαρμογές ομόσπονδης μάθησης είναι ένα πολλά υποσχόμενο πεδίο έρευνας και ανάπτυξης [45], ιδίως στα ασύρματα περιβάλλοντα IoT των εξαιρετικά περιορισμένων πόρων [46].

Βιβλιογραφία

[1] G. G. Lorentz, *Approximation of functions, AMS Chelsea Publishing*. American Mathematical Society, Providence Rhode Island, 1966, pp.168-169.

[2] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, Max Tegmark, KAN: Kolmogorov–Arnold Networks. arXiv:2404.19756v5 [cs.LG], 30 Apr 2024, pp. 3-6.

[3] "GitHub - Blealtan/efficient-kan: An efficient pure-PyTorch implementation of Kolmogorov-Arnold Network (KAN)." Accessed: Aug. 16, 2025. [Online]. Available: https://github.com/Blealtan/efficient-kan

[4] A.N. Kolmogorov. On the representation of continuous functions of several variables in the form of superpositions of continuous functions of one variable and addition. In Reports of the Academy of Sciences, vol. 114, no. 5, pp. 953–956, 1957. Translated in American Mathematical Society Translations, Series 2, vol. 28, 1963, pp. 55–59.

[5] V.I. Arnold. *On the representation of functions of several variables as a superposition of functions of a smaller number of variables*. In *Collected Works: Representations of Functions, Celestial Mechanics and KAM Theory, 1957–1965*, Springer, 2009, pp. 25–46.

[6] C. de Boor, *A Practical Guide to Splines*, 2nd ed., New York, In NY: Springer, 2001, pp. 87–117.[2]

[7] Sidharth SS , Gokul R , Anas K P , and Keerthana AR, *Chebyshev Polynomial-Based Kolmogorov–Arnold Networks: An Efficient Architecture for Nonlinear Function Approximation*. DOI: 10.48550/arXiv.2405.07200, 2024.

[8] P. L. Chebyshev, Théorie des mécanismes connus sous le nom de parallélogrammes, In Mém. Acad. Imp. Sci. St.-Pétersbourg, tom. VII, pp. 539–568, 1854. Reprinted in *Oeuvres de P. L. Tchebychef*, vol. I, New York: Chelsea, 1961, pp. 111–143.

[9] J. Demšar, “Statistical Comparisons of Classifiers over Multiple Data Sets,” *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[10] H. Zhang, Z. Huang, and Y. Wang, AC-PKAN: Attention-Enhanced and Chebyshev Polynomial-Based Physics-Informed Kolmogorov-Arnold Networks, In arXiv preprint arXiv:2505.08687, May 2025, pp:3,5

[11] Ziyao Li, fast-kan: Fast implementation of kernel attention networks (kan), [Online]. Available: https://github.com/ZiyaoLi/fast-kan, 2024.

[12] A. Afzalaghaei, fKAN: Fast Kernel Attention Networks (KAN) implementation with PyTorch. [Online]. Available: https://github.com/alirezaafzalaghaei/fKAN, 2024.

[13] A. Afzalaghaei, rKAN: Implementation of Kernel Attention Networks (KAN) with PyTorch. [Online]. Available: https://github.com/alirezaafzalaghaei/rKAN, 2024.

[14] GistNoesis, FourierKAN: A GitHub repository, [Online] Available: https://github.com/GistNoesis/FourierKAN, 2024.

[15] Sainath Dey, Mitul Goswami, Jashika Sethi, Prasant Kumar Pattnaik, Hyb-KAN ViT: Hybrid Kolmogorov-Arnold Networks Augmented Vision Transformer, In arXiv preprint arXiv:2505.04740, May 2025, pp: 3-7

[16] X. Han, X. Zhang, Y. Wu, Z. Zhang, and Z. Wu, Are KANs Effective for Multivariate Time Series Forecasting? In University of Chinese Academy of Sciences & Pengcheng Laboratory, China, 2024, pp:3-7

[17] C. J. Vaca-Rubio, L. Blanco, R. Pereira, M. Caus, Kolmogorov-Arnold Networks (KANs) for Time Series Analysis, In Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Barcelona, Spain, 2024, pp: 1-4

