



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«Σύγχρονες Τεχνολογίες Διαχείρισης Περιβαλλοντικών Κινδύνων»

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

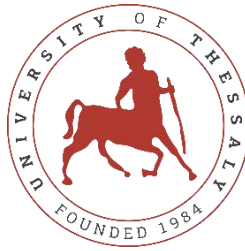
ΜΕΛΕΤΗ ΑΛΓΟΡΙΘΜΩΝ ΠΡΟΒΛΕΨΗΣ ΚΑΙ ΠΑΡΑΔΕΙΓΜΑ ΕΦΑΡΜΟΓΗΣ ΣΕ MATLAB ΓΙΑ ΤΗΝ ΠΡΟΒΛΕΨΗ ΠΛΗΜΜΥΡΩΝ

ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΦΟΙΤΗΤΗ: ΣΔΡΟΥΛΙΑΣ ΓΕΩΡΓΙΟΣ

ΑΜ: M061923013

Επιβλέπουσα Καθηγήτρια:
(*Παπαγεωργίου Ελπινίκη*)

Βόλος, Ιανουάριος 2025



UNIVERSITY OF THESSALY

INTERDEPARTMENTAL POSTGRADUATE STUDIES PROGRAMME
“Advanced Technologies for Environmental Risk Management”

POSTGRADUATE MASTER’S THESIS

Study of Prediction Algorithms and Example of Application in MATLAB for Flood Forecasting

Full Name: SDROULIAS GEORGIOS

PN: M061923013

Supervisor:
(Elpiniki Papageorgiou)

Volos, January 2025

Ονοματεπώνυμο Συγγραφέα: Σδρούλιας Γεώργιος

Ιδιότητα: Πολιτικός Μηχανικός

Copyright © Σδρούλιας Γεώργιος (2025)

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται στο συγγραφέα.

Το περιεχόμενο της διπλωματικής εργασίας εκφράζει τις απόψεις του/της συγγραφέα και ουδεμία ευθύνη για το περιεχόμενο φέρουν ο/η επιβλέπων/ουσα, και η εξεταστική επιτροπή για αυτό.

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Μεταπτυχιακής Διπλωματικής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή.

Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης.

Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στη Διπλωματική μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Μεταπτυχιακού Τίτλου των Μεταπτυχιακών Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων.

Δηλώνω, συνεπώς, ότι αυτή η Μ.Δ.Ε. προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

Ονοματεπώνυμο Συγγραφέα: Σδρούλιας Γεώργιος

Υπογραφή:

Ημερομηνία: 24/02/2025

Περιεχόμενα

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ	1
ΚΑΤΑΛΟΓΟΣ ΚΩΔΙΚΑ ΣΤΟ MATLAB	2
ΚΑΤΑΛΟΓΟΣ ΕΞΙΣΩΣΕΩΝ	3
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	4
ΠΕΡΙΛΗΨΗ	5
1. ΕΙΣΑΓΩΓΗ	6
1.1 Παρουσίαση του προβλήματος	6
1.2 Αντικείμενο της εργασίας	7
2. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ	9
2.1 Γεωγραφικά Χαρακτηριστικά της περιοχής μελέτης (Geographical Characteristics)	9
2.2 Μοντέλο Βροχής-Απορροής και Πρόβλεψη Στάθμης Νερού	9
2.3 Κατηγορίες μοντέλων πρόβλεψης στάθμης νερού	10
2.3.1 Φυσικά Μοντέλα	10
2.3.2 Εμπειρικά μοντέλα	10
2.4 Εισαγωγή στα τεχνητά νευρωνικά δίκτυα	11
2.4.1 Κόμβος Άθροισης (Summation Node)	12
2.4.2 Συνάρτηση Ενεργοποίησης	12
2.4.3 Συναρτήσεις απώλειας	13
2.5 Κατηγορίες Νευρωνικών Δικτύων:	14
2.6 Αναδρομικά νευρωνικά δίκτυα (RNN)	15
2.7 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (Long Short-Term Memory - LSTM)	15
3. ΣΥΛΛΟΓΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ	18
4. ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ	20
4.1 Εισαγωγή των δεδομένων	20
4.2 Συγχώνευση των δεδομένων σε κοινό αρχείο	20

4.3	Αντικατάσταση ελλιπών τιμών	21
4.4	Κανονικοποίηση των δεδομένων.....	22
4.5	Δημιουργία ενός κανονικοποιημένου αρχείου	22
4.6	Λογαρίθμηση των τιμών της βροχής.....	23
5.	ΣΥΝΟΨΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ.....	26
6.	ΠΡΟΕΤΟΙΜΑΣΙΑ ΔΕΔΟΜΕΝΩΝ.....	30
6.1	Δημιουργία Σετ Δεδομένων με Παράθυρα Χρόνου	30
6.2	Διαχωρισμός Δεδομένων σε Εκπαίδευση, Επικύρωση και Δοκιμή	32
7.	ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	34
7.1	Αποτελέσματα Ανάλυσης για Μοντέλο LSTM	35
7.2	Δοκιμή LSTM με 128 κρυφά στρώματα.....	38
7.3	Δοκιμή με 256 κρυφά στρώματα	39
7.4	Αρχιτεκτονική με Δύο LSTM Στρώματα	40
7.5	Αρχιτεκτονική CNN-LSTM με Δύο LSTM Στρώματα	42
7.6	Γραμμικό μοντέλο LINEAR MODEL	45
8.	ΑΝΑΚΕΦΑΛΑΙΩΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	47
9.	ΔΟΚΙΜΕΣ ΜΕ ΠΕΡΙΣΣΟΤΕΡΕΣ ΕΠΟΧΕΣ ΕΚΠΑΙΔΕΥΣΗΣ	48
10.	ΣΥΜΠΕΡΑΣΜΑΤΑ	50
11.	ΒΙΒΛΙΟΓΡΑΦΙΑ	52

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1 Συγκριτική ανάλυση των φυσικών καταστροφών κατά τις περιόδους 1980–1999 και 2000–2019. (IPCC, 2012)	6
Εικόνα 2 Αριθμός πλημμυρών κατά την περίοδο 1970-2010 (Guha-Sapir D, 2010)	7
Εικόνα 3 Γεωγραφική περιοχή μετρήσεων (Google Earth, n.d.)	9
Εικόνα 4 Σχηματική απεικόνιση ενός νευρωνικού δικτύου (Amit, 1989)	11
Εικόνα 5 Η εξασθένηση των συσχετίσεων στα RNN (Graves, 2012).....	15
Εικόνα 6 Διατήρηση της πληροφορίας στα δίκτυα LSTM. (Graves, 2012)	16
Εικόνα 7 Μετεωρολογικός σταθμός με βροχόμετρο (Ιστός Μετεωρολογικού σταθμού Οινοφύτων, χ.χ.)	18
Εικόνα 8 Αισθητήρας μέτρησης στάθμης και ταχύτητας ροής.....	19
Εικόνα 9 Ηλιακός συλλέκτης και ηλεκτρονική μονάδα επικοινωνίας.....	19
Εικόνα 10 Εμφάνιση λογαριθμισμένων δεδομένων.....	24
Εικόνα 11 Πρώτες δέκα μετρήσεις του αρχείου 'df_norm':.....	25
Εικόνα 13 Διάγραμμα στάθμης, ταχύτητας και βροχόπτωσης στις δύο γέφυρες	28
Εικόνα 14 Παράδειγμα ενός cell array.....	30
Εικόνα 15 Διαχωρισμός Δεδομένων σε Εκπαίδευση, Επικύρωση και Δοκιμή	32
Εικόνα 16 Training and Validation Loss LSTM με 50 κρυφά στρώματα, γέφυρα Φρατζή.....	35
Εικόνα 17 Διάγραμμα διασποράς LSTM με 50 κρυφά στρώματα, γέφυρα Φρατζή	36
Εικόνα 18 Πραγματικές και προβλεπόμενες τιμές LSTM με 50 κρυφά στρώματα, γέφυρα Φρατζή ..	36
Εικόνα 19 Training and Validation Loss LSTM με 50 κρυφά στρώματα, γέφυρα Κωσταλέξη	37
Εικόνα 20 Διάγραμμα διασποράς LSTM με 50 κρυφά στρώματα, γέφυρα Κωσταλέξη.....	37
Εικόνα 21 Πραγματικές και προβλεπόμενες τιμές LSTM με 50 κρυφά στρώματα, γέφυρα Κωσταλέξη	38
Εικόνα 22 Διάγραμμα διασποράς για δύο LSTM στρώματα, γέφυρα Φρατζή	41
Εικόνα 23 Διάγραμμα διασποράς για δύο2 LSTM στρώματα, γέφυρα Κωσταλέξη.....	42
Εικόνα 24 Διάγραμμα διασποράς για CNN-LSTM με δύο LSTM στρώματα, γέφυρα Φρατζή	43
Εικόνα 25 Διάγραμμα διασποράς για CNN-LSTM με δύο LSTM στρώματα, γέφυρα Κωσταλέξη.....	44
Εικόνα 26 Διάγραμμα διασποράς, γραμμικό μοντέλο γέφυρα Φρατζή.....	46
Εικόνα 27 Σύγκριση Πραγματικών και Προβλεπόμενων Τιμών, γραμμικό μοντέλο γέφυρα Φρατζή .	46

ΚΑΤΑΛΟΓΟΣ ΚΩΔΙΚΑ ΣΤΟ MATLAB

Κώδικας 1 Φόρτωση των δεδομένων	20
Κώδικας 2 Συγχώνευση των δεδομένων.....	21
Κώδικας 3 Αντικατάσταση ελλιπών τιμών	21
Κώδικας 4 Κανονικοποίηση των δεδομένων	22
Κώδικας 5 Δημιουργία αρχείου με κανονικοποιημένα δεδομένα	23
Κώδικας 6 Λογαρίθμηση των τιμών της βροχής.....	23
Κώδικας 7 Στατιστική σύνοψη	27
Κώδικας 8 Διάγραμμα στάθμης, ταχύτητας και βροχόπτωσης	29
Κώδικας 9 Δημιουργία Σειτ Δεδομένων με Παράθυρα Χρόνου	31
Κώδικας 10 Διαχωρισμός Δεδομένων σε Εκπαίδευση, Επικύρωση και Δοκιμή	33

ΚΑΤΑΛΟΓΟΣ ΕΞΙΣΩΣΕΩΝ

Εξίσωση 1 Υπολογισμός της εξόδου του κόμβου άθροισης.....	12
Εξίσωση 2 Σχέση εξόδου (u) με πόλωση (b)	12
Εξίσωση 3 Συνάρτηση Ενεργοποίησης.....	12
Εξίσωση 4 Συναρτήσεις Ενεργοποίησης	13
Εξίσωση 5 Μέσο Απόλυτο Σφάλμα	14
Εξίσωση 6 Μέσο Τετραγωνικό Σφάλμα	14
Εξίσωση 7 min-max κανονικοποίηση.....	22

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1 Στατιστική σύνοψη των στοιχείων.....	26
Πίνακας 2 LSTM με 50 κρυφά στρώματα.....	35
Πίνακας 3 LSTM με 128 κρυφά στρώματα.....	38
Πίνακας 4 LSTM με 256 hidden units.....	39
Πίνακας 5 Αποτελέσματα για όλα τα κρυφά στρώματα για τη Γέφυρα Φρατζή	39
Πίνακας 6 Αποτελέσματα για όλα τα κρυφά στρώματα για τη Γέφυρα Κωσταλέξη	39
Πίνακας 7 Αρχιτεκτονική με Δύο LSTM Στρώματα	40
Πίνακας 8 Αρχιτεκτονική CNN-LSTM με Δύο LSTM Στρώματα	42
Πίνακας 9 Σύγκριση LSTM και CNN-LSTM για τη γέφυρα Φρατζή	44
Πίνακας 10 Σύγκριση LSTM και CNN-LSTM για τη γέφυρα Κωσταλέξη	44
Πίνακας 11 Γραμμικό μοντέλο	45
Πίνακας 12 Ανακεφαλαίωση αποτελεσμάτων γέφυρα Φρατζή.....	47
Πίνακας 13 Ανακεφαλαίωση αποτελεσμάτων γέφυρα Κωσταλέξη	47
Πίνακας 14 Διαφορετικός αριθμός εποχών γέφυρα Φρατζή.....	48
Πίνακας 15 Διαφορετικός αριθμός εποχών γέφυρα Κωσταλέξη	48

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία επικεντρώνεται στην πρόβλεψη πλημμυρικών φαινομένων στην περιοχή του ποταμού Σπερχειού, με έμφαση στις γέφυρες Φραντζή και Κωσταλέξη. Στην παρούσα εργασία, το MATLAB χρησιμοποιήθηκε ως βασικό εργαλείο για την εισαγωγή, προεπεξεργασία και ανάλυση δεδομένων, καθώς και για την εφαρμογή τεχνικών μηχανικής μάθησης. Η ανάλυση πραγματοποιήθηκε χρησιμοποιώντας δεδομένα βροχοπτώσης, ταχύτητας ροής και στάθμης του ποταμού, τα οποία επεξεργάστηκαν και αναλύθηκαν στο MATLAB. Αρχικά, γίνεται περιγραφή του προβλήματος και του τρόπου συλλογής και μέτρησης των δεδομένων. Επειτα περιγράφεται η διαδικασία προεπεξεργασίας των δεδομένων όπως η κανονικοποίηση και η διαχείριση ελλειπών τιμών. Στη συνέχεια, πραγματοποιήθηκε δοκιμή διαφορετικών προσεγγίσεων πρόβλεψης, συμπεριλαμβανομένων Γραμμικών Μοντέλων και νευρωνικών δικτύων όπως το LSTM (Long Short-Term Memory), τα οποία είναι κατάλληλα για ανάλυση δεδομένων χρονοσειρών. Η εργασία επισημαίνει τη σημασία της ακριβούς πρόβλεψης των πλημμυρών για την αποτελεσματική διαχείριση φυσικών καταστροφών και την προστασία κρίσιμων υποδομών. Τα αποτελέσματα δείχνουν ότι τα νευρωνικά δίκτυα επιδεικνύουν υψηλή ικανότητα πρόβλεψης μελλοντικών τιμών στάθμης, παρέχοντας πολύτιμα δεδομένα για την αντιμετώπιση πλημμυρικών κινδύνων.

Abstract

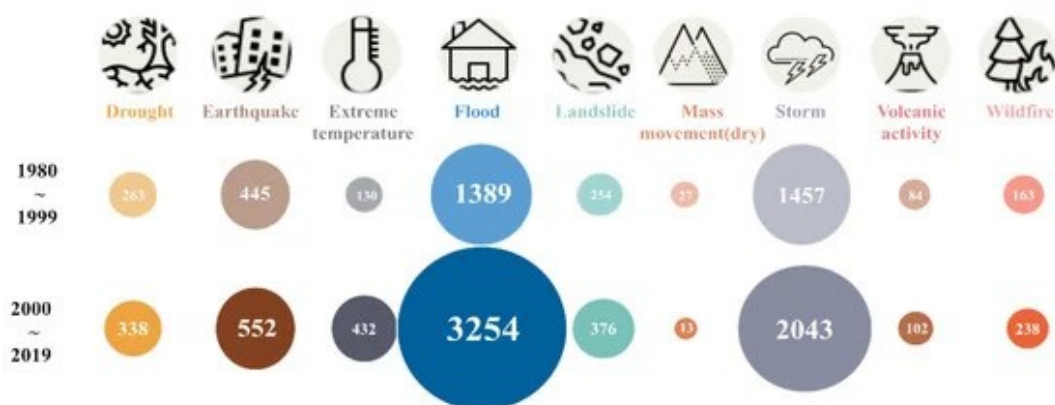
This thesis focuses on the prediction of flood events in the Sperchios River region, with emphasis on the Frantzi and Kostaleksi bridges. The study utilizes MATLAB for data preprocessing, analysis, and the application of machine learning techniques. The analysis was conducted using rainfall, flow velocity, and river level data, which were processed and analyzed in MATLAB. Initially, the method of data collection and measurement is described. Subsequently, the data preprocessing procedures, including normalization and handling of missing values, are presented. Following this, various prediction approaches were tested, including Linear Models and neural networks such as LSTM (Long Short-Term Memory), which are particularly suitable for time series analysis. The study highlights the importance of accurate flood prediction for the effective management of natural disasters and the protection of critical infrastructure. The results demonstrate that neural networks exhibit a high predictive capability for future water level values, providing valuable insights for mitigating flood risks.

1. ΕΙΣΑΓΩΓΗ

Η παρούσα διπλωματική εργασία βασίστηκε στην επιστημονική μελέτη [Applying LSTM for River's Water Level Forecasting as Potential Flood Indication: A Case Study in Greece], που δημοσιεύτηκε στο IEEE Xplore (DOI: [10.1109/IISA62523.2024.10786618]). Date Added to IEEE Xplore: 18 December 2024. Η μελέτη αυτή αποτέλεσε τη βάση για την ανάπτυξη του προτεινόμενου μοντέλου πρόβλεψης στάθμης νερού, αξιοποιώντας σύγχρονες μεθόδους μηχανικής μάθησης και ανάλυσης χρονοσειρών. Τα δεδομένα που χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία συλλέχθηκαν κατά τη διάρκεια υλοποίησης του έργου «ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΑΝΑΠΤΥΞΗ ΠΙΛΟΤΙΚΟΥ ΣΥΣΤΗΜΑΤΟΣ ΕΓΚΑΙΡΗΣ ΠΡΟΕΙΔΟΠΟΙΗΣΗΣ ΠΛΗΜΜΥΡΙΚΩΝ ΦΑΙΝΟΜΕΝΩΝ ΣΤΟ ΣΠΕΡΧΕΙΟ ΠΟΤΑΜΟ», με επιστημονικό υπεύθυνο τον Κ. Ζάρδα Γεώργιο και αναπληρώτρια επιστημονική υπεύθυνη την κ. Παπαγεωργίου Ελπινίκη. Τα δεδομένα αυτά προορίζονται αποκλειστικά για την παρούσα εργασία και δεν επιτρέπεται η χρήση, η αναπαραγωγή ή η διανομή τους από τρίτους χωρίς την έγγραφη άδεια των υπευθύνων του έργου.

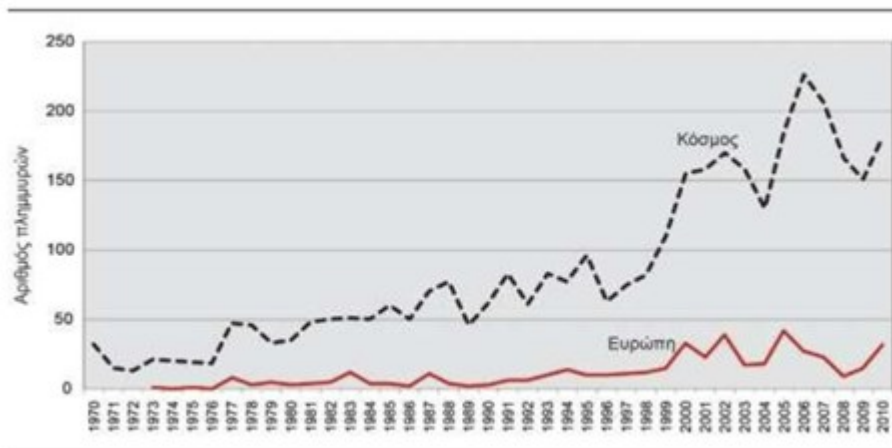
1.1 Παρουσίαση του προβήματος

Σε παγκόσμιο επίπεδο, οι επιπτώσεις των φυσικών καταστροφών αυξάνονται λόγω της κλιματικής αλλαγής που προκαλείται από την υπερθέρμανση του πλανήτη. Η συχνότητα και η ένταση αυτών των καταστροφών αναμένεται να αυξηθούν τα επόμενα χρόνια (IPCC, 2012). Η Εικόνα 1 παρουσιάζει μια συγκριτική ανάλυση των φυσικών καταστροφών μεταξύ δύο χρονικών περιόδων: 1980–1999 και 2000–2019. Οι καταστροφές κατηγοριοποιούνται σε ξηρασία, σεισμό, ακραίες θερμοκρασίες, πλημμύρα, κατολίσθηση, μαζική μετακίνηση (εδάφους), καταιγίδα, ηφαιστειακή δραστηριότητα και πυρκαγιά, με τον αριθμό των περιστατικών για κάθε κατηγορία.



