ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ

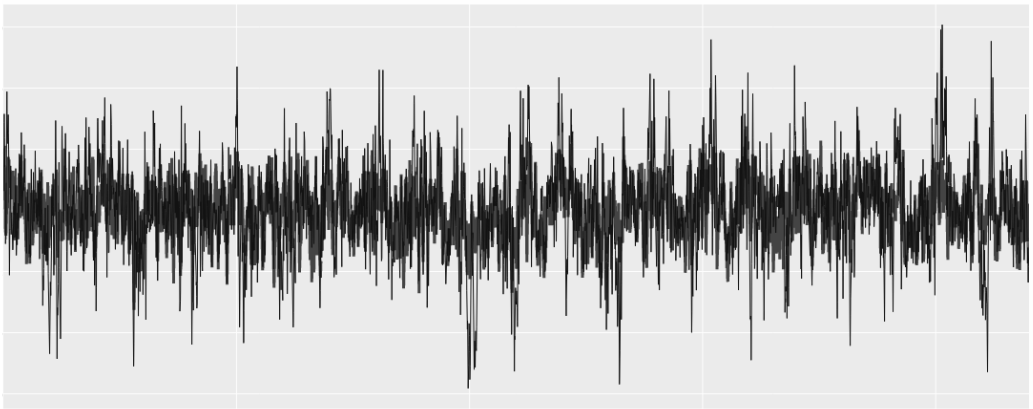
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΙΣΤΟΥ

Πρόγνωση Μετεωρολογικών Παραμέτρων με τη Χρήση Νευρωνικών Δικτύων και Μαθηματικών Μοντέλων

Μεταπτυχιακή Εργασία

Νικόλαος Γκίκας



Αθήνα, Φεβρουάριος 2023

HAROKOPIO UNIVERSITY

SCHOOL OF DIGITAL TECHNOLOGY

DEPARTMENT OF INFORMATICS AND TELEMATICS

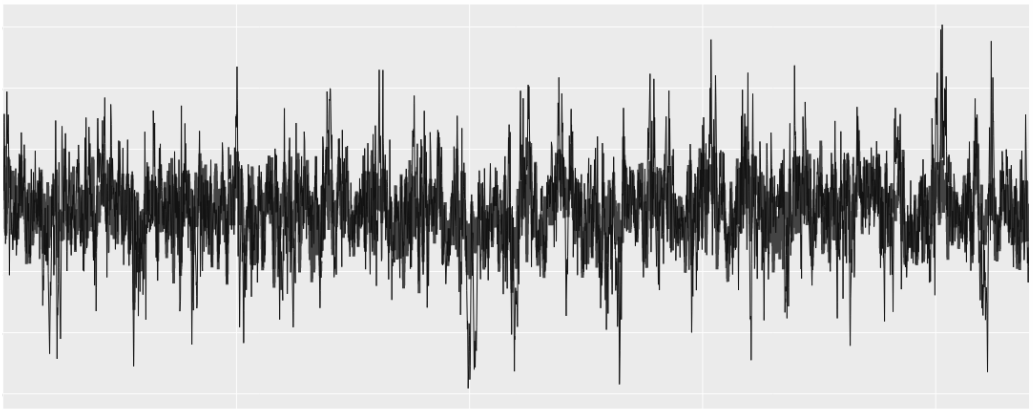
POSTGRADUATE PROGRAMME

WEB ENGINEERING

Forecasting Atmospheric Parameters Using Artificial Intelligence

Master Thesis

Nikolaos Gkikas



Athens, February 2023

ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΙΣΤΟΥ

Τριμελής Εξεταστική Επιτροπή

Βαρλάμης Ηρακλής
Αναπληρωτής Καθηγητής, Τμήμα Πληροφορικής και Τηλεματικής,
Χαροκόπειο Πανεπιστήμιο

Δίου Χρήστος
Επίκουρος Καθηγητής, Τμήμα Πληροφορικής και Τηλεματικής,
Χαροκόπειο Πανεπιστήμιο

Κατσαφάδος Πέτρος
Αναπληρωτής Καθηγητής, Τμήμα Γεωγραφίας,
Χαροκόπειο Πανεπιστήμιο

Ο Νικόλαος Γκίκας

δηλώνω υπεύθυνα ότι:

1) *Είμαι ο κάτοχος των πνευματικών δικαιωμάτων της πρωτότυπης αυτής εργασίας και από όσο γνωρίζω η εργασία μου δε συκοφαντεί πρόσωπα, ούτε προσβάλει τα πνευματικά δικαιώματα τρίτων*

2) *Η παρούσα εργασία με τίτλο «Πρόγνωση μετεωρολογικών παραμέτρων με τη χρήση νευρωνικών δικτύων και μαθηματικών μοντέλων» αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές που έχω χρησιμοποιήσει έχουν δηλωθεί κατάλληλα στις βιβλιογραφικές αναφορές και παραπομπές. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο 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.*

3) *Αποδέχομαι ότι η ΒΚΠ μπορεί, χωρίς να αλλάξει το περιεχόμενο της εργασίας μου, να τη διαθέσει σε ηλεκτρονική μορφή μέσα από τη ψηφιακή Βιβλιοθήκη της, να την αντιγράψει σε οποιοδήποτε μέσο ή/και σε οποιοδήποτε μορφότυπο καθώς και να κρατά περισσότερα από ένα αντίγραφα για λόγους συντήρησης και ασφάλειας.*

Ευχαριστίες

Χωρίς την ενθάρρυνση και την αμέριστη βοήθεια του καθηγητή μου κ. Ηρακλή Βαρλάμη, αυτή η εργασία δεν θα ολοκληρωνόταν. Τον ευχαριστώ θερμά για την υπομονή, τις συμβουλές αλλά και για την εμπιστοσύνη του καθ' όλη τη διαδικασία της συγγραφής της.

Η εμπειρία μου στο Πρόγραμμα Μεταπτυχιακών Σπουδών του Τμήματος Πληροφορικής και Τηλεματικής υπήρξε πολύτιμη για την εκπόνηση του συγκεκριμένου έργου και ιδιαίτερα σημαντική για την μετέπειτα πορεία μου. Ευχαριστώ θερμά όλους τους καθηγητές και το προσωπικό για τις γνώσεις και την υποστήριξή τους κατά τη διάρκεια των σπουδών μου.

Η ηθική και υλική συμπαράσταση της οικογένειας μου και η πίστη τους σε κάθε προσπάθεια και βήμα της ζωής μου, μού δίνει τη δύναμη να συνεχίσω. Σας ευχαριστώ από καρδιάς.

Ευχαριστώ την εταιρία Enerdia και τους συναδέλφους μου για την ενθάρρυνση, το ενδιαφέρον και την στήριξη αυτής της προσπάθειας. Ιδιαίτερες ευχαριστίες στο Θοδωρή Καραπαναγιώτη για τις πολύτιμες συμβουλές του στην συγγραφή του παρόντος έργου.

Εάν η Γνώση δεν ήταν ελεύθερα διαθέσιμη, αυτό το έργο θα ήταν αδύνατον να εκπονηθεί. Η εργασία αυτή αφιερώνεται σε όσους πίστεψαν και αγωνίστηκαν γι αυτό το Ιδανικό.

Περιεχόμενα

Συντομογραφίες

ACC	**Anomaly Correlation Coefficient**
Adam	Adaptive Moment Estimation
AI	Artificial Intelligence
ANN	Artificial Neural Network
BM	Baseline Model
BPTT	Back Propagation Through Time
CAP	Credit Assignment Path
CNN	Convolutional Neural Network
CRPS	Continuous Rank Probability Score
DL	Deep Learning
DLWP	Deep Learning Weather Prediction
ECMWF	European Centers for Environmental Prediction
ELU	Exponential Linear Unit
ENIAC	Electronic Numerical Integrator and Computer
FCNN	Fully Connected Neural Network
GCM	Global Circulation Model
GD	Gradient Descent
GFS	Global Forecasting System
GSM	Global Spectral Model
GTS	Global Telecommunications System
IFS	Integrated Forecast System
KDD	Knowledge Discovery from Data (Data Mining)
LSTM	Long Short-Term Memory
LSTM	Long Short-Term Memory
LTU	Linear Threshold Unit
MAE	Mean Absolute Error
ML	Machine Learning
MLP	Multi-Layer Perceptrons
MSE	Mean Squared Error
NAND	Not-And
NOR	Not-Or
PCNN	Partially Connected Neural Network
PE	Primitive Equations
ReLU	Rectified Linear Unit
RMSE	Root Mean Squared Error
RMSProp	Root Mean Square Propagation
RNN	Recurrent Neural Network
SGD	Stochastic Gradient Descent
SGDR	Stochastic Gradient Descent with Restarts
UKMO	United Kingdom Met. Office
WRF	Weather Research Forecasting
XOR	Exclusive Or

Περίληψη

Με την εφεύρεση των μετεωρολογικών οργάνων (βαρόμετρο, θερμόμετρο κ.α) μεταξύ 15ου και 17ου αιώνα, ξεκίνησαν οι πρώτες συστηματικές καταγραφές των μετεωρολογικών συνθηκών. Τα θεμέλια της αριθμητικής πρόγνωσης του καιρού μπήκαν το 1922, όταν οι Νορβηγοί **Vilhelm και Jacob Bjerknes**, ο **Halvor Solberg** και ο Σουηδός μετεωρολόγος **Tor Bergeron** ανέπτυξαν τη "*θεωρία των μετώπων*" στηριζόμενοι σε παρατηρήσεις ενός πυκνού δικτύου μετεωρολογικών σταθμών. Την ίδια περίπου περίοδο ο **Lewis Fry Richardson** επιχείρησε τον υπολογισμό της εξάωρης μεταβολής της ατμοσφαιρικής πίεσης σε μία περιοχή, διαδικασία που χρειάστηκε έξι εβδομάδες για να ολοκληρωθεί. Παρ'όλο που το αποτέλεσμα δεν συμβάδιζε με την πραγματικότητα, μέσω της εργασίας του υπέδειξε τον τρόπο με τον οποίο θα μπορούσε να πραγματοποιηθεί στο μέλλον, οραματιζόμενος ουσιαστικά τη χρήση των ηλεκτρονικών υπολογιστών σχεδόν τρεις δεκαετίες αργότερα. Έκτοτε η έρευνα στην αριθμητική πρόγνωση του καιρού εξελίχθηκε ραγδαία, απολύτως συνυφασμένη με την εξέλιξη της επιστήμης των υπολογιστών. Κατ'αναλογία με την αριθμητική πρόγνωση του καιρού φαίνεται ότι και οι απαρχές της Μηχανικής Εκμάθησης και της Τεχνητής Νοημοσύνης εντοπίζονται σχετικά νωρίς, αλλά για τον ίδιο λόγο έπρεπε επίσης να περιμένουν αρκετές δεκαετίες. Το 1943 οι **McCulloch και Pitts** πρότειναν ένα μαθηματικό μοντέλο προτασιακής λογικής (propositional logic) για την περιγραφή της νευρωνικής λειτουργίας των οργανισμών βάσει συγκεκριμένων παραδοχών, εμπνέοντας το έργο του ψυχολόγου **Frank Rosenblatt,** ο οποίος το 1958 περιέγραψε την πρώτη απλή μορφή μονάδα υπολογισμών (LTU), το "perceptron", ως δομική μονάδα ενός τεχνητού νευρωνικού δικτύου. Από τη δεκαετία του '80 και μετά οι τεχνικές της Μηχανικής Εκμάθησης έχουν εξελιχθεί με την εμφάνιση πολλών διαφοροποιήσεων στη δομή και τη διασύνδεση των νευρώνων ενώ η "Επιστήμη των Δεδομένων", αποτελεί αυτοτελές και γρήγορα εξελισσόμενο πεδίο έρευνας. Στην παρούσα εργασία επιχειρείται η πρόγνωση μιας βασικής μετεωρολογικής παραμέτρου, του γεωδυναμικού ύψους στα 500hpa, με τη χρήση τεχνικών μηχανικής εκμάθησης και τη δημιουργία ενός νευρωνικού δικτύου. Αν και η πρόγνωση των μελλοντικών καταστάσεων της ατμόσφαιρας αποτελεί αντικείμενο της Φυσικής της Ατμόσφαιρας, πρόσφατες εργασίες έχουν δείξει ότι το ίδιο ζητούμενο μπορεί επίσης να προσεγγιστεί από την οπτική των δεδομένων με πολλά υποσχόμενα αποτελέσματα συμβάλλοντας επίσης στην πληρέστερη κατανόηση των ατμοσφαιρικών διεργασιών.

Λέξεις - κλειδιά: *Μηχανική Εκμάθηση, Νευρωνικά Δίκτυα, Αριθμητική Πρόγνωση Καιρού*

Abstract

The invention of meteorological instruments such as the barometer and thermometer between the 15th and 17th century, marked the beginning of the systematic collection of weather data. The cornerstone of the modern weather prediction methods was laid by **Vilhelm και Jacob Bjerknes**, ο **Halvor Solberg** and **Tor Bergeron** in 1922 with the development of the "*polar front theory*", based on the observations of a dense weather stations network. Around the same period, **Lewis Fry Richardson** tried to predict by hand the 6-hour pressure tendency over a specific area, using purely arithmetic methods, but with no success. Although this procedure lasted about 6 weeks, he suggested the way of how it could be implemented fast and efficiently, foreseeing the computer era three decades later. Since the invention of computers, the evolution of Numerical Weather Prediction is in relevance with the evolution of computer science. Just like the Numerical Weather Prediction, it seems that the first papers concerning Artificial Intelligence, were also ahead of their time. In 1943, **McCulloch και Pitts** suggested a propositional logic model concerning the neuron cells' functionality based on certain assumptions. This inspired the work of the psychologist **Frank Rosenblatt** who, in 1958, described the first simple Linear Threshold Unit (LTU), called the "perceptron", as a building block of an Artificial Neural Network (ANN). Since the 80's, Machine Learning (ML) techniques have evolved with many variants concerning the structure of Neural Networks, while "Data Science" has become a standalone and rapidly growing field of research. In this study, the forecast of a basic atmospheric parameter, the Geopotential Height at 500hpa is attempted, using ML techniques and the creation of an ANN. Although forecasting the future states of the atmosphere is a subject of Atmospheric Physics, recent research has shown that the same goal could be achieved from the Data Science perspective with very promising results, contributing also to the better understanding of atmospheric processes as well.

Keywords: *Machine Learning, Artificial Neural Networks, Weather Forecasting, Numerical Weather Prediction*

Κεφάλαιο 1ο: Εισαγωγή

Είναι γνωστό πως οι καιρικές και κλιματολογικές συνθήκες, οι οποίες συνδέονται άμεσα με τη γενικότερη διαμόρφωση του φυσικού περιβάλλοντος, έπαιξαν τον σπουδαιότερο ρόλο στην ανάπτυξη και ευημερία των κοινωνιών από την απαρχή του ανθρώπινου είδους. Υπάρχουν πολλές αναφορές τόσο στη μυθολογία όσο και στην καταγεγραμμένη ιστορία σχετικά με αυτό και ο καιρός και οι μεταβολές του φαίνεται πως αποτελούσαν ένα κεντρικό ζήτημα που απασχολούσε όλους τους πολιτισμούς από την αρχή της ανθρώπινης ιστορίας έως και σήμερα. Χαρακτηριστική είναι η περίπτωση της Αρχαίας Ελλάδας, όπου όλα τα καιρικά φαινόμενα (ακόμη και οι διευθύνσεις των ανέμων) προσωποποιούνται σε θεότητες και περιγράφονται με μεγάλη λεπτομέρεια. Ωστόσο, με τις απαρχές της επιστημονικής σκέψης προέκυψαν και τα πρώτα συγγράμματα, τα οποία επιχείρησαν να ερμηνεύσουν αυτά τα φαινόμενα. Για παράδειγμα, τα "Μετεωρολογικά" του Αριστοτέλη καθώς και το βιβλίο των "Σημείων" του μαθητή του Θεοφράστου αποτελούν μία από τις πρώτες προσπάθειες καταγραφής της ανθρώπινης γνώσης εκείνης της εποχής σχετικά με τις ατμοσφαιρικές μεταβολές και αποτέλεσαν για τουλάχιστον 2000 χρόνια τη βασική πηγή πληροφόρησης, μέχρι να καταστεί εφικτή η ποσοτικοποίηση των παρατηρήσεων.

1.1: Η ανάλυση της ατμόσφαιρας ως πρόβλημα εφαρμογής των νόμων της Φυσικής

Οι πρώτες γνωστές καταγραφές ατμοσφαιρικών παραμέτρων, όπως η πίεση, η θερμοκρασία και η υγρασία, αρχίζουν μεταξύ 15^{ου} και 17^{ου} αιώνα με την εφεύρεση των αντίστοιχων επιστημονικών οργάνων, δίνοντας και τη δυνατότητα συλλογής τους με περισσότερο οργανωμένο τρόπο. Το πρώτο γνωστό δίκτυο παρατηρήσεων θεωρείται ότι δημιουργήθηκε από τον Ιταλό **Φερδινάνδο ΙΙ των Μεδίκων**, ο οποίος οργάνωσε ένα δίκτυο σε διάφορες πόλεις της βόρειας Ιταλίας, περιλαμβάνοντας σταθμούς στην Αυστρία (Innsbruck), Γερμανία (Osnabruck), Γαλλία (Παρίσι) και Πολωνία (Βαρσοβία) το 1654 (πηγή: https://en.wikipedia.org/wiki/Timeline_of_meteorology). Ωστόσο, η συλλογή των μετεωρολογικών παρατηρήσεων και η αξιοποίησή τους σε σχεδόν πραγματικό χρόνο, έπρεπε να περιμένει την εφεύρεση του τηλέγραφου το 1837. Λίγα χρόνια αργότερα, κατά τη διάρκεια του Κριμαϊκού πολέμου το 1854 και με αφορμή μια σφοδρή κακοκαιρία στην περιοχή του Εύξεινου Πόντου που προκάλεσε μεγάλες ζημιές στον αγγλικό και γαλλικό στόλο, τέθηκε από τον Γάλλο υπουργό των στρατιωτικών το ερώτημα εάν *θα ήταν εφικτό να είχε προβλεφθεί λαμβάνοντας υπόψη τις καταγραφές των μετεωρολογικών σταθμών από άλλα σημεία της Ευρώπης*. Ο διευθυντής του αστεροσκοπείου των Παρισίων La Verrier, συλλέγοντας

παρατηρήσεις από 200 και πλέον σταθμούς, κατέληξε στο συμπέρασμα ότι η συγκεκριμένη κακοκαιρία είχε προηγουμένως διασχίσει ένα μεγάλο τμήμα της Ευρώπης (πηγή: διαδικτυακός ιστότοπος Εθνικής Μετεωρολογικής Υπηρεσίας, http://emy.gr). Αυτή η παρατήρηση οδήγησε σε μια διαδικασία που αποτέλεσε και τη βασική προγνωστική μεθοδολογία για τουλάχιστον έναν αιώνα.

Η πρώτη απόπειρα πρόγνωσης με βάση τις καταγεγραμμένες παρατηρήσεις έγινε από τον **Sir Francis Beaufort**, υδρογράφο και υποναύαρχο του αγγλικού στόλου, και τον αντιναύαρχο **Robert FitzRoy**. Αφορμή για την συστηματικότερη καταγραφή και απεικόνιση των μετεωρολογικών δεδομένων μέσω της δημιουργίας μετεωρολογικών χαρτών στάθηκε ένα ακόμη δυστύχημα λόγω σφοδρής κακοκαιρίας: το ναυάγιο του Royal Charter στα ανοιχτά της Ουαλίας στις 26 Οκτωβρίου 1859 μαζί με άλλα τουλάχιστον 200 σκάφη και πλοιάρια. Αυτή η κακοκαιρία θεωρείται ίσως η εντονότερη του 19ου αιώνα στη συγκεκριμένη περιοχή. Η εργασία του FitzRoy έφερε για πρώτη φορά την ορολογία "πρόγνωση καιρού" (weather forecast) στο βιβλίο του "*The Weather Book*" που δημοσιεύτηκε το 1863. Θεωρείται αρκετά πρωτοποριακό για τα δεδομένα και το επιστημονικό υπόβαθρο της εποχής εκείνης (πηγές: https://en.wikipedia.org/wiki/Weather_forecasting και https://www.bbc.com/news/magazine-32483678). Η κατηγοριοποίηση της έντασης του ανέμου από τον Beaufort αλλά και των τύπων των νεφών από τον Διεθνή Μετεωρολογικό Οργανισμό (πρόδρομος του Παγκόσμιου Μετεωρολογικού Οργανισμού), το 1896, συνέβαλαν στην συστηματικοποίηση των καταγραφών.

Η απαρχή της αριθμητικής πρόγνωσης του καιρού, δηλαδή της πρόγνωσης μέσω της επίλυσης εξισώσεων περιγραφής μελλοντικών καταστάσεων της ατμόσφαιρας με βάση τις αρχικές συνθήκες, οφείλεται στον Lewis Fry Richardson (Richardson, L., & Lynch, P., 2007). Στο πρωτοποριακό σύγγραμμα με τίτλο "Weather Prediction by Numerical Process", το οποίο δημοσιεύθηκε το 1922, γίνεται η πρώτη προσπάθεια υπολογισμού της ατμοσφαιρικής πίεσης μιας συγκεκριμένης περιοχής 6 ώρες μετά, με βάση τα δεδομένα παρατήρησης στις 7 το πρωί της 20ης του Μαΐου 1910. Προκειμένου να παραχθεί το αποτέλεσμα, χρειάστηκαν περίπου 6 εβδομάδες υπολογισμών. Η λύση ήταν μη ρεαλιστική καθώς εκτός των άλλων προβλέφθηκε άνοδος στην πίεση κατά 145hpa (hectopascals). Ωστόσο, όπως εκ των υστέρων παρατηρεί ο Lynch, η αποτυχία δεν οφείλεται σε λανθασμένη μεθοδολογία ή υπολογισμούς, αλλά στο γεγονός ότι δεν εφαρμόστηκε κάποια τεχνική "ομαλοποίησης" (smoothing) των αρχικών

δεδομένων, με αποτέλεσμα την εισαγωγή "θορύβου" που εν τέλει οδήγησε σε πολλαπλασιασμό του σφάλματος.

Κατά την ίδια περίπου περίοδο, οι Νορβηγοί **Vilhelm και Jacob Bjerknes**, ο **Halvor Solberg** και ο Σουηδός μετεωρολόγος **Tor Bergeron** ανέπτυξαν τη "*θεωρία των μετώπων*" στο νεοσύστατο (υπό τον Vilhelm Bjerknes) τμήμα Γεωφυσικής του πανεπιστημίου του Bergen (https://earthobservatory.nasa.gov/features/Bjerknes/bjerknes_3.php, πρόσβαση: Αύγουστος 2022). Σύμφωνα με αυτή, τα καιρικά φαινόμενα λαμβάνουν χώρα σε συγκεκριμένες ζώνες που ονομάζονται "μέτωπα", τα οποία συγχρόνως αποτελούν περιοχές διαχωρισμού αερίων μαζών με διαφορετικά χαρακτηριστικά (π.χ. ψυχρότερων / θερμότερων). Για την ανάπτυξη της θεωρίας σημαντικό ρόλο έπαιξαν οι παρατηρήσεις από ένα πυκνό (για την εποχή εκείνη) δίκτυο σταθμών, που είχε εγκατασταθεί κατά τη διάρκεια του Α' Παγκοσμίου Πολέμου (Bjerknes J., & Solberg H., 1922). Όπως και ο Richardson, έτσι και ο Vilhelm Bjerknes θεωρούσε πως η μελλοντική εξέλιξη της ατμόσφαιρας θα μπορούσε να προβλεφθεί με τη χρήση ήδη γνωστών νόμων της φυσικής, αν και κάτι τέτοιο ήταν μη πρακτικό (Lynch, 2008).

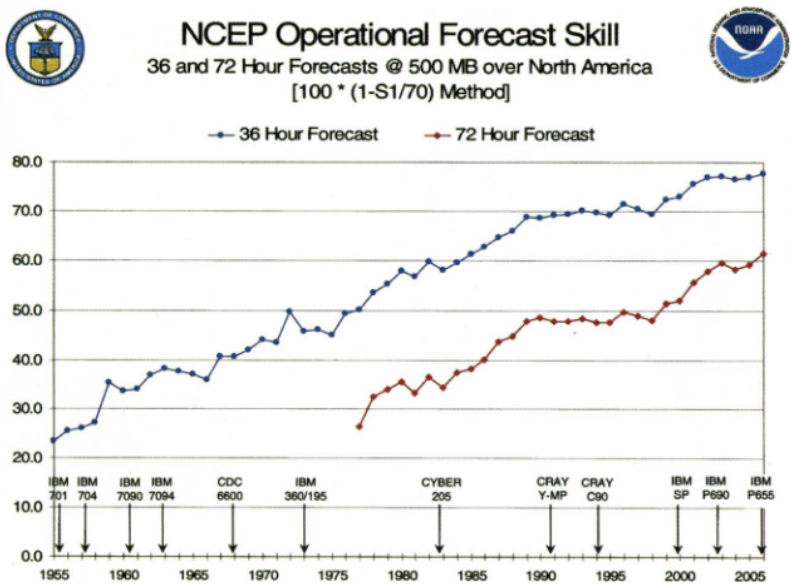
Εικόνα 1.1: *Πρόσφατη προσωπογραφία του Vilhelm Bjerknes στην προβλήτα του Bergen (εμπνευσμένη από φωτογραφίες της εποχής). Καλλιτέχνης: Rolf Groven, © Geophysical Institute Bergen, αναδημοσίευση από Lynch.*

1.2: Η εφαρμογή των νόμων της Φυσικής στους ηλεκτρονικούς υπολογιστές

Παρ'όλο που, την εποχή του μεσοπολέμου η αριθμητική πρόγνωση του καιρού ήταν μια διαδικασία που δεν είχε επιχειρησιακό νόημα, ο Richardson, εκτός από τη μαθηματική επίλυση, υπολόγισε και τον τρόπο που θα μπορούσε να γίνει πρακτικά εφαρμόσιμη. Συγκεκριμένα θεώρησε πως, προκειμένου ο απαιτούμενος χρόνος της επίλυσης να είναι ίσος με τον προγνωστικό ορίζοντα, απαιτούνται περίπου 64.000 άνθρωποι, τους οποίους ονομάζει "υπολογιστές" (computers). Η υδρόγειος χωρίζεται σε κελιά ίσου μεγέθους με βάση το γεωγραφικό πλάτος (π.χ. 2° x 2º), όπου σε κάθε ένα εξ αυτών οι συνθήκες θεωρούνται ίδιες και εντός αυτών εκτελούνται οι υπολογισμοί παράλληλα. Τη διαδικασία συντονίζει μια κεντρική ομάδα.