[18] M. Caus, C. J. Vaca-Rubio, L. Blanco, and R. Pereira, KAN4TSF: Are KAN and KAN-Based Models Effective for Time Series Forecasting? In Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), 2024, pp:1-3

[19] S. Lee, J.-K. Kim, J. Kim, T. Kim, and J. Lee, HiPPO-KAN: Efficient KAN Model for Time Series Analysis, In arXiv preprint arXiv:2410.14939, Oct. 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2410.14939

[20] I. Barasin, B. Bertalanič, M. Mohorčič, and C. Fortuna, Exploring Kolmogorov-Arnold Networks for Interpretable Time Series Classification, In Department of Communication Systems, Jožef Stefan Institute, Ljubljana, Slovenia, 2024.

[21] S. S. Costa, M. Pato, and N. Datia, An Empirical Study on the Application of KANs for Classification, In ICAAI'24 - The 8th International Conference on Advances in Artificial Intelligence, 2024, London, United Kingdom, 2024. ACM, ISBN: 979-8-4007-1801-4.

[22] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series featuRe extraction on basis of scalable hypothesis tests (tsfresh – a Python package)," *Neurocomputing*, vol. 307, Sep. 2018, pp. 72–77, doi: 10.1016/j.neucom.2018.03.067.

[23] Fulcher Ben D., and Nick S. Jones. "Hctsa : A Computational Framework for Automated Time-Series Phenotyping Using Massive Feature Extraction." Cell Systems, vol. 5, no. 5, Nov. 2017, pp. 527-531.e3, https://doi.org/10.1016/j.cels.2017.10.001.

[24] C. H. Lubba, S. S. Sethi, P. Knaute, S. R. Schultz, B. D. Fulcher, and N. S. Jones, "catch22: Canonical time-series characteristics," *Data Mining and Knowledge Discovery*, vol. 33, no. 6, pp. 1821–1852, 2019, doi: 10.1007/s10618-019-00647-x.

[25] Sepp Hochreiter and Jürgen Schmidhuber,"Long short-term memory," Neural computation, vol. 9, no. 8, 1997, pp. 1735–1780.

[26] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis : Forecasting and Control*. Hoboken, New Jersey, John Wiley & Sons, Inc., 2016, pp: 90-95

[27] Andrés Herranz González, "TimeSeries Analysis, a Complete Guide." *Kaggle.com*, 2021, www.kaggle.com/code/andreshg/timeseries-analysis-a-complete-guide.

[28] H. A. Dau, A. Bagnall, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, and E. Keogh, "The UCR time series archive," *arXiv preprint*, arXiv:1810.07758, 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1810.07758, pp: 1-4

[29] Rob J Hyndman, George Athanasopoulos, Forecasting: Principles and Practice, 2nd Edition, Monash University, Australia, 2018, ISBN-10: 0987507117

[30] D. Mindrila and M. E. Phoebe, “Tests of Significance.”, Based on Chapter 15 of The Basic Practice of Statistics (6th ed.)

[31] I. Gandin, A. Scagnetto, S. Romani, and G. Barbati, “Interpretability of time-series deep learning models: A study in cardiovascular patients admitted to Intensive care unit,” *J Biomed Inform*, vol. 121, p. 103876, pp:1-4, Sep. 2021, doi: 10.1016/J.JBI.2021.103876.

[32] D. I. Serramazza, T. Le Nguyen, and G. Ifrim, “Improving the Evaluation and Actionability of Explanation Methods for Multivariate Time Series Classification,” Aug. 2024, Accessed: Jul. 23, 2025. [Online]. Available: https://arxiv.org/pdf/2406.12507v1

[33] A. A. Ismail, M. Gunady, H. C. Bravo, and S. Feizi, “Benchmarking Deep Learning Interpretability in Time Series Predictions,” *Adv Neural Inf Process Syst*, vol. 2020-December, Oct. 2020, Accessed: Jul. 23, 2025. [Online]. Available: https://arxiv.org/pdf/2010.13924

[34] “What Is Support Vector Machine? | IBM.” Accessed: Jul. 23, 2025. [Online]. Available: https://www.ibm.com/think/topics/support-vector-machine

[35] M. S. Nidhya, S. Verma, H. B. A. Mohamed, and T. Agarwal, “Investigating Naive Bayes Algorithms for Network Time Series Analysis,” *Lecture Notes in Electrical Engineering*, vol. 1274 LNEE, pp. 227–232, 2025, doi: 10.1007/978-981-97-8043-3_36.