Εικόνα 1 Συγκριτική ανάλυση των φυσικών καταστροφών κατά τις περιόδους 1980–1999 και 2000–2019. (IPCC, 2012)

Όπως μπορούμε να διακρίνουμε από την εικόνα 1 οι πλημμύρες έχουν υπερδιπλασιαστεί, από 1.389 σε 3.254, γεγονός που υποδεικνύει μια σημαντική αύξηση. Στην εικόνα 2 μπορούμε να διακρίνουμε την αύξηση των πλημμυρών στην Ευρώπη και στον κόσμο την περίοδο 1970-2010.



Εικόνα 2 Αριθμός πλημμυρών κατά την περίοδο 1970-2010 (Guha-Sapir D, 2010)

1.2 Αντικείμενο της εργασίας

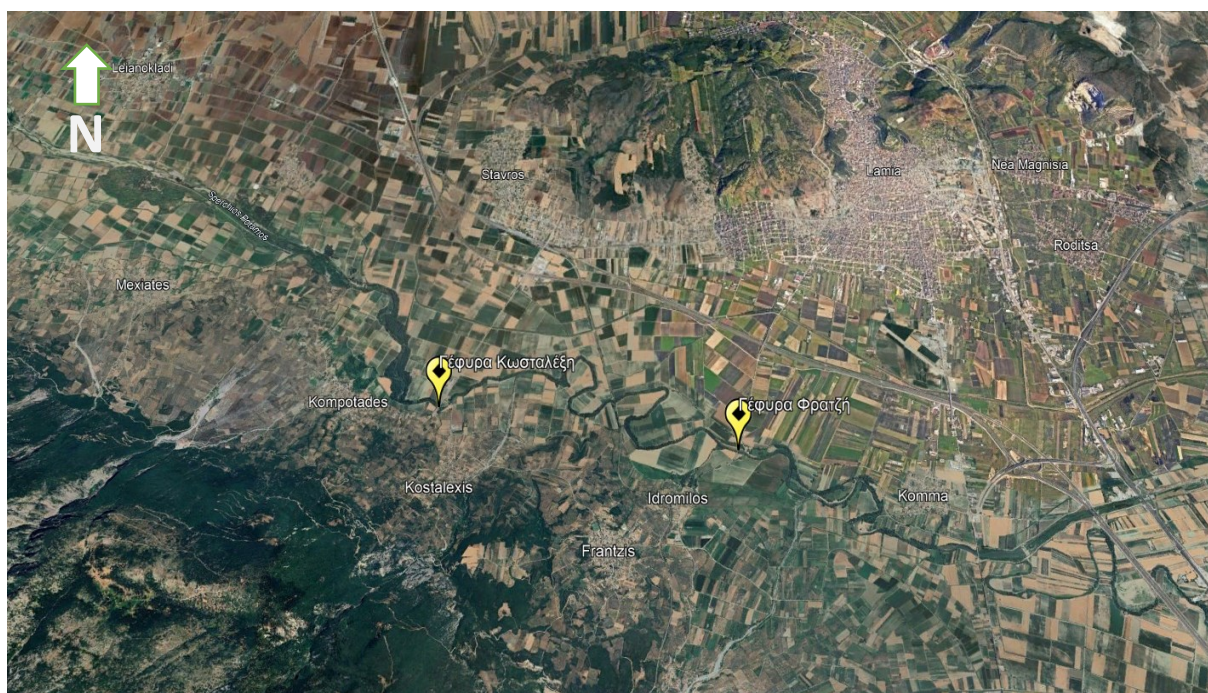
Για να μειωθούν οι ζημιές που προκαλούνται από πλημμύρες, είναι απαραίτητο αυτές να προβλέπονται γρήγορα, με ακρίβεια και να εκκενώνονται εγκαίρως οι άνθρωποι και οι περιουσίες που βρίσκονται στις πληγείσες περιοχές (Lee, 2016). Ωστόσο, η πρόβλεψη των πλημμυρών παρουσιάζει πολλές δυσκολίες, καθώς υπάρχουν πολλές μεταβλητές που πρέπει να ληφθούν υπόψη, συμπεριλαμβανομένων των χωρικών και χρονικών συσχετίσεων. Αυτό το πρόβλημα αντιμετωπίζεται με την ανάπτυξη σύγχρονων αλγορίθμων πρόβλεψης που βασίζονται στη χρήση τεχνητής νοημοσύνης και ανάλυσης δεδομένων. Η ενδελεχής παρατήρηση των φυσικών φαινομένων σε συνδυασμό με τις σύγχρονες τεχνολογικές εξελίξεις καθιστούν διαθέσιμους τεράστιους όγκους πληροφορίας. Πρόκειται για δεδομένα τα οποία αποθηκεύονται σε ψηφιακή μορφή και προσφέρονται για επιστημονική μελέτη. Η ανάπτυξη αλγορίθμων μηχανικής μάθησης επιδιώκει την εξαγωγή προτύπων από δεδομένα και παραμέτρους του παρελθόντος με σκοπό τη δυνατότητα πρόβλεψης μελλοντικών καταστάσεων. Στο πλαίσιο της παρούσας έρευνας, δόθηκε έμφαση σε τρεις βασικές παραμέτρους: τη βροχόπτωση, την ταχύτητα ροής και τη στάθμη του νερού σε ένα ποταμό. Τα δεδομένα αυτά εισήχθησαν και προεπεξεργάστηκαν στο MATLAB ώστε να είναι κατάλληλα για χρήση σε αλγορίθμους μηχανικής μάθησης. Ένας από τους κύριους στόχους της έρευνας ήταν η σύγκριση διαφορετικών αρχιτεκτονικών μοντέλων πρόβλεψης, ώστε να αναδειχθεί η βέλτιστη προσέγγιση στη πρόβλεψη της στάθμης του νερού. Αρχικά, εξετάστηκαν σύνθετα μοντέλα, όπως τα Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (LSTM) με γνώμονα τις καλές δυνατότητες τους στην ανάλυση δεδομένων χρονοσειρών. Επιπλέον, εξετάστηκαν απλούστερα μοντέλα, όπως τα γραμμικά, προκειμένου να διαπιστωθεί αν μπορούν να παράγουν εξίσου αξιόπιστα αποτελέσματα με τα πιο σύνθετα μοντέλα. Ακόμη εξετάστηκαν πιο εξελιγμένες αρχιτεκτονικές, όπως τα υβριδικά μοντέλα CNN-LSTM, προκειμένου να βελτιωθεί περαιτέρω η ακρίβεια των προβλέψεων. Στο τέλος γίνεται παρουσίαση και ανάλυση των αποτελεσμάτων σε πίνακες με στόχο το σχολιασμό και την εύρεση του αποδοτικότερου μοντέλου με τη βέλτιστη αρχιτεκτονική. Τα δεδομένα της έρευνας συλλέχθηκαν από τον ποταμό Σπερχειό της Στερεάς Ελλάδας και πιο συγκεκριμένα από δύο γέφυρες που βρίσκονται κοντά στους

οικισμούς Φραντζής και Κωσταλέξης νοτιοδυτικά της Λαμίας. Η περιοχή του Σπερχειού ποταμού, αποτελεί σημείο αυξημένου κινδύνου για πλημμύρες λόγω της γεωγραφικής και κλιματικής της θέσης, χρειάζεται ένα σύστημα που να προβλέπει το ύψος του νερού και να προειδοποιεί εγκαίρως τους κατοίκους και τις αρμόδιες αρχές.

2. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

2.1 Γεωγραφικά Χαρακτηριστικά της περιοχής μελέτης (Geographical Characteristics)

Ο Σπερχειός ποταμός περνά τρία χιλιόμετρα νότια από την πόλη της Λαμίας, ξεκινάει την πορεία του από τα ανατολικά της κεντρικής Πίνδου και τον ορεινό όγκο του Τυμφρηστού, κινείται από τα δυτικά προς τα ανατολικά και μετά από 82,5 χλμ. εκβάλλει στον Μαλιακό κόλπο. Στην πορεία του δημιουργεί την μεγάλη κοιλάδα της Φθιώτιδας (κοιλάδα του Σπερχειού), που απλώνεται ανάμεσα στους ορεινούς όγκους της Οίτης στα νότια και της Όθρυος στα βόρεια. Η λεκάνη απορροής του ποταμού καλύπτει μια έκταση 1.517 τ.χλμ. Αυτές οι τοποθεσίες επελέγησαν λόγω της σημασίας τους για τη διαχείριση των υδάτων και της εγγύτητάς τους σε ευάλωτες περιοχές. Τα γεωγραφικά χαρακτηριστικά, όπως η κλίση της κοίτης και η μορφολογία της περιοχής, επηρεάζουν τη συμπεριφορά της ροής και τις μετρήσεις.



Εικόνα 3 Γεωγραφική περιοχή μετρήσεων (Google Earth, n.d.)

2.2 Μοντέλο Βροχής-Απορροής και Πρόβλεψη Στάθμης Νερού

Το μοντέλο βροχής-απορροής (Rainfall-runoff modelling) αποτελεί έναν από τους βασικούς πυλώνες της υδρολογικής επιστήμης, με μακρά ιστορία και σημαντικές εφαρμογές. Η βασική του λειτουργία αφορά την περιγραφή της σχέσης μεταξύ της απορροής ενός ποταμού και των μετεωρολογικών παρατηρήσεων, όπως η βροχόπτωση, η θερμοκρασία και άλλοι παράγοντες, σε μια υδρολογική λεκάνη. Η υδρολογική λεκάνη ή λεκάνη απορροής ενός ποταμού είναι η περιοχή από την οποία απορρέει όλο το επιφανειακό νερό σε μια κοινή έξοδο. Η πρόβλεψη της στάθμης του νερού σχετίζεται άμεσα με τα μοντέλα βροχής-απορροής, καθώς η στάθμη είναι το αποτέλεσμα της απορροής που προκύπτει από μετεωρολογικές συνθήκες και φυσικές διεργασίες όπως η εξάτμιση, η διήθηση και η τήξη

του χιονιού. Η αποτύπωση και ανάλυση αυτής της σχέσης είναι καίριας σημασίας για τη διαχείριση υδατικών πόρων και την πρόληψη πλημμυρών, ειδικά υπό το πρίσμα ακραίων καιρικών φαινομένων. Οι υδρολογικές διαδικασίες χαρακτηρίζονται από μη γραμμικές αλληλεπιδράσεις και εξαρτώνται σε μεγάλο βαθμό από τις καταστάσεις του συστήματος που αναπαριστούν τη λεκάνη απορροής. Αυτό έχει ως αποτέλεσμα την ανάπτυξη υδρολογικών μοντέλων, τα οποία βασίζονται σε ένα σύστημα εισόδου, μαθηματικών εξισώσεων και εξόδου. Η είσοδος στα μοντέλα αυτά αποτελείται από υδρομετεωρολογικές παραμέτρους, όπως η βροχόπτωση. Μαθηματικές εξισώσεις χρησιμοποιούνται για τη μετατροπή της βροχής σε απορροή και τελικά η έξοδος των μοντέλων είναι η εκτίμηση της απορροής του ποταμού, η οποία συνδέεται άμεσα με τη στάθμη του νερού. (Beven, 2012)

2.3 Κατηγορίες μοντέλων πρόβλεψης στάθμης νερού

2.3.1 Φυσικά Μοντέλα

Τα φυσικά μοντέλα βασίζονται στην κατανόηση της φυσικής επιστήμης που σχετίζεται με τις υδρολογικές διαδικασίες. Αυτά τα μοντέλα βασίζονται στη μηχανική των ρευστών και χρησιμοποιούν θεμελιώδεις φυσικούς νόμους, όπως η διατήρηση της μάζας, της ορμής και της ενέργειας, για τη διατύπωση εξισώσεων που περιγράφουν την κίνηση του νερού σε ένα συγκεκριμένο σύστημα. Κάποια από αυτά είναι τα MIKE HYDRO River, HEC-HMS, SOBEK, EFDC και άλλα. Παρά την υψηλή ακρίβεια που προσφέρουν στις προβλέψεις της στάθμης του νερού, τα φυσικά μοντέλα έχουν σημαντικούς περιορισμούς. Απαιτούν μεγάλη ποικιλία δεδομένων, όπως τοπογραφικά, μετεωρολογικά και υδρολογικά δεδομένα, τα οποία συχνά δεν είναι διαθέσιμα σε κατάλληλη χωρική και χρονική ανάλυση. Επιπλέον, η προσομοίωσή τους είναι ιδιαίτερα απαιτητική, καθιστώντας τα ακατάλληλα για βραχυπρόθεσμες ή προβλέψεις σε πραγματικό χρόνο. Ένας ακόμη περιορισμός αυτών των μοντέλων είναι η δυσκολία στην πλήρη αναπαράσταση των μη γραμμικών αλληλεπιδράσεων των υδρολογικών διαδικασιών, ιδιαίτερα υπό ακραίες συνθήκες, όπως πλημμύρες ή παρατεταμένες ξηρασίες. Αυτό μπορεί να οδηγήσει σε μείωση της ακρίβειας των προβλέψεων. Επιπρόσθετα, η ανάπτυξη και βαθμονόμηση αυτών των μοντέλων απαιτεί εις βάθος γνώση των χαρακτηριστικών των λεκανών απορροής, καθώς τα φυσικά και χωρικά χαρακτηριστικά τους διαφέρουν σημαντικά από περιοχή σε περιοχή. Παρά τους περιορισμούς αυτούς, τα φυσικά μοντέλα παραμένουν πολύτιμα εργαλεία για την κατανόηση και τη μελέτη της υδρολογικής συμπεριφοράς σε λεπτομερή κλίμακα. Ωστόσο, λόγω του υψηλού υπολογιστικού κόστους και της ανάγκης για εκτεταμένα δεδομένα, η χρήση τους είναι συχνά περιορισμένη σε πειραματικές ή μικρές λεκάνες απορροής. (Cho, 2022) (Beven, 2012)

2.3.2 Εμπειρικά μοντέλα

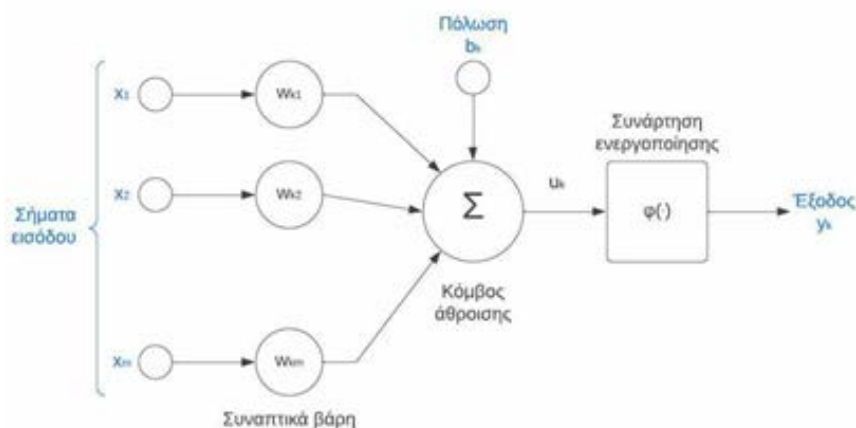
Τα εμπειρικά μοντέλα, τα οποία είναι καθοδηγούμενα από δεδομένα (data-driven) επιδιώκουν να προσδιορίσουν μη γραμμικές στατιστικές σχέσεις μεταξύ της εισόδου και της εξόδου ενός συστήματος, χωρίς να απαιτούν βαθιά κατανόηση των φυσικών διεργασιών που συνδέουν αυτές τις μεταβλητές. Αυτή η ιδιότητα τα έχει καταστήσει δημοφιλή για την πρόβλεψη υδρολογικών παραμέτρων σε περιπτώσεις όπου η πολυπλοκότητα ή η έλλειψη δεδομένων καθιστά δύσκολη τη χρήση φυσικών μοντέλων. Τα περισσότερα εμπειρικά μοντέλα ακολουθούν λογική μαύρου κουτιού (black box) με την

έννοια ότι η γνώση για τις εσωτερικές διαδικασίες είναι ελάχιστη (Beven, 2012). Η μαθηματική συνάρτηση που χρησιμοποιείται για τη μετατροπή της βροχής σε στάθμη νερού είναι είτε εμπειρική είτε εξαρτάται από τεχνικές μηχανικής μάθησης (Machine Learning), οι οποίες δεν βασίζονται σε φυσικές διεργασίες. Τα τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks - ANN) αποτελούν μία από τις δημοφιλέστερες μεθόδους, χάρη στη δυνατότητά τους να προσαρμόζονται σε πολύπλοκα αρχικά δεδομένα. (Daniell, 1991)

2.4 Εισαγωγή στα τεχνητά νευρωνικά δίκτυα

Τα τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks, ANNs), γνωστά και ως νευρωνικά δίκτυα, είναι εμπνευσμένα από τη λειτουργία των νευρώνων στον ανθρώπινο εγκέφαλο. Ο εγκέφαλος αποτελεί ένα εξαιρετικά σύνθετο βιολογικό σύστημα, στο οποίο οι νευρώνες συνεργάζονται για να εκτελέσουν υπολογισμούς με ταχύτητα και αποτελεσματικότητα που ξεπερνά τις δυνατότητες των σύγχρονων ψηφιακών υπολογιστών. Παρόμοια, τα τεχνητά νευρωνικά δίκτυα αποτελούνται από τεχνητούς «νευρώνες», δηλαδή απλές μονάδες επεξεργασίας που λαμβάνουν είσοδο από διαφορετικές πηγές, αλληλοεπιδρούν μεταξύ τους και έχουν τη δυνατότητα αποθήκευσης εμπειρικής γνώσης. Σύμφωνα με τον (Haykin, 2016) τα νευρωνικά δίκτυα μοιάζουν με τον ανθρώπινο εγκέφαλο σε δύο κύρια σημεία:

- Απόκτηση γνώσης μέσω της διαδικασίας μάθησης.
Το δίκτυο μαθαίνει από το περιβάλλον του, αξιοποιώντας τη γνώση που αποκτά μέσω εμπειρικών δεδομένων και προσαρμόζοντας τη λειτουργία του.
- Αποθήκευση γνώσης μέσω των συναπτικών βαρών.
Η δύναμη των συνδέσεων μεταξύ των νευρώνων, που ονομάζονται συναπτικά βάρη, χρησιμοποιείται ως μέσο αποθήκευσης της αποκτηθείσας γνώσης, επιτρέποντας στο δίκτυο να την αξιοποιεί για μελλοντική χρήση.



Εικόνα 4 Σχηματική απεικόνιση ενός νευρωνικού δικτύου (Amit, 1989)

Τα συναπτικά βάρη αποτελούν το κεντρικό στοιχείο αποθήκευσης της γνώσης στα τεχνητά νευρωνικά δίκτυα. Κάθε βάρος πολλαπλασιάζεται με το αντίστοιχο σήμα εισόδου και συμβάλλει στον υπολογισμό της εξόδου του νευρώνα. Η βελτιστοποίησή τους πραγματοποιείται μέσω αλγορίθμων μηχανικής μάθησης, οι οποίοι τροποποιούν τα βάρη για να επιτευχθεί η επιθυμητή έξοδος. Σε αντίθεση με τις συνάψεις του ανθρώπινου εγκεφάλου, οι οποίες έχουν θετικές τιμές, στα τεχνητά νευρωνικά δίκτυα τα βάρη μπορούν να πάρουν τόσο θετικές όσο και αρνητικές τιμές.

2.4.1 Κόμβος Άθροισης (Summation Node)

Ο κόμβος άθροισης λειτουργεί ως γραμμικός συνδυαστής των σταθμισμένων σημάτων εισόδου. Αν τα σήματα εισόδου είναι $x_1, x_2, \dots, x_n$ και τα αντίστοιχα συναπτικά βάρη είναι $w_1, w_2, \dots, w_n$ τότε η έξοδος u του κόμβου άθροισης δίνεται από τη σχέση:

Εξίσωση 1 Υπολογισμός της εξόδου του κόμβου άθροισης.

$$u = \sum_{i=1}^n w_i \cdot x_i = w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n \cdot x_n$$

Το μοντέλο παρουσιάζει επίσης μια εξωτερικά εφαρμοζόμενη πόλωση (Bias), η οποία αποτελεί ακόμα μια εκπαιδευόμενη παράμετρο του νευρώνα και καθορίζει τη θετική ή αρνητική μετατόπιση της συνάρτησης ενεργοποίησης. Αυτή προστίθεται γραμμικά στην έξοδο του κόμβου άθροισης.