Εικόνα 1.2: *Το "όραμα του Richardson" για την πραγματοποίηση της αριθμητικής πρόγνωσης του καιρού σχεδιασμένο από τον ιρλανδό καλλιτέχνη Stephen Conlin, υπό την καθοδήγηση του John Byrne, καθηγητή, επικεφαλής του τμήματος της Επιστήμης των Υπολογιστών στο Trinity College του Δουβλίνο. Το ημισφαίριο πάνω στο οποίο πραγματοποιούνται οι υπολογισμοί βρίσκεται στο εσωτερικό ενός πολυόροφου κτιρίου, κάθε επίπεδο του οποίου είναι χωρισμένο σε τμήματα όπου εργάζονται οι "υπολογιστές". Ο συντονισμός της διαδικασίας πραγματοποιείται στο κέντρο του ημισφαιρίου. (Πηγή: eumetsoc)*

Το "όραμα του Richardson" είναι ο πρόδρομος της αριθμητικής πρόγνωσης του καιρού, όπως υλοποιήθηκε για πρώτη φορά λίγες δεκαετίες αργότερα. Πράγματι, κάθε περιοχή εφαρμογής είναι συνήθως χωρισμένη σε ίσα τμήματα με βάση το σύστημα γεωγραφικών συντεταγμένων, κάτι το οποίο ορίζει την οριζόντια ανάλυση της πρόγνωσης με την προσθήκη συγκεκριμένων κατακόρυφων επιπέδων (χωρική ανάλυση). Οι "υπολογιστές" αντιστοιχούν σε επεξεργαστικές διεργασίες οι οποίες εκτελούνται παράλληλα πάνω στο δημιουργηθέν πλέγμα σε συγκεκριμένες χρονικές στιγμές, οι οποίες ορίζουν τη χρονική ανάλυση του βήματος της πρόγνωσης. Η CPU του υπολογιστή αναλαμβάνει τον συντονισμό και εκτέλεση των διεργασιών.



Σχήμα 1.1: *Η διαχρονική εξέλιξη της επιχειρησιακής προγνωστικής ικανότητας του Εθνικού Κέντρου Περιβαλλοντικών Προγνώσεων (NCEP) στις προγνώσεις 36 και 72 ωρών, όπως αποτυπώνονται από τον δείκτη S1 (τροποποιημένο) (Πηγή: Harper, Uccelini, Kalnay, Carey, Morone, 2007).*

Το 1950 πραγματοποιήθηκε η πρώτη αριθμητική πρόγνωση με τη χρήση του **ENIAC** (Electronic Numerical Integrator and Computer) (Lynch, 2008). Εκτιμάται πως η υπολογιστική του ισχύς ήταν τέτοια, ώστε μία ώρα λειτουργίας του αντιστοιχούσε στην εργασία 2400 ανθρώπων (πηγή: The History of Computing Project). Ειδικότερα για τον υπολογισμό του γεωδυναμικού ύψους και του στροβιλισμού στη στάθμη των 500hpa πάνω από τη Β. Αμερική τις επόμενες 24 ώρες σε ένα πλέγμα 15 x 13 σημείων (Charney et al., 1950) χρειάστηκε περίπου τον ίδιο χρόνο υπολογισμού, ενώ ο Von Neumann εκτίμησε ότι η αντίστοιχη ανθρώπινη εργασία θα χρειαζόταν περίπου 5 χρόνια. Το αποτέλεσμα θεωρήθηκε ρεαλιστικό, επιβεβαιώνοντας την προσπάθεια του Richardson. Τελικά η πρώτη επιχειρησιακή πρόγνωση

πραγματοποιήθηκε στη Σουηδία (Staff Members, Institute of Meteorology, University of Stockholm, 1954), ενώ την ίδια χρονιά ακολούθησαν και οι Ηνωμένες Πολιτείες (Kalnay, 2002). Η κατασκευή των ημιαγωγών (transistors), στη σύγχρονη μορφή τους το 1959, οι οποίοι σταδιακά αντικατέστησαν τις λυχνίες των πρώτων υπολογιστών, συνετέλεσε τα μέγιστα στη βελτίωση της απόδοσης, στη σμίκρυνση του μεγέθους και της κατανάλωσης των ηλεκτρονικών υπολογιστών. Τις επόμενες δεκαετίες βελτιώθηκε με ανάλογο τρόπο και η αριθμητική πρόγνωση της ατμόσφαιρας, τόσο ως προς την ακρίβειά της όσο και ως προς τους χρόνους υπολογισμού (σχήμα 1.1).

Όλα τα στάδια της μελέτης των ατμοσφαιρικών διεργασιών, από την συλλογή των παρατηρήσεων μέχρι τη διατύπωση της θεωρίας και την επαλήθευσή της, δημιούργησαν τις θεωρητικές βάσεις για τη διαδικασία της ατμοσφαιρικής προσομοίωσης. Η πραγματοποίησή της όμως κατέστη δυνατή μόνο λόγω της εξέλιξης των ηλεκτρονικών υπολογιστών. Σήμερα η αριθμητική πρόγνωση εκτελείται σε ένα πυκνό πλέγμα που καλύπτει όλη την υφήλιο, συνήθως σε τέσσερις ημερήσιους προγνωστικούς κύκλους (00, 06, 12, 18 UTC) φτάνοντας σε χρονικό ορίζοντα 10 ή 15 ημερών. Μερικές τυπικές αναλύσεις είναι οι 0.25° x 0.25° από το Εθνικό Κέντρο Περιβαλλοντικών Προγνώσεων των ΗΠΑ (NCEP), ενώ οι αντίστοιχες από το Ευρωπαϊκό Κέντρο Μεσοπρόθεσμων Προγνώσεων (ECMWF) πραγματοποιούνται σε ένα πλέγμα περίπου 9 x 9 km (https://www.ecmwf.int/en/forecasts/documentation-and-support). Το αποτέλεσμα είναι η παραγωγή ενός πολύ μεγάλου όγκου δεδομένων. Ενδεικτικά, λαμβάνοντας υπόψιν τη χωρική ανάλυση του προγνωστικού μοντέλου του NCEP για τα ανασυντεθειμένα αρχικά δεδομένα (reanalysis πεδία), σε ανάλυση 1° x 1º, τα σημεία που καλύπτονται στον πλανήτη είναι 65.160 (σχήμα 1.2) με 610 προγνωστικές παραμέτρους (features) για κάθε σημείο[1], δηλαδή 48.218.400 διαφορετικές τιμές, μόνο για μία χρονική στιγμή. Εάν υπολογίσουμε με βάση την προγνωστική ανάλυση 0.25° x 0.25º, ο παραπάνω αριθμός γίνεται περίπου 16 φορές μεγαλύτερος.

[1] *Σχετικές πληροφορίες σε αυτόν τον σύνδεσμο:* *https://www.nco.ncep.noaa.gov/pmb/products/gfs/gfs.t00z.pgrb2.1p00.anl.shtml* ,
και γενικότερα για την πρόσβαση στα προγνωστικά δεδομένα του παγκόσμιου προγνωστικού συστήματος GFS: *https://www.nco.ncep.noaa.gov/pmb/products/gfs/* .
Πρόσβαση στα reanalysis (FNL) πεδία: *https://rda.ucar.edu/*

	Date	Latitude	Longitude	Value
0	20190623	90.0	0.0	5424.714375
1	20190623	90.0	1.0	5424.714375
2	20190623	90.0	2.0	5424.714375
3	20190623	90.0	3.0	5424.714375
4	20190623	90.0	4.0	5424.714375
...	...	...	...	...
65155	20190623	-90.0	355.0	4839.354375
65156	20190623	-90.0	356.0	4839.354375
65157	20190623	-90.0	357.0	4839.354375
65158	20190623	-90.0	358.0	4839.354375
65159	20190623	-90.0	359.0	4839.354375

[65160 rows x 4 columns]

Σχήμα 1.2: *Ενδεικτική εκτύπωση του dataframe που δημιουργείται μετά την ανάγνωση ενός και μόνο reanalysis αρχείου, για μία ημερομηνία (στήλη 2) και μία προγνωστική παράμετρο (Geopotential Height στα 500hpa - στήλη 5), για κάθε συνδυασμό γεωγραφικου πλάτους και μήκους (στήλη 3 & 4). Καθώς η πρώτη στήλη αναφέρει τον αριθμό εγγραφής, υπολογίζονται συνολικά 65160.*

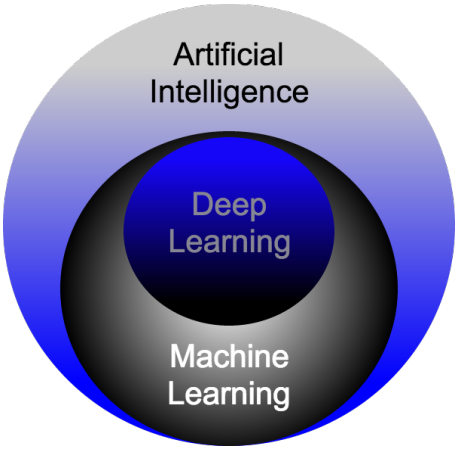
1.3 Η μελέτη της ατμόσφαιρας, ως ζητούμενο διαχείρισης και εξαγωγής γνώσης από δεδομένα μεγάλου όγκου

Η αριθμητική πρόγνωση και ανάλυση της ατμόσφαιρας κατέστη δυνατή λόγω της παράλληλης έρευνας και εξέλιξης της επιστήμης των υπολογιστών σε θέματα μεταξύ άλλων, όπως η αύξηση της υπολογιστικής ισχύος, της μνήμης και του αποθηκευτικού χώρου. Είναι όμως προφανές πως ταυτόχρονα είναι και ένα ζητούμενο συλλογής, διαχείρισης και αποθήκευσης μεγάλου όγκου δεδομένων, τα οποία αναφέρονται και ως “Big Data”. Εφόσον, όπως διαπιστώθηκε, ένα πολύ σημαντικό στάδιο στην μοντελοποίηση της ατμόσφαιρας είναι ο όγκος και η ακρίβεια της αρχικής πληροφορίας, μια διαφορετική προσέγγιση θα μπορούσε να προκύψει, εάν το ζητούμενο της πρόγνωσης των μελλοντικών καταστάσεων της ατμόσφαιρας προσεγγιστεί από την οπτική των δεδομένων.

Η διαχείριση ενός μεγάλου όγκου δεδομένων και η προσπάθεια εύρεσης συσχετίσεων με σκοπό την εξαγωγή συμπερασμάτων και χρήσιμης πληροφορίας (**Data Mining**) αποτελεί ένα ευρύ και ανεξάρτητο πεδίο έρευνας με εφαρμογές σε πολλές επιστήμες. Έτσι λοιπόν, **εξόρυξη δεδομένων** (Data Mining ή Knowledge Discovery from Data ή απλώς KDD), ονομάζεται η διαδικασία της ανακάλυψης μοτίβων και συσχετίσεων προερχόμενες από ένα μεγάλο όγκο δεδομένων (Han, Kamber, Pei, Jaiwei, Micheline, Jian, 2011). Για την εξεύρεση χρήσιμης πληροφορίας χρησιμοποιούνται εργαλεία της στατιστικής, της τεχνητής νοημοσύνης (όπως για παράδειγμα τα νευρωνικά δίκτυα και η μηχανική εκμάθηση), καθώς και της διαχείρισης βάσεων δεδομένων. Εντός αυτών τα δεδομένα οργανώνονται κατά τέτοιο τρόπο, ώστε να μπορούν στη συνέχεια να αντληθούν, να καθαριστούν και να χρησιμοποιηθούν στην μετέπειτα ανάλυση.

Η **Μηχανική Εκμάθηση** (Machine Learning ή ML), ορίζεται ως "η επιστήμη του προγραμματισμού των ηλεκτρονικών υπολογιστών κατά τέτοιο τρόπο, ώστε να είναι σε θέση να μάθουν από τα δεδομένα" (Aurélien Géron, 2017). Η διαδικασία αυτή έχει ως στόχο τη δημιουργία ενός μοντέλου και μπορεί να χωριστεί σε 3 μεγάλες προσεγγίσεις: Εκμάθηση με επίβλεψη (**supervised**), χωρίς επίβλεψη (**unsupervised**) και μέσω ενίσχυσης/επιβράβευσης (**reinforcement learning**). Η διαφορά μεταξύ της supervised και της unsupervised ML είναι ότι στην μεν πρώτη τα δεδομένα που χρησιμοποιούνται ως είσοδοι στο μοντέλο έχουν προηγουμένως κατηγοριοποιηθεί (labeled data), ενώ στη δεύτερη η κατηγοριοποίηση είναι ένα από τα ζητούμενα. Σχετικά με την προσέγγιση του reinforcement learning, το λογισμικό αλληλεπιδρά με το περιβάλλον του προσπαθώντας να βελτιώσει συγκεκριμένες μετρικές (scores), που λειτουργούν ως "επιβράβευση".

Ενώ το βασικό αντικείμενο της μηχανικής εκμάθησης είναι η μελέτη των δεδομένων με στόχο την εξαγωγή συμπερασμάτων, η λήψη αποφάσεων που βασίζεται σε αυτά αποτελεί αντικείμενο της "**τεχνητής νοημοσύνης"** (Artificial Intelligence - AI). Είναι ωστόσο αρκετά δύσκολο να διατυπωθεί ένας σαφής ορισμός για το τί ακριβώς είναι η AI, διότι ο ίδιος ο ορισμός του "τί είναι η νοημοσύνη" είναι επίσης δύσκολος (IBM Cloud Education, 2020). Όπως αναφέρει ο McCarthy (2004), *"τεχνητή νοημοσύνη είναι η επιστήμη και η μηχανική της δημιουργίας έξυπνων μηχανών, ειδικότερα των υπολογιστικών προγραμμάτων. Σχετίζεται με το παρόμοιο έργο της χρήσης των υπολογιστών για την κατανόηση της ανθρώπινης νοημοσύνης, όμως η AI δεν χρειάζεται να οριοθετηθεί μέσω μεθόδων που να είναι βιολογικά παρατηρήσιμες"*. Σε πρακτικό επίπεδο, μπορεί να γίνει περισσότερο κατανοητή η έννοια της Τεχνητής Νοημοσύνης μέσω της σχέσης της με τη Μηχανική και τη Βαθιά Εκμάθηση. Ενώ ένα πρόγραμμα / μοντέλο, το οποίο είναι σε θέση να αναγνωρίζει έναν ποδηλάτη μεταξύ άλλων αντικειμένων στο δρόμο, αποτελεί αντικείμενο της ML, ένα ενσωματωμένο σύστημα σε ένα αυτοκίνητο, το οποίο σε πραγματικό χρόνο θα λαμβάνει αποφάσεις αποτρέποντας πιθανές συγκρούσεις με ποδηλάτες, αποτελεί έργο της AI. Εκτός των κλασικών ML μεθόδων (K-nearest neighbor, support vector machines, decision tree, random forest, naive Bayes, linear regression, association rules, k-means clustering κ.ά.), η περίπτωση χρήσης **"Τεχνητού Νευρωνικού Δικτύου"** (Artificial Neural Network - ANN), για την αναγνώριση του ποδηλάτη, αποτελεί αντικείμενο του πεδίου της **"Βαθιάς Εκμάθησης"** (Deep Learning - DL) (Sarker I.H., 2021). Η σχέσεις μεταξύ DL, ML και AI, απεικονίζονται στο σχήμα 1.3.



Σχήμα 1.3: *Σχηματική απεικόνιση των σχέσεων μεταξύ Βαθιάς Μηχανικής Εκμάθησης, Μηχανικής Εκμάθησης και Τεχνητής Νοημοσύνης.*

Οι αναλογίες στη διαδικασία της εκμάθησης μεταξύ των ηλεκτρονικών υπολογιστών και του βιολογικού εγκεφάλου είναι πολλές και φαίνεται πως οι υπολογιστές τείνουν να μιμηθούν τη βιολογία. Η **ταξινόμηση** των εξωτερικών ερεθισμάτων αποτελεί ένα βασικό στοιχείο της διαδικασίας. Ωστόσο, μια ακόμη σημαντική διεργασία, που λαμβάνει χώρα παράλληλα και σε συνδυασμό με την πρώτη, είναι η **αναγνώριση μοτίβων** (pattern recognition) που εμφανίζονται στα δεδομένα. Λαμβάνοντας υπόψη το μοτίβο που τυχόν εμφανίζεται ως προς το χρόνο και τη συχνότητα εκδήλωσης, τον τρόπο ή τη συσχέτισή τους με άλλα, είναι δυνατόν τα δεδομένα (ή ερεθίσματα) να ταξινομηθούν για την περαιτέρω επεξεργασία τους ή αντιστρόφως, η αναγνώριση του μοτίβου που εμφανίζεται, δύναται να πραγματοποιηθεί, αφού πρώτα τα δεδομένα έχουν ταξινομηθεί.

Αν η ατμόσφαιρα μελετηθεί από την σκοπιά των δεδομένων, τότε, σε αυτήν την περίπτωση, η εκπαίδευση ενός AI μοντέλου μπορεί να πραγματοποιηθεί πάνω σε παρελθοντικά δεδομένα με τη χρήση ML ή / και DL τεχνικών για την ανακάλυψη συσχετισμών μεταξύ διαφόρων παραμέτρων. Τότε θα είναι εφικτή και η προσομοίωση των επόμενων καταστάσεων για μία ή περισσότερες παραμέτρους σε μία η περισσότερες περιοχές ταυτόχρονα.

Οι αρχικές εργασίες προς αυτήν την κατεύθυνση εντοπίζονται κατά τη διάρκεια της δεκαετίας του ‘90 και είχαν ως στόχο τη σημειακή βελτίωση διαφόρων προσομοιώσεων. Ενδεικτικά, η χρήση νευρωνικού δικτύου με δεδομένα εισαγωγής τον άνεμο, τη θερμοκρασία, την υγρασία, το ηλεκτρικό πεδίο και τους δείκτες ατμοσφαιρικής αστάθειας έδειξε σαφώς

βελτιωμένα αποτελέσματα σε σχέση με προηγούμενες μεθόδους στην πρόγνωση κεραυνικής δραστηριότητας στην περιοχή του Kennedy Space Center στη νότια Florida (Frankel et al. 1995). Στην εργασία των Marzban & Stumpf (1996), ένα αντίστοιχο νευρωνικό δίκτυο το οποίο εκπαιδεύτηκε σε δεδομένα 23 μεταβλητών προερχόμενα από Doppler Radars για την ανίχνευση ανεμοστρόβιλων, παρουσίασε σημαντική βελτίωση του δείκτη Critical Success Index (CSI). Οι Kuligowski & Barros (1998) έδειξαν πως η εκπαίδευση ενός νευρωνικού δικτύου οπισθοδιάδοσης (backpropagation) 25 παραμέτρων εισαγωγής και 11 κρυφών επιπέδων μπορούσε να βελτιώσει το δείκτη σφάλματος Root Mean Square Error (RMSE) στην πρόγνωση του υετού τεσσάρων θέσεων στις Ηνωμένες Πολιτείες έως και 10% σε περιπτώσεις επεισοδίων που ξεπερνούν τα 25mm, ενώ η εφαρμογή στατιστικών μεθόδων διόρθωσης των αποτελεσμάτων βελτίωσε το δείκτη κατά 5%. Η εξέλιξη της αρχιτεκτονικής των νευρωνικών δικτύων από το τέλος της δεκαετίας του ‘90 και ειδικότερα η ανάπτυξη Recurrent Neural Network (RNN) αρχιτεκτονικών και η Long Short-Term Memory (LSTM) παραλλαγή τους (Hochreiter & Schmidhuber, 1997), όπως επίσης και η Convolutional Neural Network (CNN) (Krizhevsky et al., 2012) έφεραν νέες δυνατότητες και προοπτικές χρήσης τους στην πρόγνωση του καιρού λίγα χρόνια αργότερα. Κατά κανόνα, τα LSTM δίκτυα χρησιμοποιούνται για την πρόγνωση μιας χρονοσειράς τιμών σε σημειακό επίπεδο, όπως ο υετός, έχοντας ως δεδομένα εισόδου άλλες μετρήσεις (θερμοκρασία, άνεμος, ατμοσφαιρική πίεση, σχετική υγρασία κ.ά.), λαμβάνοντας υπόψιν τους συσχετισμούς τους διαχρονικά λόγω της ικανότητας “μνήμης” που διαθέτουν (Hoang et al, 2020). Από την άλλη πλευρά, εάν το ζητούμενο είναι η πρόγνωση κάποιας παραμέτρου σε πολλαπλά σημεία, τα οποία ορίζουν μια περιοχή, η προσέγγιση μπορεί να γίνει μέσω της χρήσης ενός CNN μοντέλου, το οποίο χρησιμοποιείται περισσότερο για την κατηγοριοποίηση εικόνων μέσω της αναγνώρισης μοτίβων σε αυτές. Κατ’αυτόν τον τρόπο τα διάφορα μετεωρολογικά πεδία αντιμετωπίζονται ως δεδομένα εικόνας, με στόχο την πρόβλεψη της χωρικής εξέλιξής τους.

Η διαφορετική “data-driven” προσέγγιση στην αριθμητική πρόγνωση του καιρού άρχισε πολύ πρόσφατα να προκαλεί το δημόσιο ενδιαφέρον, σε βαθμό όπου πλέον τίθεται το ερώτημα εάν έχει έρθει η στιγμή να αντικαταστήσει τις κλασικές μεθόδους της αριθμητικής πρόγνωσης του καιρού μέσω της επίλυσης των διαφορετικών εξισώσεων που περιγράφουν τις βασικές κινήσεις των αερίων στην ατμόσφαιρα.

Πιο συγκεκριμένα, γίνεται αναφορά[2] στην εργασία των Weyn et al. (2020), οι οποίοι έδειξαν πως μόνο με τη χρήση ενός CNN (Convolution Neural Network), το οποίο είχε εκπαιδευτεί σε δεδομένα και μοτίβα των προηγούμενων 40 ετών και έχοντας ως δεδομένο εισόδου την προσπίπτουσα ηλιακή ακτινοβολία, δύναται να πραγματοποιηθεί η προσομοίωση της ατμόσφαιρας σε βασικές παραμέτρους (γεωδυναμικό ύψος των 500hpa, θερμοκρασία κ.ά.) για διάφορα χρονικά διαστήματα για έως και ένα έτος μετά. Από την αξιολόγηση του **"καιρικού μοντέλου βαθιάς μηχανικής εκμάθησης"** (Deep Learning Weather Prediction - DLWP Model) ως προς τους δείκτες Root Mean Square Error (RMSE) και Anomaly Correlation Coefficient (ACC), προέκυψε πως, η συγκεκριμένη προσέγγιση μπορεί να επιτύχει καλύτερα αποτελέσματα σε σχέση με ένα μοντέλο παγκόσμιας κυκλοφορίας (Global Circulation Model - GCM) χαμηλής ανάλυσης, όχι όμως εάν συγκριθεί με τις αντίστοιχες τιμές αξιολόγησης ενός μοντέλου υψηλής ανάλυσης. Από την άλλη πλευρά όμως, ο υπολογιστικός χρόνος της DLWP προσομοίωσης είναι κατά ~7.000 φορές μικρότερος σε σχέση με ένα παρόμοιας οριζόντιας ανάλυσης GC μοντέλο, γεγονός που ανοίγει το δρόμο στη δυνατότητα προγνώσεων πολλαπλού δείγματος (ensembles), όπως ακριβώς συμβαίνει και με τα κλασικά GCM's αλλά με πολύ μικρότερο υπολογιστικό κόστος.