[36] J. Lines, S. Taylor, and A. Bagnall, “Time series classification with HIVE-COTE: The hierarchical vote collective of transformation-based ensembles,” *ACM Trans Knowl Discov Data*, vol. 12, no. 5, Jul. 2018, doi: 10.1145/3182382.

[37] “Daily minimum temperatures.” Accessed: Jul. 23, 2025. [Online]. Available: https://www.kaggle.com/datasets/suprematism/daily-minimum-temperatures

[38] “GitHub - KindXiaoming/pykan: Kolmogorov Arnold Networks.” Accessed: Aug. 16, 2025. [Online]. Available: https://github.com/KindXiaoming/pykan

[39] “GitHub – Sidhu2690/ChebyshevKAN: Chebyshev Polynomial-Based Kolmogorov-Arnold Networks”. Accessed: May 2024, [Online]. Available: https://github.com/sidhu2690/ChebyshevKAN

[40] H. Xiao, Z. Xinfeng, W. Yiling, Z. Zhenduo, and W. Zhe, “GitHub - 2448845600/EasyTSF: Experiment ASsistance for Your Time-Series Forecasting, EasyTSF.” Accessed: 2024. [Online]. Available: https://github.com/2448845600/EasyTSF

[41] S. Willson, “KAN Time-Series Experiment.” Accessed: Sep. 04, 2025. [Online]. Available: https://www.kaggle.com/code/somdudewillson/kan-time-series-experiment

[42] I. Barašin, B. Bertalanič, M. Mohorčič, and C. Fortuna, “Exploring Kolmogorov-Arnold Networks for Interpretable Classification,” *arXiv preprint arXiv:2411.14904*, 2024. [Online]. Available: https : //github.com/irina − b1/KANT S.

[43] Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. International Journal of Forecasting, 37(1), 388–427. https://doi.org/10.1016/J.IJFORECAST.2020.06.008

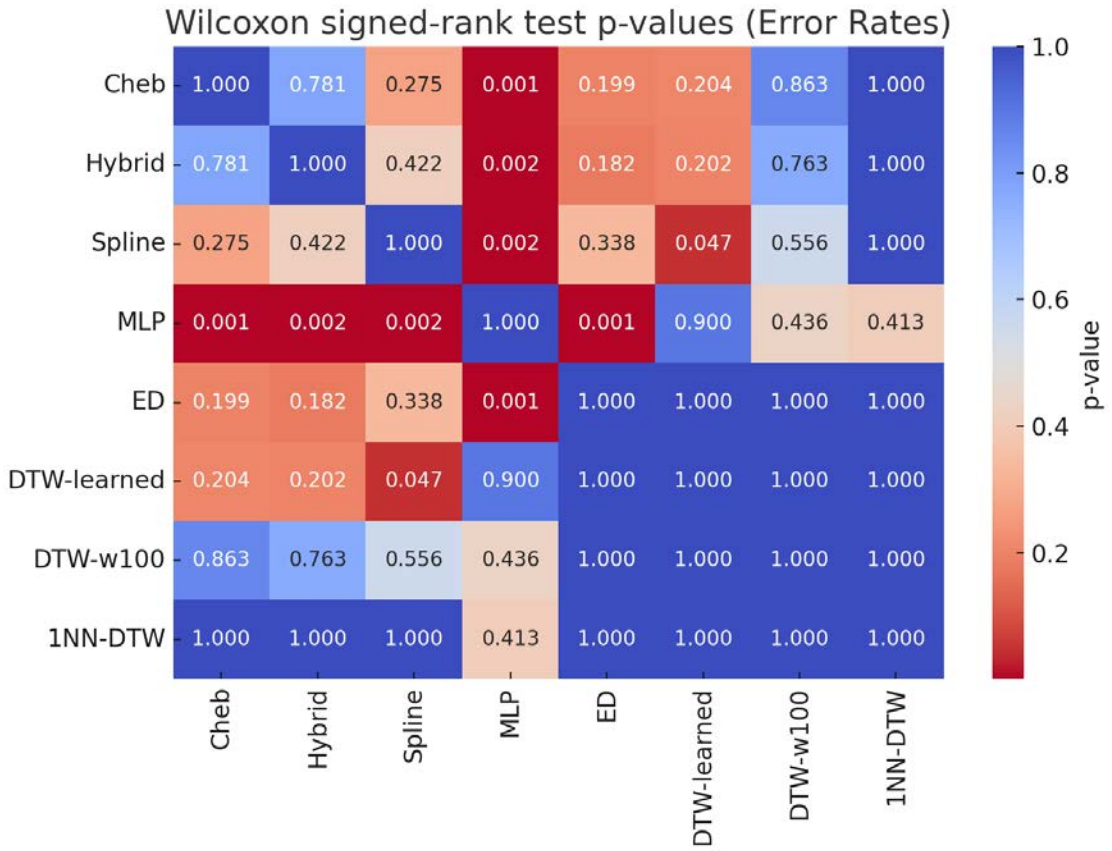
[44] Y. Peng, Y. Wang, F. Hu, M. He, Z. Mao, X. Huang, and J. Ding., Predictive modeling of flexible EHD pumps using Kolmogorov -Arnold networks, Biomimetic Intelligence and Robotics 4, 4 (2024), 100184, pp: 1-3.