Εξίσωση 2 Σχέση εξόδου (u) με πόλωση (b)

$$u = \sum_{i=1}^n w_i \cdot x_i + b$$

2.4.2 Συνάρτηση Ενεργοποίησης

Η συνάρτηση ενεργοποίησης (Activation Function) αποτελεί ένα μη γραμμικό στοιχείο παραμόρφωσης των τιμών και κανονικοποίησης των πεπερασμένων τιμών της εξόδου σε ένα ορισμένο εύρος πλάτους. Πιο συγκεκριμένα, για συνάρτηση ενεργοποίησης ϕ , η έξοδος του νευρώνα υπολογίζεται ως εξής:

Εξίσωση 3 Συνάρτηση Ενεργοποίησης

$$y = \phi(u) = \phi \left(\sum_{i=1}^n w_i \cdot x_i + b \right)$$

Οι συναρτήσεις ενεργοποίησης αποτελούν βασικό στοιχείο των τεχνητών νευρωνικών δικτύων, εισάγοντας μη γραμμικότητα και επιτρέποντας στα δίκτυα να αναγνωρίζουν σύνθετες σχέσεις μεταξύ εισόδων και εξόδων. Η μη γραμμικότητα που εισάγουν καθιστά το

δίκτυο ικανό να διαχειρίζεται πολύπλοκα προβλήματα, ενώ χωρίς αυτές το δίκτυο θα λειτουργούσε ως γραμμικός μετασχηματιστής, ανεξάρτητα από τα επίπεδα και τους νευρώνες του. Συχνά, οι συναρτήσεις αυτές αναφέρονται και ως συναρτήσεις περιορισμού (squashing functions), επειδή περιορίζουν το εύρος εξόδου του νευρώνα σε προκαθορισμένα όρια. Κάθε συνάρτηση ενεργοποίησης είναι κατάλληλη για συγκεκριμένους τύπους προβλημάτων, ενώ η επιλογή της επηρεάζει σημαντικά την απόδοση του μοντέλου. Παρακάτω παρουσιάζονται ορισμένες από τις πιο διαδεδομένες συναρτήσεις ενεργοποίησης (S. R. Dubey, 2022):

Εξίσωση 4 Συναρτήσεις Ενεργοποίησης

- **Sigmoid:** $\phi(u) = \frac{1}{1+e^{-u}}$
- **ReLU:** $\phi(u) = \max(0, u)$
- **Tanh:** $\phi(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}$

2.4.3 Συναρτήσεις απώλειας

Η συνάρτηση απώλειας (Loss Function) είναι κρίσιμη για την εκπαίδευση μοντέλων μηχανικής μάθησης, καθώς μετράει τη διαφορά μεταξύ της προβλεπόμενης τιμής από το μοντέλο και της πραγματικής τιμής της εξόδου. Δηλαδή στη συγκεκριμένη έρευνα υπολογίζει τη διαφορά της προβλεπόμενης στάθμης νερού με την πραγματική. Ο στόχος της εκπαίδευσης ενός μοντέλου είναι η ελαχιστοποίηση αυτής της διαφοράς, ώστε το μοντέλο να γίνει όσο το δυνατόν πιο ακριβές στην πρόβλεψη των εξόδων. Καθορίζοντας την απώλεια, μπορούμε να ποσοτικοποιήσουμε την απόδοση του μοντέλου και να επιλέξουμε το καταλληλότερο μοντέλο για ένα πρόβλημα. Ένα σύνθετο πρόβλημα όπως αυτό της πρόβλεψης χρονοσειρών, παρουσιάζει πολυάριθμους παράγοντες που οδηγούν στο συμπέρασμα του ότι δεν είναι δυνατό να οριστεί μια συγκεκριμένη συνάρτηση απώλειας. Πολλοί παράγοντες επηρεάζουν την επιλογή της, όπως η παρουσία ακραίων τιμών στα δεδομένα, οι ασάφειες στις κατανομές των δεδομένων, οι απαιτήσεις του επιλεγμένου μοντέλου, οι υπολογιστικοί πόροι που απαιτούνται, καθώς και οι απαιτήσεις για την απόδοση του μοντέλου (A. Jadon, 2022).

Υπάρχουν διάφορες συναρτήσεις απώλειας που χρησιμοποιούνται ανάλογα με τον τύπο του μοντέλου και το πρόβλημα που προσπαθεί να επιλύσει. Για παράδειγμα, σε προβλήματα παλινδρόμησης, δύο από τις πιο σημαντικές συναρτήσεις απώλειας είναι το Μέσο Απόλυτο Σφάλμα (Mean Absolute Error - MAE) και το Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error - MSE), που περιγράφονται ως εξής:

- Μέσο Απόλυτο Σφάλμα (MAE):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

Εξίσωση 5 Μέσο Απόλυτο Σφάλμα

Το MAE παρέχει ένα απλό και ομοιόμορφο μέτρο απόδοσης του μοντέλου, καθώς υπολογίζει την απόλυτη διαφορά μεταξύ της πραγματικής και της προβλεπόμενης εξόδου. Δεν είναι ιδιαίτερα ευαίσθητο στις ακραίες τιμές (outliers), πράγμα που το καθιστά χρήσιμο όταν οι ακραίες τιμές δεν πρέπει να επηρεάζουν υπερβολικά την εκπαίδευση του μοντέλου

- Μέσο Τετραγωνικό Σφάλμα (MSE)

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Εξίσωση 6 Μέσο Τετραγωνικό Σφάλμα

Το MSE μετρά τη μέση τετραγωνική διαφορά μεταξύ των πραγματικών και των προβλεπόμενων τιμών. Είναι ευαίσθητο σε ακραίες τιμές, επειδή το τετράγωνο των διαφορών αυξάνει την επίδραση των μεγάλων σφαλμάτων. Αυτό το καθιστά κατάλληλο για περιπτώσεις όπου τα μεγάλα σφάλματα είναι πιο σημαντικά.

2.5 Κατηγορίες Νευρωνικών Δικτύων:

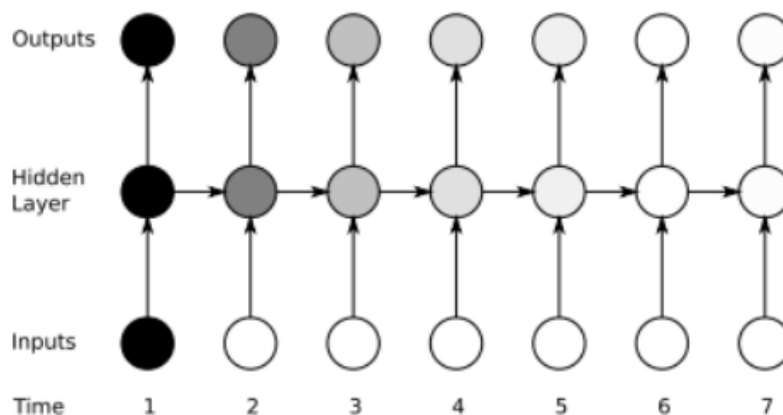
Ένας μεμονωμένος τεχνητός νευρώνας αποτελεί ένα καλά δομημένο μοντέλο, όμως δεν περιλαμβάνει σε καμία περίπτωση τις απαραίτητες ικανότητες για την επίλυση περίπλοκων προβλημάτων υπολογιστικής νοημοσύνης. Όταν όμως συνδυάζονται δύο ή περισσότεροι νευρώνες, δημιουργείται ένα νευρωνικό δίκτυο. Ο τρόπος με τον οποίο συνδέονται οι νευρώνες μεταξύ τους ονομάζεται αρχιτεκτονική του νευρωνικού δικτύου. Αυτή η σύνδεση είναι κρίσιμη για την αποτελεσματικότητα του δικτύου στην επίλυση διαφόρων προβλημάτων. Ο τρόπος με τον οποίο συνδέονται οι νευρώνες μεταξύ τους χωρίζει τα νευρωνικά δίκτυα σε κατηγορίες οι οποίες παρουσιάζονται παρακάτω.

- Πρόσθιας Τροφοδότησης (Feedforward Networks): Σε αυτά τα δίκτυα, οι νευρώνες ενός επιπέδου συνδέονται με τους νευρώνες του επόμενου επιπέδου με μια μόνο κατεύθυνση. Δεν υπάρχουν ανατροφοδοτήσεις ή βρόχοι, και τα δεδομένα προχωρούν από το ένα επίπεδο στο άλλο χωρίς να επιστρέφουν.
- Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks RNN): Αυτά τα δίκτυα περιλαμβάνουν βρόχους ή ανατροφοδότηση, επιτρέποντας τη ροή του σήματος πίσω σε προηγούμενα επίπεδα.

2.6 Αναδρομικά νευρωνικά δίκτυα (RNN)

Τα αναδρομικά νευρωνικά δίκτυα (RNN) έχουν αναδειχθεί σε ένα από τα πιο δημοφιλή εργαλεία μηχανικής μάθησης, κυρίως λόγω της ικανότητάς τους να συνδυάζουν πληροφορίες από προηγούμενα δεδομένα και να τις ενσωματώνουν κατά την εκπαίδευση των τωρινών δειγμάτων. Ωστόσο, ένα σημαντικό μειονέκτημα αυτών των δικτύων είναι η αδυναμία τους να διαχειριστούν αποτελεσματικά τις συσχετίσεις μεταξύ δεδομένων που βρίσκονται σε μεγάλη χρονική απόσταση. Το πρόβλημα αυτό προκύπτει λόγω της περιορισμένης «μνήμης» που διαθέτουν τα RNN κατά τη διαδικασία εκπαίδευσης, η οποία περιορίζει τον αριθμό των προηγούμενων βημάτων που λαμβάνονται υπόψη. Στην πράξη, αυτό σημαίνει ότι η εκπαίδευση δεδομένων απαιτεί μεγάλο χρόνο και υπολογιστικούς πόρους, ενώ συχνά το σφάλμα αυξάνεται σε υψηλές τιμές, προκαλώντας μεγάλες διακυμάνσεις στα βάρη του συστήματος σε κάθε επανάληψη της διαδικασίας εκμάθησης.

Η Εικόνα 5 απεικονίζει την ευαισθησία κάθε βήματος, με εντονότερη σκίαση να σημαίνει μεγαλύτερη εξάρτηση, από την είσοδο μέχρι τα τελευταία βήματα όπου εξαφανίζεται εντελώς.



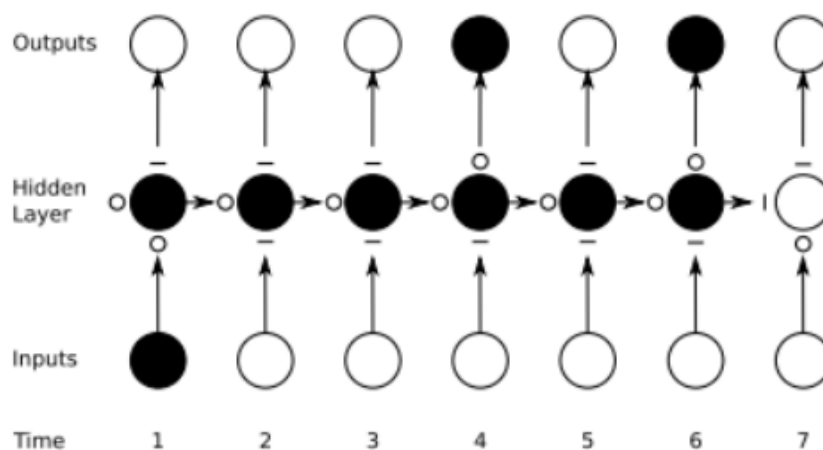
Εικόνα 5 Η εξασθένιση των συσχετίσεων στα RNN (Graves, 2012)

Για την αντιμετώπιση αυτού του προβλήματος, έχουν δοκιμαστεί διάφορες προσεγγίσεις, όπως η εισαγωγή χρονικής καθυστέρησης, η εφαρμογή της διάδοσης του σφάλματος σε διακριτούς χρόνους, ή η χρήση σταθερών στα δίκτυα. Ωστόσο, η τεχνική που έχει επικρατήσει και έχει αναδείξει τα καλύτερα αποτελέσματα είναι τα Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (Long Short-Term Memory - LSTM). Τα LSTM εισήχθησαν το 1997 από τους Sepp Hochreiter και Jürgen Schmidhuber, και η βασική τους λειτουργία επικεντρώνεται στη συγκράτηση πληροφοριών στη μνήμη για μεγάλα χρονικά διαστήματα. Αυτό τα καθιστά ιδιαίτερα αποτελεσματικά στη διαχείριση χρονοσειρών και άλλων εφαρμογών που απαιτούν μακροπρόθεσμες συσχετίσεις. (Schmidhuber, 1997).

2.7 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (Long Short-Term Memory - LSTM)

Η αρχιτεκτονική των LSTM μοιάζει αρκετά με αυτή των RNN, με την έννοια ότι απαρτίζονται και αυτά από ένα αριθμό μονάδων συνδεδεμένων αλυσιδωτά μεταξύ τους σε κάθε

επίπεδο. Η διαφορά βρίσκεται στο ότι οι μονάδες (modules ή blocks) που βρίσκονται στα εσωτερικά επίπεδα του δικτύου ενσωματώνουν έναν μηχανισμό που επιτρέπει την αποθήκευση και διαχείριση πληροφοριών για μεγαλύτερα χρονικά διαστήματα χωρίς απώλειες. Αυτό γίνεται μέσω ενός αλγορίθμου που είναι δομημένος σε επίπεδα. Αποτελούνται από κελιά μνήμης (cell), πύλες εισόδου, πύλες εξόδου και πύλες επιλεκτικής συγκράτησης (forget gate). Τα κελιά μνήμης επιλέγουν ποιες πληροφορίες θα αποθηκεύσουν με το να συνδυάζουν την προηγούμενη κατάσταση, την τρέχουσα μνήμη και την είσοδο. Η χρήση αυτών των πυλών διασφαλίζει την αποθήκευση και πρόσβαση στις πληροφορίες ακόμα και με την πάροδο πολλών βημάτων. Στην Εικόνα 6 παρουσιάζεται η διατήρηση της πληροφορίας καθώς ο χρόνος προχωρά, σε ένα δίκτυο LSTM με ένα κρυφό επίπεδο. Ο συμβολισμός “O” υποδεικνύει ότι η εκάστοτε πύλη είναι ανοιχτή, ενώ ο συμβολισμός “-” υποδηλώνει ότι η πύλη είναι κλειστή. Όπως φαίνεται στην παρακάτω εικόνα, η μονάδα μνήμης έχει τη δυνατότητα να διατηρεί την πληροφορία του πρώτου βήματος, όταν η πύλη εισόδου παραμένει κλειστή και η πύλη επιλεκτικής συγκράτησης είναι ανοιχτή. Αυτό αναδεικνύει τη λειτουργική ικανότητα των LSTM να διαχειρίζονται και να συγκρατούν κρίσιμες πληροφορίες για μεγάλα χρονικά διαστήματα, διασφαλίζοντας την αποδοτική επεξεργασία δεδομένων χρονοσειρών.



Εικόνα 6 Διατήρηση της πληροφορίας στα δίκτυα LSTM. (Graves, 2012)

Σύμφωνα με τη μελέτη του (Le, 2019) στη λεκάνη απορροής του ποταμού Ντα στο Βιετνάμ χρησιμοποιήθηκε μοντέλο LSTM για την πρόβλεψη πλημμυρών με δεδομένα ημερήσιων απορροών και βροχόπτωσης ως σύνολα εισόδου. Η έξοδος του μοντέλου ρυθμίστηκε για να προβλέπει την ποσότητα απορροής για 1, 2 και 3 ημέρες στο μέλλον. Η αξιολόγηση της απόδοσης πραγματοποιήθηκε χρησιμοποιώντας τους δείκτες Nash–Sutcliffe efficiency (NSE) και root mean squared error (RMSE). Τέλος, η τιμή NSE των δεδομένων της πρώτης ημέρας ήταν περίπου 99,1%, η τιμή NSE της δεύτερης ημέρας ήταν 94,9%, και η τιμή NSE της τρίτης ημέρας ήταν 87,2%, οπότε ήταν δυνατό να επιτευχθεί ακρίβεια άνω του 87,2% για όλα τα δεδομένα εξόδου. Μέσω αυτής της μελέτης, επιβεβαιώθηκε ότι το μοντέλο LSTM, απέδωσε αρκετά καλά για την πρόβλεψη της απορροής ποταμών. Υβριδικά μοντέλα που ενσωματώνουν διαφορετικές αρχιτεκτονικές βαθιάς μάθησης, όπως τα μοντέλα CNN-LSTM, έχουν δείξει σημαντικές βελτιώσεις στην αποτύπωση σύνθετων χρονικών και χωρικών εξαρτήσεων στα υδρολογικά δεδομένα. Αυτή η υβριδοποίηση εντοπίζει αποτελεσματικά τα χωρικά και χρονικά χαρακτηριστικά στα δεδομένα βροχόπτωσης, που είναι απαραίτητα για

ακριβή πρόβλεψη. Για παράδειγμα, οι (Li, 2022) ανέπτυξαν ένα μοντέλο CNN-LSTM για την πρόβλεψη της ροής των ποταμών, το οποίο υπερέβη τα μοντέλα που χρησιμοποιούσαν είτε μόνο CNN είτε μόνο LSTM. Η εφαρμογή των LSTM μοντέλων στην πρόβλεψη πλημμυρών έχει αποδειχθεί ιδιαίτερα αποτελεσματική, όπως φαίνεται και από τη μελέτη στη λεκάνη απορροής του ποταμού Jinhua στην Κίνα (Jun Li, 2024). Στη συγκεκριμένη έρευνα, χρησιμοποιήθηκαν υπολογισμένα και προσομοιωμένα δεδομένα για την εκπαίδευση και αξιολόγηση του μοντέλου, επιτυγχάνοντας υψηλή ακρίβεια στην πρόβλεψη των υδρολογικών μεταβολών. Εξετάστηκαν διάφοροι λόγοι σύνθεσης των δεδομένων (υπολογισμένα προς προσομοιωμένα), με τον βέλτιστο να είναι 2:1, καθώς εξισορροπεί τη ρεαλιστική μεταβλητότητα των υπολογισμένων δεδομένων με τις συστηματικές τάσεις των προσομοιώσεων. Τα αποτελέσματα έδειξαν ότι η σωστή ρύθμιση των υπερπαραμέτρων του LSTM, όπως το μέγεθος των batches, ο αριθμός των εποχών και η λειτουργία ενεργοποίησης (ReLU), μπορεί να βελτιώσει σημαντικά την ικανότητα πρόβλεψης.

3. ΣΥΛΛΟΓΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Η συλλογή και μέτρηση των δεδομένων αποτελεί κρίσιμο βήμα για την πρόβλεψη των πλημμυρικών φαινομένων. Τα δεδομένα που χρησιμοποιήθηκαν για την ανάλυση αφορούν τις εξής τρεις παραμέτρους:

- Βροχόπτωση
- Ταχύτητα ροής
- Στάθμη του ποταμού

Μέθοδοι Συλλογής Δεδομένων

Για τη συλλογή των δεδομένων χρησιμοποιήθηκαν μετεωρολογικοί σταθμοί με βροχομετρητές για τη μέτρηση της βροχόπτωσης καθώς και αισθητήρες, οι οποίοι ήταν τοποθετημένοι στις δύο γέφυρες, για τη μέτρηση της ταχύτητας και της στάθμης του νερού. Αρχικά η συλλογή των δεδομένων βροχόπτωσης πραγματοποιήθηκε μέσω ενός δικτύου μετεωρολογικών σταθμών που έχουν τοποθετηθεί σε επιλεγμένες περιοχές στην υδρολογική λεκάνη του ποταμού Σπερχειού. Οι σταθμοί αυτοί είναι εξοπλισμένοι με βροχόμετρα (rain gauges), τα οποία χρησιμοποιούνται για την καταγραφή των ποσοτήτων βροχόπτωσης σε συγκεκριμένα χρονικά διαστήματα. Τα βροχόμετρα αυτά καταγράφουν τη βροχόπτωση με ακρίβεια σε χιλιοστά ανά ώρα ή λεπτό και έχουν τη δυνατότητα να συλλέγουν και να αποθηκεύουν τα δεδομένα σε χρονοσειρές με τη χρήση συστημάτων καταγραφής και απομακρυσμένης μετάδοσης.