Ωστόσο, έχουν προηγηθεί κι άλλες εργασίες προς την ίδια κατεύθυνση, καθώς AI τεχνικές έχουν χρησιμοποιηθεί και για την επεξεργασία και βελτίωση των αποτελεσμάτων των αριθμητικών προσομοιώσεων (post-processing). Για παράδειγμα, οι Chapman et al. (2019) εκμεταλλεύτηκαν τη δυνατότητα των CNN νευρωνικών δικτύων ως προς την ανάλυση πολύπλοκων σχέσεων σε μεγάλο όγκο δεδομένων με στόχο τη βελτίωση των προσομοιώσεων στο πεδίο της μεταφοράς των υδρατμών του παγκόσμιου προγνωστικού συστήματος GFS που αναπτύσσει το Εθνικό Κέντρο Περιβαλλοντικών Προγνώσεων των Η.Π.Α. Λαμβάνοντας υπόψιν την απόκλιση του GFS σε σχέση με τα πραγματικά δεδομένα της ροής της υγρασίας,

---

[2] *Ενδεικτικά:*

University of Washington News:

https://www.washington.edu/news/2020/12/15/a-i-model-shows-promise-to-generate-faster-more-accurate-weather-forecasts/

AGU Science News: https://eos.org/research-spotlights/the-ai-forecaster-machine-learning-takes-on-weather-prediction

NVIDIA: https://developer.nvidia.com/blog/global-ai-weather-forecaster-makes-predictions-in-seconds/

όπως συλλέγονται από δορυφορικές παρατηρήσεις σε ~5.000 δείγματα, το νευρωνικό μοντέλο ήταν σε θέση να βελτιώσει τον δείκτη RMSE κατά 9-17% στο σύνολο των προγνωστικών ωρών (έως 168). Μία ακόμη post-processing χρήση της τεχνητής νοημοσύνης στην αριθμητική πρόγνωση του καιρού είναι η εργασία των Rasp & Lerch (2018), στην οποία προτείνεται μια νέα τεχνική βελτίωσης των προγνωστικών αποτελεσμάτων που βασίζονται στην ανάλυση των ensembles με τη χρήση νευρωνικών δικτύων. Τα μοντέλα εκπαιδεύτηκαν πάνω σε μετρήσεις ενός μεγάλου πλήθους μετεωρολογικών σταθμών για χρονικό διάστημα 8 ετών, μαθαίνοντας συσχετίσεις με διάφορες προγνωστικές παραμέτρους τόσο συνολικά όσο και κατά περίπτωση. Η συγκεκριμένη τεχνική έδειξε βελτίωση του δείκτη Continuous Ranked Probability Score (**CRPS**) στην παράμετρο της θερμοκρασίας στα 2μ. σε σχέση με τις προηγούμενες στατιστικές μεθόδους υπολογισμού της βέλτιστης πρόγνωσης μεταξύ όλων των μελών των ensembles. Στα συμπεράσματα της έρευνας αναφέρεται πως, εκτός από τις πολλές δυνητικές επιχειρησιακές εφαρμογές που προκύπτουν από αυτήν την τεχνική, η συγκεκριμένη ανάλυση μπορεί να είναι εξαιρετικά χρήσιμη για την εξόρυξη γνώσης σε σχέση με παράγοντες που ενδεχομένως επηρεάζουν την προγνωστική απόκλιση. Αυτή η γνώση μπορεί να αποδειχθεί εξαιρετικά χρήσιμη για την βελτίωση των αριθμητικών προσομοιώσεων, ενώ ταυτόχρονα *"αμφισβητεί την κοινή πεποίθηση που χαρακτηρίζει τα νευρωνικά δίκτυα ως μαύρα κουτιά"*.

1.4: Σκοπός της εργασίας

Η συγκεκριμένη διπλωματική εργασία αποσκοπεί στη δημιουργία ενός μοντέλου μηχανικής εκμάθησης, το οποίο έχοντας εκπαιδευτεί σε παρελθοντικές χρονοσειρές δεδομένων καιρού όπως δίδονται από τα αρχικά (analysis) πεδία του παγκόσμιου προγνωστικού συστήματος (GFS), θα είναι σε θέση να προβλέπει τις επόμενες τιμές σε μια περιοχή - στόχο (ορίστηκε η Αθήνα). Εξετάζονται πολλές εναλλακτικές υλοποιήσεις στην αρχιτεκτονική του μοντέλου και πραγματοποιείται η αξιολόγηση των αποτελεσμάτων βάση κατάλληλων στατιστικών συναρτήσεων.

Ακολουθεί η διάρθρωση της εργασίας: Στο 2ο κεφάλαιο γίνεται αναφορά στο θεωρητικό υπόβαθρο της αριθμητικής πρόγνωσης του καιρού και στη συνέχεια παρουσιάζονται οι τεχνικές πρόγνωσης μελλοντικών καταστάσεων με τη χρήση LSTM, RNN και CNN δικτύων. Βάσει των παραπάνω, στο 3° κεφάλαιο αναπτύσσεται η μεθοδολογία της συγκεκριμένης εργασίας, τα εργαλεία που χρησιμοποιήθηκαν αλλά και η αρχιτεκτονική του δικτύου, το οποίο έχοντας ως δεδομένα εισόδου παρατηρηθείσες (analysis) παραμέτρους στη

στάθμη των 500hpa, θα μπορεί να προβλέπει την μελλοντική ατμοσφαιρική κατάσταση. Τα αποτελέσματα των προσομοιώσεων και των υπολογισμών αναλύονται στο 4ο κεφάλαιο. Τέλος, στο 5ο κεφάλαιο αναγράφονται η σύνοψη και τα συμπεράσματα.

Κεφάλαιο 2ο: Η αριθμητική πρόγνωση του καιρού και η μελέτη των φαινομένων μέσω της χρήσης των νευρωνικών δικτύων

2.1: Η αριθμητική πρόγνωση του καιρού

Οι ατμοσφαιρικές διεργασίες διέπονται από τους νόμους του κλάδου της δυναμικής των ρευστών και της θερμοδυναμικής. Η θεωρία του Bjerknes (1922) κατάφερε να τις περιγράψει με σαφέστερο τρόπο, προσφέροντας το υπόβαθρο ώστε να πραγματοποιηθούν προγνώσεις μεγαλύτερης ακρίβειας. Εντούτοις, μέχρι την κατασκευή των ηλεκτρονικών υπολογιστών, η προγνωστική διαδικασία συνέχισε να εκτελείται με τον ίδιο τρόπο. Οι μετεωρολόγοι, λαμβάνοντας υπόψη τις επίγειες παρατηρήσεις και έχοντας ως οδηγό τη γενικότερη έκφραση αυτής της θεωρίας, υπέθεταν στη συνέχεια την κίνηση των συστημάτων. Από την άλλη πλευρά, ο Richardson (1922), χρειάστηκε αρκετές εβδομάδες προκειμένου να υπολογίσει μία μόνο παράμετρο σε ένα σημείο στην επιφάνεια της Γης, κάτι το οποίο καθιστούσε αυτή τη μέθοδο πρακτικά μη εφαρμόσιμη εκείνη την περίοδο για επιχειρησιακούς σκοπούς.

2.1.1: Οι πρώτες προσπάθειες

Η πρώτη προσπάθεια επιχειρησιακής πρόγνωσης σε καθημερινή βάση πραγματοποιήθηκε από την επιστημονική ομάδα του Carl-Gustav Rossby το 1954 (Harper et al., 2007). Ακολούθησαν οι ΗΠΑ με την συνεργασία του Ναυτικού, της Αεροπορίας και της Μετεωρολογικής Υπηρεσίας το 1957-1958 (Lynch, 2008) και στη συνέχεια κι άλλες χώρες (όπως Ιαπωνία και Αυστραλία). Για την επίτευξη της ατμοσφαιρικής προσομοίωσης, αρχικά χρησιμοποιήθηκαν βαροτροπικά μοντέλα μίας παραμέτρου, στα οποία λαμβάνεται υπόψιν μόνο η κυκλοφορία σε ένα ατμοσφαιρικό επίπεδο (στα 500hpa), σύμφωνα με την εξίσωση:

$$\frac{\partial \nabla^2 \psi}{\partial t} + V_\psi \cdot \nabla(\nabla^2 \psi + f) = 0 \qquad (2.1)$$

, *όπου:*

ψ *: η εξίσωση ροής του γεωστροφικού ανέμου,*

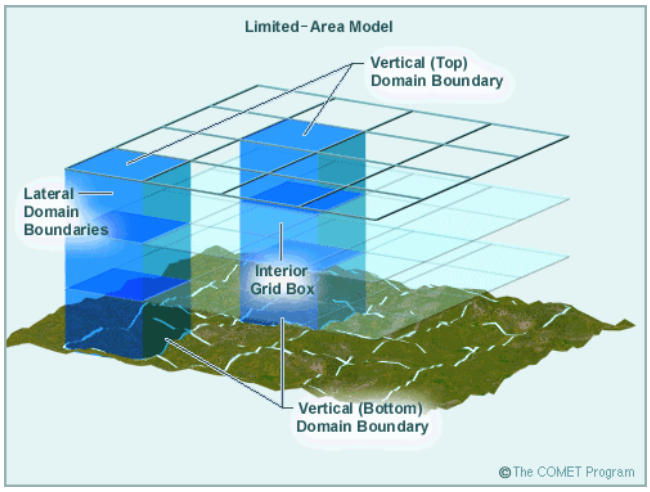
V_ψ: *ο μη αποκλίνων άνεμος,*

f : *η παράμετρος coriolis.*

Σε αυτά τα μοντέλα γίνεται η παραδοχή ότι η πυκνότητα της ατμόσφαιρας παραμένει σταθερή σε κάθε ισοβαρική επιφάνεια (άρα η βαροκλινικότητα είναι μηδενική). Για το λόγο αυτό δε μπορούν να ληφθούν υπόψιν οι θερμικές διεργασίες, οι οποίες όμως συνεισφέρουν σε ένα μεγάλο βαθμό στη δημιουργία νέων καιρικών συστημάτων. Ωστόσο κατά τη δεκαετία του '60 και του '70, η πολυπλοκότητα των υπολογισμών αυξήθηκε με την εφαρμογή των βασικών εξισώσεων περιγραφής των κινήσεων της ατμόσφαιρας (Primitive Equations - PE), οι οποίες αποτελούν τη βάση των υπολογισμών και στα σύγχρονα συστήματα πρόγνωσης.

2.1.2: Αριθμητική Πρόγνωση Καιρού: Η σύγχρονη προσέγγιση

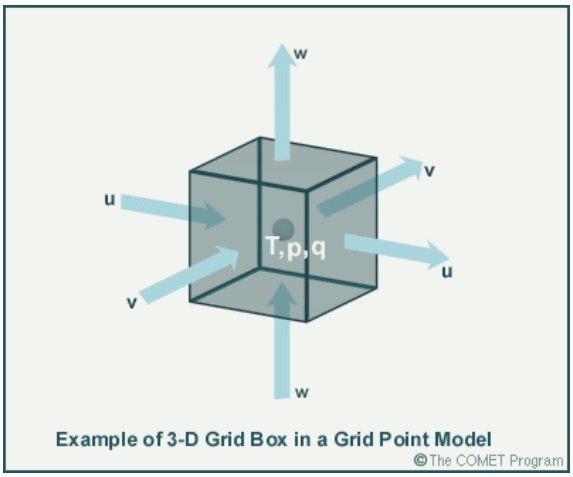
Στην σύγχρονη προσέγγιση της αριθμητικής πρόγνωσης του καιρού ο πλανήτης είναι συνήθως χωρισμένος σε ίσα τμήματα με βάση το σύστημα γεωγραφικών συντεταγμένων, κάτι το οποίο ορίζει την **οριζόντια ανάλυση** της πρόγνωσης ακολουθώντας την ιδέα του Richardson. Η χωρική ανάλυση προκύπτει με την προσθήκη συγκεκριμένων κατακόρυφων επιπέδων (συνήθως άνω των 25) από την επιφάνεια της θάλασσας έως τα όρια της ατμόσφαιρας (σχήμα 2.1).



Σχήμα 2.1: *Γραφική αναπαράσταση του πλέγματος ενός περιοχικού μοντέλου πρόγνωσης της ατμόσφαιρας. Διακρίνεται η οριζόντια ανάλυση, τα κατακόρυφα επίπεδα καθώς και τα 3-διαστάσεων κελιά που δημιουργούνται. Πηγή: The COMET Program (https://www.comet.ucar.edu/)*

Οι υπολογισμοί εκτελούνται πάνω στο τριών διαστάσεων πλέγμα σε συγκεκριμένες χρονικές στιγμές που ορίζουν την **χρονική ανάλυση** του βήματος της πρόγνωσης. Σε κάθε ένα εκ των πλεγμάτων θεωρούμε ομοιογενείς συνθήκες, οι οποίες σύμφωνα με τον Bjerknes μπορούν να οριστούν από επτά παραμέτρους: **Θερμοκρασία** (T), **Πίεση** (P), **Υγρασία** (q),

Πυκνότητα (ρ) και των **τριών συνιστωσών του διανύσματος του ανέμου** στο χώρο (u, v, z) (σχήμα 2.2)



Σχήμα 2.2: *Γραφική αναπαράσταση ενός κελιού (x,y,z) στο οποίο εκτελούνται οι υπολογισμοί των βασικών ατμοσφαιρικών παραμέτρων. Πηγή: The COMET program.*

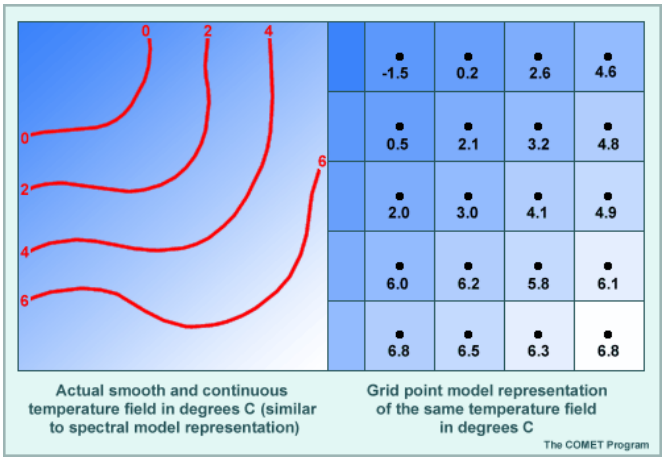
Σε κάθε ένα από αυτά υποθέτουμε πως η αλλαγή σε κάποια προγνωστική παράμετρο Α, σε χρονικό διάστημα t, είναι ίση με τις αθροιστικές επιπτώσεις όλων των διεργασιών που αναγκάζουν την Α να αλλάξει. Η γενική μαθηματική έκφραση της παραπάνω πρότασης θα μπορούσε να είναι η εξής:

$$A^{forecast} = A^{initial} + F(A)\Delta t \Leftrightarrow \frac{\Delta A}{\Delta t} = F(A) \qquad (2.2)$$

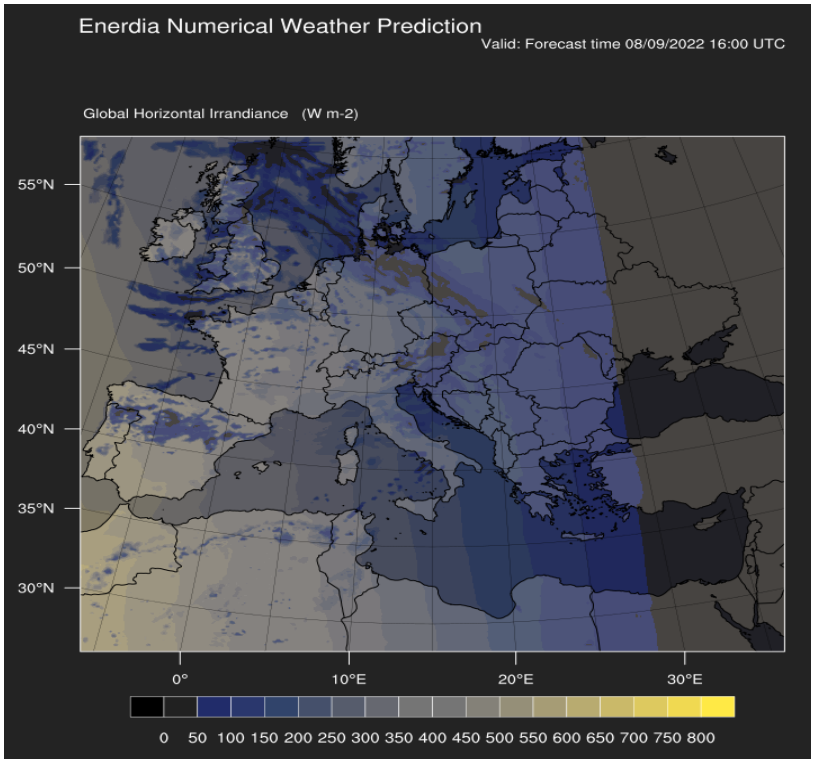
Όπως επισημαίνει η Kalnay (2002), « *οι βασικές εξισώσεις [περιγραφής της ατμόσφαιρας] είναι οι νόμοι διατήρησης της μάζας, της ενέργειας, της ορμής, της υγρασίας (σε όλες τις φάσεις), καθώς και ο νόμος των ιδανικών αερίων, εφαρμοσμένοι σε ξεχωριστά "πακέτα αέρα". Για την επίλυσή τους περιλαμβάνουν τα κύματα βαρύτητας και ήχου, συνεπώς στη χωρική και χρονική τους διακριτοποίηση απαιτείται η χρήση μικρότερων χρονικών βημάτων ή άλλου είδους τεχνικές ώστε να τις επιβραδύνουν* ».

Το αποτέλεσμα των υπολογισμών συνήθως αναπαρίσταται σε επίπεδα δύο διαστάσεων, εκκινώντας από την επιφάνεια της θάλασσας έως και τα ανώτερα στρώματα της ατμόσφαιρας. Κάθε σημείο του πλέγματος χαρακτηρίζεται από μία τιμή για κάθε παράμετρο, όμως στη διαδικασία της οπτικοποίησης τα δεδομένα παρουσιάζονται αφού πρώτα έχουν εφαρμοστεί τεχνικές παρεμβολής (σχήμα 2.3). Στο σχήμα 2.4 απεικονίζεται η πρόγνωση της εισερχόμενης ηλιακής ακτινοβολίας στην Ευρώπη από το μοντέλο WRF-ARW V4.0 σε

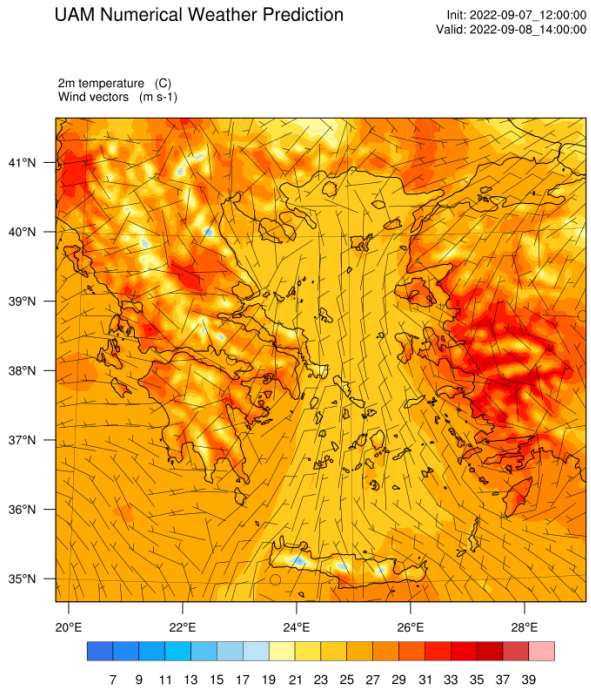
ανάλυση 12km, το οποίο έχει αναπτυχθεί στις υπολογιστικές εγκαταστάσεις της εταιρείας Enerdia. Στα σχήματα 2.5 & 2.6 παρουσιάζονται τα προγνωστικά αποτελέσματα του συστήματος WRF-ARW, το οποίο αναπτύσσεται σε συνεργασία με το δήμο Διστόμου-Αράχωβας-Αντίκυρας για τη θερμοκρασία στα 2μ και τη διεύθυνση και ταχύτητα του ανέμου στα 10μ. στην Ελλάδα και στην ευρύτερη περιοχή του Παρνασσού, σε οριζόντιες αναλύσεις 5χλμ και 1χλμ αντίστοιχα.



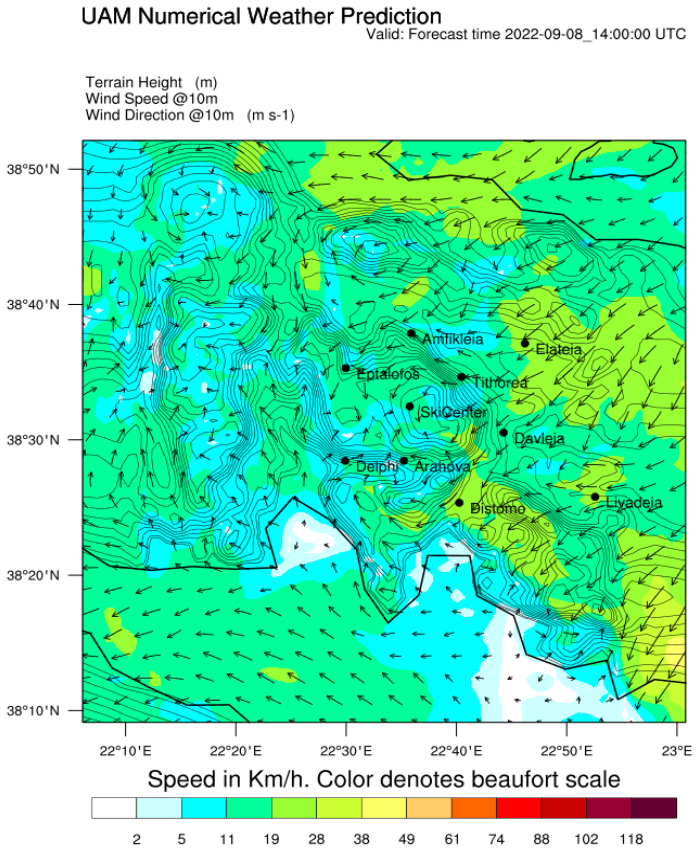
Σχήμα 2.3: *Εικονικό παράδειγμα αποτύπωσης της πρόγνωσης της θερμοκρασίας στα 2μ, όπως υπολογίζεται στο πλέγμα του μοντέλου (δεξί τμήμα) και όπως οπτικοποιείται στην τελική παρουσίαση (αριστερό τμήμα). Πηγή: The COMET program.*



Σχήμα 2.4: *Πρόγνωση εισερχόμενης ηλιακής ακτινοβολίας στην Ευρώπη από το μοντέλο WRF-ARW V4.0 της εταιρίας Enerdia, για τις 8 Σεπτεμβρίου 2022, 16UTC σε ανάλυση 12km. Πηγή: https://portal.ergacis.com*



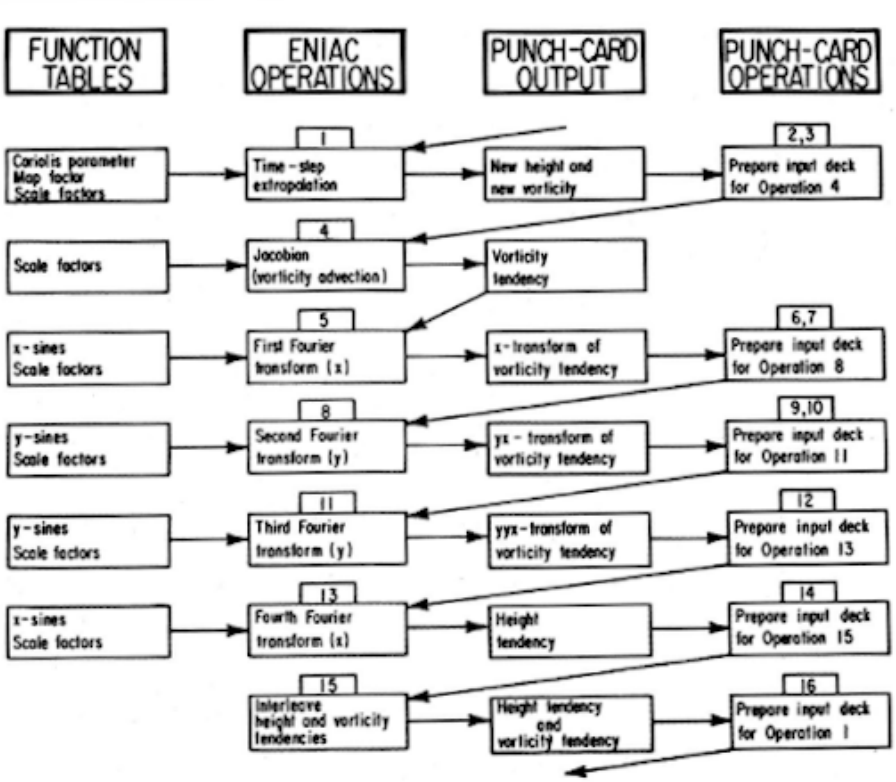
Σχήμα 2.5: *Πρόγνωση της θερμοκρασίας στα 2μ στην Ελλάδα από το μοντέλο UAM/WRF, για τις 8 Σεπτεμβρίου 2022, 14UTC, σε ανάλυση 5km. Πηγή: http://umeteo.com/arahovameteo/meteomaps.html*



Σχημα 2.6: *Πρόγνωση του ανέμου στα 10μ στην ευρύτερη περιοχή του Παρνασσού από το μοντέλο UAM/WRF, για τις 8 Σεπτεμβρίου 2022, 14UTC, σε ανάλυση 1km.*

Πηγή: http://umeteo.com/arahovameteo/meteomaps.html

Βάσει των παραπάνω, θα μπορούσε να θεωρηθεί ότι η αριθμητική πρόγνωση του καιρού είναι στην ουσία ένα πρόγραμμα (ή ένα σύνολο προγραμμάτων) τα οποία "τρέχουν" σε κάποιον υπερυπολογιστή με στόχο την επίλυση συγκεκριμένων εξισώσεων για κάθε σημείο στο χώρο. Τελικός σκοπός είναι η σύνθεση των αποτελεσμάτων σε ένα ενιαίο πεδίο για την παραγωγή των προσομοιώσεων. Η διαδικασία εκτέλεσης των υποπρογραμμάτων ακολουθεί μια συγκεκριμένη σειρά, ώστε τα επιμέρους αποτελέσματα να αποτελέσουν τις πηγές εισόδου για άλλα υποπρογράμματα κ.ο.κ. μέχρι την τελική πρόγνωση. Εξετάζοντας τα πρώτα προγράμματα που εκτελέστηκαν στις απαρχές της αριθμητικής πρόγνωσης (όπου χρησιμοποιήθηκαν βαροτροπικά μοντέλα ενός επιπέδου με στόχο την πρόγνωση μίας παραμέτρου - το γεωδυναμικό ύψος στα 500hpa), η διαδικασία αυτή φαίνεται σχετικά απλή με τα σημερινά δεδομένα. Στο σχήμα 2.7, παρουσιάζεται ο κώδικας ροής του πρώτου αριθμητικού μοντέλου που εκτελέστηκε στον ENIAC.