[45] Y. Ma, Z. Yang, L.-D. Ibanez, “Enhancing Federated Learning with Kolmogorov-Arnold Networks: A Comparative Study Across Diverse Aggregation Strategies”, *arXiv preprint arXiv:2505.07629v1 , 2025.*

[46] E. Fragkou, E. Chini, M. Papadopoulou, D. Papakostas, D.Katsaros, S. Dustdar, “Distributed Federated Deep Learning in Clustered Internet of Things Wireless Networks with Data Similarity-based Client Participation”, *IEEE Internet Computing magazine*, vol. 28, no. 6, pp. 53-61, 2024.

Παράρτημα Α

Πρόσθετα Αποτελέσματα



Σχήμα Α1: Heatmap με p-values από το Wilcoxon signed-rank test για τα μοντέλα.

	Ακρίβεια			
Αριθμός Δειγμάτων	**HybridKAN** (128,128,64,64,32)	**SplineKAN** (128,64)	**ChebyshevKAN (256,128,64)**	**N datasets**
<200	0.747	0.716	**0.751**	8
200–1000	0.670	**0.673**	0.649	21
>1000	0.757	0.749	**0.767**	11

Πίνακας Α1: Αξιολόγηση Ακρίβειας σε σχέση με τον αριθμό δειγμάτων

	Μέση ακρίβεια (Accuracy)			
Grid size	**SplineKAN (40,40)**	**SplineKAN (128,64)**	**HybridKAN (128,64,64,32,32)**	**HybridKAN (128,128,64,64,32)**
1	0.6790	0.6936	0.707	0.7107
3	0.6928	0.7045	**0.712**	0.7083
5	**0.6973**	0.7068	0.704	0.7119
7	0.6939	**0.7093**	0.709	0.7116
10	0.6915	0.7068	0.706	**0.7144**
35	0.6700	0.6920	0.697	0.6976

Πίνακας Α.2: Μέση ακρίβεια για τα μεγέθη Grid στην κατηγοριοποίηση

	Μέση ακρίβεια (Accuracy)			
Chebyshev Degree	**ChebyshevKAN (128,64)**	**ChebyshevKAN (256,128,64)**	**HybridKAN (128,64,64,32,32)**	**HybridKAN (128,128,64,64,32)**
1	**0.7073**	0.7099	0.7086	0.7113
3	0.7059	0.7092	0.7133	0.7142
4	0.7059	**0.7104**	**0.7176**	0.7086
5	0.7040	0.7041	0.7108	**0.7166**
7	0.7046	0.7075	0.7072	0.7098
10	0.7066	0.7073	0.7049	0.7066
35	0.6845	0.6849	0.6612	0.6742

Πίνακας Α.3: Μέση Ακρίβεια για τα μεγέθη Chebyshev στην κατηγοριοποίηση