Εικόνα 7 Μετεωρολογικός σταθμός με βροχόμετρο (Ιστός Μετεωρολογικού σταθμού Οινόφυτων, χ.χ.)

Η ταχύτητα ροής και η στάθμη του ποταμού Σπερχειού μετρήθηκαν χρησιμοποιώντας αισθητήρες και υπερηχητικά όργανα μέτρησης που τοποθετήθηκαν σε συγκεκριμένες θέσεις στις γέφυρες Φραντζή και Κωσταλέξη. Οι αισθητήρες τοποθετήθηκαν πάνω από την επιφάνεια του νερού, κοντά στα θεμέλια των γεφυρών και καταγράφουν τη ροή και τη στάθμη του νερού με τη χρήση υπερηχητικών κυμάτων. Τα υπερηχητικά παλμικά κύματα που εκπέμπονται από τους αισθητήρες ανακλώνται στο νερό και μετρούνται με βάση τον χρόνο επιστροφής, επιτρέποντας τον υπολογισμό της ταχύτητας ροής και της στάθμης. Αυτή η μέθοδος είναι ιδανική για ποτάμια με μεταβλητό βάθος και ταχύτητα. Οι αισθητήρες αποτελούνται από έναν ηλιακό συλλέκτη ο οποίος παρέχει ενέργεια για την αυτονομία του συστήματος και μια ηλεκτρονική μονάδα επικοινωνίας για τη συλλογή και μετάδοση δεδομένων, προφανώς για παρακολούθηση από απόσταση. Οι μετρήσεις γίνονται σε συνεχή βάση (π.χ. κάθε 10 λεπτά), αποθηκεύονται σε καταγραφικά δεδομένων και αποστέλλονται σε κεντρικό υπολογιστικό σύστημα για ανάλυση.



Εικόνα 8 Αισθητήρας μέτρησης στάθμης και ταχύτητας ροής



Εικόνα 9 Ηλιακός συλλέκτης και ηλεκτρονική μονάδα επικοινωνίας

4. ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Η προεπεξεργασία των δεδομένων αποτελεί κρίσιμο στάδιο στη διαδικασία ανάλυσης, καθώς προετοιμάζει τα αρχικά δεδομένα να καταστούν αξιοποιήσιμα από τα συστήματα μηχανικής μάθησης. Η ποιότητα των δεδομένων επηρεάζει άμεσα την ακρίβεια και την αξιοπιστία των αποτελεσμάτων, γι' αυτό και οι διαδικασίες καθαρισμού και μετασχηματισμού των δεδομένων είναι απαραίτητες. Συχνά, τα αρχικά δεδομένα περιέχουν ελλείψεις, θόρυβο (noise) - δηλαδή τυχαία κυμαινόμενα δεδομένα χωρίς ουσιαστικό περιεχόμενο, λανθασμένες ή ακραίες τιμές, και είναι απαραίτητο να απομακρύνονται. Με την προεπεξεργασία, τα δεδομένα καθαρίζονται, μετασχηματίζονται και συνδυάζονται με τρόπο που διευκολύνει την εξαγωγή γνώσης. Μεταξύ των τεχνικών που εφαρμόζονται περιλαμβάνονται η απομάκρυνση ακραίων τιμών, η κανονικοποίηση αριθμητικών δεδομένων σε κλίμακες κατάλληλες για ανάλυση, καθώς και η μείωση δεδομένων μέσω επιλογής των πιο σημαντικών χαρακτηριστικών. Η διαδικασία αυτή στοχεύει στη βελτίωση της ποιότητας των δεδομένων, διευκολύνοντας τα μοντέλα μηχανικής μάθησης να εντοπίσουν μοτίβα και σχέσεις που συμβάλλουν στην εξαγωγή αξιόπιστων αποτελεσμάτων. Έτσι, η προεπεξεργασία διασφαλίζει ότι τα δεδομένα είναι κατάλληλα διαμορφωμένα για να αποτελέσουν τη βάση της εκπαίδευσης των αλγορίθμων. (S. A. Alasadi and W. S. Bhaya, 2017)

4.1 Εισαγωγή των δεδομένων

Αρχικά, πραγματοποιήθηκε η φόρτωση των δεδομένων στο Matlab από τα αρχεία Excel των δύο γεφυρών (Frantzi και Kostaleksi). Τα δεδομένα πρέπει να είναι σε μια δομή πίνακα που αποτελείται από στήλες με φυσική σημασία. Κάθε στήλη είναι ένα χαρακτηριστικό (feature) και αντιστοιχεί σε φυσικό μέγεθος (π.χ. βροχόπτωση, στάθμη κτλ.) ενώ κάθε γραμμή αποτελεί ένα στιγμιότυπο (instance) με χρονική σημασία. Στη συγκεκριμένη εργασία τα δεδομένα περιλαμβάνουν τρία χαρακτηριστικά (features):

- Βροχόπτωση (Rain)
- Ταχύτητα Ροής (Velocity)
- Στάθμη Νερού (Distance)

Matlab:

```
% Φόρτωση δεδομένων από αρχεία Excel  
frantzi_data = readtable('frantzi.xlsx');  
kostaleksi_data = readtable('kostaleksi.xlsx');
```

Κώδικας 1 Φόρτωση των δεδομένων

4.2 Συγχώνευση των δεδομένων σε κοινό αρχείο

Τα δεδομένα από τις δύο γέφυρες ενώθηκαν χρονικά ώστε να υπάρχει συνολική εικόνα για τις δύο περιοχές. Η χρονική στήλη μετατράπηκε σε μορφή datetime για το συγχρονισμό των

δεδομένων.

Matlab:

```
% Μετατροπή σε datetime
frantzi_data.DateTime = datetime(frantzi_data.DateTime, 'InputFormat', 'yyyy-MM-dd HH:mm:ss');
kostaleksi_data.DateTime = datetime(kostaleksi_data.DateTime, 'InputFormat', 'yyyy-MM-dd HH:mm:ss');
% Συγχώνευση των δεδομένων
combined_data = synchronize(frantzi_data, kostaleksi_data, 'union', 'fillwithmissing');
```

Κώδικας 2 Συγχώνευση των δεδομένων

4.3 Αντικατάσταση ελλιπών τιμών

Οι ημιτελείς χρονοσειρές δεδομένων αποτελούν ένα κοινό και αναπόφευκτο πρόβλημα που εμφανίζεται στα περισσότερα σύνολα δεδομένων και μπορεί να προκύψει από διάφορες αιτίες. Για παράδειγμα, ορισμένες μετρήσεις ενδέχεται να μην ήταν διαθέσιμες τη χρονική στιγμή της καταγραφής λόγω τεχνικών προβλημάτων. Επιπλέον, αστοχίες λογισμικού ή υλικού, κατά την αποθήκευση ή μεταφορά των δεδομένων, μπορούν να οδηγήσουν στην ύπαρξη κενών τιμών. Η διαχείριση των χαμένων τιμών είναι κρίσιμη, καθώς μπορεί να επηρεάσει αρνητικά την ακρίβεια και την αποτελεσματικότητα των αλγορίθμων μηχανικής μάθησης. Η λανθασμένη ή ακατάλληλη διαχείριση αυτών των τιμών μπορεί να προκαλέσει αποπροσανατολισμό στον αλγόριθμο και να οδηγήσει σε λανθασμένα αποτελέσματα. (Hu, 2001) Στην παρούσα εργασία, οι ελλιπείς τιμές που εντοπίστηκαν στις μεταβλητές διαχειρίστηκαν με τη μέθοδο της γραμμικής παρεμβολής. Η μέθοδος αυτή βασίζεται στην εκτίμηση των χαμένων τιμών μέσω γραμμικής σχέσης μεταξύ των παρακείμενων τιμών, διατηρώντας τη συνέχεια των δεδομένων. Η γραμμική παρεμβολή εξασφαλίζει ότι οι ελλιπείς τιμές αντικαθίστανται με τρόπο που διατηρεί τη γραμμικότητα και τη συνοχή της χρονοσειράς.

Matlab:

```
% Διαχείριση ελλιπών τιμών με γραμμική παρεμβολή
combined_data.Rain = fillmissing(combined_data.Rain, 'linear');
combined_data.Velocity = fillmissing(combined_data.Velocity, 'linear');
combined_data.WaterLevel = fillmissing(combined_data.WaterLevel, 'linear');
```

Κώδικας 3 Αντικατάσταση ελλιπών τιμών

4.4 Κανονικοποίηση των δεδομένων

Η κανονικοποίηση (normalization) των δεδομένων αποτελεί μία βασική διαδικασία προεπεξεργασίας που μετασχηματίζει τις μεταβλητές σε μία συγκεκριμένη κλίμακα τιμών, συνήθως εντός του εύρους [0, 1]. Η ανάγκη για κανονικοποίηση προκύπτει όταν οι τιμές μιας μεταβλητής παρουσιάζουν μεγάλες αποκλίσεις, όπως για παράδειγμα τιμές της τάξης του 0.01 και του 1000. Αυτή η ανισορροπία μπορεί να επηρεάσει αρνητικά τη λειτουργία και την απόδοση των αλγορίθμων μηχανικής μάθησης. Επιπλέον, η κανονικοποίηση είναι εξίσου σημαντική όταν χρησιμοποιούνται πολλές διαφορετικές μεταβλητές με διαφορετικές κλίμακες τιμών. Σε τέτοιες περιπτώσεις, οι μεγάλες αποκλίσεις μεταξύ των μεταβλητών μπορεί να οδηγήσουν τους αλγορίθμους να δίνουν υπερβολικό βάρος στις μεταβλητές με τις μεγαλύτερες τιμές, αγνοώντας ουσιαστικά τη συνεισφορά των υπολοίπων. Μέσω της κανονικοποίησης, οι τιμές όλων των μεταβλητών μετασχηματίζονται σε μία κοινή κλίμακα, επιτρέποντας την καλύτερη σύγκριση μεταξύ τους και διασφαλίζοντας ότι οι αλγόριθμοι αποδίδουν ίση σημασία σε κάθε μεταβλητή. Αυτό είναι ιδιαίτερα σημαντικό για τα νευρωνικά δίκτυα, στα οποία οι τιμές εισόδου πρέπει να κυμαίνονται εντός συγκεκριμένων ορίων, καθώς βοηθά στη σταθερότητα και ταχύτερη σύγκλιση του αλγορίθμου κατά τη διαδικασία της εκπαίδευσης. Για την προετοιμασία των δεδομένων προς χρήση στα μοντέλα πρόβλεψης, εφαρμόστηκε min-max normalization ώστε οι τιμές να κυμαίνονται μεταξύ 0 και 1. Η διαδικασία αυτή περιγράφεται από την παρακάτω εξίσωση:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

Εξίσωση 7 min-max κανονικοποίηση

Όπου x είναι η αρχική τιμή, $x_{\min}$ και $x_{\max}$ οι ελάχιστες και μέγιστες τιμές της μεταβλητής, αντίστοιχα, και x' η κανονικοποιημένη τιμή.

Matlab:

```
% Κανονικοποίηση των δεδομένων  
combined_data.Rain = (combined_data.Rain - rain_min) / (rain_max - rain_min);  
combined_data.Velocity = (combined_data.Velocity - velocity_min) / (velocity_max - velocity_min);  
combined_data.WaterLevel = (combined_data.WaterLevel - water_min) / (water_max - water_min);
```

Κώδικας 4 Κανονικοποίηση των δεδομένων

4.5 Δημιουργία ενός κανονικοποιημένου αρχείου

Τα κανονικοποιημένα δεδομένα αποθηκεύτηκαν σε νέο αρχείο με όνομα 'df_norm' για περαιτέρω ανάλυση και χρήση στα νευρωνικά δίκτυα.

Matlab:

```
% Δημιουργία πίνακα με κανονικοποιημένα δεδομένα

df_norm = table(combined_data.DateTime, combined_data.Rain, combined_data.Velocity,
combined_data.WaterLevel, ...

'VariableNames', {'DateTime', 'dist_fr_norm', 'dist_kost_norm', 'vel_norm', 'rain_norm'});
```

Κώδικας 5 Δημιουργία αρχείου με κανονικοποιημένα δεδομένα

4.6 Λογαρίθμηση των τιμών της βροχής

Τα δεδομένα βροχόπτωσης τείνουν να είναι ασύμμετρα, με πολλές τιμές κοντά στο μηδέν και λίγες πολύ υψηλές τιμές. Η λογαρίθμηση μετατρέπει τα δεδομένα σε μια πιο κανονική κατανομή βοηθώντας τα μοντέλα να εκπαιδευτούν καλύτερα. Οι μεγάλες τιμές βροχόπτωσης (ακραία καιρικά φαινόμενα) μπορεί να προκαλέσουν προβλήματα στα μοντέλα πρόβλεψης, καθώς επηρεάζουν δυσανάλογα τη διαδικασία εκπαίδευσης. Ο λογάριθμος συμπίεζει τις μεγάλες τιμές, μειώνοντας το εύρος των δεδομένων. (C. M. Teng, 1999)

Matlab:

```
% 1. Δημιουργία νέας στήλης με λογαριθμοποιημένη βροχόπτωση
epsilon = 1e-6; % Μικρή σταθερά για αποφυγή log(0)
df_norm.rain_log = log(df_norm.rain_norm + epsilon);

% 2. Δημιουργία Ιστογράμματος για αρχική βροχή
figure;

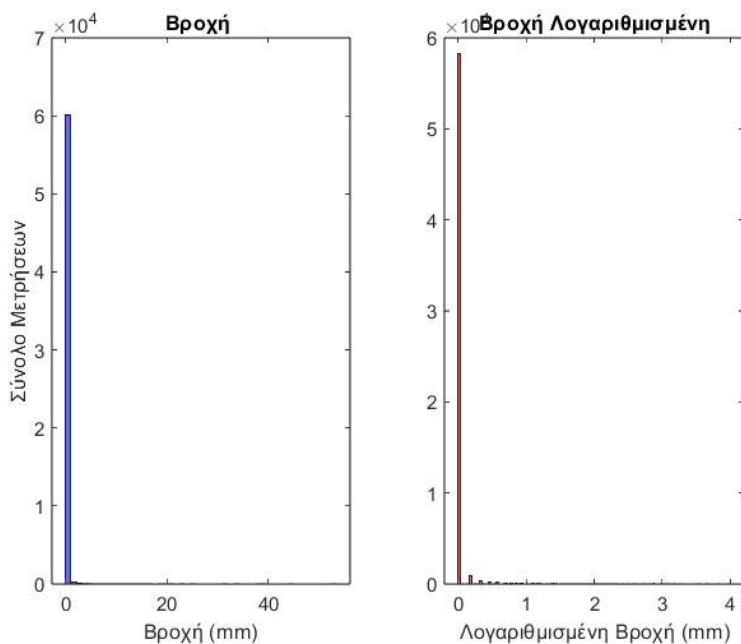
% Ιστόγραμμα αρχικής βροχόπτωσης
subplot(1, 2, 1);
histogram(df_norm.rain_norm, 50, 'FaceColor', 'b');
title('Βροχή μετά από διόρθωση Outliers');
xlabel('Τιμές Βροχής (rain\_norm)');
ylabel('Σύνολο Μετρήσεων');

% Ιστόγραμμα λογαριθμοποιημένης βροχής
subplot(1, 2, 2);
histogram(df_norm.rain_log, 100, 'FaceColor', 'r');
title('Βροχή λογαριθμισμένη');
xlabel('log(Τιμές Βροχής + ε)');
ylabel('Σύνολο Μετρήσεων');

% Αποθήκευση του γραφήματος
saveas(gcf, 'log_rain.png');
```

Κώδικας 6 Λογαρίθμηση των τιμών της βροχής

Η εικόνα 10 παρουσιάζει δύο γραφήματα που συγκρίνουν την κατανομή των τιμών βροχής πριν και μετά από λογαριθμικό μετασχηματισμό. Στο αριστερό γράφημα, η κατανομή δείχνει ότι οι περισσότερες τιμές βροχής είναι συγκεντρωμένες πολύ κοντά στο μηδέν. Οι υψηλές τιμές είναι σπάνιες, γεγονός που προκαλεί "συμπίεση" των δεδομένων στα χαμηλότερα επίπεδα του άξονα. Αυτό δείχνει έντονη ασυμμετρία στην κατανομή, κάτι που μπορεί να δημιουργήσει προβλήματα στους αλγόριθμους μηχανικής μάθησης. Στο δεξί γράφημα, μετά τον λογαριθμικό μετασχηματισμό, οι τιμές είναι περισσότερο κατανομημένες, με τα μεγαλύτερα ύψη βροχής να αυξάνονται και τα μικρότερα να συμπιέζονται. Αυτή η προσέγγιση μειώνει την διαφορά και εξισορροπεί την κατανομή των δεδομένων βροχόπτωσης, διευκολύνοντας τους αλγόριθμους να αναγνωρίσουν μοτίβα στα δεδομένα.



Εικόνα 10 Εμφάνιση λογαριθμισμένων δεδομένων

Πλέον το αρχείο 'df_norm' είναι ένας πίνακας timetable 61894x4. Ακολουθεί μία εικόνα με τις πρώτες 10 μετρήσεις.

	DateTime	1 dist_fr_norm	2 dist_kost_norm	3 vel_norm	4 rain_norm	5	6	7
1	01-Oct-2020 17:30:...	0.1545	0.0970	0.0736	0			
2	01-Oct-2020 18:00:...	0.1525	0.0970	0.0585	0			
3	01-Oct-2020 18:30:...	0.1525	0.0970	0.0585	0			
4	01-Oct-2020 19:00:...	0.1525	0.0970	0.0792	0			
5	01-Oct-2020 19:30:...	0.1525	0.0970	0.0717	0			
6	01-Oct-2020 20:00:...	0.1505	0.0970	0.0660	0			
7	01-Oct-2020 20:30:...	0.1505	0.0957	0.0660	0			
8	01-Oct-2020 21:00:...	0.1485	0.0957	0.0755	0			
9	01-Oct-2020 21:30:...	0.1485	0.0957	0.0736	0			
10	01-Oct-2020 22:00:...	0.1465	0.0957	0.0755	0			

Εικόνα 11 Πρώτες δέκα μετρήσεις του αρχείου 'df_norm':

Σύνοψη των Βημάτων:

1. Φόρτωση των δεδομένων από αρχεία Excel.
2. Συγχώνευση των δεδομένων και ευθυγράμμιση με βάση τον χρόνο.
3. Αντικατάσταση των ελλειπών τιμών με γραμμική παρεμβολή.
4. Κανονικοποίηση των δεδομένων με χρήση min-max normalization.
5. Αποθήκευση των κανονικοποιημένων δεδομένων σε κοινό αρχείο "df_norm"
6. Λογαρίθμηση των τιμών της βροχής

5. ΣΥΝΟΨΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Αφού ολοκληρώθηκε η επεξεργασία των δεδομένων έγινε μία στατιστική σύνοψη των στοιχείων του πίνακα `df_norm` που περιγράφουν τα δεδομένα της στάθμης, της ταχύτητας, και της βροχόπτωσης στις δύο γέφυρες. Τα στοιχεία παρουσιάζονται σε μορφή πίνακα και διαγραμμάτων καθώς γίνεται και αναφορά στον κώδικα που χρησιμοποιήθηκε για την εξαγωγή των στοιχείων.

Πίνακας 1 Στατιστική σύνοψη των στοιχείων

Row	count	mean	std	min	max
dist_fr_norm	61894	0.2066	0.1199	0.0416	0.9406
dist_kost_norm	61894	0.1106	0.0796	0.0262	0.9214
vel_norm	61894	0.0899	0.0405	0.0377	0.8038
rain_norm	61894	0.0055	0.0371	0	1.0428

Από τον Πίνακα 1 μπορούμε να συμπεράνουμε ότι η διακύμανση της στάθμης στις δύο γέφυρες, καθώς και η ταχύτητα της ροής, υποδεικνύουν ότι η περιοχή υπόκειται σε σχετικά σταθερές υδρολογικές συνθήκες, με κάποιες περιόδους έντονων βροχοπτώσεων. Οι χαμηλές μέσες τιμές της βροχόπτωσης, σε συνδυασμό με τις ιδιαίτερα υψηλές μέγιστες τιμές, υπογραμμίζουν τη σημασία της καταγραφής αυτών των γεγονότων για την πρόβλεψη πλημμυρών.