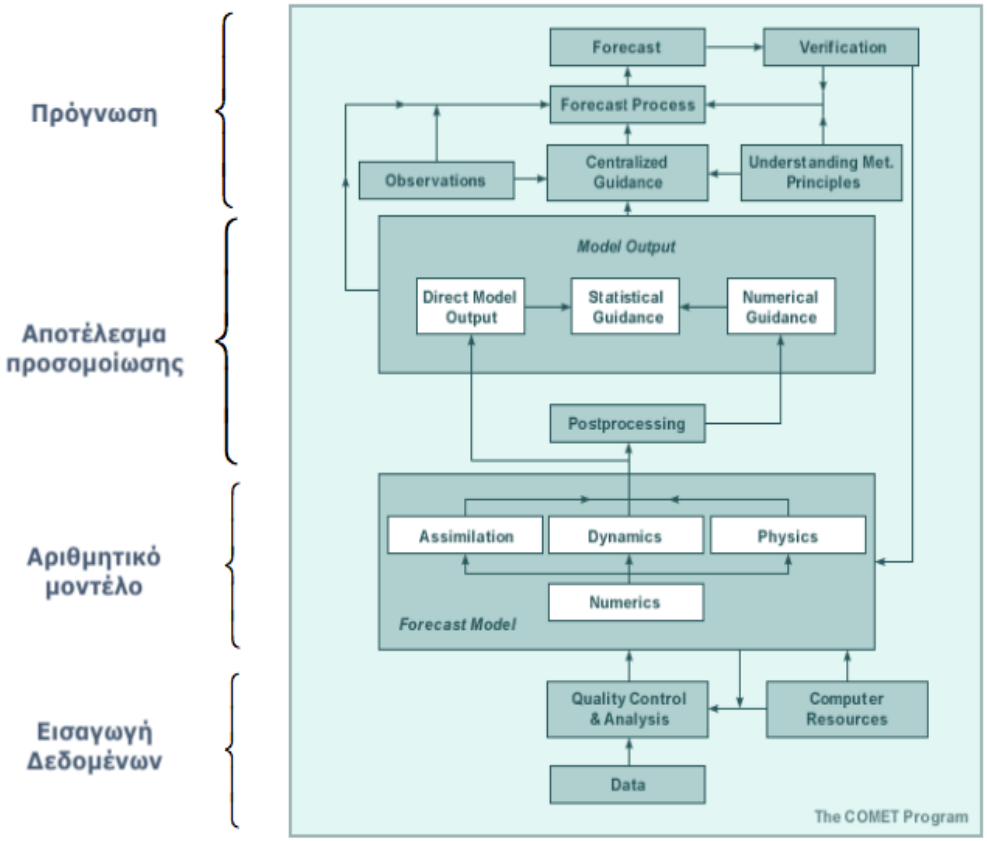


Σχήμα 2.7: *Το διάγραμμα ροής του κώδικα της πρώτης αριθμητικής πρόγνωσης καιρού που πραγματοποιήθηκε στον ENIAC (Lynch, 2008)*

Στην σύγχρονη προσέγγιση η διαδικασία είναι πολύ πιο περίπλοκη περιλαμβάνοντας στάδια που είτε προηγούνται του υπολογισμού των ατμοσφαιρικών παραμέτρων (pre-processing), είτε έπονται (post-processing). Ως pre-processing στάδιο θεωρείται για παράδειγμα η αφομοίωση των δεδομένων (Data Assimilation), τα οποία προέρχονται από διάφορες πηγές όπως δορυφορικές εικόνες, μετεωρολογικοί σταθμοί, ραδιοβολήσεις, παρατηρήσεις από αεροσκάφη, πλοία κ.ά. πηγές και η δημιουργία ενός ενιαίου κανονικοποιημένου πεδίου που μπορεί να τροφοδοτήσει το μοντέλο. Μια αποτελεσματική και ισχυρή προσέγγιση στη διαδικασία του Data Assimilation, είναι η τετραδιάστατη παραμετρική αφομοίωση δεδομένων (4D-Var), η οποία περιλαμβάνει την «adjoint τεχνική» και έχει το πλεονέκτημα της αφομοίωσης διαφόρων παρατηρήσεων που κατανέμονται στο χώρο και στο χρόνο στο αριθμητικό μοντέλο, διατηρώντας παράλληλα τη δυναμική και φυσική συνοχή με το μοντέλο (Κατσαφάδος, Μαυροματίδης, 2015). Η συγκεκριμένη τεχνική χρησιμοποιείται πλέον από πολλά παγκόσμια προγνωστικά συστήματα όπως ECMWF, GFS, UKMO κ.ά. Από την άλλη πλευρά, ένα παράδειγμα "post-processing" επεξεργασίας των αποτελεσμάτων της προσομοίωσης αποτελούν τα "φίλτρα Kalman", τα οποία αποτελούνται από ένα σετ εξισώσεων που υλοποιούν αποτελεσματικά τη μέθοδο των ελαχίστων τετραγώνων. Στην

πράξη, ένα "φίλτρο Kalman" αποτελεί μια στατιστικά βέλτιστη εκτίμηση των διαδικασιών ενός δυναμικού συστήματος, συνδυάζοντας πραγματικές παρατηρήσεις και πρόσφατες προγνώσεις με την εφαρμογή κατάλληλων βαρών για την επίτευξη ενός αποτελέσματος με μικρότερες αποκλίσεις (Galanis et. al, 2006).

Στο σχήμα 2.8 παρουσιάζεται το διάγραμμα ροής όλων των διαδικασιών που λαμβάνουν χώρα για την επίτευξη της πρόγνωσης του καιρού.



Σχήμα 2.8: *Διάγραμμα ροής όλων των διαδικασιών που λαμβάνουν χώρα για την επίτευξη της πρόγνωσης του καιρού.*

Τροποποίηση από το πρόγραμμα COMET (https://www.comet.ucar.edu/ *)*

2.1.3: Παγκόσμια προγνωστικά συστήματα

Από τον Αύγουστο του 1980, ξεκίνησε η πλανητική προσομοίωση με το 12 επιπέδων (Global Spectral Model - GSM) μοντέλο της Εθνικής Μετεωρολογικής Υπηρεσίας των ΗΠΑ, σε οριζόντια ρομβοειδή (rhomboidal) ανάλυση R30, η οποία αντιστοιχεί περίπου σε 2.25° x 3.75° μοίρες (γ.πλάτος x γ.μήκος) με προγνωστικό ορίζοντα τις 48h (White et. al, 2019). Την 1η Αυγούστου 1979 το Integrated Forecast System (IFS) του Ευρωπαϊκού Κέντρου

Μεσοπρόθεσμων Προγνώσεων (ECMWF), ένα ακόμη φασματικό μοντέλο, πέρασε στην επιχειρησιακή του φάση, με καθημερινές προγνώσεις σε ανάλυση T106, η οποία αντιστοιχεί περίπου σε 1.12° μοίρες με προγνωστικό ορίζοντα τις 10 ημέρες (ECMWF, 2019). Ακολούθησαν η Μετεωρολογική Υπηρεσία του Ηνωμένου Βασιλείου (United Kingdom Met. Office - UKMO) (Ιστότοπος UK Met. Office), η Καναδική Μετεωρολογική Υπηρεσία κ.ά. Σήμερα η ανάλυση των προγνωστικών συστημάτων έχει αυξηθεί τόσο, ώστε να καλύπτει τον πλανήτη με ακρίβεια λίγων χιλιομέτρων. Στον πίνακα που ακολουθεί παρουσιάζονται τα χαρακτηριστικά μερικών από τα πλέον γνωστά παγκόσμια μοντέλα.

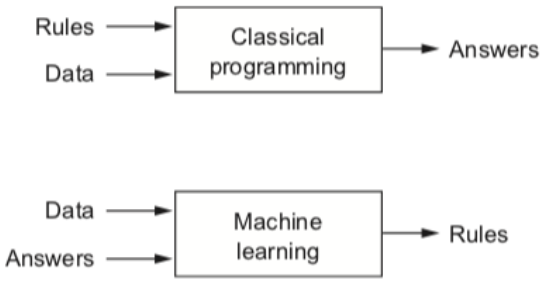
	ECMWF	GFS	UKMO	ICON
Ανάλυση (km)	9	27	10	13
Κατακόρυφα επίπεδα	137	127	70	90
Προγνωστικός ορίζοντας (h)	240	384	240	168
Ανανέωση / ημέρα	4	4	2	4
Χρονικό βήμα (h)	1 (μέχρι τις 78h)	1 (μέχρι τις 120h)	1 (μέχρι τις 48h)	1 (μέχρι τις 72h)

Πίνακας 2.1: *Γενικά χαρακτηριστικά μερικών γνωστών παγκόσμιων μοντέλων.*

2.2: Τα νευρωνικά δίκτυα

Στην τυπική περίπτωση επίλυσης κάποιου προβλήματος μέσω του προγραμματισμού, όπως για παράδειγμα η προσομοίωση της ατμόσφαιρας, απαιτούνται συνήθως δεδομένα εισόδου στο πρόγραμμα (οι αρχικές συνθήκες) και ένα σύνολο κανόνων (οι εξισώσεις που περιγράφουν τις ατμοσφαιρικές διεργασίες), ώστε να παραχθεί ένα αποτέλεσμα. Η βελτίωση τους επιτυγχάνεται με την είσοδο ακριβέστερων δεδομένων ή / και με την εφαρμογή βελτιωμένων κανόνων. Στη διαδικασία της μηχανικής εκμάθησης η παραπάνω λογική αντιστρέφεται σε σχέση με το τί αποτελεί είσοδο αλλά και έξοδο ενός προγράμματος / μοντέλου. Σε αυτήν την περίπτωση τα δεδομένα “data” και η παρατηρηθείσα τιμή τους “answers” σε διαφορετικό χρονικό ή χωρικό σημείο είναι οι τιμές εισόδου και ζητούμενο αποτελούν οι ίδιοι οι νόμοι που διέπουν το υπό μελέτη φαινόμενο (σχήμα 2.9). Για παράδειγμα, ένα μοντέλο θα μπορούσε να τροφοδοτηθεί με τα δεδομένα αισθητήρων κίνησης ενός οχήματος κατά τη διάρκεια ενός φρεναρίσματος σε διάφορες περιπτώσεις, ώστε να ελεγχθεί η

ασφάλεια των ελαστικών. Παρ'όλο που αυτό μπορεί να περιγραφεί μέσω των νόμων της μηχανικής με παράγοντες όπως η μάζα, η επιτάχυνση και η τριβή, το αποτέλεσμα στην πραγματική ζωή μπορεί να είναι τελείως διαφορετικό από το αναμενόμενο ακόμη και με θεωρητικά ίδιες τις αρχικές συνθήκες. Χαρακτηριστική επίσης είναι η περίπτωση του προβλήματος του διπλού εκκρεμούς. Αν και οι νόμοι που διέπουν την κίνηση είναι επίσης γνωστοί, το σύστημα καταλήγει σε μια χαοτική μη προβλέψιμη κίνηση για μεγάλες γωνίες μετά από κάποιο χρονικό σημείο. Ένα κατάλληλα σχεδιασμένο μοντέλο μηχανικής εκμάθησης, μπορεί εάν εκτεθεί σε ένα μεγάλο όγκο πειραματικών δεδομένων να μάθει συσχετισμούς και μοτίβα που προηγουμένως δεν ήταν γνωστά.

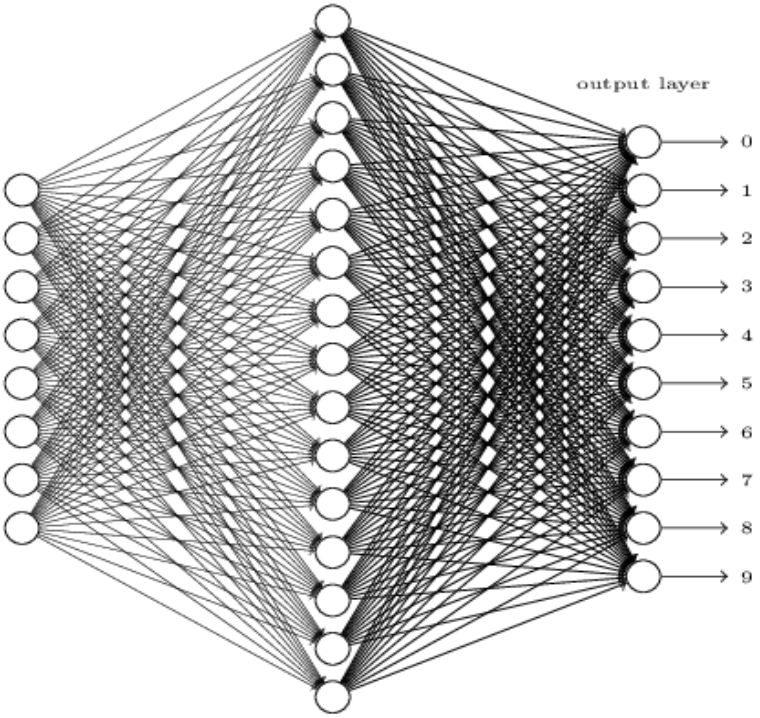


Σχήμα 2.9: *Η λογική διαφορά στην επίλυση ενός κλασικού προγραμματιστικού προβλήματος κι ενός προβλήματος μηχανικής εκμάθησης.*

Τυπικά η απαρχή των νευρωνικών δικτύων θεωρείται η εργασία των McCulloch και Pitts, ***"A logical calculus of the ideas immanent in nervous activity"*** (1943). Σε αυτήν προτείνεται ένα μαθηματικό μοντέλο προτασιακής λογικής (propositional logic) για την περιγραφή της νευρωνικής λειτουργίας των οργανισμών με βάση κάποιες παραδοχές όπως ότι: 1) *η μορφή του δικτύου δεν αλλάζει*, 2) *η λειτουργία του κάθε νευρώνα αντιμετωπίζεται ως μια ενιαία διαδικασία*, 3) *η μόνη καθυστέρηση μεταφοράς του σήματος εντοπίζεται στην περιοχή των συνάψεων*, 4) *η ανασταλτική δραστηριότητα κάποιας σύναψης αποτρέπει σε κάθε περίπτωση την ενεργοποίηση του νευρώνα*, 5) *για την ενεργοποίηση ενός νευρώνα ένας συγκεκριμένος αριθμός συνάψεων θα πρέπει επίσης να ενεργοποιηθεί, γεγονός ανεξάρτητο από την προηγούμενη κατάσταση του νευρώνα*. Η συγκεκριμένη εργασία έδωσε το κίνητρο για την αντίστοιχη του ψυχολόγου Frank Rosenblatt, ***"The perceptron: A probabilistic model for information storage and organization in the brain"*** (1958), όπου χρησιμοποιώντας σχεδόν τις ίδιες παραδοχές, θεώρησε ως δεδομένα εισόδου (inputs) αριθμούς που ανήκουν στο σύνολο Z, σε αντίθεση με την προτασιακή λογική του δικτύου των McCulloch και Pitts όπου τα δεδομένα είναι δυαδικής (0-1) μορφής. Στην περίπτωση των **perceptrons** το αποτέλεσμα καθορίζεται από την ύπαρξη βαρών για κάθε είσοδο (w: weights), ενώ για την ενεργοποίηση του νευρώνα απαιτείται η

ύπαρξη της απόκλισης (b: bias), η οποία επίσης εκφράζεται μέσω κάποιου αριθμού. Για το λόγο αυτό, αυτή η απλή μορφή μονάδας υπολογισμών καλείται και Linear Threshold Unit **(LTU)**.

Κάθε νευρωνικό δίκτυο αποτελείται τουλάχιστον από ένα επίπεδο εισόδου και ένα επίπεδο εξόδου (input, output layer), ενώ μπορεί ενδιάμεσα να υπάρχουν κι άλλα επίπεδα νευρώνων τα οποία καλούνται "κρυφά" (hidden layers). Όσον αφορά την αρχιτεκτονική τους, διακρίνονται τρείς κύριοι τύποι: 1. *Τα Artificial Neural Networks* (**ANN's**), 2. τα *Recurrent Neural Networks* (**RNN's**) και 3. τα *Convolutional Neural Networks* (**CNN's**). Ένας ακόμη διαχωρισμός αφορά στον τρόπο διασύνδεσης των νευρώνων μεταξύ των διαφόρων επιπέδων, όπου υπάρχουν δύο κατηγορίες: Τα *πλήρως διασυνδεδεμένα* (**Fully Connected Neural Networks, FCNN's**), όπου κάθε νευρώνας ενός επιπέδου, διασυνδέεται με όλους τους υπόλοιπους στο επόμενο επίπεδο, και τα *μερικώς διασυνδεδεμένα* (**Partially Connected Neural Networks, PCNN's**). Η **"βαθιά εκμάθηση" (Deep Learning)**, αναφέρεται σε σύγχρονα νευρωνικά δίκτυα που αποτελούνται από τουλάχιστον ένα κρυφό επίπεδο (Patterson & Gibson, 2017). Αυτός ο όρος φαίνεται πως χρησιμοποιήθηκε αρχικά το 2006 από τον Geoffrey Hinton, γνωστικό ψυχολόγο και επιστήμονα των υπολογιστών, στην εργασία του με τίτλο "*A fast learning algorithm for Deep Belief Nets*" (2006). Σε αυτήν περιγράφεται μια προσέγγιση για την εκμάθηση μιας "μηχανής Boltzmann", η οποία όμως αναφέρεται ως "βαθιά", λόγω του αριθμού των νευρώνων που χρησιμοποιήθηκαν. Ο Jurgen Schmidhuber (2014) ορίζει ως "πολύ βαθιά δίκτυα" όσα η συντομότερη διαδρομή συνάψεων μεταξύ των κόμβων (Credit Assignment Path - CAP) είναι μεγαλύτερη των δέκα. Στο σχήμα 2.10 απεικονίζεται το παράδειγμα ενός υποθετικού πλήρους διασυνδεδεμένου νευρωνικού δικτύου βαθιάς εκμάθησης με οκτώ νευρώνες εισόδου, ένα κρυφό επίπεδο δεκαπέντε νευρώνων και δέκα νευρώνες εξόδου.



Σχήμα 2.10*: Παράδειγμα ενός πλήρους διασυνδεδεμένου νευρωνικού δικτύου βαθιάς εκμάθησης (CAP = 2) με οκτώ νευρώνες εισόδου, ένα κρυφό επίπεδο δεκαπέντε νευρώνων και δέκα νευρώνες εξόδου. Πηγή: Nielsen, 2019, Online βιβλίο "Neural Networks and Deep Learning", Κεφάλαιο 1.* [*http://neuralnetworksanddeeplearning.com/*](http://neuralnetworksanddeeplearning.com/)

2.2.1: Αρχή Λειτουργίας

Έστω ο παρακάτω μαθηματικός κανόνας, ο οποίος θα μπορούσε να εφαρμοστεί στην περίπτωση ενός perceptron:

$$\text{Εάν: } \sum_i \quad wi\ xi\ + b \leq 0 \rightarrow 0$$

$$\text{Εάν: } \sum_i \quad wi\ xi\ + b > 0 \rightarrow 1 \tag{2.3}$$

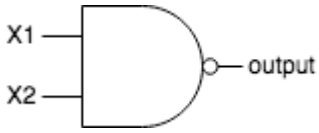
όπου w: κάποιο βάρος και b: bias.

Θέτοντας ως w_1, w_2 = -1 και b = 2 $\forall\ x1, x2 \in [0,1], x1, x2 \in N$, δημιουργείται ο ακόλουθος πίνακας τιμών:

x_1	x_2	Έξοδος	Αιτιολόγηση
0	**0**	**1**	**0·(-1) + 0·(-1) + 2 > 0**
0	**1**	**1**	**0·(-1) + 1·(-1) + 2 > 0**
1	**0**	**1**	**1·(-1) + 0·(-1) + 2 > 0**
1	**1**	**0**	**1·(-1) + 1·(-1) + 2 = 0**

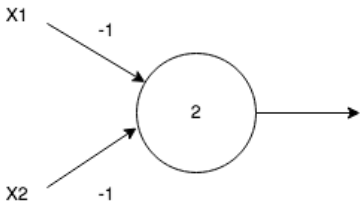
Πίνακας 2.1*: Πίνακας αποτελεσμάτων του μαθηματικού κανόνα (2.3) για όλους τους πιθανούς συνδυασμούς τιμών των μεταβλητών x_1, x_2.*

Όπως προκύπτει όμως, ο παραπάνω πίνακας τιμών ταυτίζεται με τον πίνακα αληθείας της λογικής πύλης (Not-AND) **NAND**, η οποία έχει επίσης δύο εισόδους και μία έξοδο (σχήμα 2.11) και δίνει αποτέλεσμα "0" μόνο στην περίπτωση που και οι δύο είσοδοι είναι "1".



Σχήμα 2.11*: Σχηματική αναπαράσταση της λογικής πύλης NAND*

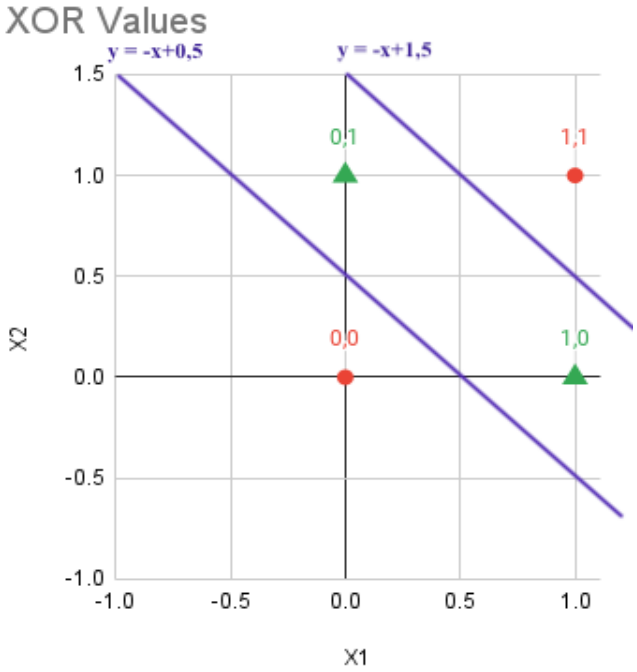
Η συγκεκριμένη πύλη και η NOR (Not-OR) παρουσιάζουν ειδικό ενδιαφέρον, καθώς θεωρούνται ως "οικουμενικές" και "συναρτησιακά πλήρεις" (**functional complete**), διότι με την αποκλειστική επαναλαμβανόμενη χρήση κάποιας εξ αυτών και την κατάλληλη διασύνδεσή τους μπορούν να παραχθούν όλες οι υπόλοιπες λογικές πράξεις. Η αναπαράσταση ενός τεχνητού νευρώνα "perceptron", ο οποίος εκτελεί την ίδια λογική πράξη φαίνεται στο σχήμα 2.12, όπου τα βάρη εισόδου ισοδυναμούν με -1 και η απόκλιση (bias) με 2.



Σχήμα 2.12*: Αναπαράσταση του μαθηματικού κανόνα ο οποίος υλοποιεί την λογική πύλη NAND σε μορφή perceptron, με δύο τιμές εισόδου X1, X2, βάρη w_1=w_2 = -1 και bias (b) = 2.*

Αν και φαινομενικά ένα perceptron και μια λογική πύλη (όπως η NAND) εκτελούν ακριβώς την ίδια διεργασία, τα δύο βασικά πλεονέκτημα των LTU's είναι ότι 1) μπορούν να δεχθούν περισσότερες των δύο εισόδων και 2) κάθε μία είσοδος μπορεί να συνδυαστεί με κάποιο "βάρος", γεγονός που, σε συνδυασμό και με τον παράγοντα b, προσδίδει στα νευρωνικά δίκτυα την ικανότητα εκμάθησης.

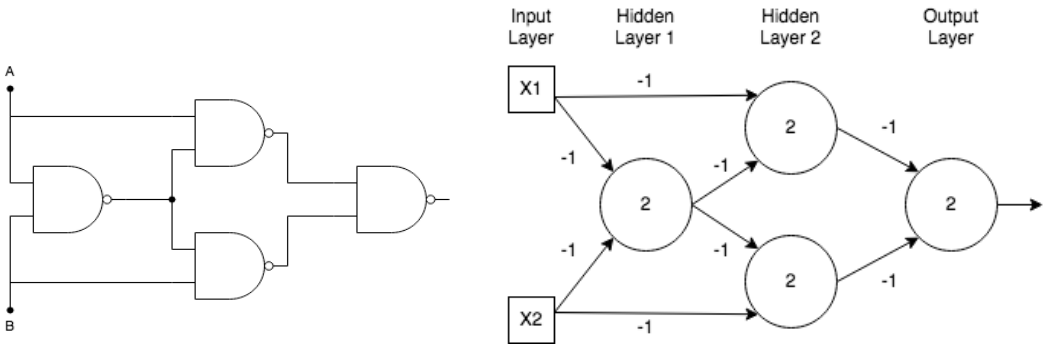
Το 1969 οι Minsky & Papert στο βιβλίο τους "*Perceptrons: an introduction to computational geometry"* διαπιστώνουν κάποιες αδυναμίες των **perceptrons**, μία εκ των οποίων είναι το γεγονός ότι δε μπορούν να επιλύσουν απλά προβλήματα κατηγοριοποίησης. Ένα εξ αυτών είναι και η εύρεση των τιμών εξόδου της πύλης XOR (η οποία δίνει έξοδο 1, όταν μόνο μία εκ των δύο εισόδων είναι 1). Το ζητούμενο λοιπόν ανάγεται στον προσδιορισμό ενός μαθηματικού κανόνα αντίστοιχου του (2.3), ο οποίος όμως θα περιλαμβάνει τις κατάλληλες τιμές w_1, w_2 και b προκειμένου να παραχθούν τα επιθυμητά αποτελέσματα της πύλης XOR.



Σχήμα 2.13: *Γραφική απεικόνιση της αδυναμίας υλοποίησης της λογικής πύλης XOR, μέσω της εφαρμογής γραμμικών συναρτήσεων.*

Έτσι λοιπόν, εάν οι τέσσερις συνδυασμοί εισόδων απεικονιστούν ως σημεία στο επίπεδο (0,0 - 0,1 - 1,0 - 1,0) προκύπτει το γράφημα του σχήματος 2.13. Με πράσινο τρίγωνο σημειώνονται αυτές που παράγουν έξοδο 1 και με κόκκινο κύκλο όσες παράγουν έξοδο 0. Από

τον τρόπο λειτουργίας των LTU's όμως προκύπτει ότι η εξίσωση κατηγοριοποίησης είναι πάντα γραμμική, άρα και ο διαχωρισμός των τιμών σε κάθε perceptron θα γίνεται με επίσης γραμμικό τρόπο. Είναι φανερό ότι στην περίπτωση της XOR κάτι τέτοιο είναι μαθηματικά αδύνατον, διότι απαιτούνται δύο διαφορετικές συναρτήσεις διαχωρισμού, οι οποίες θα ορίζουν τα επίπεδα όπου βρίσκονται οι τιμές εξόδου 0 (με εισόδους 0,0 και 1,0) και 1 (με εισόδους 0,1 και 1,0). Από την άλλη πλευρά όμως, εφόσον η πύλη NAND παρουσιάζει συναρτησιακή πληρότητα και μπορεί να υλοποιηθεί μέσω ενός perceptron με τα κατάλληλα βάρη και το κατάλληλο bias, όπως παρουσιάζεται και στο σχήμα 2.12, τελικά είναι εφικτό να υπάρξει υλοποίηση και της πύλης XOR, αλλά με τον συνδυασμό τεσσάρων NAND πυλών ή μέσω διαδοχικών perceptrons όπως φαίνεται στο σχήμα 2.14. Η αρχιτεκτονική στην οποία διαδοχικά perceptrons τοποθετούνται σε επίπεδα ονομάζεται **MLP** (Multi-Layer Perceptrons).



Σχήμα 2.14: *Η υλοποίηση της πύλης XOR μέσω του κατάλληλου συνδυασμού NAND πυλών (αριστερά) και η μετάφρασή της σε ένα μικρό νευρωνικό δίκτυο (δεξιά) με βάρη -1 και bias 2.*

Το νευρωνικό δίκτυο που τελικά υλοποιεί την XOR διαθέτει δύο κρυφά επίπεδα μεταξύ των εισόδων x_1 και x_2 και της εξόδου που είναι δυαδικής μορφής. Εάν όμως η τιμές των bias και των βαρών δεν ήταν εξ αρχής γνωστές, γνωρίζοντας ποιά είναι η ζητούμενη τιμή σε σχέση με τα δεδομένα εισόδου, είναι δυνατόν με την εφαρμογή κατάλληλων τεχνικών εκπαίδευσης να καθοριστούν τελικά οι τιμές των βαρών και της απόκλισης σε κάθε νευρώνα ώστε να παραχθεί το ίδιο αποτέλεσμα.

2.2.2: Γενικές Αρχές Εκπαίδευσης

Η κριτική που ασκήθηκε στον τρόπο λειτουργίας των perceptrons ως μεθόδου εύρεσης λύσεων αφορούσε μόνο τη δομική μονάδα ενός ANN (perceptron) και δεν έλαβε υπόψιν το γεγονός ότι ο συνδυασμός πολλών με κατάλληλο τρόπο θα μπορούσε να έχει το ίδιο αποτέλεσμα. Επιπροσθέτως, σε αντίθεση με τις λογικές πύλες, όπου η αρχιτεκτονική μίας (ή

ενός συνδυασμού αυτών) ανταποκρίνεται μόνο σε ένα συγκεκριμένο πρόβλημα, στην περίπτωση των ANNs αυτό δεν συμβαίνει: Κάποιο νευρωνικό δίκτυο μπορεί να εκπαιδευτεί για την επίλυση πολλών διαφορετικών προβλημάτων, χωρίς όμως να απαιτείται προηγούμενη γνώση σχετικά με τους νόμους που διέπουν τα δεδομένα. Η προβληματική συνεπώς ανάγεται στον σωστό καθορισμό των παραγόντων *w*, *b* σε ένα MLP νευρωνικό δίκτυο. Σε απλά ζητούμενα όπως η (εμπειρική) επίλυση της XOR, αυτό μπορεί να είναι εφικτό, σε δυσκολότερα προβλήματα όμως, όπου απαιτείται η δημιουργία ενός μεγαλύτερου νευρωνικού δικτύου με μεγαλύτερο όγκο δεδομένων, ο καθορισμός αυτών των παραγόντων είναι πρακτικά αδύνατος με το χέρι.

Η έννοια της "εκπαίδευσης" ενός ANN τύπου MLP περιγράφηκε τελικά πολλά χρόνια αργότερα στο σύγγραμμα των Rumelhart et al. (1986a), στο οποίο αναφέρεται η **"backpropagation"** τεχνική. Ο όρος προέρχεται από το γεγονός ότι η διαδικασία εκκινεί από το τέλος προς την αρχή. Εφόσον τα βάρη των ενδιάμεσων επιπέδων για κάθε νευρώνα είναι ήδη γνωστά, αρχικά υπολογίζεται το συνολικό σφάλμα στην έξοδο του δικτύου και στη συνέχεια ο αλγόριθμος διατρέχει το δίκτυο προς την αρχή, με σκοπό να βρεθεί ο βαθμός συνεισφοράς κάθε νευρώνα σε αυτό. Στη βασική της ιδέα χρησιμοποιείται η συνάρτηση του μέσου τετραγωνικού σφάλματος (MSE), η οποία μπορεί να οριστεί ως εξής για το σύνολο N των νευρώνων ενός δικτύου:

$$E = MSE = \frac{1}{N}\sum_{j=1}^{N} (y_j - \hat{y}_j)^2 \qquad (2.4)$$

όπου,

$\hat{\boldsymbol{y}}_{\boldsymbol{j}}$***:*** *η τιμή εξόδου του νευρώνα j για αυτό το στιγμιότυπο,*

$\boldsymbol{y_j}$***:*** *η επιθυμητή τιμή εξόδου του νευρώνα j για αυτό το στιγμιότυπο*.

Αυτό που επιχειρείται είναι η εύρεση των κατάλληλων τιμών *w*, *b*, ώστε η τιμή του σφάλματος να ελαχιστοποιηθεί μέσα από διαδοχικές δοκιμές *k,* ο αριθμός των οποίων είναι παραμετροποιήσιμος. Συνεπώς ορίζεται ένας χώρος 3 διαστάσεων εφόσον τελικά:

$$f(w, b) = \frac{1}{N}\sum_{j=1}^{N} (y_j - \hat{y}_j)^2 \qquad (2.5)$$

Έστω λοιπόν η περίπτωση δύο διαδοχικών δοκιμών *k, k* + 1. Ως προς την μεταβολή της παραμέτρου του βάρους (w), για τη σύνδεση μεταξύ νευρώνα i, j, θα ισχύει:

$$w_{ij}^{k+1} = w_{ij}^{k} + \Delta w, \quad (2.6)$$

όπου Δ εκφράζει μια μικρή ποσότητα.

Εξετάζοντας κάθε νευρώνα χωριστά, η εξίσωση (2.5) μπορεί να απλοποιηθεί:

$$f(w,b) \; = \; Error^2 \text{ , διότι } Error \; = \; (y_j - \widehat{y_j}) \quad (2.7)$$

Παραγωγίζοντας ως προς τον *w* μπορούμε να έχουμε (κανόνας αλυσίδας):

$$\frac{df}{dw} = \frac{df}{dError} \cdot \frac{dError}{dw} \Rightarrow \frac{df}{dw} = 2 \cdot E \cdot \frac{dError}{dw} \quad (2.8)$$

και αντιστοίχως ως προς τον b:

$$\frac{df}{db} = 2 \cdot E \cdot \frac{dError}{db} \quad (2.9)$$

Όμως έχουμε ορίσει πως $Error \; = \; (y_j - \widehat{y_j})$, το οποίο μπορεί να γίνει:

$$Error = y_j - (wx_i + b) \quad (2.10)$$

διότι από τον τρόπο που ορίζεται η συνάρτηση ενεργοποίησης κάθε νευρώνα *j* σε ένα LTU, η τιμή εξόδου είναι μια γραμμική συνάρτηση της μορφής: $\widehat{y_i} = wx_i + b$, όπου x_i είναι η τιμή εισόδου από τον νευρώνα *i.* Άρα η 2.8 γίνεται:

$$\frac{df}{dw} = 2 \cdot E \cdot \frac{d}{dw}(y_j - (w_{ij}x_i + b)) \quad (2.11)$$

όμως το b είναι το bias, το οποίο εφ'όσον παραγωγίζουμε ως προς w θεωρείται σταθερό, ενώ το y_j είναι η ζητούμενη (σωστή) τιμή εξόδου κάθε νευρώνα και άρα θα πρέπει επίσης να θεωρηθεί σταθερή (αναγράφονται ως C1, C2), οπότε η (2.11) γίνεται:

$$\frac{df}{dw} = 2 \cdot E \cdot \frac{d}{dw}(c_1 \; - (w_{ij}x_i + c_2)) \; = \; -2 \cdot E \cdot x_i = \; -2 \cdot (y_j - \widehat{y_j}) \cdot x_i \quad (2.12)$$

Λόγω της (2.12), η (2.6) μπορεί να γίνει:

$$w_{ij}^{k+1} = w_{ij}^{k} - 2 \cdot (y_j - \widehat{y_j}) \cdot x_i \qquad (2.13)$$

και προσθέτοντας τον ρυθμό εκμάθησης ($\boldsymbol{\eta}$) στον οποίο ενσωματώνεται και ο παράγοντας 2, μπορεί επίσης να γραφτεί και ως εξής σε επίπεδο δικτύου:

$$\boldsymbol{w_{ij}^{k+1} = w_{ij}^{k} - \eta \cdot [\sum \quad (y_j - \widehat{y_j}) \cdot x_i]} \qquad (2.14)$$

όπου,

w_{ij}: *συμβολίζεται η σύνδεση μεταξύ των νευρώνων i (είσοδος) και j (έξοδος),*

x_i: *η τιμή εισόδου από τον νευρώνα i,*

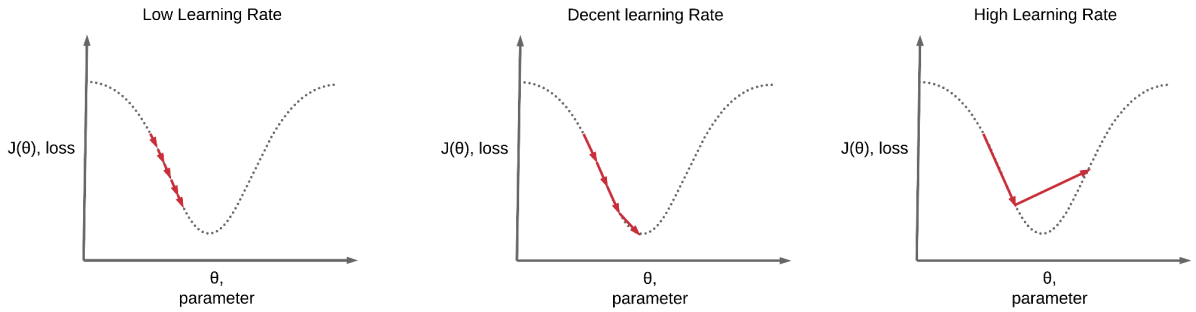
$\hat{y}_j$: *η τιμή εξόδου του νευρώνα j,*

y_j: *η επιθυμητή τιμή εξόδου του νευρώνα j για αυτό το στιγμιότυπο και*

η: *ο ρυθμός εκμάθησης.*

Η παράγωγος της συνάρτησης κόστους (2.8), ονομάζεται και "κλίση καθόδου" (**gradient descent**) και καθορίζεται από τα βάρη των συνδέσεων. Μέσω διαδοχικών επαναλήψεων εφαρμόζονται αλλαγές σε αυτά και κατόπιν υπολογίζεται η νέα παράγωγος. Έχοντας ως στόχο την ελαχιστοποίηση της τιμής, η οποία θα πρέπει να τείνει στο μηδέν μετά από k επαναλήψεις, μπορούν να εξαχθούν τα βέλτιστα βάρη για κάθε σύνδεση μεταξύ όλων των νευρώνων του δικτύου. Σαν αποτέλεσμα, η ικανότητα εκμάθησης καθορίζεται από το γεγονός ότι για κάθε είσοδο x_i στον νευρώνα *i* η οποία παράγει μια λανθασμένη έξοδο, ενισχύονται τα βάρη των εισόδων που θα μπορούσαν να συνεισφέρουν σε ένα ορθότερο αποτέλεσμα (Geron, 2017).

Ο **ρυθμός εκμάθησης (η)**, είναι μια σημαντική παράμετρος που εκφράζει το μέγεθος των αλλαγών που εφαρμόζονται σε κάθε επανάληψη. Εάν είναι μικρός, τότε απαιτείται μεγάλος αριθμός επαναλήψεων, ώστε ο αλγόριθμος να προσεγγίσει τη βέλτιστη λύση, εάν όμως είναι μεγάλος, τότε οι επαναλήψεις είναι μεν λίγες αλλά αυξάνεται η πιθανότητα να μην μπορέσει ποτέ να φτάσει στην επιθυμητή λύση. Στο σχήμα 2.15 απεικονίζεται με γραφικό τρόπο η επίδραση του ρυθμού εκμάθησης στην εύρεση του πιθανού ελαχίστου της συνάρτησης κόστους.



Σχήμα 2.15: *Η επιρροή του ρυθμού εκμάθησης στην εύρεση του ελαχίστου της συνάρτησης κόστους (ελάχιστη τιμή της κλίσης καθόδου). Πολύ μικρές, ή πολύ μεγάλες τιμές ενδέχεται να οδηγήσουν σε διαφορετικό αποτέλεσμα. Πηγή:*

https://www.deeplearningwizard.com/deep_learning/boosting_models_pytorch/lr_scheduling/ .

2.2.2.1: Κατηγορίες Αλγορίθμων Βελτιστοποίησης

Για μικρά νευρωνικά δίκτυα[3], τα οποία καλούνται να εκπαιδευτούν σε έναν σχετικά διαχειρίσιμο όγκο δεδομένων, η εύρεση της ελάχιστης κλίσης καθόδου μέσα από τη διαδοχικούς υπολογισμούς του συνολικού σφάλματος του δικτύου E_{net} (διαδικασία που γενικά αποκαλείται "**Batch Gradient Descent**", ή απλώς **GD)** είναι υπολογιστικά εύκολη. Όσο όμως αυξάνεται ο όγκος των δεδομένων αλλά και το μέγεθος του νευρωνικού δικτύου (που σημαίνει αύξηση του πλήθους των συνδέσεων), αυξάνονται πολλαπλασιαστικά και οι επαναλήψεις του υπολογισμού της συνάρτησης κόστους με αποτέλεσμα η διαδικασία να γίνεται εξαιρετικά αργή για αποδεκτές τιμές της παραμέτρου η.

Η μέθοδος του "**Stochastic Gradient Descent**" **(SGD)**, που αρχικά είχε παρουσιαστεί στη δημοσίευση των Robbins & Monro (1951), εφαρμόζεται ευρέως στην εκπαίδευση των ANN's. Σε αυτήν την περίπτωση χρησιμοποιείται ένα τυχαίο δείγμα των δεδομένων κάθε φορά για να υπολογιστεί η συνάρτηση κόστους και να καθοριστούν οι παράμετροι. Κατά τον τρόπο

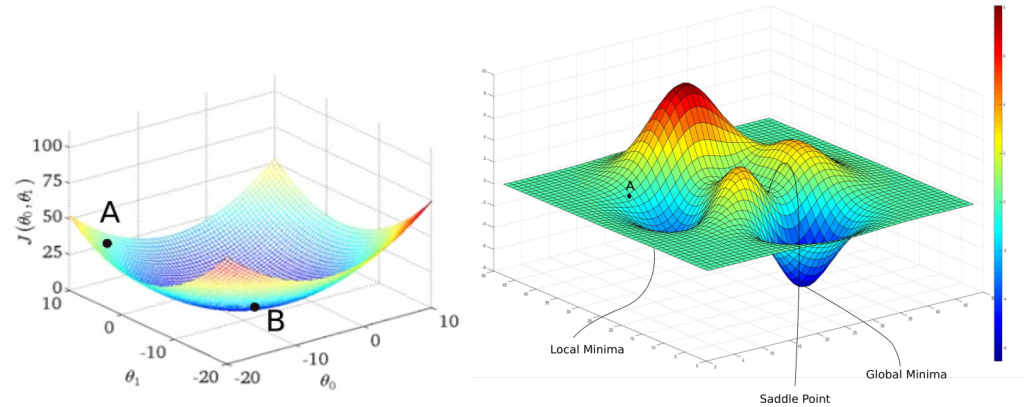
---

[3] *Δεν υπάρχει σαφής ορισμός του τί εστί "μικρό" ή "μεγάλο" όταν αναφερόμαστε στα νευρωνικά δίκτυα καθώς αυτό αποτελεί συνάρτηση του όγκου των δεδομένων πάνω στα οποία καλούνται να εκπαιδευτούν αλλά και της υπολογιστικής ισχύος. Θεωρώντας έναν συμβατικό Η/Υ, τεσσάρων πυρήνων επεξεργασίας και 8GB μνήμη με dataset της τάξεως μεγέθους εκατοντάδων χιλιάδων εγγραφών, με περίπου 10 attributes που να περιγράφουν κάθε καταχώρηση, ένα "μικρό" νευρωνικό δίκτυο τύπου MLP, μπορεί να έχει 10 εισόδους, 2 κρυφά επίπεδα που αποτελούνται έως 5 νευρώνες το καθένα και (αναλόγως του ζητούμενου), από 1 έως 10 εξόδους.*

αυτό μπορεί πολύ γρήγορα να προσεγγιστεί μια "καλή" λύση, χωρίς ωστόσο να αποτελεί και τη βέλτιστη, κάτι που όμως τις περισσότερες φορές είναι αποδεκτό. Ο όρος "stochastic" αναφέρεται στην τυχαιότητα της επιλογής κάθε φορά. Μια ενδιάμεση μέθοδος που εκμεταλλεύεται τα πλεονεκτήματα των προηγούμενων δύο είναι η "**Mini-Batch Gradient Descent**", όπου η επιλογή είναι τυχαία αλλά χρησιμοποιούνται μεγαλύτερα πακέτα δειγμάτων σε σχέση με την SGD.

Στις περισσότερες περιπτώσεις τα νευρωνικά δίκτυα δεν αποτελούνται από LTU's, συνεπώς οι μετασχηματισμοί που λαμβάνουν χώρα εντός των νευρώνων είναι μη γραμμικοί. Τότε το επίπεδο που ορίζεται από τη συνάρτηση σφάλματος δεν είναι συμμετρικό, τέτοιο ώστε να θυμίζει το σχήμα ενός κυπέλλου, όπως λ.χ. στο σχήμα 2.15, αλλά περισσότερο ακανόνιστο μοιάζοντας με την εικόνα κάπου γεωφυσικού χάρτη, όπου παρουσιάζονται όρη που το σφάλμα μεγιστοποιείται και καταβυθίσεις, οι οποίες ορίζουν τα τοπικά ελάχιστα (σχήμα 2.16). Εκκινώντας από ένα τυχαίο σημείο η διαδικασία του Batch Gradient Descent είναι εξαιρετικά πιθανό να οδηγήσει τον αλγόριθμο στην εύρεση ενός τοπικού ελαχίστου αντί για το ολικό. Από την άλλη πλευρά, φαίνεται πως η SGD και η Mini-Batch Gradient Descent μέθοδοι, λόγω της τυχαιότητας που εμπεριέχουν, ακολουθούν μια περισσότερο ακανόνιστη πορεία, μειώνοντας τελικά την πιθανότητα εγκλωβισμού σε κάποιο τοπικό ελάχιστο. Όπως όμως έχει ήδη ειπωθεί, αδυνατούν να επιτύχουν την ακρίβεια της GD, διαφορά που στην περίπτωση αυτή, λόγω και του πολύπλοκου "τοπίου" της συνάρτησης κόστους, μπορεί να είναι αρκετά μεγάλη ώστε να είναι αποδεκτή.

Η **"simulated annealing"** τεχνική η οποία παρουσιάστηκε από τους Kirkpatrick et al. (1983), για την επίλυση του "προβλήματος του περιπλανώμενου πωλητή" μπορεί να βοηθήσει στο να βελτιωθεί η απόδοση της SGD και της Mini-Batch Gradient Descent, εφαρμόζοντας ένα πλάνο μεταβολής του ρυθμού εκμάθησης κατά τη διάρκεια της εκπαίδευσης του νευρωνικού δικτύου. Οι παραλλαγές είναι αρκετές, όπου ο ρυθμός μπορεί να ακολουθεί μια πτωτική πορεία ή να τον αναγκάζει σε περιοδικές αυξομειώσεις. Κατ' αυτόν τον τρόπο δύναται να εξερευνηθεί καλύτερα το "τοπίο" μιας αρκετά πολύπλοκης συνάρτησης σφάλματος.



Σχήμα 2.16: *Δύο διαφορετικά τοπία συναρτήσεων κόστους. Στην απλή εκδοχή (αριστερά) παρατηρείται μόνο ένα ελάχιστο (σημείο B), ενώ σε μια περισσότερο πολύπλοκη περίπτωση διαπιστώνονται ένα τοπικό και ένα ολικό ελάχιστο. Πηγή:*

https://blog.paperspace.com/intro-to-optimization-in-deep-learning-gradient-descent/

2.2.2.1.1: Ο αλγόριθμος SGD

Στα σύγχρονα προβλήματα, τα οποία προσεγγίζονται με μη γραμμικό τρόπο, η χρήση των SGD/Mini-Batch Gradient Decent κυριαρχεί. Ωστόσο (όπως συμβαίνει και με την GD), παρουσιάζουν ευαισθησία στην τιμή του ρυθμού εκμάθησης (η), κάτι που η "simulated annealing" τεχνική (και οι παραλλαγές της) προσπαθούν να επιλύσουν μέσω της μεταβολής του κατά τη διάρκεια της εκπαίδευσης. Όμως σε επεκτάσεις της **SGD** που έχουν προταθεί, ο ρυθμός εκμάθησης επίσης τροποποιείται αλλά με διαφορετικό τρόπο:

Η εξίσωση (2.14) :

$$w_{ij}^{k+1} = w_{ij}^{k} - \eta \cdot [\sum \quad (y_j - \hat{y}_j) \cdot x_i]$$

θα μπορούσε να γραφτεί και ως εξής:

$$w := w - \eta \cdot \frac{1}{n}\sum_{i=1}^{n} \quad \nabla Q_i(w) \;=\; w - \eta \nabla Q(w) \qquad (2.15)$$

ως γενίκευση του κανόνα της **ανανέωσης των βαρών** (w) μετά από n δοκιμές. Η Q συμβολίζει την αντικειμενική συνάρτηση, της οποίας την τιμή θέλουμε να ελαχιστοποιήσουμε για το σύνολο των νευρώνων i, εξαρτώμενη από την τιμή των w. Ο όρος ∇ υποδηλώνει τη συνολική κλίση (gradient) της Q, όταν οι παράγοντες της αντικειμενικής συνάρτησης είναι πολλοί. Στη συνέχεια παρουσιάζονται μερικές παραλλαγές του αλγορίθμου SGD.

AdaGrad

Για παράδειγμα, στον αλγόριθμο **AdaGrad** (Adaptive Gradient) (Duchi et al., 2011) το η δεν αποτελεί μια συνολική τιμή αλλά αφορά κάθε παράμετρο, έτσι ώστε:

$$w_j := w_j - \frac{\eta}{\sqrt{G_{j,j}}} \cdot g_j \qquad (2.16)$$

με το $\{G_{j,j}\}$ να συμβολίζει τα στοιχεία ενός διανύσματος που αποτελεί τη διαγώνιο του πίνακα βαρών διαστάσεων j x j, με τα οποία πολλαπλασιάζεται ο ρυθμός εκμάθησης.

Εάν λοιπόν το g_τ ισούται με το $\nabla Q_i(w)$ όπως παρουσιάστηκε στην (2.15), $g_\tau = \nabla Q_i(w)$ και τ οι επαναλήψεις, τότε: $G_{j,j} = \sum_{\tau=1}^{t} g_{\tau,j}^2$.