Matlab για Στατιστική σύνοψη των στοιχείων:

```

% Υπολογισμός βασικών στατιστικών για το df_norm (εξαιρείται η DateTime)
stats = table();

% Υπολογισμός για κάθε κανονικοποιημένη στήλη
stats.count = [sum(~ismissing(df_norm.dist_fr_norm)); ...
               sum(~ismissing(df_norm.dist_kost_norm)); ...
               sum(~ismissing(df_norm.vel_norm)); ...
               sum(~ismissing(df_norm.rain_norm))];

stats.mean = [mean(df_norm.dist_fr_norm, 'omitnan'); ...
              mean(df_norm.dist_kost_norm, 'omitnan'); ...
              mean(df_norm.vel_norm, 'omitnan'); ...
              mean(df_norm.rain_norm, 'omitnan')];

stats.std = [std(df_norm.dist_fr_norm, 'omitnan'); ...
             std(df_norm.dist_kost_norm, 'omitnan'); ...
             std(df_norm.vel_norm, 'omitnan'); ...
             std(df_norm.rain_norm, 'omitnan')];

stats.min = [min(df_norm.dist_fr_norm, [], 'omitnan'); ...
             min(df_norm.dist_kost_norm, [], 'omitnan'); ...
             min(df_norm.vel_norm, [], 'omitnan'); ...
             min(df_norm.rain_norm, [], 'omitnan')];

stats.max = [max(df_norm.dist_fr_norm, [], 'omitnan'); ...
             max(df_norm.dist_kost_norm, [], 'omitnan'); ...
             max(df_norm.vel_norm, [], 'omitnan'); ...
             max(df_norm.rain_norm, [], 'omitnan')];

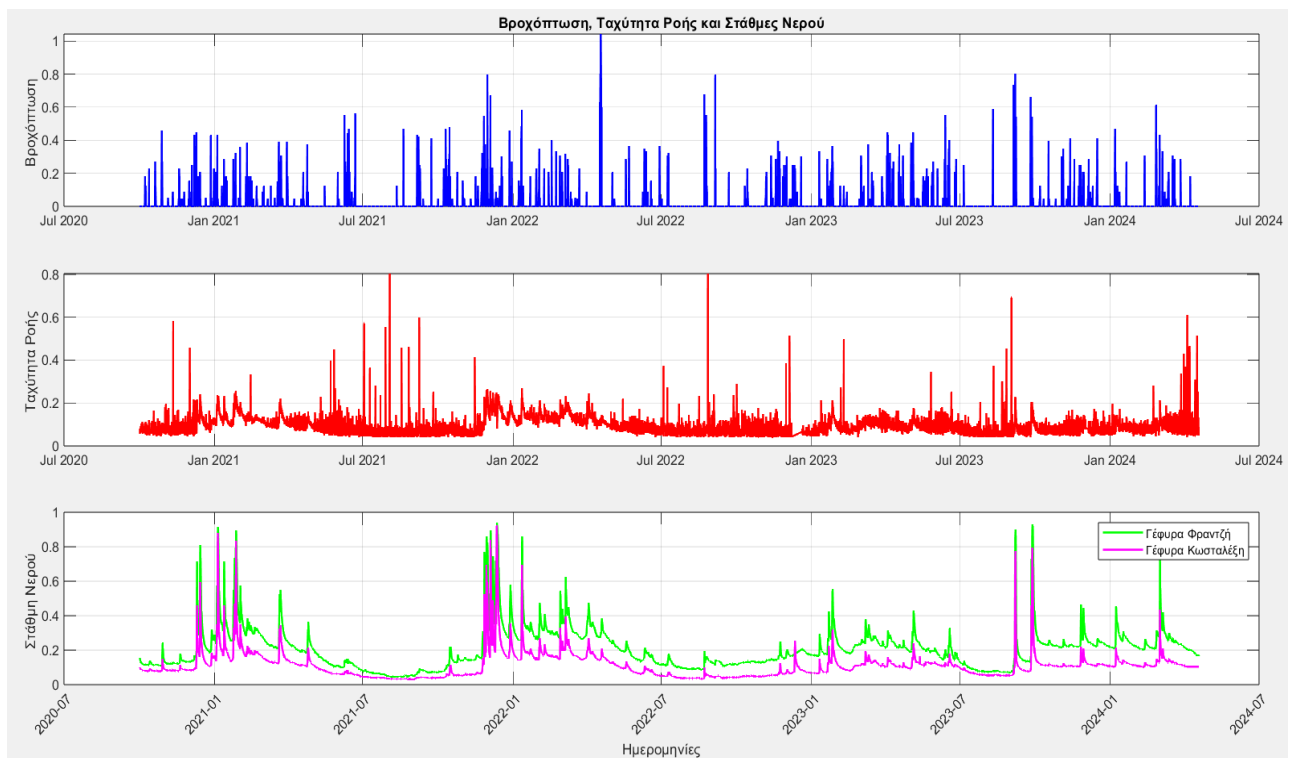
% Εμφάνιση των στατιστικών
disp(stats);

```

Κώδικας 7 Στατιστική σύνοψη

Στη συνέχεια θα δημιουργήσουμε τρία γραφήματα που παρουσιάζουν τις μεταβολές στην πάροδο του χρόνου για τη βροχόπτωση, την ταχύτητα ροής και τη στάθμη νερού για τη Γέφυρα Φραντζή και τη Γέφυρα Κωσταλέξη (Εικόνα 13). Στο πρώτο γράφημα, που αφορά τη βροχόπτωση, παρατηρούνται έντονες αυξήσεις κατά συγκεκριμένες χρονικές περιόδους, ενώ σε άλλες οι τιμές παραμένουν χαμηλές. Αυτές οι αυξήσεις εμφανίζονται κυρίως κατά τους χειμερινούς μήνες, γεγονός που συνάδει με τα εποχιακά μοτίβα των καιρικών φαινομένων. Στο δεύτερο γράφημα, που αφορά την ταχύτητα ροής, η τάση είναι παρόμοια με αυτή της βροχόπτωσης, κάτι που υποδηλώνει συσχέτιση μεταξύ των δύο μεταβλητών. Οι κορυφές στην ταχύτητα ροής φαίνεται να ακολουθούν τις έντονες βροχοπτώσεις, δείχνοντας ότι οι βροχές επηρεάζουν άμεσα τη ροή του ποταμού. Στο τρίτο γράφημα

απεικονίζονται οι στάθμες νερού στις δύο γέφυρες, με την πράσινη γραμμή να αντιστοιχεί στη Γέφυρα Φραντζή και τη μωβ στη Γέφυρα Κωσταλέξη. Οι στάθμες παρουσιάζουν έντονες αυξήσεις κατά τις ίδιες περιόδους με τις αιχμές στη βροχόπτωση και την ταχύτητα ροής. Η Γέφυρα Φραντζή εμφανίζει υψηλότερες στάθμες σε σχέση με την Κωσταλέξη, πιθανότατα λόγω διαφορετικών υδρολογικών χαρακτηριστικών, όπως η γεωμορφολογία ή η τοποθεσία. Συνοψίζοντας, υπάρχει ξεκάθαρη συσχέτιση ανάμεσα στη βροχόπτωση, την ταχύτητα ροής και τη στάθμη νερού, με τα δεδομένα να παρουσιάζουν παρόμοια μοτίβα που σχετίζονται με τα καιρικά φαινόμενα.



Εικόνα 12 Διάγραμμα στάθμης, ταχύτητας και βροχόπτωσης στις δύο γέφυρες

Matlab για Διάγραμμα στάθμης, ταχύτητας και βροχόπτωσης στις δύο γέφυρες:

```

% Δημιουργία Διαγράμματος Στάθμης, Ταχύτητας και Βροχόπτωσης
figure;

% Πρώτο Υποδιάγραμμα: Βροχόπτωση
subplot(3,1,1);
plot(df_norm.DateTime, df_norm.rain_norm, 'b-', 'LineWidth', 1);
xlabel('Ημερομηνία');
ylabel('Βροχόπτωση (mm)');
title('Διάγραμμα Βροχόπτωσης');
grid on;

% Δεύτερο Υποδιάγραμμα: Ταχύτητα
subplot(3,1,2);
plot(df_norm.DateTime, df_norm.vel_norm, 'g-', 'LineWidth', 1);
xlabel('Ημερομηνία');
ylabel('Ταχύτητα Ροής (m/s)');
title('Διάγραμμα Ταχύτητας Ροής');
grid on;

% Τρίτο Υποδιάγραμμα: Στάθμη Νερού
subplot(3,1,3);
plot(df_norm.DateTime, df_norm.dist_fr_norm, 'r-', 'LineWidth', 1, 'DisplayName', 'Γέφυρα Φραντζή');
hold on;
plot(df_norm.DateTime, df_norm.dist_kost_norm, 'k-', 'LineWidth', 1, 'DisplayName', 'Γέφυρα Κωσταλέξη');
xlabel('Ημερομηνία');
ylabel('Στάθμη Νερού (m)');
title('Διάγραμμα Στάθμης Νερού');
legend('show');
grid on;

```

Κώδικας 8 Διάγραμμα στάθμης, ταχύτητας και βροχόπτωσης

6. ΠΡΟΕΤΟΙΜΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

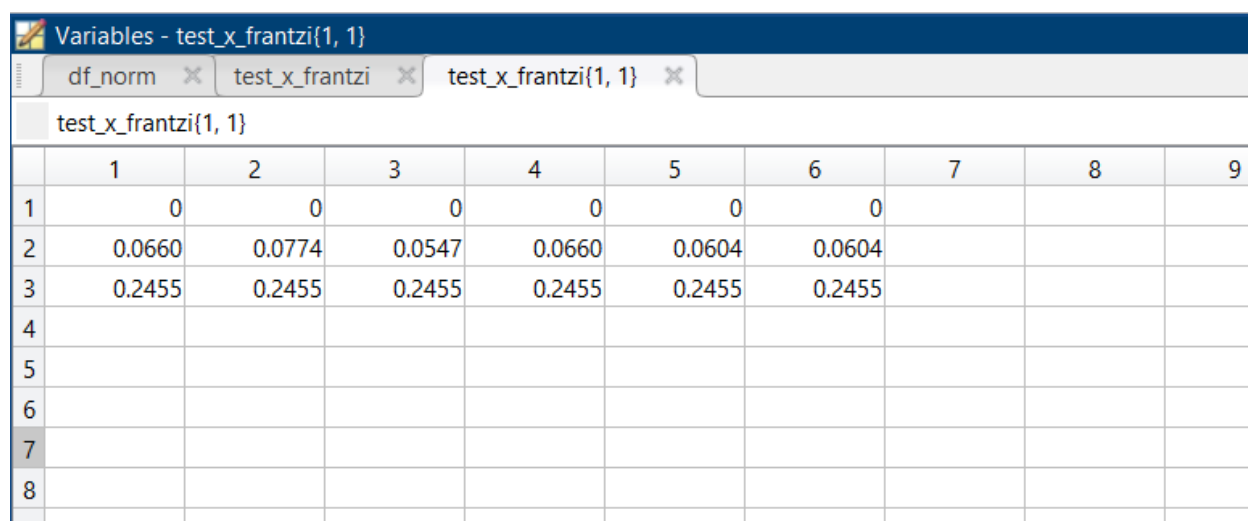
Σε αυτή την ενότητα τα δεδομένα θα προετοιμαστούν για την είσοδο στους αλγόριθμους μηχανικής μάθησης. Η προετοιμασία των δεδομένων περιλαμβάνει τα εξής βήματα:

6.1 Δημιουργία Σετ Δεδομένων με Παράθυρα Χρόνου

Αρχικά δημιουργούμε διαδοχικά παράθυρα δεδομένων με 6 χρονικά βήματα (time steps) σε cell array. Κάθε δείγμα δεδομένων περιέχει ένα παράθυρο 6 χρονικών στιγμών με 3 χαρακτηριστικά (βροχή, ταχύτητα, απόσταση). Το cell array διευκολύνει την εισαγωγή δεδομένων στο LSTM, καθώς το δίκτυο απαιτεί διαδοχικά δεδομένα σε αυτήν τη μορφή.

Σε αυτή τη διαδικασία:

- Κάθε στοιχείο του cell array περιέχει ένα πίνακα των χαρακτηριστικών για το αντίστοιχο χρονικό παράθυρο.
- Το μοντέλο χρησιμοποιεί αυτές τις ακολουθίες για να εντοπίσει μοτίβα και συσχετίσεις στα δεδομένα με την πάροδο του χρόνου.



	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	0			
2	0.0660	0.0774	0.0547	0.0660	0.0604	0.0604			
3	0.2455	0.2455	0.2455	0.2455	0.2455	0.2455			
4									
5									
6									
7									
8									
9									

Εικόνα 13 Παράδειγμα ενός cell array

Ανάλυση του Παραδείγματος

1. Cell Array:

- Το test_x_frantzi είναι ένα cell array που περιέχει πολλαπλά κελιά (το καθένα αντιπροσωπεύει ένα παράθυρο χρόνου).
- Στην εικόνα, βλέπουμε το περιεχόμενο του πρώτου κελιού:
 - Κάθε σειρά αντιστοιχεί σε ένα χαρακτηριστικό (feature).
 - Κάθε στήλη αντιπροσωπεύει ένα χρονικό βήμα (time step).

2. Δομή των Δεδομένων:

- Πρώτη σειρά: Τιμές βροχόπτωσης (rain_norm).

- Δεύτερη σειρά: Τιμές ταχύτητας (vel_norm).
- Τρίτη σειρά: Τιμές στάθμης στη γέφυρα Φράντζη (dist_fr_norm).

3. Χρονικά Βήματα:

- Οι στήλες (1 έως 6) αναπαριστούν τις τιμές των χαρακτηριστικών για διαδοχικά χρονικά βήματα.

Matlab:

```
% Ορίζουμε τις παραμέτρους
window_size = 6; % Μήκος παραθύρου (time steps)
% Δεδομένα από το df_norm για τις δύο γέφυρες
data_frantzi = [df_norm.rain_norm, df_norm.vel_norm, df_norm.dist_fr_norm]; % 3 χαρακτηριστικά
data_kostaleksi = [df_norm.rain_norm, df_norm.vel_norm, df_norm.dist_kost_norm]; % 3
χαρακτηριστικά
% Προετοιμασία κελιών
X_frantzi = {}; Y_frantzi = [];
X_kost = {}; Y_kost = [];
% Δημιουργία παραθύρων για τη γέφυρα Frantzi
for i = 1:(size(data_frantzi, 1) - window_size)
    X_frantzi{i, 1} = data_frantzi(i:i+window_size-1, :); % [3 x window_size]
    Y_frantzi(i, 1) = df_norm.dist_fr_norm(i + window_size); % Επόμενη χρονική στιγμή
end
% Δημιουργία παραθύρων για τη γέφυρα Kostaleksi
for i = 1:(size(data_kostaleksi, 1) - window_size)
    X_kost{i, 1} = data_kostaleksi(i:i+window_size-1, :); % [3 x window_size]
    Y_kost(i, 1) = df_norm.dist_kost_norm(i + window_size);
end
% Μετατροπή σε cell array
X_frantzi = X_frantzi(:); Y_frantzi = Y_frantzi(:);
X_kost = X_kost(:); Y_kost = Y_kost(:);
```

Κώδικας 9 Δημιουργία Σετ Δεδομένων με Παράθυρα Χρόνου

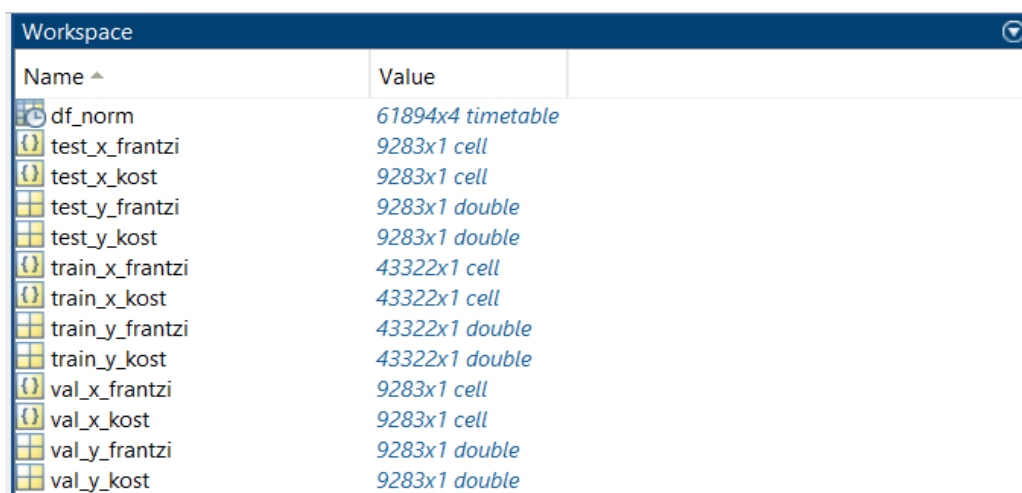
6.2 Διαχωρισμός Δεδομένων σε Εκπαίδευση, Επικύρωση και Δοκιμή

Διαχωρίζουμε τα δεδομένα σε τρία σύνολα:

- Εκπαίδευσης (Training Set): 70% των δεδομένων.
- Επικύρωσης (Validation Set): 15% των δεδομένων.
- Δοκιμής (Test Set): 15% των δεδομένων.

Ο διαχωρισμός των δεδομένων σε σύνολα εκπαίδευσης, επικύρωσης και δοκιμής είναι ένα βασικό βήμα στη διαδικασία εκπαίδευσης ενός μοντέλου μηχανικής μάθησης. Ο σωστός διαχωρισμός επιτρέπει την αντικειμενική αξιολόγηση της απόδοσης του μοντέλου, εξασφαλίζοντας τη δυνατότητά του να γενικεύει σε νέα, “αόρατα” δεδομένα. Αυτό είναι ιδιαίτερα σημαντικό, καθώς χωρίς αυτή τη διαδικασία, ένα μοντέλο μπορεί να φαίνεται αποδοτικό κατά την εκπαίδευση αλλά να αποτυγχάνει να ανταποκριθεί σε πραγματικές συνθήκες, φαινόμενο που είναι γνωστό ως υπερπροσαρμογή. Με τον διαχωρισμό, μπορούμε να αξιολογήσουμε την απόδοση του μοντέλου σε ένα ανεξάρτητο σύνολο δεδομένων, το οποίο παρέχει μια εκτίμηση της πραγματικής του απόδοσης. Τα δεδομένα εκπαίδευσης χρησιμοποιούνται για να εκπαιδεύσουν το μοντέλο, τα δεδομένα επικύρωσης για να βελτιστοποιήσουν τις παραμέτρους και τις ρυθμίσεις του, και τα δεδομένα δοκιμής για να εκτιμήσουμε πώς θα αποδίδει το μοντέλο σε πραγματικά σενάρια. Επιπλέον, ο διαχωρισμός των δεδομένων μας προστατεύει από τη διαρροή πληροφοριών από το σύνολο δοκιμών προς τη διαδικασία εκπαίδευσης. Η διαρροή δεδομένων μπορεί να οδηγήσει σε υπερβολικά αισιόδοξες εκτιμήσεις απόδοσης, δίνοντας την ψευδαίσθηση ότι το μοντέλο είναι καλύτερο από ό,τι είναι στην πραγματικότητα. Διατηρώντας τα σύνολα αυστηρά διαχωρισμένα, εξασφαλίζεται ότι το μοντέλο αξιολογείται σε δεδομένα που δεν έχει δει ποτέ πριν. Αξιολογώντας τα μοντέλα σε κοινό σύνολο δοκιμών, μπορούμε να επιλέξουμε αυτό που προσφέρει την καλύτερη απόδοση και να κάνουμε τις απαραίτητες αλλαγές στην αρχιτεκτονική του για να βελτιώσουμε τα αποτελέσματα.

Πλέον το workspace στο Matlab περιέχει τα εξής δεδομένα:



Name	Value
df_norm	61894x4 timetable
test_x_frantzi	9283x1 cell
test_x_kost	9283x1 cell
test_y_frantzi	9283x1 double
test_y_kost	9283x1 double
train_x_frantzi	43322x1 cell
train_x_kost	43322x1 cell
train_y_frantzi	43322x1 double
train_y_kost	43322x1 double
val_x_frantzi	9283x1 cell
val_x_kost	9283x1 cell
val_y_frantzi	9283x1 double
val_y_kost	9283x1 double

Εικόνα 14 Διαχωρισμός Δεδομένων σε Εκπαίδευση, Επικύρωση και Δοκιμή

Πλέον τα δεδομένα είναι έτοιμα για είσοδο στους αλγόριθμους μηχανικής μάθησης. Στο επόμενο κεφάλαιο γίνεται ανάλυση των αποτελεσμάτων και υπολογισμός του σφάλματος χρησιμοποιώντας το RMSE (Root Mean Squared Error) για να αξιολογήσουμε την ακρίβεια του μοντέλου στις προβλέψεις.