RMSProp

Κάτι παρόμοιο συμβαίνει και στον αλγόριθμο **RMSProp** (Root Mean Square Propagation). Σε αυτήν την περίπτωση ο ρυθμός εκμάθησης διαιρείται με έναν παράγοντα που περιλαμβάνει τον κυλιόμενο μέσο όρο της κλίσης για κάθε βάρος. Η εξίσωση (2.15) ανανέωσης των βαρών τροποποιείται ως εξής:

$$w := w - \frac{\eta}{\sqrt{u(w,t)}} \cdot \nabla Q_i(w) \qquad (2.17)$$

όπου $u(w,t) := \gamma u(w,t-1) + (1-\gamma)(\nabla Q_i(w))^2$ *και* γ *ο παράγοντας "λήθης" (forgetting factor).*

Adam

Επέκταση του RMSProp αποτελεί ο αλγόριθμος **Adam** (Adaptive Moment Estimation), όπου στην περίπτωση αυτή συνυπολογίζεται ο κυλιόμενος μέσος όρος της κλίσης για δύο χρονικές στιγμές:

$$w^{(t+1)} \leftarrow w^t - \eta \frac{\widehat{m_w}}{\sqrt{\widehat{u_w}} + \varepsilon} \qquad (2.18)$$

με τα $\widehat{m_w}$ και $\widehat{u_w}$ να συμβολίζουν την πρώτη και τη δεύτερη χρονική στιγμή και ορίζονται ως εξής:

$\widehat{m_w} = \frac{m_w^{(t+1)}}{1-\beta_1^t}$ και $\widehat{u_w} = \frac{u_w^{(t+1)}}{1-\beta_2^t}$

ενώ οι παράγοντες $m_w^{(t+1)}$ και $u_w^{(t+1)}$ ως εξής:

$$m_w^{(t+1)} \leftarrow \beta_1 \cdot \quad m_w^{(t)} + (1-\beta_1) \cdot \nabla_w Q^{(t)}$$

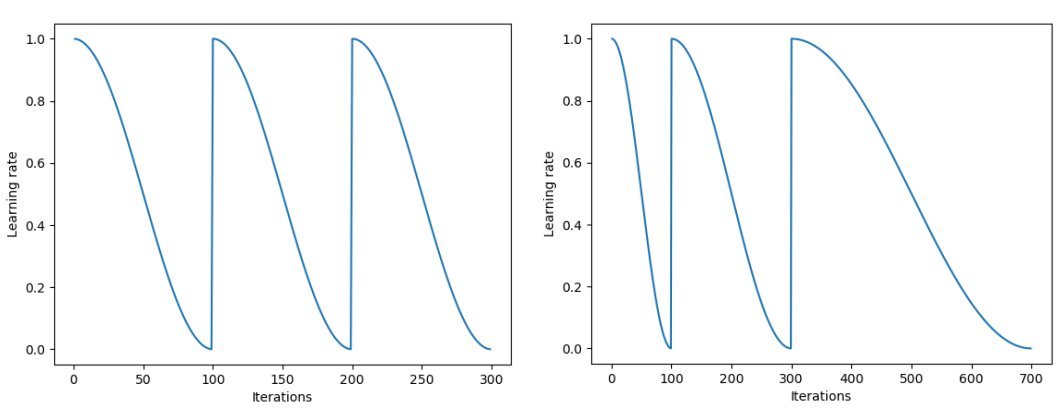
και

$$u_w^{(t+1)} \leftarrow \beta_2 \cdot \quad u_w^{(t)} + (1-\beta_2) \cdot (\nabla_w Q^{(t)})^2$$

Το *t* δείχνει τον τρέχοντα κύκλο εκπαίδευσης (ο πρώτος είναι 0) και οι παράγοντες β_1 και β_2 είναι οι "παράγοντες λήθης" για τις κλίσεις κατά την πρώτη και δεύτερη χρονική στιγμή. Το γεγονός ότι ο αλγόριθμος Adam αποτελεί επέκταση του RMSProp προκύπτει από τον παράγοντα $u_w^{(t+1)}$, που πρόκειται για τον $u(w,t)$ στην περίπτωση της RMSProp κάτι που τελικά ορίζει το "momentum". Η προσθήκη του $m_w^{(t+1)}$ είναι ένας εκθετικά μειούμενος μέσος όρος των προηγούμενων κλίσεων, παρόμοιος του "momentum". Η συνύπαρξη των δύο μειώνει δραστικά την "ταχύτητα" προσέγγισης κάποιου ελάχιστου και άρα την πιθανότητα αυτό να αποτελεί τοπικό ελάχιστο.

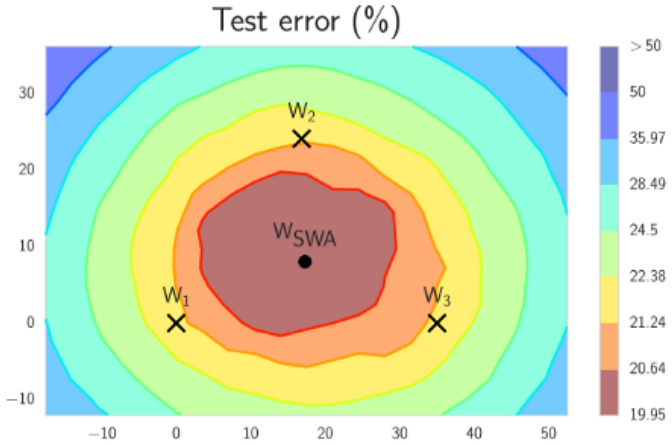
Άλλοι αλγόριθμοι βελτιστοποίησης

Οι **Adadelta** (Zeiler, 2012), **Adamax** (Kingma et al, 2014), **Nadam** (Dozat, 2015) είναι μερικοί ακόμη αλγόριθμοι εύρεσης του ελάχιστου σφάλματος (optimizers), οι οποίοι αποτελούν παραλλαγές του Adam. Από την άλλη πλευρά, ο **Ftrl** (McMahan et al., 2013) αποτελεί μια παραλλαγή της βασικής SGD μεθόδου αλλά αναπτύχθηκε για να εφαρμοστεί σε διαδικτυακά δεδομένα που αφορούν σε προτεινόμενες διαφημίσεις μέσω της πλατφόρμας της Google. Οι Loshchilov & Hutter (2016) ανέπτυξαν μια ακόμη παραλλαγή της SGD, η οποία ονομάστηκε **SGDR** (Stochastic Gradient Descent with Restarts), όπου, κατ' αναλογία με την "simulated annealing" τεχνική, ο ρυθμός εκμάθησης μειώνεται ακολουθώντας την συνάρτηση συνημιτόνου και στη συνέχεια εκκινεί πάλι από το μέγιστο για να ακολουθήσει την ίδια πορεία. Προτείνεται επίσης η σταδιακή αύξηση της περιόδου σε κάθε επανάληψη ως εναλλακτική της βασικής μεθόδου (σχήμα 2.17).



Σχήμα 2.17: *Η συνημιτονοειδής μεταβολή του ρυθμού εκμάθησης στην SGDR τεχνική με σταθερή περίοδο (αριστερά) και με αυξανόμενη (δεξιά).*

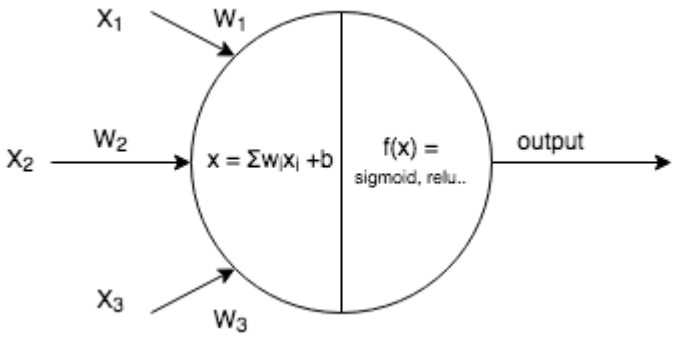
Η μέθοδος SGDR μπορεί να επεκταθεί ακόμη περισσότερο εφαρμόζοντας μια τεχνική **ensembling** που βασίζεται στην αποθήκευση των βαρών του νευρωνικού δικτύου κάθε φορά που η συνάρτηση κόστους προσεγγίζει κάποιο ελάχιστο. Τα τελικά βάρη του δικτύου είναι ο μέσος όρος των βαρών των τοπικών ελαχίστων κάτι που ονομάζεται **Stochastic Weight Averaging**. Σε σχέση με την πολλαπλή εκπαίδευση ενός νευρωνικού δικτύου (κλασικό ensembling), διαδικασία μεγάλου υπολογιστικού κόστους, η συγκεκριμένη, κατά τους ερευνητές, φαίνεται πως επιτυγχάνει παρόμοια αποτελέσματα. Εξάλλου, η εμπειρική παρατήρηση οδηγεί στο συμπέρασμα πως ο μέσος όρος των βαρών που εμφανίζονται σε τοπικά ελάχιστα βρίσκεται σε ένα άλλο σημείο του "τοπίου" της συνάρτησης κόστους, το οποίο μπορεί να δώσει μια περισσότερο γενικευμένη και εν τέλει ρεαλιστική εικόνα (σχήμα 2.18) (Izmailov et al., 2019).



Σχήμα 2.18: *Η τοπολογία των τοπικών ελαχίστων (x) και του ολικού ελάχιστου μιας συνάρτησης κόστους στην περίπτωση της εφαρμογής της μεθόδου Stochastic Weight Averaging*

2.2.2.2: Συναρτήσεις ενεργοποίησης

Στον ορισμό των perceptrons θεωρήθηκε ότι, ενώ οι είσοδοι και τα βάρη είναι φυσικοί αριθμοί, η έξοδος είναι δυαδικής μορφής, με συνέπεια ο τεχνητός νευρώνας να ενεργοποιείται μόνο όταν ισχύει μια συγκεκριμένη συνθήκη, η οποία διαμορφώνεται από τα βάρη των εισόδων και μια ορισμένη απόκλιση (**binary step activation**). Η συγκεκριμένη υλοποίηση για $n \geq 1$ τεχνητούς νευρώνες περιγράφηκε από τους McCulloch & Pitts (1943) καθώς και Rosenblatt (1958), για να επιλύσει προβλήματα αναγνώρισης εικόνων (μια ειδική περίπτωση προβλήματος κατηγοριοποίησης). Μια διαφορετική προσέγγιση θα μπορούσε να μην περιλαμβάνει κάποια ειδική συνθήκη εξόδου, συνεπώς η έξοδος κάθε νευρώνα είναι απλώς το άθροισμα των εισόδων και κάποιου bias για όλες τις τιμές εισόδου ή κατά περίπτωση (**linear activation**). Εναλλακτικά, το αποτέλεσμα της εξόδου θα μπορούσε να διαμορφώνεται από κάποια μη γραμμική συνάρτηση, συνήθως λογαριθμικής / εκθετικής μορφής (**non-linear activation**) (σχήμα 2.19). Οι Rumelhart et al. (1986a) χρησιμοποίησαν τελικά τη λογιστική συνάρτηση, ώστε να είναι δυνατόν να οριστεί η κλίση της συνάρτησης κόστους, δηλαδή η παράγωγός της, και άρα να μπορεί μέσω της backpropagation τεχνικής να εκπαιδευτεί το δίκτυο[4].

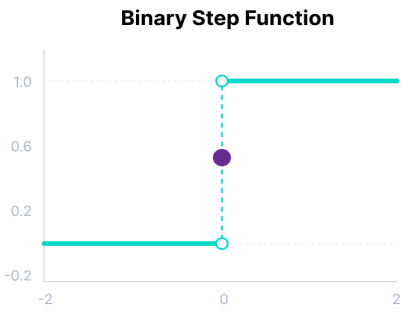


Σχήμα 2.19: *Σχηματική αναπαράσταση νευρώνα 3 εισόδων x με βάρη w και 1 εξόδου, όπου εκτός από την γραμμική συνάρτηση καθορισμού του x υπάρχει μια επιπλέον συνάρτηση ενεργοποίησης του νευρώνα f(x).*

[4] *Κατά τον Schmidhuber (2014), η πρώτη αναφορά στην BP τεχνική έγινε από τον Werbos το 1981 στην εργασία του "Applications of advances in nonlinear sensitivity analysis", ενώ κατά τους Russell & Norvig στο βιβλίο τους "Artificial Intelligence: A Modern Approach", 1995, η χρήση της ιδέας μιας σιγμοειδούς συνάρτησης ανήκει στους Bryson & Ho στο βιβλίο "Applied Optimal Control", το 1969.*

Στα υποκεφάλαια 2.2.2.2.1 - 2.2.2.2.5 παρουσιάζονται μερικές ευρέως χρησιμοποιούμενες συναρτήσεις ενεργοποίησης.

2.2.2.2.1: Binary Step

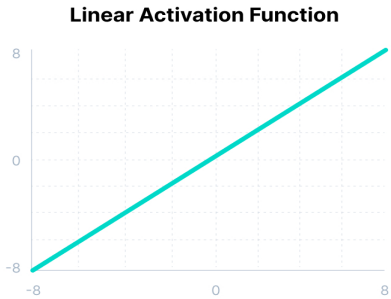


Σχήμα 2.20: *Γραφική αναπαράσταση της συνάρτησης ενεργοποίησης δυαδικού βήματος (binary step). Στον οριζόντιο άξονα (x) οι τιμές εισόδου και στον κάθετο άξονα (y) οι τιμές εξόδου.*

Πηγή: https://www.v7labs.com/blog/neural-networks-activation-functions

Ο μαθηματικός κανόνας (2.3) αποτελεί την έκφραση ενσωμάτωσης των εισόδων σε κάθε νευρώνα, μέσω της οποίας παράγεται μια έξοδος binary μορφής (0 ή 1) (σχήμα 2.20), που αποτελεί είσοδο για έναν ή περισσότερους νευρώνες. Η συνάρτηση κόστους (παράδειγμα η MSE (2.4)) υπολογίζει τη διαφορά μεταξύ πραγματικής και επιθυμητής εξόδου, η παράγωγος της οποίας ορίζει την κλίση καθόδου. Όμως, για δυαδικές τιμές, η παράγωγος της συγκεκριμένης συνάρτησης είναι 0, ενώ για x = 0 δεν ορίζεται και άρα, σε τέτοιου είδους δίκτυα, δεν είναι δυνατόν να εφαρμοστεί η τεχνική της οπισδοδιάδοσης (backpropagation).

2.2.2.2.2: Linear Activation

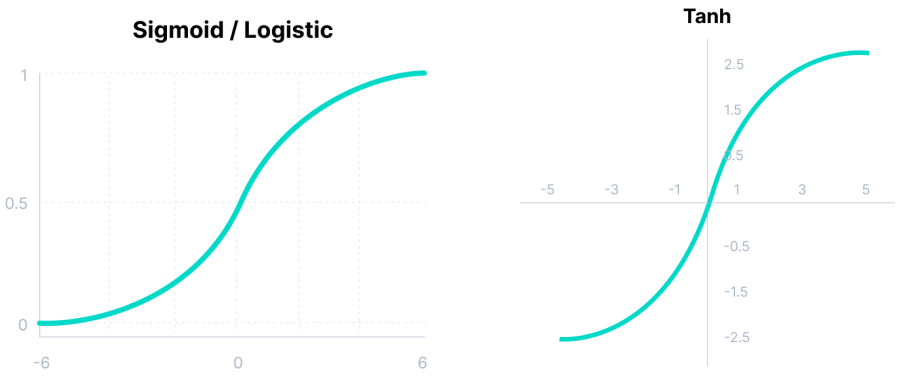


Σχήμα 2.21: *Γραφική αναπαράσταση της γραμμικής συνάρτησης ενεργοποίησης. Στον οριζόντιο άξονα (x) οι τιμές εισόδου και στον κάθετο άξονα (y) οι τιμές εξόδου.*

Πηγή: https://www.v7labs.com/blog/neural-networks-activation-functions

Στην περίπτωση όπου η έξοδος κάθε νευρώνα είναι μόνο το άθροισμα των εισόδων πολλαπλασιασμένο με κάποιο βάρος και με την προσθήκη ενός bias, τότε η συνάρτηση ενεργοποίησης εκφράζεται από τον γενικό τύπο $f(x) = x$, της οποίας η παράγωγος είναι ένας σταθερός αριθμός που δεν έχει κάποια σχέση με τις τιμές εισόδου. Κατά συνέπεια, δεν είναι επίσης εφικτό να εκπαιδευτεί ένα τέτοιο δίκτυο με την backpropagation τεχνική. Επιπροσθέτως, η ύπαρξη επιπέδων χάνει το νόημά της, διότι το τελικό αποτέλεσμα είναι απλώς το άθροισμα των παραμέτρων όλων των επιμέρους συναρτήσεων.

2.2.2.2.3: Sigmoid (logistic, tanh)



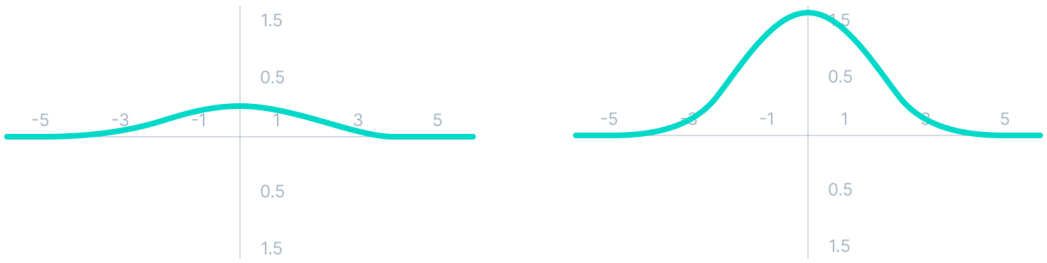
Σχήμα 2.22: *Γραφική αναπαράσταση δύο σιγμοειδών συναρτήσεων ενεργοποίησης (λογιστική - αριστερά και συνάρτηση εφαπτομένης - δεξιά). Στον οριζόντιο άξονα (x) οι τιμές εισόδου και στον κάθετο άξονα (y) οι τιμές εξόδου.*

Πηγή: https://www.v7labs.com/blog/neural-networks-activation-functions

Η βασική σιγμοειδής συνάρτηση ορίζεται ως: $f(x) = \frac{1}{1+e^{-x}}$ με τη γραφική παράστασή της να απεικονίζεται στο σχήμα 2.22, ενώ η παράγωγός της είναι η $f'(x) = f(x)(1 - f(x))$.Έχει ως τιμές εξόδου το διάστημα [0, 1], κάτι που, για τον υπολογισμό πιθανότητας, αποτελεί πλεονέκτημα. Από την άλλη πλευρά, η **tanh** ορίζεται ως: $f(x) = \frac{(e^{x}-e^{-x})}{(e^{x}+e^{-x})}$, ενώ η παράγωγός της είναι η $f'(x) = 1 - f(x)^2$, οι τιμές εξόδου της οποίας είναι το διάστημα [-1, 1]. Το πλεονέκτημά της είναι το γεγονός ότι η κεντρική τιμή της κυμαίνεται γύρω από το 0, ενώ οι έντονα θετικές τιμές απεικονίζονται κοντά στο 1 και οι έντονα αρνητικές κοντά στο -1. Αυτό το εύρος τιμών την καθιστά κατάλληλη για προβλήματα κατηγοριοποίησης.

Το γεγονός ότι η παράγωγος των σιγμοειδών συναρτήσεων είναι κάποιος αριθμός >0 τις καθιστά κατάλληλες για την εκπαίδευση των νευρωνικών δικτύων και γι αυτόν τον λόγο χρησιμοποιούνται κυρίως στα ενδιάμεσα επίπεδα των νευρώνων. Διαπιστώνεται όμως ότι, για

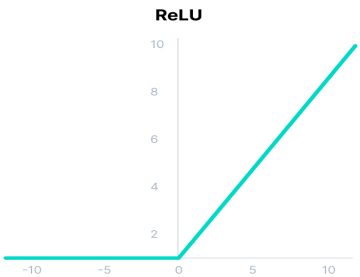
βαθιά/πολύ βαθιά νευρωνικά δίκτυα, οι συγκεκριμένες συναρτήσεις ενεργοποίησης, μπορεί να ευθύνονται για το πρόβλημα της "**απόσβεσης των κλίσεων**" (Vanishing Gradients Problem). Ο όρος αναφέρεται στην σταδιακή απομείωση του σφάλματος που εμφανίζεται κατά την εφαρμογή της μεθόδου της SGD, καθώς κινείται αντίστροφα από την έξοδο του δικτύου προς την είσοδο, επιχειρώντας να διορθώσει τα βάρη των συνδέσεων κάθε νευρώνα, για την συνολική βελτίωση της συνάρτησης κόστους. Πολλές φορές το σφάλμα είναι τόσο μικρό τη στιγμή που προσεγγίζουμε τα αρχικά επίπεδα του δικτύου, έχοντας σαν αποτέλεσμα την πολύ μικρή επίδραση (Brownlee, 2019). Με άλλα λόγια, η πληροφορία για την εμφάνιση του σφάλματος σταδιακά χάνεται και το δίκτυο δεν μπορεί να εκπαιδευτεί περαιτέρω. Το πρόβλημα μπορεί να εντοπιστεί στις τιμές της κλίσης των σιγμοειδών συναρτήσεων που εκφράζονται μέσω της παραγώγου (σχήμα 2.23). Για τιμές εισόδου μικρότερες του -3 και μεγαλύτερες του 3 η κλίση τους είναι σχεδόν μηδενική, ενώ στη λογιστική η μέγιστη τιμή της κλίσης είναι, $f'(0) = 0{,}25$, γεγονός που μεγιστοποιεί το πρόβλημα.



Σχήμα 2.23: *Γραφική αναπαράσταση της κλίσης (παράγωγος) της σιγμοειδούς συνάρτησης (αριστερά) και της εφαπτομενικής (δεξιά) για διάφορες τιμές του x. Διαπιστώνεται ότι για τιμές < -3 και > 3 η κλίση προσεγγίζει το 0.*

Πηγή: https://www.v7labs.com/blog/neural-networks-activation-functions

2.2.2.2.4: Συναρτήσεις ReLU (Leaky, Parametric, ELU)



Σχήμα 2.24: *Γραφική αναπαράσταση της συνάρτησης ReLU. Στον οριζόντιο άξονα (x) οι τιμές εισόδου και στον κάθετο άξονα (y) οι τιμές εξόδου.*

Πηγή: https://www.v7labs.com/blog/neural-networks-activation-functions

Όπως προδίδει το όνομά της (Rectified Linear Unit), πρόκειται για μια παραλλαγή της γραμμικής συνάρτησης αλλά με τη διαφορά ότι:

$$\text{Εάν: } x < 0, \text{ τότε: } f(x) = 0$$
$$\text{Εάν: } x \geq 0, \text{ τότε: } f(x) = x\ .$$

Αυτό ορίζει αντιστοίχως και την παράγωγο:

$$\text{Εάν: } x < 0, \text{ τότε: } f'(x) = 0$$
$$\text{Εάν: } x \geq 0, \text{ τότε: } f'(x) = 1$$

Το γεγονός ότι η παράγωγος (κλίση) εξαρτάται από την τιμή του x προσδίδει την ικανότητα εκπαίδευσης του δικτύου έναντι της linear. Σε σχέση με οποιαδήποτε άλλη συνάρτηση ενεργοποίησης, η relu κοστίζει πολύ λιγότερο υπολογιστικά, καθώς μόνο ένας συγκεκριμένος αριθμός νευρώνων τελικά ενεργοποιείται (εφ' όσον για αρνητικές τιμές η τιμή εξόδου είναι πάντα 0). Αυτό ακριβώς το χαρακτηριστικό όμως αποτελεί ταυτόχρονα και την κύρια αδυναμία της διότι τελικά οι αρνητικές τιμές δεν λαμβάνονται υπόψιν και ένας συγκεκριμένος αριθμός νευρώνων δεν ενεργοποιείται ποτέ, κάτι που αναφέρεται και ως "*dying relu problem*". Για να ξεπεραστεί αυτό το πρόβλημα, μπορούν να εφαρμοστούν κάποιες παραλλαγές της, οι οποίες όμως εμφανίζουν άλλα μειονεκτήματα.

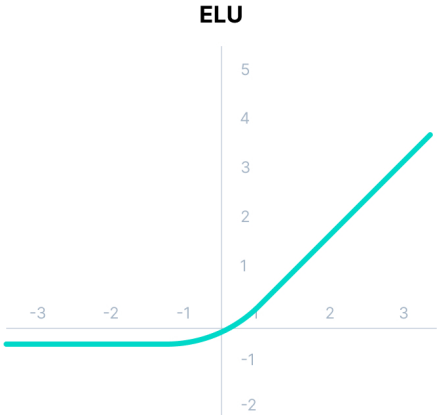
Για παράδειγμα η **Leaky ReLU** όπου:

$$\text{Εάν: } x \leq 0, \text{ τότε: } f(x) = 0.1 \cdot x$$
$$\text{Εάν: } x > 0, \text{ τότε: } f(x) = x\ ,$$

προσδίδει μια μικρή κλίση για μηδενικές ή αρνητικές τιμές ξεπερνώντας το "*dying relu*" πρόβλημα, όμως η μικρή κλίση της μπορεί να καθυστερήσει την εκπαίδευση του μοντέλου, ενώ ταυτόχρονα είναι υπολογιστικά πολύ πιο κοστοβόρα. Στην περίπτωση που ο μικρός αριθμός (≈ 0.1) αντικατασταθεί με μια παράμετρο 'α', τότε προκύπτει η **parametric relu**, η οποία μπορεί να επανακαθορίζεται κατά τη διαδικασία της εκπαίδευσης, ώστε να λάβει τη βέλτιστη τιμή. Από την άλλη πλευρά, η **ELU** (Exponential Linear Units) είναι μια εναλλακτική προσέγγιση, όπου η παράμετρος 'α' εισάγεται σε μια εκθετική συνάρτηση έτσι, ώστε σταδιακά και για πολύ αρνητικές τιμές η τιμή εξόδου να σταθεροποιείται σε έναν αριθμό.