Matlab:

```
% Καθορισμός ποσοστών
train_ratio = 0.70;
val_ratio = 0.15;
test_ratio = 0.15;

% Αριθμός παραδειγμάτων
num_samples = length(X_frantzi);

% Υπολογισμός δειγμάτων
num_train = floor(train_ratio * num_samples);
num_val = floor(val_ratio * num_samples);
num_test = num_samples - num_train - num_val;

% Διαχωρισμός για τη γέφυρα Frantzi
train_x_frantzi = X_frantzi(1:num_train);
train_y_frantzi = Y_frantzi(1:num_train);
val_x_frantzi = X_frantzi(num_train+1:num_train+num_val);
val_y_frantzi = Y_frantzi(num_train+1:num_train+num_val);
test_x_frantzi = X_frantzi(num_train+num_val+1:end);
test_y_frantzi = Y_frantzi(num_train+num_val+1:end);

% Διαχωρισμός για τη γέφυρα Kostaleksi
train_x_kost = X_kost(1:num_train);
train_y_kost = Y_kost(1:num_train);
val_x_kost = X_kost(num_train+1:num_train+num_val);
val_y_kost = Y_kost(num_train+1:num_train+num_val);
test_x_kost = X_kost(num_train+num_val+1:end);
test_y_kost = Y_kost(num_train+num_val+1:end);
```

Κώδικας 10 Διαχωρισμός Δεδομένων σε Εκπαίδευση, Επικύρωση και Δοκιμή

7. ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Για την εκπαίδευση και ανάλυση των αποτελεσμάτων, αρχικά ορίζεται και εκπαιδεύεται το μοντέλο LSTM. Η εκπαίδευση πραγματοποιήθηκε με 10 εποχές (epochs). Μια εποχή αντιπροσωπεύει τον κύκλο κατά τον οποίο όλα τα δεδομένα εκπαίδευσης περνούν μία φορά από το μοντέλο για την ενημέρωση των παραμέτρων του. Για παράδειγμα, αν το σύνολο δεδομένων περιλαμβάνει 1000 δείγματα και το μέγεθος του batch είναι 10, τότε κάθε εποχή αποτελείται από 100 βήματα (iterations). Αυτή η διαδικασία εξασφαλίζει ότι κάθε δείγμα εκπαίδευσης χρησιμοποιείται για την προσαρμογή των παραμέτρων του μοντέλου.

Το μοντέλο LSTM που χρησιμοποιήθηκε στη μελέτη περιλαμβάνει συγκεκριμένα στρώματα που σχεδιάστηκαν για την καλύτερη ανάλυση των χρονοσειρών. Στη δομή του περιλαμβάνονται:

- Ένα LSTM επίπεδο, το οποίο έχει σχεδιαστεί για να αναγνωρίζει μακροπρόθεσμες εξαρτήσεις στα δεδομένα.
- Ένα πλήρως συνδεδεμένο επίπεδο (fully connected layer), που εξάγει την τελική πρόβλεψη της στάθμης του νερού.
- Μια συνάρτηση παλινδρόμησης (regression layer), η οποία υπολογίζει το σφάλμα πρόβλεψης.

Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε ο Adam optimizer (Adaptive Moment Estimation), για την προσαρμογή του ρυθμού εκμάθησης. Ο Adam optimizer συνδυάζει τα πλεονεκτήματα των αλγορίθμων Stochastic Gradient Descent (SGD) και RMSProp, καθιστώντας την εκπαίδευση γρήγορη και αποδοτική. Με την ολοκλήρωση της εκπαίδευσης, πραγματοποιούνται προβλέψεις χρησιμοποιώντας το τεστ δοκιμής. Τέλος υπολογίζεται το μέσο τετραγωνικό σφάλμα (Root Mean Squared Error), το μέσο απόλυτο σφάλμα (Mean Absolute Error) και ο συντελεστής προσδιορισμού (R^2) για να αξιολογηθεί η ακρίβεια του μοντέλου στις προβλέψεις.

Ο συντελεστής προσδιορισμού είναι μια στατιστική μέτρηση που μετρά την αναλογία της διακύμανσης στην εξαρτημένη μεταβλητή που είναι προβλέψιμη από τις ανεξάρτητες μεταβλητές. Κυμαίνεται από 0 έως 1, με υψηλότερη τιμή να υποδηλώνει καλύτερη εφαρμογή. Μια τιμή $R^2=1$ υποδηλώνει ότι το μοντέλο προβλέπει τέλεια την εξαρτημένη μεταβλητή, ενώ μια τιμή $R^2=0$ υποδηλώνει ότι το μοντέλο αποτυγχάνει να εξηγήσει καμία από τις μεταβλητές στην εξαρτημένη μεταβλητή. Για να υπολογίσουμε το R^2 , συγκρίνουμε το άθροισμα των τετραγώνων διαφορών μεταξύ των παρατηρούμενων τιμών και των προβλεπόμενων τιμών (SSR) με το συνολικό άθροισμα των τετραγωνικών διαφορών μεταξύ των παρατηρούμενων τιμών και του μέσου όρου τους (SST). Ο τύπος για τον συντελεστή προσδιορισμού είναι ο εξής:

$$R^2 = 1 - (SSR/SST)$$

7.1 Αποτελέσματα Ανάλυσης για Μοντέλο LSTM

Αρχικά και για τις δύο γέφυρες χρησιμοποιείται η εξής αρχιτεκτονική:

Πίνακας 2 LSTM με 50 κρυφά στρώματα

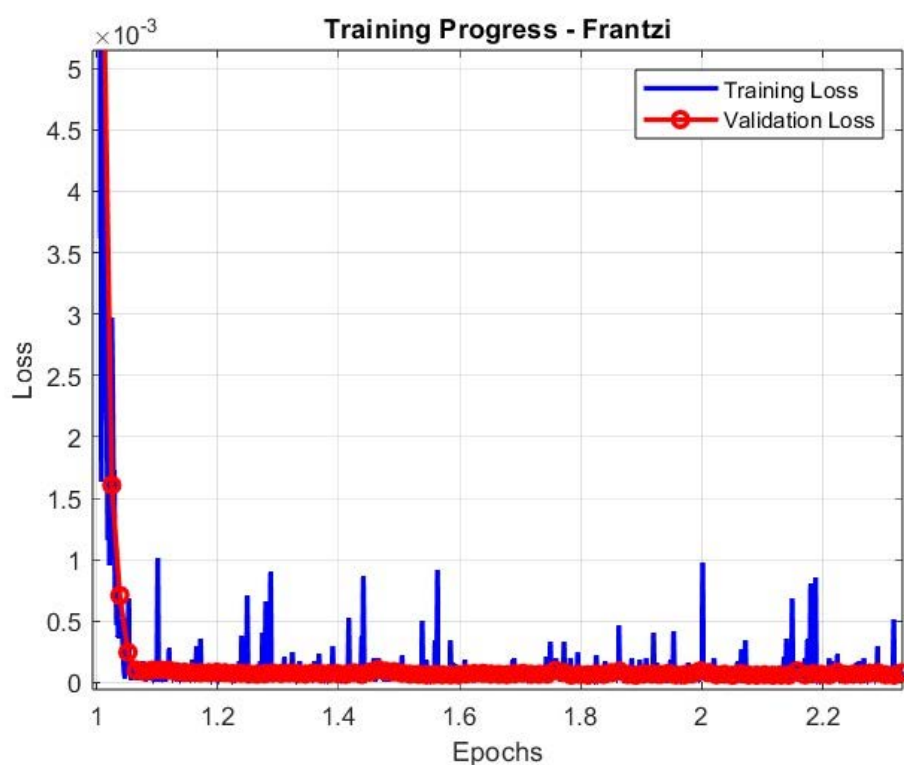
	Όνομα Στρώματος	Τύπος Στρώματος	Περιγραφή
1	sequenceinput	Sequence Input	Sequence input με 3 διαστάσεις (n_features).
2	lstm	LSTM	LSTM με 50 κρυφά στρώματα (hidden units).
3	fc	Fully Connected	Fully connected στρώμα με 1 έξοδο.
4	regressionoutput	Regression Output	Μέθοδος εξόδου με mean squared error.

Αποτελέσματα για τη γέφυρα Φρατζή (50 κρυφά στρώματα):

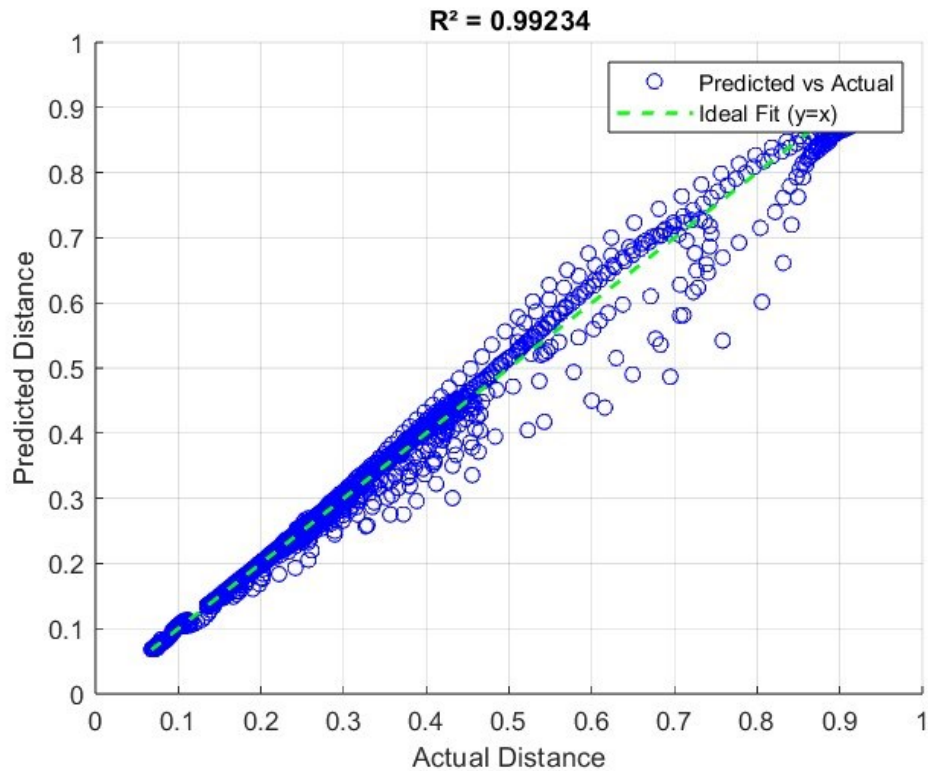
RMSE: 0.009020

MAE: 0.002381

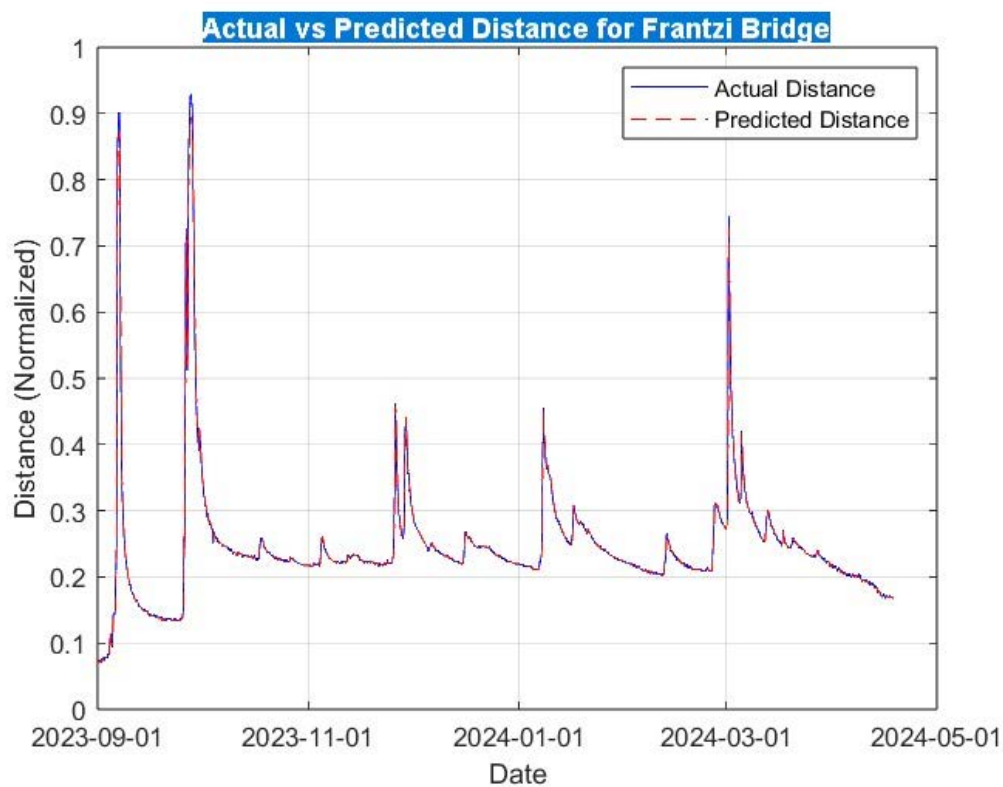
R²: 0.992342



Εικόνα 15 Training and Validation Loss LSTM με 50 κρυφά στρώματα, γέφυρα Φρατζή



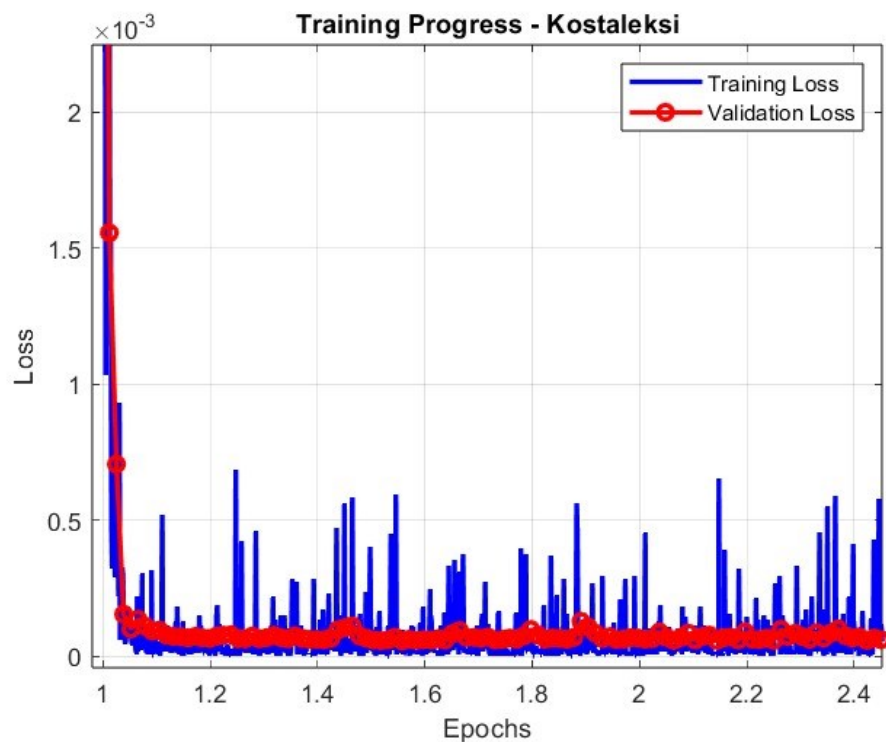
Εικόνα 16 Διάγραμμα διασποράς LSTM με 50 κρυφά στρώματα, γέφυρα Φρατζή



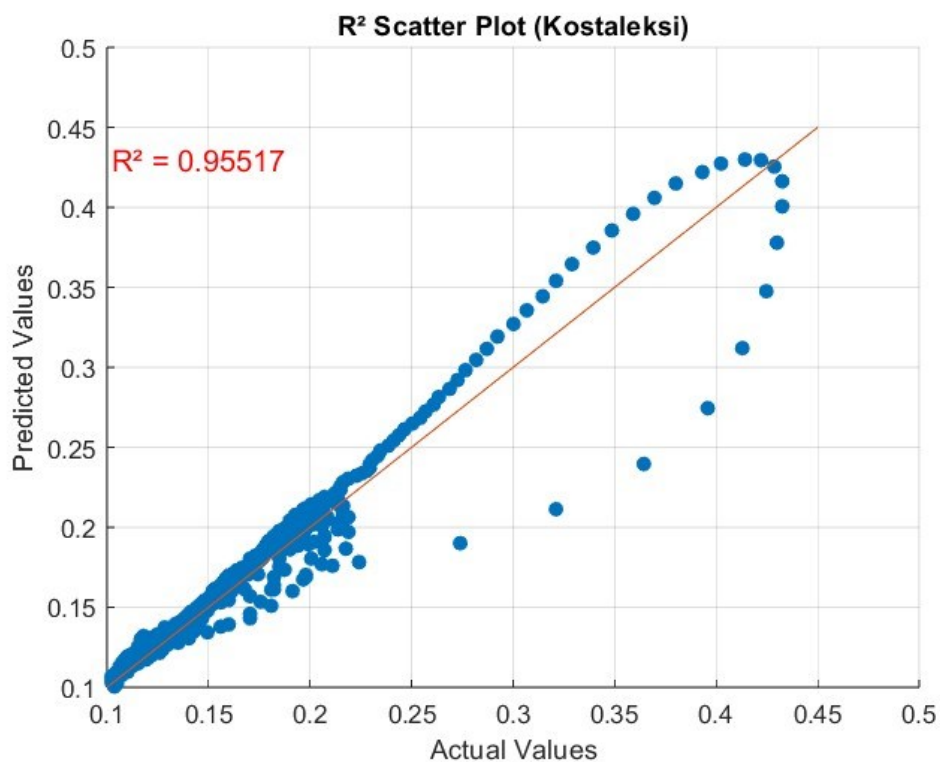
Εικόνα 17 Πραγματικές και προβλεπόμενες τιμές LSTM με 50 κρυφά στρώματα, γέφυρα Φρατζή

Αποτελέσματα για τη γέφυρα Κωσταλέξη (50 κρυφά στρώματα):

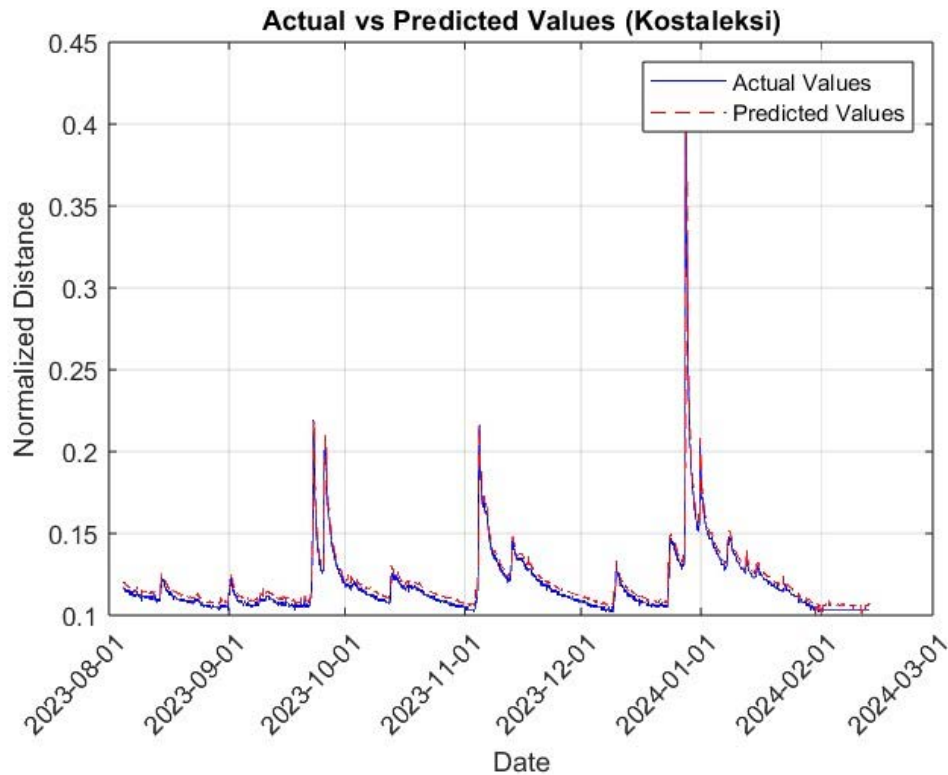
RMSE: 0.004803
MAE: 0.003460
 R^2 : 0.955174



Εικόνα 18 Training and Validation Loss LSTM με 50 κρυφά στρώματα, γέφυρα Κωσταλέξη



Εικόνα 19 Διάγραμμα διασποράς LSTM με 50 κρυφά στρώματα, γέφυρα Κωσταλέξη



Εικόνα 20 Πραγματικές και προβλεπόμενες τιμές LSTM με 50 κρυφά στρώματα, γέφυρα Κωσταλέξη

Στη συνέχεια γίνεται δοκιμή με περισσότερα κρυφά στρώματα στο LSTM Layer:

7.2 Δοκιμή LSTM με 128 κρυφά στρώματα.

Πίνακας 3 LSTM με 128 κρυφά στρώματα

	Όνομα Στρώματος	Τύπος Στρώματος	Περιγραφή
1	sequenceinput	Sequence Input	Sequence input με 3 διαστάσεις (n_features).
2	lstm	LSTM	LSTM με 128 κρυφά στρώματα (hidden units).
3	fc	Fully Connected	Fully connected στρώμα με 1 έξοδο.
4	regressionoutput	Regression Output	Μέθοδος εξόδου με mean squared error.

Αποτελέσματα για τη γέφυρα Φραντζή (128 κρυφά στρώματα):

RMSE for Frantzi: 0.007447

MAE for Frantzi: 0.005621

R² for Frantzi: 0.976122

Αποτελέσματα για τη γέφυρα Κωσταλέξη (128 κρυφά στρώματα):

RMSE: 0.004566

MAE: 0.002745

R²: 0.959495

7.3 Δοκιμή με 256 κρυφά στρώματα

Πίνακας 4 LSTM με 256 hidden units

	Όνομα Στρώματος	Τύπος Στρώματος	Περιγραφή
1	sequenceinput	Sequence Input	Sequence input με 3 διαστάσεις (n_features).
2	lstm	LSTM	LSTM με 256 κρυφά στρώματα (hidden units).
3	fc	Fully Connected	Fully connected στρώμα με 1 έξοδο.
4	regressionoutput	Regression Output	Μέθοδος εξόδου με mean squared error.

Αποτελέσματα για τη Γέφυρα Φρατζή (256 κρυφά στρώματα):

RMSE: 0.005765

MAE: 0.002214

R²: 0.985691

Αποτελέσματα για τη Γέφυρα Κωσταλέξη (256 κρυφά στρώματα):

RMSE: 0.003785

MAE: 0.000983

R²: 0.972169

Αποτελέσματα για LSTM με διαφορετικά κρυφά στρώματα για τη γέφυρα Φρατζή:

Πίνακας 5 Αποτελέσματα για όλα τα κρυφά στρώματα για τη Γέφυρα Φρατζή

Hidden Units	RMSE	MAE	R ²
50	0.009020	0.002381	0.992342
128	0.007447	0.005621	0.976122
256	0.005765	0.002214	0.985691

ΑΝΑΛΥΣΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

- Με 50 κρυφά στρώματα, το R² είναι 0.992, δείχνοντας πολύ καλή προσαρμογή.
- Με 128 κρυφά στρώματα, το R² μειώθηκε ελαφρώς σε 0.976, με αύξηση του MAE.
- Με 256 κρυφά στρώματα, το R² βελτιώθηκε ξανά σε 0.986, ενώ το RMSE μειώθηκε, δείχνοντας ότι η προσαρμογή είναι πιο σταθερή.

Αποτελέσματα LSTM με διαφορετικά κρυφά στρώματα για τη γέφυρα Κωσταλέξη:

Πίνακας 6 Αποτελέσματα για όλα τα κρυφά στρώματα για τη Γέφυρα Κωσταλέξη

Hidden Units	RMSE	MAE	R ²
50	0.004803	0.003460	0.955174
128	0.004566	0.002745	0.959495
256	0.003785	0.000983	0.972169

ΑΝΑΛΥΣΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

- Με 50 κρυφά στρώματα, το R^2 είναι 0.955, κάτι που δείχνει καλή προσαρμογή, αν και χαμηλότερη από τη γέφυρα Φρατζή.
- Με 128 κρυφά στρώματα, το R^2 αυξήθηκε ελαφρώς σε 0.959, με σημαντική μείωση του MAE.
- Με 256 κρυφά στρώματα, το R^2 βελτιώθηκε αισθητά σε 0.972, με το χαμηλότερο MAE (0.000983), κάτι που δείχνει ότι η υψηλότερη πολυπλοκότητα βοηθά σε αυτήν την περίπτωση.

Γενικές Παρατηρήσεις για τις δοκιμές με διαφορετικό αριθμό κρυφών στρωμάτων

Η γέφυρα Φρατζή φαίνεται να επιτυγχάνει την καλύτερη προσαρμογή με 50 κρυφά στρώματα, ενώ για την Κωσταλέξη, τα 256 κρυφά στρώματα δίνουν τα καλύτερα αποτελέσματα.

7.4 Αρχιτεκτονική με Δύο LSTM Στρώματα

Πίνακας 7 Αρχιτεκτονική με Δύο LSTM Στρώματα

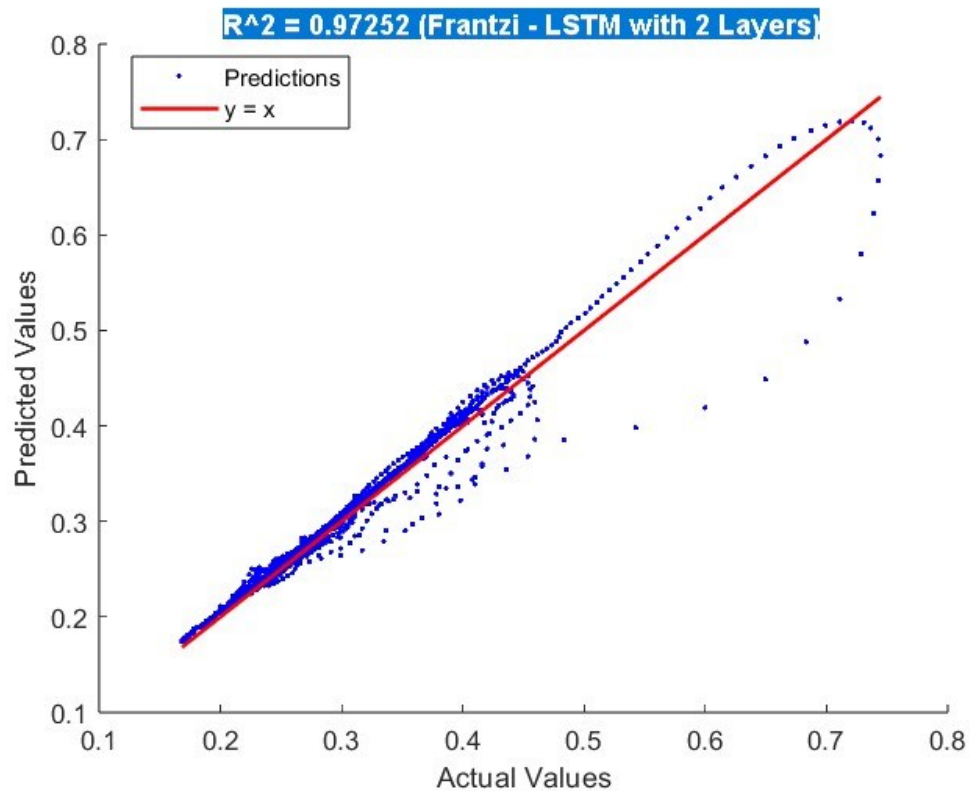
	Όνομα Στρώματος	Τύπος Στρώματος	Περιγραφή
1	sequenceinput	Sequence Input	Sequence input με 3 διαστάσεις.
2	lstm	LSTM	LSTM με 128 hidden units.
3	lstm	LSTM	LSTM με 64 hidden units.
4	fc	Fully Connected	Fully connected στρώμα με 1 έξοδο.
5	regressionoutput	Regression Output	Μέθοδος εξόδου με mean squared error.

Αποτελέσματα για τη γέφυρα Φρατζή (με δύο LSTM Στρώματα):

RMSE: 0.007989

MAE: 0.005095

R^2 : 0.972519



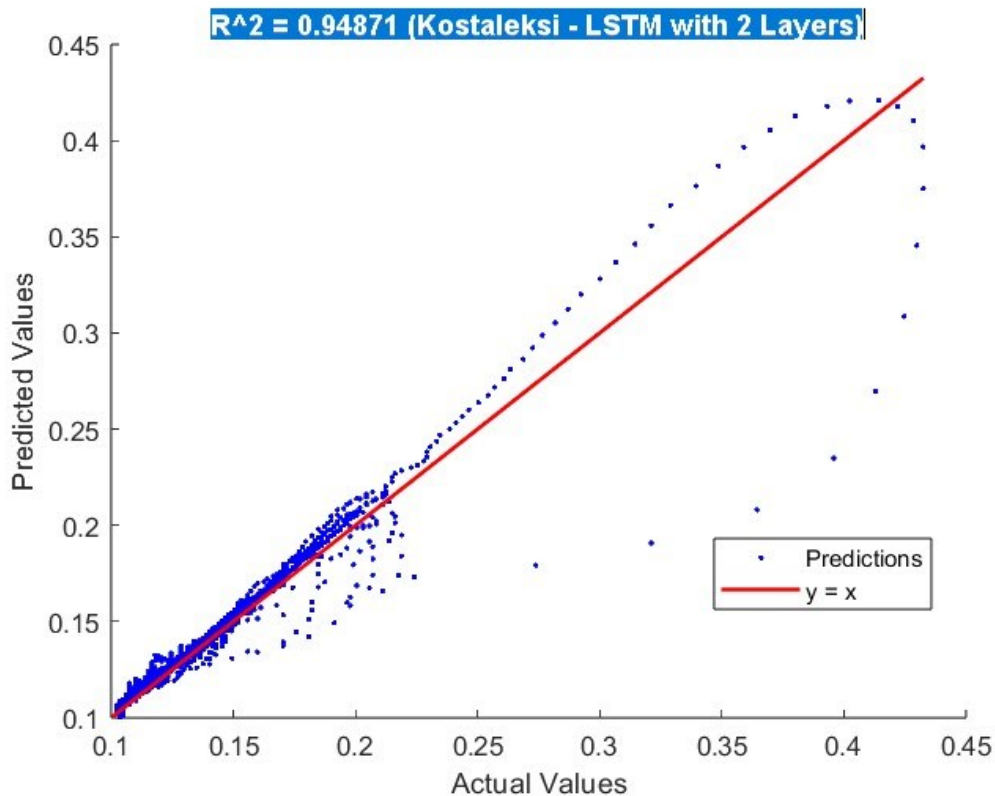
Εικόνα 21 Διάγραμμα διασποράς για δύο LSTM στρώματα, γέφυρα Φραντζή

Αποτελέσματα για τη γέφυρα Κωσταλέξη (με δύο LSTM Στρώματα):

RMSE: 0.005138

MAE: 0.002861

R^2 : 0.948713



Εικόνα 22 Διάγραμμα διασποράς για δύο2 LSTM στρώματα, γέφυρα Κωσταλέξη

7.5 Αρχιτεκτονική CNN-LSTM με Δύο LSTM Στρώματα

Πίνακας 8 Αρχιτεκτονική CNN-LSTM με Δύο LSTM Στρώματα

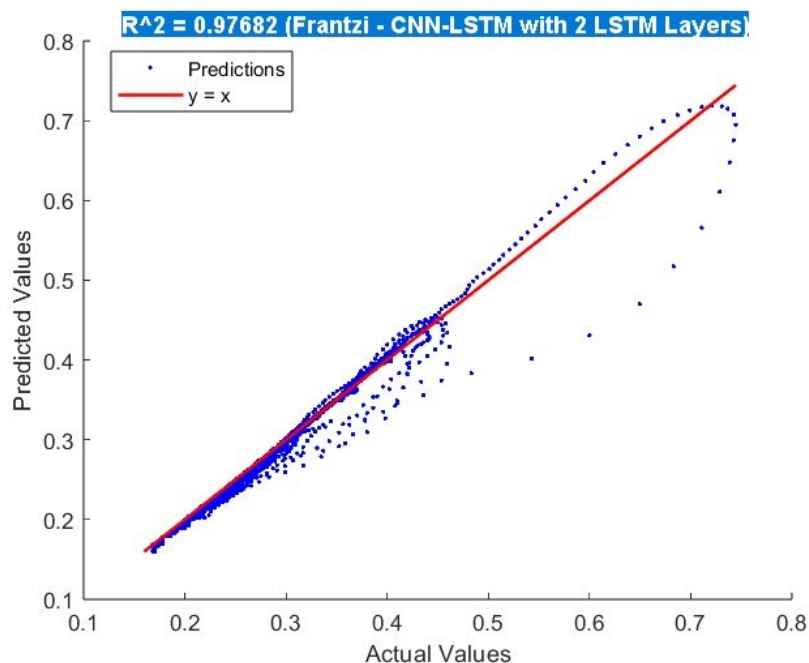
	Όνομα Στρώματος	Τύπος Στρώματος	Περιγραφή
1	sequenceinput	Sequence Input	Sequence input με 3 διαστάσεις (n_features).
2	conv1d	1-D Convolution	64 φίλτρα, μέγεθος 3, βήμα 1, padding "same".
3	relu1	ReLU	Ενεργοποίηση ReLU.
4	maxpool	1-D Max Pooling	Μέγεθος παραθύρου 2, βήμα 1, padding [0 0].
5	lstm1	LSTM	LSTM με 128 μονάδες (hidden units).
6	lstm2	LSTM	LSTM με 64 μονάδες (hidden units).
7	fc	Fully Connected	Fully connected στρώμα με 1 έξοδο.
8	regressionoutput	Regression Output	Μέθοδος εξόδου με mean squared error για πρόβλεψη συνεχών τιμών (π.χ., στάθμη)

Αποτελέσματα για τη γέφυρα Φρατζή (CNN-LSTM με 2 LSTM στρώματα):

RMSE: 0.007338

MAE: 0.004561

R²: 0.976817



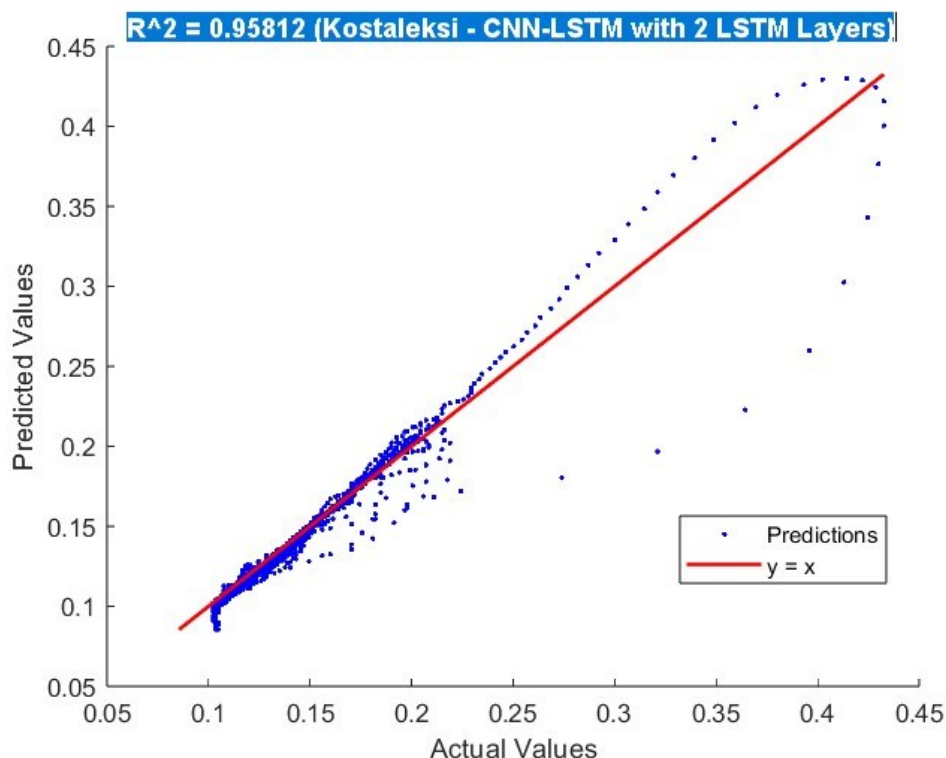
Εικόνα 23 Διάγραμμα διασποράς για CNN-LSTM με δύο LSTM στρώματα, γέφυρα Φρατζή

Αποτελέσματα για τη γέφυρα Κωσταλέξη (CNN-LSTM με 2 LSTM στρώματα):

RMSE: 0.004643

MAE: 0.002528

R²: 0.958120



Εικόνα 24 Διάγραμμα διασποράς για CNN-LSTM με δύο LSTM στρώματα, γέφυρα Κωσταλέξη

Σύγκριση Αποτελεσμάτων ανάμεσα σε LSTM και CNN-LSTM για τις γέφυρες Φρατζή και Κωσταλέξη

Για τη γέφυρα Φρατζή:

Πίνακας 9 Σύγκριση LSTM και CNN-LSTM για τη γέφυρα Φρατζή

Τύπος Δικτύου	RMSE	MAE	R^2
2 LSTM στρώματα	0.007989	0.005095	0.972519
CNN-LSTM	0.006616	0.003513	0.981150

Ανάλυση για τη Γέφυρα Φρατζή:

- Το μοντέλο CNN-LSTM με 2 LSTM στρώματα είχε καλύτερη ακρίβεια (υψηλότερο $R^2 = 0.981150$) από το LSTM με 2 στρώματα ($R^2 = 0.972519$).
- Το CNN-LSTM παρουσίασε μικρότερη RMSE (0.006616) σε σύγκριση με το LSTM (0.007989), δείχνοντας καλύτερη πρόβλεψη.

Για τη γέφυρα Κωσταλέξη:

Πίνακας 10 Σύγκριση LSTM και CNN-LSTM για τη γέφυρα Κωσταλέξη

Τύπος Δικτύου	RMSE	MAE	R^2
2 LSTM στρώματα	0.005765	0.002214	0.985691
CNN-LSTM	0.004643	0.002528	0.958120

- Αντίστροφη τάση παρατηρείται εδώ. Το LSTM με 2 στρώματα είχε καλύτερη επίδοση ($R^2 = 0.985691$) σε σχέση με το CNN-LSTM ($R^2 = 0.958120$).
- Το LSTM επίσης πέτυχε μικρότερη RMSE (0.005765 έναντι 0.004643) και MAE, αποδεικνύοντας ότι ήταν πιο ακριβές για τη συγκεκριμένη γέφυρα.

7.6 Γραμμικό μοντέλο LINEAR MODEL

Τα γραμμικά μοντέλα αποτελούν μια από τις πιο βασικές και δημοφιλείς μεθόδους στη στατιστική και τη μηχανική μάθηση για την πρόβλεψη και την ανάλυση δεδομένων. Στόχος τους είναι να μοντελοποιήσουν τη σχέση μεταξύ μιας εξαρτημένης μεταβλητής (dependent variable) και μιας ή περισσότερων ανεξάρτητων μεταβλητών (independent variables) μέσω μιας γραμμικής συνάρτησης. Έγινε δοκιμή με το πιο απλό μοντέλο, που περιέχει ένα νευρώνα και γραμμική συνάρτηση μεταφοράς.