Πιο συγκεκριμένα:

$$\text{Εάν: } x \geq 0 \text{ τότε, } f(x) = x$$
$$\text{Εάν: } x < 0 \text{ τότε, } f(x) = a \cdot (e^x - 1)$$



Σχήμα 2.25: *Γραφική αναπαράσταση της συνάρτησης ReLU. Στον οριζόντιο άξονα (x) οι τιμές εισόδου και στον κάθετο άξονα (y) οι τιμές εξόδου.*

Πηγή: https://www.v7labs.com/blog/neural-networks-activation-functions

Η εισαγωγή της εκθετικής συνάρτησης για τις αρνητικές τιμές του x εξομαλύνει τις τιμές εξόδου σε ένα μικρό εύρος (κάτι που αποτελεί το βασικό πλεονέκτημα της tanh), ενώ στις θετικές η έξοδος είναι απλώς γραμμική, ώστε να μην εμφανίζεται το πρόβλημα της απόσβεσης των κλίσεων. Αν και η **ELU** αποτελεί μια ισχυρή εναλλακτική έναντι της βασικής ReLU, το μειονέκτημά της είναι ότι αυξάνει αρκετά τον χρόνο υπολογισμού, ενώ η τιμή του παράγοντα 'α' είναι εξ'αρχής καθορισμένη. Πολλές ακόμη παραλλαγές μπορούν να προκύψουν, εάν στη θέση του δεύτερου σκέλους της εξίσωσης (για x < 0) προστεθούν διάφορες ακόμη συναρτήσεις, αναλόγως των ιδιαίτερων αναγκών της εκπαίδευσης.

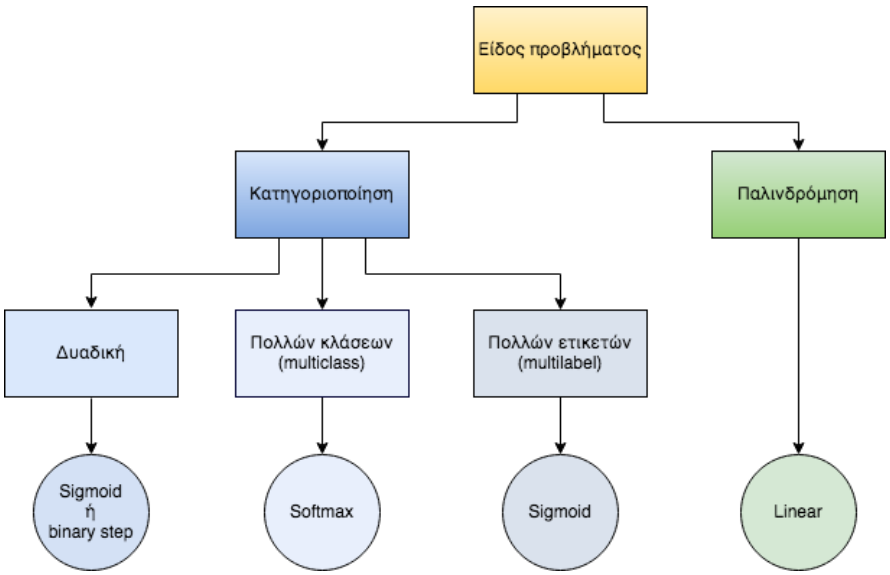
2.2.2.2.5: Softmax

Πολύ συχνά, ιδίως σε προβλήματα κατηγοριοποίησης, το ζητούμενο είναι η πιθανότητα εμφάνισης ενός γεγονότος. Για παράδειγμα, εάν αυτό που απαιτείται είναι η αναγνώριση κάποιου αντικειμένου ανάμεσα σε πολλά άλλα σε μια εικόνα, τότε η επιλογή κάποιας σιγμοειδούς συνάρτησης ενεργοποίησης είναι ενδεχομένως η καταλληλότερη για τα ενδιάμεσα επίπεδα των νευρώνων. Για κάθε ένα από τα αντικείμενα μπορεί στο τέλος να δοθεί μια πιθανότητα (π.χ. 0.5 για το Α, 0.4 για το Β, 0.7 για το Γ, 0.8 για το Δ), ωστόσο το άθροισμά

τους ξεπερνάει το 100%. Η softmax ορίζει τη **σχετική πιθανότητα** για κάθε μία από τις τιμές εξόδου, ώστε το άθροισμά τους να είναι πάντα 1. Στην απλή μαθηματική της μορφή μπορεί να περιγραφεί από τον ακόλουθο τύπο για κάθε έξοδο x_i και j το σύνολο των x (στο παραπάνω παράδειγμα είναι 4):

$$softmax(x_i) = \frac{e^{x_i}}{\sum_j \quad e^{x_j}}$$

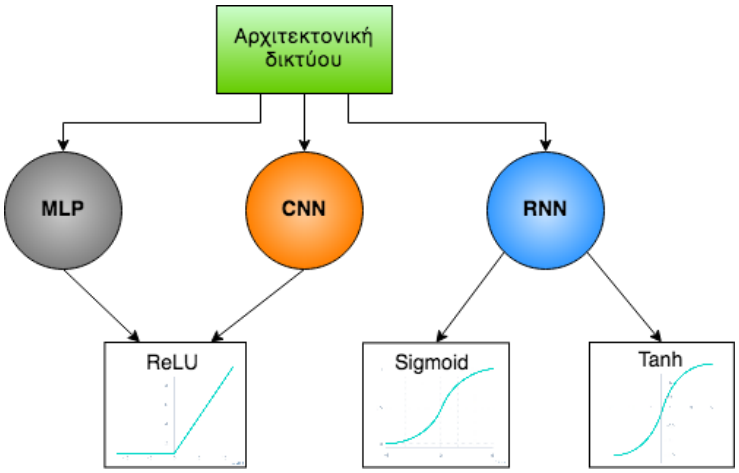
Τα σύγχρονα, πολλών επιπέδων νευρωνικά δίκτυα χρησιμοποιούν κάποια μη-γραμμική συνάρτηση ενεργοποίησης στα ενδιάμεσα επίπεδα, η οποία συνήθως είναι ίδια για όλους τους κόμβους / νευρώνες, ενώ στο επίπεδο εξόδου μπορεί να γίνεται χρήση binary step ή linear activation. Το είδος των συναρτήσεων που θα χρησιμοποιηθούν σχετίζεται πάντα με το είδος του προβλήματος προς επίλυση και η σωστή επιλογή τους αποτελεί παράγοντα καθοριστικής σημασίας για την καλή εκπαίδευση του δικτύου και την βελτίωση των αποτελεσμάτων. Σχετικά με το επίπεδο εξόδου, η συνηθέστερη χρήση διαφόρων συναρτήσεων ενεργοποίησης για συγκεκριμένους τύπους προβλημάτων, απεικονίζεται στο σχήμα 2.26:



Σχήμα 2.26: *Η συνηθέστερη χρήση διαφόρων συναρτήσεων ενεργοποίησης στο επίπεδο εξόδου ενός νευρωνικού δικτύου, ανάλογα με το είδος του προβλήματος*

Όσον αφορά τα ενδιάμεσα επίπεδα, οι συναρτήσεις ενεργοποίησης συσχετίζονται με την αρχιτεκτονική του δικτύου. Στις περιπτώσεις MLP γίνεται χρήση της ReLU (ή κάποιας παραλλαγής της), καθώς είναι συχνά επιθυμητή η γραμμική έξοδος κυρίως για θετικές τιμές,

αναλόγως και στα δίκτυα convolutional αρχιτεκτονικής. Στις περιπτώσεις RNN δικτύων είναι προτιμότερη η χρήση κάποιας σιγμοειδούς συνάρτησης. Στο σχήμα 2.27 παρουσιάζεται γραφικά η χρήση των συναρτήσεων ενεργοποίησης σε σχέση με μερικούς βασικούς τύπους της αρχιτεκτονικής των δικτύων.



Σχήμα 2.27: *Η συνηθέστερη χρήση διαφόρων συναρτήσεων ενεργοποίησης στα ενδιάμεσα επίπεδα ενός νευρωνικού δικτύου αναλόγως της αρχιτεκτονικής του.*

2.2.3: Αρχιτεκτονική νευρωνικών δικτύων

Όπως περιγράφηκε και στο υποκεφάλαιο 2.2.1, τα "perceptrons" είναι η πρώτη κατηγορία νευρώνων που δημιουργήθηκε. Στην ουσία τους αποτελούν έναν συνδυασμό γραμμικών συναρτήσεων έχοντας τη δυνατότητα να υλοποιήσουν απλές εφαρμογές εκμάθησης. Καθώς όμως οι ανάγκες έγιναν πιο σύνθετες και η έρευνα προχώρησε, η συγκεκριμένη αρχιτεκτονική εξελίχθηκε και εξειδικεύτηκε κατά περίπτωση.

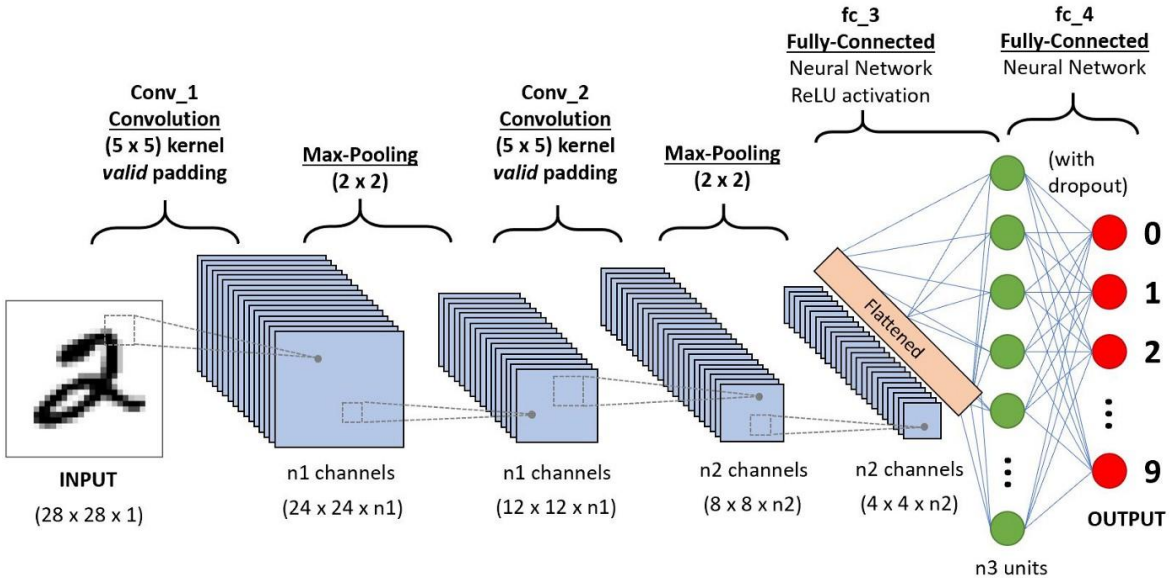
2.2.3.1: Convolutional Neural Network (CNN)

Η ονομασία φαίνεται πως προέρχεται από την εργασία των LeCun et al. (1998), όπου επιχειρείται η επίλυση του προβλήματος της αναγνώρισης χειρόγραφων γραμμάτων. Εκτός των άλλων, περιγράφεται η αρχιτεκτονική LeNet-5, η οποία προκύπτει σαν αποτέλεσμα μακροχρόνιας έρευνας ήδη από το 1988-9. Φαίνεται όμως πως το συγκεκριμένο ζητούμενο είχε απασχολήσει τους ερευνητές και τα προηγούμενα χρόνια. Για παράδειγμα, στην εργασία των Zhang et al. (1988) επιχειρείται η αναγνώριση αλφαβήτου, ενώ η εργασία των Homma et al. (1988) περιγράφει τη διαδικασία της φωνητικής αναγνώρισης, όπου γίνεται για πρώτη φορά

η χρήση της λέξης “convolutional” για την περιγραφή των συνδέσεων μεταξύ των νευρώνων. Ήδη όμως από τον Fukushima το 1980 μερικά από τα βασικά χαρακτηριστικά των CNN’s, όπως η εξαγωγή χαρακτηριστικών (feature extraction), η ομαδοποίηση επιπέδων (pooling layers) αλλά και η “convolution” τεχνική, είχαν ήδη αναπτυχθεί και δοκιμαστεί σε ένα δίκτυο-πρόγονο του CNN, το οποίο ονομάστηκε “neocognitron”. Σήμερα η CNN αρχιτεκτονική χρησιμοποιείται ευρέως, καθώς θεωρείται κατάλληλη εκτός των άλλων για την αναγνώριση και κατηγοριοποίηση εικόνων αλλά και για την αναγνώριση φωνητικών και γραπτών κειμένων (Natural Language Processing).

Βασική δομή

Το κύριο χαρακτηριστικό τους είναι ότι αποτελούνται από τρία βασικά επίπεδα, η σειρά των οποίων έχει ως εξής: 1. *Convolutional Layer*, 2. *Pooling Layer*, 3. *Fully-Connected Layer*. Συνήθως υπάρχουν πολλά διαδοχικά Convolutional - Pooling επίπεδα στη σειρά (σχήμα 2.28)



Σχήμα 2.28: *Σχηματική αναπαράσταση της αρχιτεκτονικής ενός CNN δικτύου. Διακρίνονται τα διαδοχικά Convolution και Max-Pooling επίπεδα μέχρι να καταλήξει σε ένα πλήρως διασυνδεδεμένο τελικό δίκτυο (Fully Connected Neural Network - FCNN) που αποτελείται από την έξοδο των προηγούμενων επιπέδων και την τελική.*

Πηγή: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Η ύπαρξη των convolutional / pooling επιπέδων έχει ως στόχο να μειώσει τον όγκο των υπό επεξεργασία δεδομένων, εξάγοντας ταυτόχρονα τα σημαντικότερα χαρακτηριστικά τους. Έχοντας πλέον αναλυθεί σε απλούστερες μορφές, τροφοδοτούν το τελικό FCNN, το οποίο με βάση αυτήν την πληροφορία καλείται να πραγματοποιήσει την κατηγοριοποίηση.

Πιο αναλυτικά, στο **Convolutional** επίπεδο εφαρμόζεται η μέθοδος του "filtering", μέσω της οποίας γίνεται η εξαγωγή γενικότερων / σημαντικότερων χαρακτηριστικών. Για παράδειγμα, αν θεωρηθεί μια τυχαία εικόνα σε τόνους του γκρι 3 x 3 pixels, θα μπορούσε να δημιουργηθεί ένας αντίστοιχος υποθετικός πίνακας [1] με τιμές από 0-255. Προκειμένου να εφαρμοστεί το "φιλτράρισμα", υποθέτουμε πίνακα [2] δυαδικών τιμών έστω 2 x 2 pixels. Σαν αποτέλεσμα πολλαπλασιασμού κάθε ομάδας 2 x 2 στοιχείων του πίνακα [1] με τον πίνακα [2] προκύπτει ένας νέος πίνακας διαστάσεων 2 x 2, όπως παρακάτω:

12	55	28
2	102	36
100	18	0

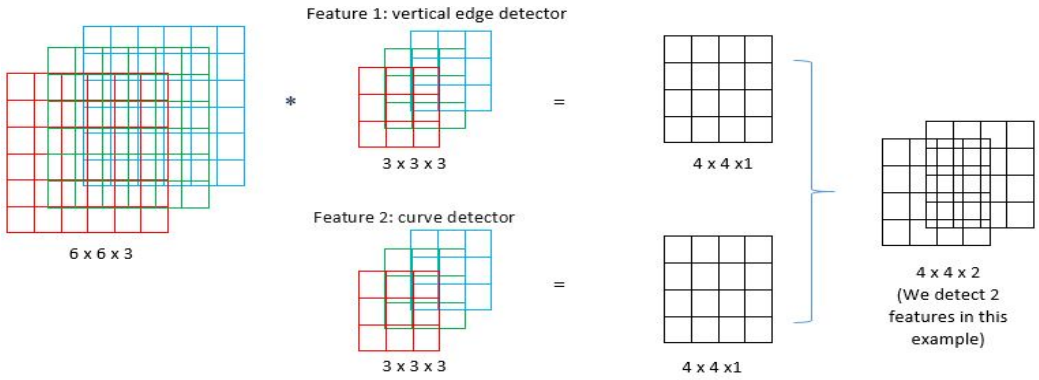
X

1	0
1	1

=

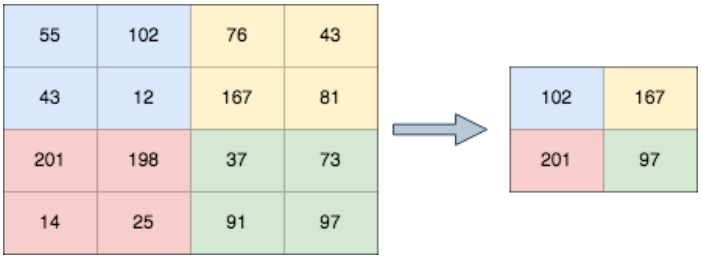
116 (12*1+55*0+2*1+102*1)	**193** (55*1+28*0+102*1+36*1)
120 (2*1+102*0+100*1+18*1)	**120** (102*1+36*0+18*1+0*1)

Στην πραγματικότητα τα δεδομένα εισόδου, τα φίλτρα αλλά και ο τρόπος που πραγματοποιείται η διαδικασία είναι πιο σύνθετος διότι οι εικόνες μπορεί να είναι 3 επιπέδων (RGB), τα φίλτρα μεγαλύτερων διαστάσεων για κάθε ένα από αυτά τα επίπεδα, ενώ μπορεί να είναι περισσότερα του ενός για την εξαγωγή διαφορετικών γνωρισμάτων (σχήμα 2.29).



Σχήμα 2.29: *Παράδειγμα CNN δικτύου με πολλαπλά φίλτρα.*
Πηγή: https://www.analyticsvidhya.com/blog/2022/01/convolutional-neural-network-an-overview/

Στη συνέχεια, κατά τη διαδικασία του **pooling**, στόχος είναι η ανίχνευση του σημαντικότερου στοιχείου κάθε ενός από τα features που προέκυψαν στο προηγούμενο στάδιο είτε μέσω της εύρεσης της μέγιστης τιμής (max pooling) είτε με την εύρεση του μέσου όρου των τιμών. Συνεπώς, χρησιμοποιώντας το παράδειγμα που αναπτύχθηκε στην προηγούμενη παράγραφο, για max pooling θα προέκυπτε η τιμή 193 ενώ για average pooling, η τιμή 137,25. Εάν στον τελικό πίνακα [m x n], ισχύει (m | n) > 1, τότε εφαρμόζεται εκ νέου η διαδικασία του convolution και στη συνέχεια του pooling, μέχρι να προκύψει μονοδιάστατος πίνακας [m x n], (m | n) = 1, οι τιμές του οποίου θα χρησιμεύσουν ως είσοδος για τον τελικό FCNN. Στο σχήμα 2.30, ένα παράδειγμα εφαρμογής Max-Pooling σε έναν πίνακα 4x4 που οδηγεί σε ένα νέο πίνακα 2x2.



Σχήμα 2.30: *Η διαδικασία του Max-Pooling σε ένα CNN δίκτυο, που οδηγεί από ένα πίνακα διαστάσεων 4x4 σε πίνακα διαστάσεων 2x2.*

Είναι εμφανές ότι στο convolutional επίπεδο του δικτύου τα φίλτρα επιτελούν το έργο των νευρώνων, δεχόμενα ως είσοδο έναν πολυδιάστατο πίνακα και παράγοντας ως έξοδο έναν άλλο πίνακα, μέσω μιας συνάρτησης ενεργοποίησης. Κατ' αντιστοιχία στο pooling επίπεδο, η συναρτήσεις max ή average pooling μπορούν επίσης να θεωρηθούν ως νευρώνες με αντίστοιχες συναρτήσεις ενεργοποίησης.

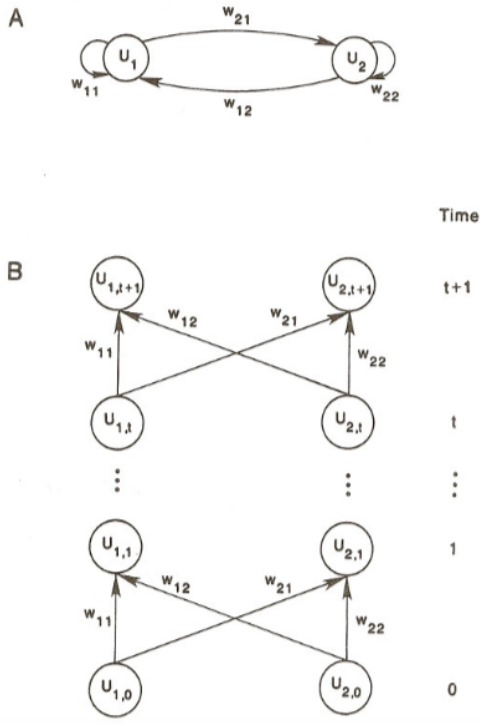
2.2.3.2: Recurrent Neural Networks (RNNs)

Όπως ήδη αναφέρθηκε, η ιδέα των ANNs ξεκίνησε μέσα από την παρατήρηση και μελέτη της λειτουργίας των βιολογικών νευρώνων. Οι πρώτες έρευνες σε αυτόν τον τομέα, ήδη από το τέλος της δεκαετίας του '50, εστιάστηκαν στην ανάλυση και κατηγοριοποίηση εικόνων (Géron, 2017). Ωστόσο, η ανθρώπινη νοημοσύνη (όπως και άλλων θηλαστικών) δεν αντιλαμβάνεται μέσω των αισθητηρίων οργάνων το περιβάλλον με στατικό τρόπο αλλά ως μια αλληλουχία πληροφοριών από όλα τα αισθητήρια όργανα. Ο συσχετισμός όλων των αυτών των αλληλουχιών δημιουργεί την έννοια της αντίληψης και κατανόησης και μέσω αυτής της

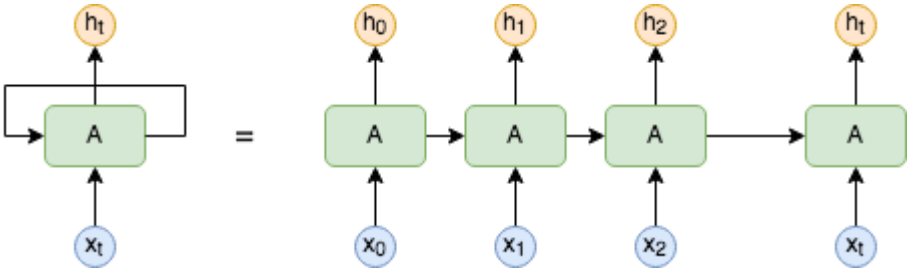
διαδικασίας μπορεί στη συνέχεια ο εγκέφαλος να πραγματοποιήσει μια προσομοίωση (πρόβλεψη) του μέλλοντος. Είναι λοιπόν φανερό πως σε όλη τη διαδικασία υπεισέρχεται η έννοια του χρόνου και πιο συγκεκριμένα της χρονικής μεταβολής μιας ή πολλών παρατηρούμενων τιμών και ο συσχετισμός τους. Τα RNNs προσπαθούν να μιμηθούν αυτήν την διαδικασία επιτρέποντας στην πληροφορία να παραμείνει (Olah C., 2015), με την πρώτη αναφορά σε αυτήν την αρχιτεκτονική να γίνεται στην εργασία των Rumelhart et al., το 1986.

Βασική δομή

Ο όρος "recurrent" ("επανάληψη") προέρχεται από το ότι η έξοδος κάθε νευρώνα επιστρέφει στην είσοδο για t φορές/χρονικές στιγμές. Όπως αναφέρουν οι Rumelhart et al, το ίδιο αποτέλεσμα θα μπορούσε να επιτευχθεί, εάν απλώς σε ένα πλήρως διασυνδεδεμένο ANN/MLP επαναλάβουμε την ίδια δομή πολλές φορές, ώστε να προσομοιωθεί η χρονική εξέλιξη. Στο σχήμα 2.31 παρουσιάζεται η σύγκριση μεταξύ των δύο δικτύων, ενώ η αναπαράσταση της ανάπτυξης ενός νευρώνα στον αντίστοιχο ενός τυπικού ANN παρουσιάζεται στο σχήμα 2.32.