Γέφυρα Φρατζή

Πίνακας 11 Γραμμικό μοντέλο

Intercept	Εκτίμηση (Estimate)	Σφάλμα (SE)	t-Στατιστικό (tStat)	p-Τιμή (pValue)
	0.022528	0.00038321	58.787	0
x1	-0.004479	0.0039941	-1.1214	0.26212
x2	0.43928	0.0055651	78.935	0
x3	1.271	0.0026546	478.8	0

Number of observations: 49515, Error degrees of freedom: 49511

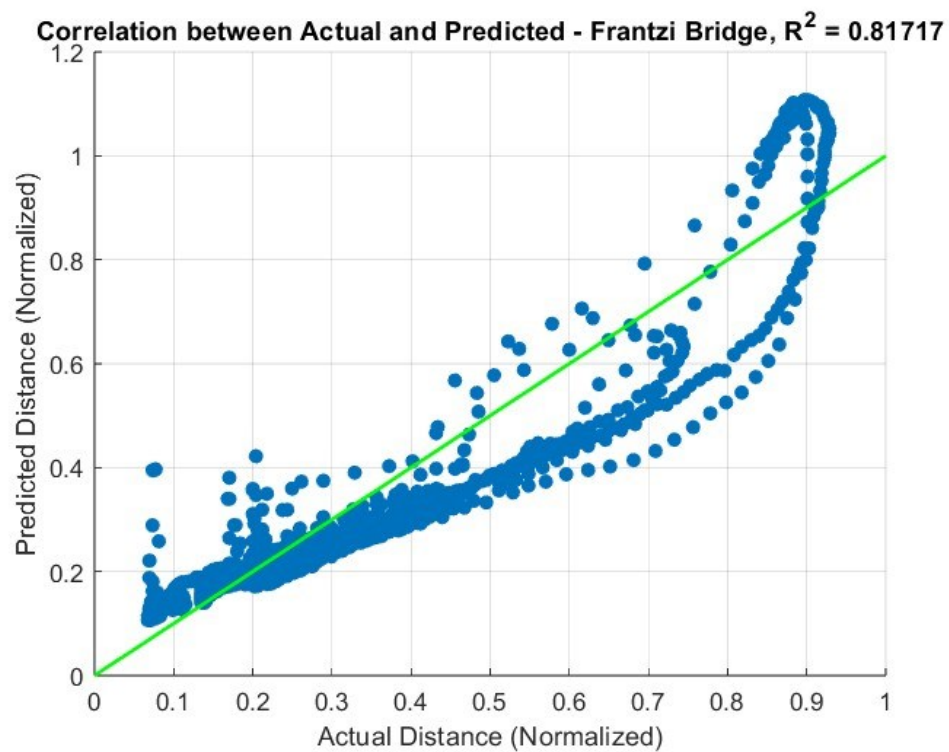
Εμφάνιση των αποτελεσμάτων

MSE = 0.0023616 m²

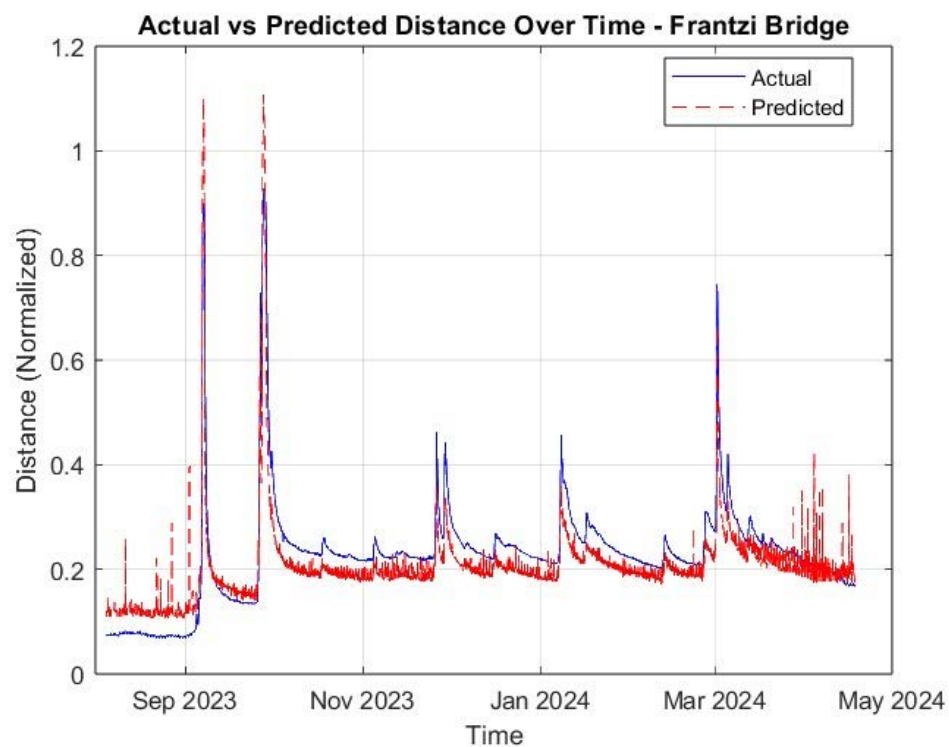
MAE = 0.039163 m

MAPE = 21.0478 %

$R^2 = 0.81717$



Εικόνα 25 Διάγραμμα διασποράς, γραμμικό μοντέλο γέφυρα Φρατζή



Εικόνα 26 Σύγκριση Πραγματικών και Προβλεπόμενων Τιμών, γραμμικό μοντέλο γέφυρα Φρατζή

8. ΑΝΑΚΕΦΑΛΑΙΩΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Γέφυρα Φρατζή:

Πίνακας 12 Ανακεφαλαίωση αποτελεσμάτων γέφυρα Φρατζή

Μοντέλο	RMSE	MAE	R ²
50 Hidden Units	0.00902	0.00238	0.99234
128 Hidden Units	0.00745	0.00562	0.97612
256 Hidden Units	0.00577	0.00221	0.98569
CNN-LSTM	0.00662	0.00351	0.98115
2 LSTM στρώματα	0.00799	0.00510	0.97252

Ανάλυση

- Η καλύτερη απόδοση επιτυγχάνεται με την αρχιτεκτονική των 256 Hidden Units, με το μικρότερο RMSE (0.00577) και το μεγαλύτερο R² (0.98569).
- Η αρχιτεκτονική με CNN-LSTM αποδίδει επίσης καλά, αλλά υπολείπεται ελαφρώς σε ακρίβεια σε σχέση με τα 256 Hidden Units.

Γέφυρα Κωσταλέξη:

Πίνακας 13 Ανακεφαλαίωση αποτελεσμάτων γέφυρα Κωσταλέξη

Μοντέλο	RMSE	MAE	R ²
50 Hidden Units	0.00480	0.00346	0.95517
128 Hidden Units	0.00457	0.00275	0.95950
256 Hidden Units	0.00379	0.00098	0.97217
CNN-LSTM	0.00457	0.00275	0.95950
2 LSTM στρώματα	0.00514	0.00286	0.94871

Ανάλυση

- Η καλύτερη απόδοση επιτυγχάνεται επίσης με την αρχιτεκτονική των 256 Hidden Units, με το μικρότερο RMSE (0.00379) και το μεγαλύτερο R² (0.97217).
- Η προσθήκη δεύτερου LSTM στρώματος (2 LSTM Layers) μειώνει την απόδοση σε σχέση με τα 256 Hidden Units.

9. ΔΟΚΙΜΕΣ ΜΕ ΠΕΡΙΣΣΟΤΕΡΕΣ ΕΠΟΧΕΣ ΕΚΠΑΙΔΕΥΣΗΣ

Η εκπαίδευση ενός LSTM μοντέλου απαιτεί τη σωστή επιλογή των εποχών, ώστε το μοντέλο να μάθει τα μοτίβα των δεδομένων χωρίς να υποεκπαιδεύεται (underfitting) ή να υπερπροσαρμόζεται (overfitting). Δοκιμάζοντας περισσότερες εποχές, μπορούμε να εντοπίσουμε το ιδανικό σημείο εκπαίδευσης, παρακολουθώντας τη μεταβολή των δεικτών απόδοσης (RMSE, MAE, R^2) και αξιολογώντας πότε το μοντέλο σταματά να βελτιώνεται. Αν η απόδοση αυξάνεται, περισσότερες εποχές μπορεί να οδηγήσουν σε καλύτερα αποτελέσματα. Αντίθετα, αν παρατηρηθεί υπερπροσαρμογή, ίσως χρειάζεται κανονικοποίηση, dropout ή μεγαλύτερος όγκος δεδομένων. Οι δοκιμές με διαφορετικά πλήθη εποχών είναι απαραίτητες για την εύρεση της βέλτιστης ισορροπίας μεταξύ ακρίβειας και γενίκευσης του μοντέλου. Πραγματοποιήθηκαν δοκιμές για την καλύτερη αρχιτεκτονική με τα 256 Hidden Units με 100 και 200 εποχές εκπαίδευσης (MaxEpochs) και για τις δύο γέφυρες.

Γέφυρα Φρατζή:

Πίνακας 14 Διαφορετικός αριθμός εποχών γέφυρα Φρατζή

Max Epochs	RMSE	MAE	R^2
10	0.00577	0.00221	0.98569
100	0.005301	0.001753	0.987904
200	0.005694	0.003236	0.986043

Γέφυρα Κωσταλέξη:

Πίνακας 15 Διαφορετικός αριθμός εποχών γέφυρα Κωσταλέξη

Max Epochs	RMSE	MAE	R^2
10	0.00379	0.00098	0.97217
100	0.004575	0.003102	0.959328
200	0.003579	0.001017	0.975108

ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΜΕ ΔΙΑΦΟΡΕΤΙΚΟ ΑΡΙΘΜΟ ΜΕΓΙΣΤΩΝ ΕΠΟΧΩΝ ΕΚΠΑΙΔΕΥΣΗΣ

Γέφυρα Φρατζή:

- 10 εποχές: Το R^2 δείχνει καλή συσχέτιση (0.98569), αλλά το RMSE (0.00577) και το MAE (0.00221) είναι υψηλότερα από τις 100 και 200 εποχές, γεγονός που υποδηλώνει ότι το μοντέλο δεν έχει μάθει πλήρως τα μοτίβα στα δεδομένα.
- 100 εποχές: Παρατηρούμε το καλύτερο RMSE (0.005301) και MAE (0.001753), ενώ το R^2 είναι το υψηλότερο (0.987904), γεγονός που δείχνει ότι το μοντέλο έχει εκπαιδευτεί επαρκώς χωρίς υπερπροσαρμογή.
- 200 εποχές: Το R^2 μειώνεται ελαφρώς (0.986043), ενώ το MAE αυξάνεται (0.003236), κάτι που δείχνει πιθανή υπερπροσαρμογή, όπου το μοντέλο αρχίζει να μαθαίνει "θόρυβο" αντί για πραγματικά μοτίβα.

Γέφυρα Κωσταλέξη:

- 10 εποχές: Έχει χαμηλό MAE (0.00098) και σχετικά υψηλό R^2 (0.97217), αλλά το RMSE (0.00379) δείχνει ότι το μοντέλο μπορεί να μην έχει μάθει αρκετά.
- 100 εποχές: Το R^2 μειώνεται (0.959328) και το RMSE αυξάνεται (0.004575), γεγονός που δείχνει ότι το μοντέλο δεν βελτιώνεται σημαντικά με 100 εποχές.
- 200 εποχές: Το RMSE (0.003579) και το MAE (0.001017) είναι τα καλύτερα, ενώ το R^2 φτάνει το (0.975108), υποδηλώνοντας ότι το μοντέλο γενικεύει καλύτερα τα δεδομένα.

Για τη Γέφυρα Φρατζή, οι 100 εποχές είναι η βέλτιστη επιλογή, καθώς μετά από αυτό το σημείο δεν υπάρχει σημαντική βελτίωση και αρχίζει να εμφανίζεται υπερπροσαρμογή. Για την Κωσταλέξη, οι 200 εποχές αποδίδουν καλύτερα, καθώς το μοντέλο έχει τη χαμηλότερη τιμή σφάλματος και τη βέλτιστη απόδοση.

10. ΣΥΜΠΕΡΑΣΜΑΤΑ

Βασικό συμπέρασμα της διπλωματικής εργασίας είναι ότι τα νευρωνικά δίκτυα LSTM αποτελούν μια αξιόπιστη επιστημονική μέθοδο για τον τομέα της υδρολογίας και τα μοντέλα πρόβλεψης πλημμυρών. Η διπλωματική ανέδειξε σημαντικά αποτελέσματα που επικυρώνουν το θεωρητικό υπόβαθρο και ενισχύουν την κατανόηση του τρόπου λειτουργίας των αναδρομικών νευρωνικών δικτύων σε πραγματικά δεδομένα. Επιβεβαιώθηκε η ικανότητά τους να διαχειρίζονται ιστορικά δεδομένα αλλά και να αντιλαμβάνονται τις χρονικές τους εξαρτήσεις. Σημαντικό στοιχείο μελέτης αποτέλεσε και η σύγκρισή των αναδρομικών νευρωνικών δικτύων με των πιο απλών γραμμικών μοντέλων αλλά και πιο σύνθετων, αναδεικνύοντας την υπεροχή τους στα προβλήματα πρόγνωσης με ανάλυση δεδομένων χρονοσειρών.

Αναγνωρίζοντας τη μεγάλη δυσκολία πρόβλεψης της στάθμης του νερού, λόγω των μη γραμμικών σχέσεων και της συμπεριφοράς που διέπουν τα υδρολογικά συστήματα, πραγματοποιήθηκε μια διαδικασία δοκιμών και βελτιστοποίησης για να επιτευχθεί η σύγκλιση σε ένα μοντέλο που παρέχει το χαμηλότερο δυνατό σφάλμα πρόβλεψης. Από την εργασία έγινε κατανοητός ο τρόπος λειτουργίας των νευρωνικών δικτύων και ο λόγος που υπερσχύουν έναντι των άλλων μεθόδων μηχανικής μάθησης στο κομμάτι της πρόβλεψης φυσικών καταστροφών. Το μοντέλο LSTM, χάρη στην ικανότητά του να διαχειρίζεται και να συγκρατεί πληροφορίες για μεγάλα χρονικά διαστήματα, αποδείχθηκε ιδιαίτερα αποτελεσματικό στην πρόβλεψη της στάθμης του ποταμού, αξιοποιώντας την εξάρτηση της πληροφορίας σε χρονικά βήματα. Όλα τα LSTM μοντέλα παρουσίασαν ικανοποιητική απόδοση, όμως το μοντέλο με τα 256 κρυφά στρώματα αναδείχτηκε ως η καλύτερη επιλογή για τις προβλέψεις της στάθμης του Σπερχειού ποταμού. Το CNN-LSTM, παρότι πιο σύνθετο, προσφέρει καλή ακρίβεια, αλλά δεν υπερτερεί των μοντέλων LSTM με σταθερό αριθμό κρυφών στρωμάτων. Οι δοκιμές με διαφορετικό αριθμό εποχών έδειξαν ότι η επίδοση του μοντέλου διαφέρει μεταξύ των γεφυρών. Για τη Γέφυρα Φρατζή, οι 100 εποχές ήταν η βέλτιστη επιλογή, καθώς προσέφεραν την καλύτερη ισορροπία μεταξύ σφάλματος και γενίκευσης. Αντίθετα, για τη Γέφυρα Κωσταλέξη, η βέλτιστη απόδοση επιτεύχθηκε στις 200 εποχές, όπου το μοντέλο εμφάνισε το χαμηλότερο σφάλμα και την υψηλότερη ακρίβεια. Αυτά τα ευρήματα επιβεβαιώνουν τη σημασία του σωστού προσδιορισμού των υπερπαραμέτρων για τη βέλτιστη απόδοση κάθε μοντέλου.

Στο πρώτο σκέλος της εργασίας παρήχθησαν σημαντικές γνώσεις για την ανάλυση δεδομένων με τη χρήση του MATLAB, τόσο στο κομμάτι της εισαγωγής των δεδομένων στα μοντέλα μηχανικής μάθησης, όσο και στην προεπεξεργασία τους. Τα δεδομένα που συλλέχθηκαν για τις γέφυρες Φραντζή και Κωσταλέξη έχουν καλή ακρίβεια και καλύπτουν μεγάλο εύρος χρόνου, αλλά απαιτείται η προσεκτική επεξεργασία τους για την εξάλειψη των κενών και των λανθασμένων τιμών. Οι διαδικασίες κανονικοποίησης και κάλυψης των κενών μέσω της γραμμικής παρεμβολής φαίνεται να έχουν συμβάλει σημαντικά στη βελτίωση της ακρίβειας του μοντέλου. Από τα κοινά διαγράμματα που παρουσιάζουν τις μεταβολές των δεδομένων στην πάροδο του χρόνου έγινε κατανοητό ότι υπάρχει ξεκάθαρη συσχέτιση ανάμεσα στη βροχόπτωση, την ταχύτητα ροής και τη στάθμη νερού, με τα δεδομένα να παρουσιάζουν παρόμοια μοτίβα. Ο συνδυασμός των τριών δεδομένων εισόδου (βροχόπτωση, ταχύτητα ροής, στάθμη) επιβεβαιώθηκε ως επαρκής για την

πρόβλεψη, αλλά η εισαγωγή επιπλέον μεταβλητών, όπως η διατομή του ποταμού και η υγρασία εδάφους, θα μπορούσε να αυξήσει την ακρίβεια.

Η παρούσα μελέτη ανοίγει νέες προοπτικές για μελλοντική έρευνα. Αρχικά, θα μπορούσαν να πραγματοποιηθούν επιπλέον δοκιμές και διαδικασίες βελτιστοποίησης των παραμέτρων και της αρχιτεκτονικής του δικτύου. Επιπλέον, επιπρόσθετες εξωτερικές μεταβλητές αλλά και αύξηση του αριθμού των διαθέσιμων δεδομένων προς επεξεργασία, θα είχαν τη δυνατότητα μείωσης του θορύβου και καλύτερης επεξήγησης των μη γραμμικών σχέσεων της πρόβλεψης. Στην έρευνα αξιοποιήθηκαν κυρίως μετεωρολογικά δεδομένα ενώ θα ήταν εξίσου αποτελεσματικό ένα σύνολο μεταβλητών που αναλύουν τα χωρικά στοιχεία του ποταμού όπως το εμβαδόν της διατομής, η κλίση και το βάθος. Τέλος η ενσωμάτωση του μοντέλου σε ένα σύστημα πρόγνωσης πραγματικού χρόνου θα μπορούσε να βελτιώσει σημαντικά τη διαχείριση κρίσεων και να προστατεύσει ανθρώπινες ζωές και περιουσίες.

11. ΒΙΒΛΙΟΓΡΑΦΙΑ

1. A. Jadon, A. P. (2022). A Comprehensive Survey of Regression Based Loss Functions for Time Series Forecasting.
2. Amit, D. J. (1989). Modeling Brain Functions: The World of Attractor Neural Networks. New York.
3. Beven, K. (2012). *Rainfall-Runoff Modelling: The Primer: Second Edition*.
4. bigblue.academy. (2023). <https://bigblue.academy/gr/neuronika-diktua>.
5. C. M. Teng, “ . N. (1999). Correcting Noisy Data.
6. Cho, M. (2022). Water Level Prediction Model Applying a Long Short-Term Memory (LSTM)–Gated Recurrent Unit (GRU) Method for Flood Prediction .
7. Daniell, T. M. (1991). Neural networks. Applications in hydrology and water resources engineering.
8. Graves, A. (2012). Supervised Sequence Labelling with Recurrent Neural Networks.
9. Guha-Sapir D, V. F. (2010). Annual Disaster Statistical Review .
10. Haykin, S. S. (2016). Neural Networks and Learning Machines.
11. Hu, J. W.-B. (2001). A comparison of several approaches to missing attribute values in data mining.
12. IPCC, I. P. (2012). Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Geneva, Switzerland, Switzerland.
13. Jun Li, G. W. (2024). Ανάκτηση από Optimizing flood predictions by integrating LSTM and physical-based models with mixed historical and simulated data.
14. Le. (2019). *Application of long short-term memory (LSTM) neural network for flood forecasting*. Ανάκτηση από
15. Lee. (2016). Development of Rainfall-Flood Damage Estimation Function using Nonlinear Regression Equation. J. Soc. Disaster Inf. 2016, .
16. Li, X. (2022). *Hybrid cnn-lstm models for river flow prediction*. Ανάκτηση από <https://doi.org/10.2166/ws.2022.170>.
17. S. A. Alasadi and W. S. Bhaya. (2017). Review of data preprocessing techniques in data mining.
18. S. R. Dubey, S. K. (2022). Activation functions in deep learning.
19. Schmidhuber, S. H. (1997). “Long Short-Term Memory,” Neural Comput.,.
20. Yaseen. (2015). *Artificial intelligence based models for stream-flow forecasting*.
21. Ιστός Μετεωρολογικού σταθμού Οινόφυτων. (χ.χ.). Ανάκτηση από <https://oinofytaweather.gr/stathmos.html>
22. Πλέρου, Α. (χ.χ.). Τεχνητά νευρωνικά δίκτυα προσομοίωσης του ανθρωπίνου εγκεφάλου.