Σχήμα 2.31: *Α. Η σύνδεση "recurrent" μορφής μεταξύ δύο νευρώνων όπου τα βάρη ανανεώνονται διαρκώς για κάθε χρονική στιγμή. Β. Η αντίστοιχη υλοποίηση με διαδοχικούς διασυνδεδεμένους νευρώνες, τα ζεύγη των οποίων αντιστοιχούν σε κάθε χρονική στιγμή. Πηγή: Rumelhart et al., 1986*



Σχήμα 2.32: *Σχηματική αναπαράσταση του χρονικού αναπτύγματος ενός recurring νευρώνα A για t χρονικές στιγμές με είσοδο X και έξοδο h.*

Η μεταβλητή t υποδηλώνει τον αριθμό των επαναλήψεων που αντιστοιχεί στον αριθμό των νευρώνων σε ένα κλασικό ANN δίκτυο και η μεταβλητή x για κάθε χρονική στιγμή t αντιπροσωπεύει τις διαδοχικές εισόδους μιας χρονοσειράς δεδομένων. Αν και δεν υπάρχει κάποιος σαφώς διατυπωμένος κανόνας για το πόσο μεγάλος πρέπει να είναι ο αριθμός των επαναλήψεων (αναφέρονται και ως hidden units), το λογικό συμπέρασμα είναι πως θα πρέπει να είναι τουλάχιστον ίσος με το συνολικό μέγεθος της υπό μελέτης σειράς / ή του αθροίσματος των πιθανών συνδυασμών της που διερευνώνται (περισσότερα αναφέρονται στο 3ο κεφάλαιο). Ο τρόπος μετάδοσης της πληροφορίας για κάθε επανάληψη είναι παρόμοιος με τα υπόλοιπα νευρωνικά δίκτυα, με την συνάρτηση να συμπεριλαμβάνει την είσοδο, κάποιο βάρος (w) και μια απόκλιση (b), τροποποιημένη όμως ως εξής (χρησιμοποιώντας τον συμβολισμό του σχήματος 2.32):

$$h_t = x_t \cdot w_t \ + h_{(t-1)} \cdot w_h \ + b \qquad (2.19)$$

Επισημαίνεται η ύπαρξη δύο εισόδων, με την πρώτη να αποτελεί τη νέα πληροφορία που εισέρχεται στο δίκτυο (x_t) και τη δεύτερη την πληροφορία από την προηγούμενη κατάστασή του $(h_{(t-1)})$. Όπως ακριβώς συμβαίνει και στα υπόλοιπα δίκτυα, η εκπαίδευση των RNN έχει ως στόχο τον σωστό καθορισμό των βαρών μέσω της ελαχιστοποίησης της συνάρτησης κόστους. Επειδή όμως αυτό λαμβάνει χώρα για διαδοχικές χρονικές στιγμές, η μέθοδος ονομάζεται "**Backpropagation Through Time**" (BPTT).

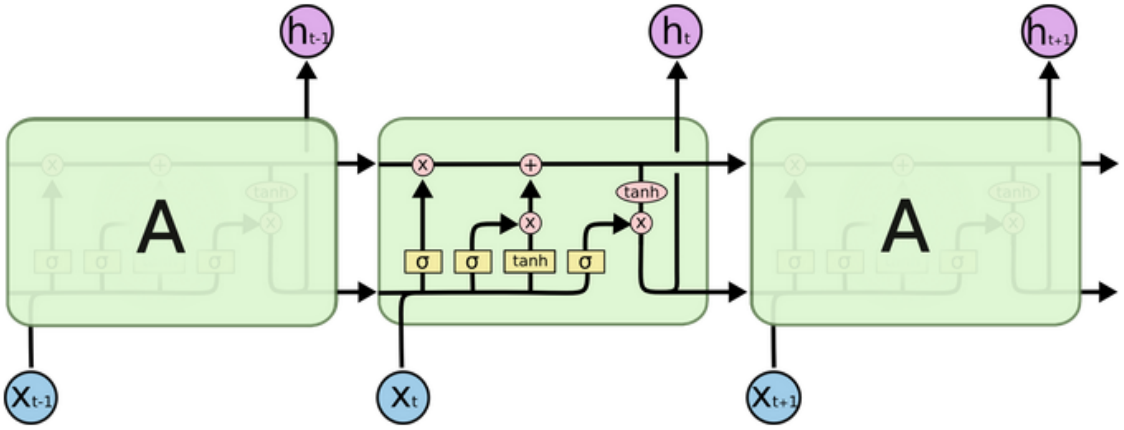
Το πρόβλημα της απόσβεσης των κλίσεων (Vanishing Gradient Descent) είναι κοινό για την εκπαίδευση όλων των δικτύων πάνω σε μεγάλο όγκο δεδομένων, ειδικότερα μάλιστα όσον αφορά τα RNNs, των οποίων το βάθος λόγω της αρχιτεκτονικής τους είναι στην πραγματικότητα αρκετά μεγάλο. Επιπροσθέτως, δεδομένα όπως οι λέξεις ενός κειμένου, η

πορεία της θερμοκρασίας ή της ατμοσφαιρικής πίεσης, η εξέλιξη της τιμής μιας μετοχής στο χρόνο κ.ά. παρουσιάζουν συχνά συσχετίσεις σε βάθος χρόνου και με τρόπο συχνά ακανόνιστο. Καθώς ένας RNN νευρώνας έχει την τάση να "θυμάται" περισσότερο την πιο πρόσφατη πληροφορία, τέτοιου είδους χαρακτηριστικά είναι αδύνατον να εξαχθούν. Το 1997 οι Hochreiter & Schmidhuber παρουσίασαν μια επέκταση των RNN δικτύων τα οποία ονομάστηκαν LSTMs (Long Short-Term Memory), λόγω της ικανότητας διατήρησης της πληροφορίας για αρκετά μεγάλο πλήθος χρονικών βημάτων.

2.2.3.4: LSTM

Τα LSTM δίκτυα αποτελούν στην ουσία μια εξειδίκευση των RNN δικτύων με την προσθήκη μερικών ακόμη χαρακτηριστικών όπως η "πύλη λήθης", η "πύλη εισόδου", η "πύλη εξόδου" αλλά και μιας παράλληλης ροής πληροφορίας η οποία ονομάζεται "κατάσταση νευρώνα".

Στο σχήμα 2.33 παρουσιάζονται τα δομικά χαρακτηριστικά ενός LSTM νευρώνα.



Σχήμα 2.33: *Σχηματική απεικόνιση ενός LSTM νευρώνα. Πηγή: https://colah.github.io/posts/2015-08-Understanding-LSTMs/*

Ως "κατάσταση νευρώνα" (cell state) ορίζεται η παράλληλη ροή πληροφορίας, η οποία - όπως απεικονίζεται στο σχήμα 2.33 - σημειώνεται στην πάνω πλευρά. Επιπλέον, κάθε είσοδος (X_t) που αφορά στα νέα δεδομένα που εισέρχονται σε κάθε χρονική στιγμή υπόκειται σε διάφορους μετασχηματισμούς που συμβολίζονται με [σ], [tanh], (x), (+) τροποποιώντας την κατάστασή του. Η νέα κατάσταση αλληλεπιδρά με τον κλασικό RNN υπολογιστικό κύκλο, για να δημιουργήσει το τελικό αποτέλεσμα. Στη συνέχεια εξηγούνται επιγραμματικά τα νέα χαρακτηριστικά που έχουν προστεθεί δημιουργώντας τελικά μια νέα μαθηματική έκφραση:

1. **Πύλη Λήθης (Forget Gate):** Δέχεται ως δεδομένα εισόδου το αποτέλεσμα της προηγούμενης κατάστασης του νευρώνα ($h_{(t-1)}$) και τη νέα τιμή (x_t), τα οποία πολλαπλασιάζονται με κάποιο βάρος w και προστίθεται το bias. Το αποτέλεσμα εισάγεται σε μια σιγμοειδή συνάρτηση $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$, όπου εάν $f_t \to 1$, τότε το δεδομένο θεωρείται σημαντικό, ενώ εάν $f_t \to 0$, τότε δεν θα πρέπει να παραμείνει. Η πληροφορία σχετικά με το νέο δεδομένο μεταφέρεται στη συνέχεια στην ροή της κατάστασης του νευρώνα [σημείο (x)].
2. **Πύλη Εισόδου (Input Gate):** Κατ' αναλογία με την Forget Gate, δέχεται επίσης ως δεδομένα εισόδου τα $h_{(t-1)}$ και x_t και "αποφασίζει" για την σημαντικότητά τους μέσω του συνδυασμού μιας tanh συνάρτησης που δημιουργεί ένα διάνυσμα των υπό ανανέωση τιμών και μιας σιγμοειδούς που καθορίζει τον βαθμό στον οποίο αυτό θα συμβεί. Εάν i_t το αποτέλεσμα της σιγμοειδούς και $\underline{C_t}$ το διάνυσμα των τιμών, τότε το τελικό αποτέλεσμα είναι: $i_t \cdot \underline{C_t}$, το οποίο φτάνει στη ροή κατάστασης του νευρώνα. Αν C_{t-1} η προηγούμενη κατάσταση του νευρώνα, τότε η νέα κατάσταση C_t είναι το άθροισμα του αποτελέσματος της Forget και της Input Gate: $C_t = f_t \cdot C_{t-1} + i_t \cdot \underline{C_t}$.
3. **Πύλη Εξόδου (Output Gate):** Η έξοδος ενός RNN νευρώνα αποτελεί απλώς το άθροισμα της προηγούμενης κατάστασής του πολλαπλασιασμένης με κάποιο βάρος και της νέας τιμής εισόδου επίσης πολλαπλασιασμένης με κάποιο άλλο βάρος. Στην περίπτωση του LSTM η συγκεκριμένη έξοδος τροποποιείται μέσα από μια σιγμοειδή συνάρτηση / φίλτρο, όπου γίνεται ο διαχωρισμός σε σημαντικά και λιγότερο σημαντικά γνωρίσματα. Ταυτόχρονα, η τελική κατάσταση του νευρώνα C_t τροποποιείται μέσω μιας tanh συνάρτησης (για να λάβει τιμές μεταξύ -1 και 1). Ο πολλαπλασιασμός (x) της τροποποιημένης τελικής κατάστασης του νευρώνα C_t με την τροποποιημένη έξοδό του o_t διαμορφώνει την τελική έξοδο h_t , άρα:
   $h_t = o_t \cdot tanh(C_t)$
   όπου: $o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$ *και* W_o *το αθροιστικό βάρος στη συνάρτηση εξόδου ενός τυπικού RNN νευρώνα με εισόδους* h_{t-1} *και* x_t *, όπως έχει ήδη παρουσιαστεί και* b_o *το bias.*

Με βάση τα όσα αναλύθηκαν στα σημεία 1, 2 και 3, η τελική έξοδος h_t του LSTM ορίζεται ως εξής:

$$h_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \cdot tanh \cdot \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \cdot C_{t-1} + i_t \cdot \underline{C_t}) \qquad (2.20)$$

Στο πλαίσιο αυτής της εργασίας, μεταξύ άλλων επιλέχθηκε η LSTM αρχιτεκτονική ως η καταλληλότερη για την δημιουργία νευρωνικού δικτύου ικανού, ώστε να ανιχνεύσει και να εκπαιδευτεί σε πολύπλοκες συσχετίσεις, οι οποίες εμφανίζονται σε μεγάλες χρονοσειρές μετεωρολογικών δεδομένων.

Κεφάλαιο 3°: Μεθοδολογία

Στην παρούσα εργασία διερευνάται η δυνατότητα πρόγνωσης των διακυμάνσεων στο βασικό προγνωστικό πεδίο των 500hpa στην περιοχή των Αθηνών με τη χρήση νευρωνικού δικτύου λαμβάνοντας υπόψιν την αντίστοιχη διακύμανση σε άλλα σημεία του πλανήτη.

3.1 Συλλογή Δεδομένων

Η ισοβαρική στάθμη των 500hpa επιλέχθηκε διότι βρίσκεται στο μέσον της ατμόσφαιρας και αποτελεί βασική μεταβλητή για την πρόγνωση του καιρού. Ταυτόχρονα η ακρίβεια της προσομοίωσής της από τα αριθμητικά μοντέλα αξιολογείται συστηματικά ως προς τους δείκτες RMSE, BIAS, Anomaly Correlation Coefficient κ.ά[5]. Γενικότερα σε κάθε περίπτωση μελέτης και ανάλυσης χρονοσειρών, τα στοιχεία της πληρότητας και της μέγιστης δυνατής ακρίβειας είναι προϋπόθεση για την εξαγωγή χρήσιμων συμπερασμάτων. Στην περίπτωση των μετεωρολογικών δεδομένων αυτό δεν είναι πάντα εφικτό, ιδίως όσον αφορά τη μέση και ανώτερη τροπόσφαιρα όπου η μοναδική πηγή καταγραφών είναι οι ραδιοβολήσεις που προέρχονται από τα μετεωρολογικά μπαλόνια. Σε αυτήν την περίπτωση όμως η συλλογή της πληροφορίας δεν λαμβάνει χώρα πάντα σε συγκεκριμένα χρονικά διαστήματα και επιπλέον η χωρική διασπορά τους είναι ακανόνιστη. Από την άλλη πλευρά, παρ' όλο που για το κατώτατο στρώμα της τροπόσφαιρας (2μ ή 10μ πάνω από το έδαφος) υπάρχουν συστηματικές και αξιόπιστες καταγραφές πολλών ετών (πχ θερμοκρασία), τα δεδομένα αυτά παρουσιάζουν έντονη εξάρτηση από το ανάγλυφο, την ώρα της ημέρας αλλά και πολλούς ακόμη παράγοντες με αποτέλεσμα να καθίσταται δύσκολη η κανονικοποίησή τους, εισάγοντας αρκετούς μη συστηματικούς παράγοντες σφαλμάτων που ενδέχεται να επηρεάσουν τη μελέτη.

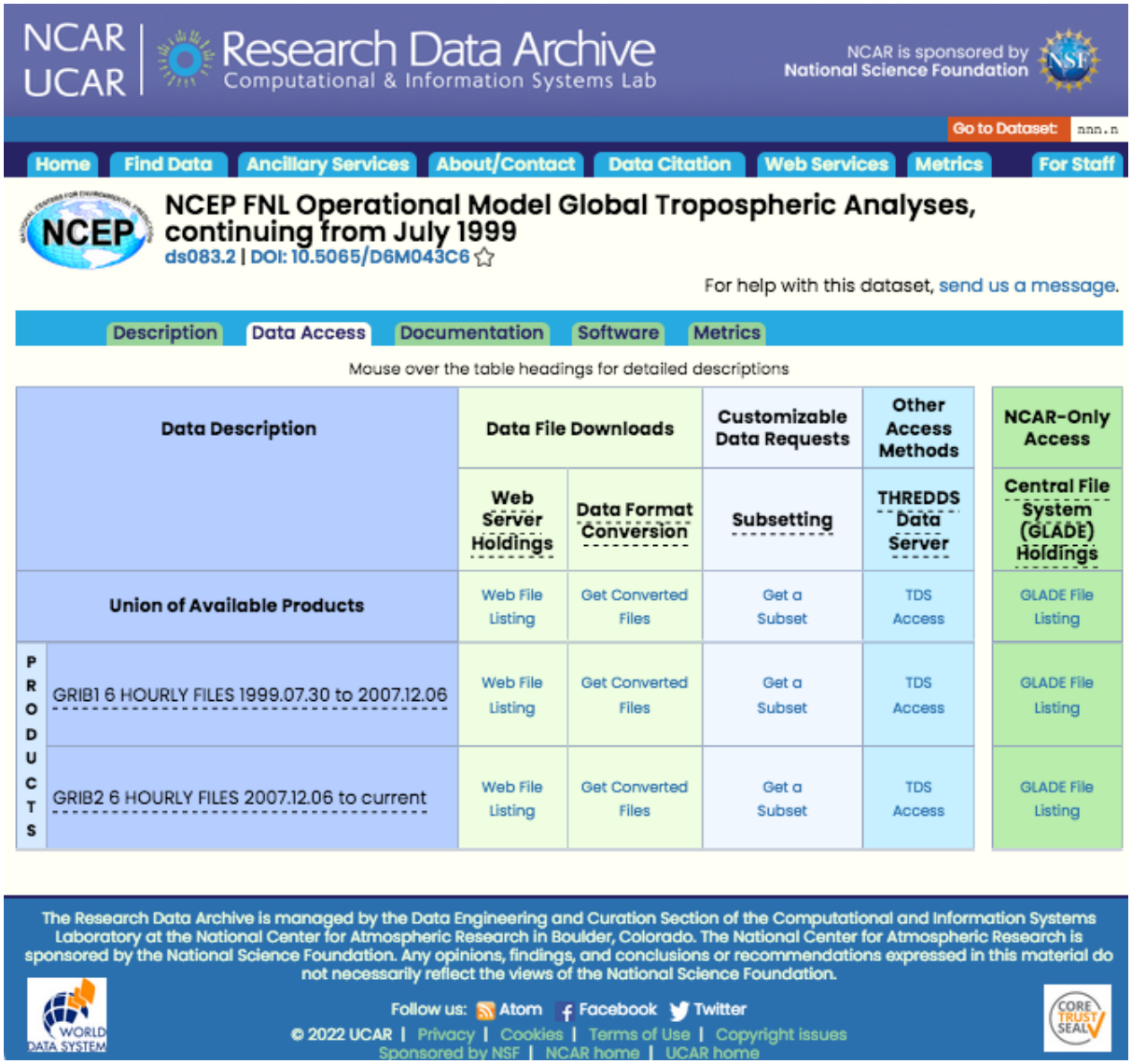
Όπως περιγράφεται στο κεφάλαιο 2, ένα βασικό στοιχείο της διαδικασίας δημιουργίας των ατμοσφαιρικών προσομοιώσεων είναι η ενσωμάτωση όλων των διαθέσιμων παρατηρήσεων (στις οποίες συμπεριλαμβάνονται και οι ραδιοβολήσεις) σε κάθε προγνωστικό κύκλο του μοντέλου, ο οποίος είναι συνήθως κάθε 6 ώρες. Οι παρατηρήσεις προέρχονται από το Παγκόσμιο Σύστημα Τηλεπικοινωνιών (**Global Telecommunications System - GTS**), το *«συντονισμένο παγκόσμιο σύστημα των τηλεπικοινωνιακών εγκαταστάσεων και ρυθμίσεων για*

[5] *Στην κοινή ιστοσελίδα των NOAA/NWS των ΗΠΑ παρουσιάζονται σε καθημερινή βάση τα αποτελέσματα των αξιολογήσεων μερικών γνωστών παγκόσμιων προγνωστικών συστημάτων με τη χρήση στατιστικών δεικτών:* https://www.emc.ncep.noaa.gov/users/verification/global/gfs/ops/

τη γρήγορη συλλογή, ανταλλαγή και διάδοση παρατηρήσεων και επεξεργασμένης πληροφορίας μέσα στο πλαίσιο του World Weather Watch» (WMO, 2019, Basic Documents No.2, Volume I - General Meteorological Standards and Recommended Practices). Αυτά τα δεδομένα κανονικοποιούνται, ώστε να δημιουργηθεί ένα αρχικό πεδίο ίδιας ανάλυσης σε όλα τα ατμοσφαιρικά επίπεδα, το οποίο αποτελεί την πηγή εισόδου για τις ατμοσφαιρικές προσομοιώσεις κάθε 6 ώρες. Συνεπώς τα δεδομένα αυτά μπορούν να θεωρηθούν πλήρη, ενώ η μεγάλη ακρίβειά τους επιβεβαιώνεται και από το γεγονός ότι χρησιμοποιούνται ευρέως ως αρχικά πεδία προσομοίωσης και σε άλλα μοντέλα (πχ περιοχικά), όπως για παράδειγμα το WRF[6].

Μεταξύ διαφόρων προγνωστικών συστήματων, ως "ground truth" δεδομένα θεωρήθηκαν τα αρχικά πεδία του Παγκόσμιου Προγνωστικού Συστήματος (GFS), του Εθνικού Κέντρου Περιβαλλοντικών Προγνώσεων (NCEP) των Η.Π.Α., τα οποία είναι ελεύθερα διαθέσιμα για διάφορες χρονικές περιόδους στην ιστοσελίδα https://rda.ucar.edu/. Προτιμήθηκε η "ds083.2" χρονοσειρά (National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, 2000), η οποία περιλαμβάνει όλα τα analysis πεδία (FNL) από τις 30 Ιουλίου 1999 - 18UTC ανά 6 ώρες έως και την τρέχουσα ημέρα και ανανεώνεται σε καθημερινή βάση (εικόνα 3.1). Εφ' όσον στην παρούσα εργασία το ενδιαφέρον εστιάζεται στην ημερήσια διακύμανση της βασικής μετεωρολογικής παραμέτρου του γεωδυναμικού ύψους στα 500hpa, επιλέχθηκε μόνο ο 00 UTC προγνωστικός κύκλος για κάθε ημέρα. Η οριζόντια ανάλυση τους είναι 1° x 1º, δηλαδή ένα σύνολο 61560 σημείων σε όλο τον πλανήτη, ενώ βάσει της διαθεσιμότητάς του, το χρονικό εύρος μελέτης ορίστηκε από την 1/1/2017 έως και τις 31/8/2020. **Συνεπώς, πρόκειται για ένα αρχικό dataset 61560 σημείων (features) x 1339 ημερών**. Τα δεδομένα αποθηκεύονται και διατίθενται υπό τη μορφή grib1/grib2. Πρόκειται για αρχεία binary μορφής και παρουσιάζουν συγκεκριμένη κωδικοποίηση, συνεπώς για να διαβαστούν θα πρέπει να γίνει η χρήση διαφόρων εργαλείων για τα οποία γίνεται λόγος στο υποκεφάλαιο 3.4.

[6] *Στην ιστοσελίδα του μοντέλου WRF (Weather Research Forecast), ως αρχική πηγή προσομοιώσεων προτείνονται μεταξύ άλλων τα reanalysis πεδία του NCEP (GFS-FNL). Περισσότερες πληροφορίες: https://www2.mmm.ucar.edu/wrf/users/download/free_data.html*



Εικόνα 3.1: *Το περιβάλλον άντλησης των reanalysis (FNL) δεδομένων του παγκόσμιου προγνωστικού συστήματος GFS από την ιστοσελίδα* https://rda.ucar.edu/datasets/ds083.2/#!access *, όπου απεικονίζεται η βάση δεδομένων "ds083.2" καθώς και οι διαθέσιμες επιλογές της.*

3.2: Επεξεργασία Δεδομένων

3.2.1 Διαχωρισμός

Για την ανάπτυξη ενός νευρωνικού δικτύου το οποίο επιχειρεί πρόγνωση βασισμένη σε χρονοσειρές τα δεδομένα διαχωρίζονται συνήθως σε δύο κατηγορίες: Στα δεδομένα εκπαίδευσης (**train**) και στα δεδομένα αξιολόγησης (**test**). Οι εξαρτημένες μεταβλητές (πρόγνωσης) ορίζονται ως $\boldsymbol{Y_j}$, ενώ ως $\boldsymbol{x_i}$ ορίζονται οι ανεξάρτητες μεταβλητές, οι οποίες στην

ML/AI προσέγγιση αναφέρονται και ως “features”. Ορίζοντας ως **d_in** τον αριθμό ημερών εκπαίδευσης και **d_out** τον αριθμό ημερών αξιολόγησης, προκύπτουν:

α) οι *ανεξάρτητες* μεταβλητές *εκπαίδευσης* $\boldsymbol{x_i}$ ($i \in N*$), οι οποίες δημιουργούν ένα πίνακα δεδομένων “**trainx**” διαστάσεων [i x d_in],

β) οι *εξαρτημένες* μεταβλητές *εκπαίδευσης* $\boldsymbol{Y_j}$ ($j \in N*$), οι οποίες δημιουργούν έναν πίνακα δεδομένων “**trainY**” διαστάσεων [j x d_in],

γ) οι *ανεξάρτητες* μεταβλητές *αξιολόγησης* $\boldsymbol{x_i}$ ($i \in N*$), οι οποίες δημιουργούν έναν πίνακα δεδομένων “**testx**” διαστάσεων [i x d_out],

δ) οι *εξαρτημένες* μεταβλητές *αξιολόγησης* $\boldsymbol{Y_j}$ ($j \in N*$), οι οποίες δημιουργούν έναν πίνακα δεδομένων “**testY**”, διαστάσεων [j x d_out].

Για τον αριθμό των train / test ημερών (*d_in*, *d_out* αντίστοιχα) ισχύει γενικά *d_in* > *d_out*. Αν και ένας συνηθισμένος διαχωρισμός είναι 60% και 40% ή 70% και 30%, αυτό δεν είναι περιοριστικό και εξαρτάται πάντα από τις ανάγκες της έρευνας. Στην συγκεκριμένη περίπτωση μελέτης ένα έτος μπορεί να θεωρηθεί αρκετό για την αξιολόγηση του μοντέλου καθώς καλύπτει όλη την εποχική μεταβλητότητα. Επιπροσθέτως, εφόσον η φύση της ατμόσφαιρας είναι χαοτική, ένας μεγάλος αριθμός ετών εκμάθησης συνεπάγεται και περισσότερους πιθανούς συνδυασμούς συσχετίσεων στους οποίους θα μπορούσε να εκπαιδευτεί το νευρωνικό δίκτυο. Από την άλλη πλευρά όμως, η παρουσία ενός σχετικά μεγάλου δείγματος στα δεδομένα αξιολόγησης (πέραν του έτους) καθιστά τους στατιστικούς δείκτες περισσότερο αξιόπιστους. Επιδιώκοντας την ισορροπία μεταξύ των παραπάνω συνθηκών, ως test(x,Y) dataset, επιλέχθηκε τελικά το διάστημα 1/1/2020 έως 31/8/2020 (244 ημέρες), στις οποίες πραγματοποιείται η στατιστική αξιολόγηση των αποτελεσμάτων του μοντέλου. Τα δεδομένα αυτά αποτελούν το ~18,2% του συνολικού όγκου, ενώ τα δεδομένα εκπαίδευσης train(x,Y) αποτέλεσαν το ~81,8%. Έτσι λοιπόν, οι αρχικές μεταβλητές έχουν ως εξής:

dataset_days = 1339,

d_in = 1095,

d_out = 244.

Επίσης, ο ***αρχικός*** αριθμός των features (πλήθος ανεξάρτητων μεταβλητών x) ταυτίζεται με τον αριθμό των σημείων στα δεδομένα εκπαίδευσης.

Αν και το αρχικό πλήθος ***i*** των x είναι εξ αρχής γνωστό, ο αριθμός ***j*** των εξαρτημένων μεταβλητών Y, δηλαδή τα σημεία στα οποία επιχειρείται η πρόγνωση, καθορίζεται από την φύση του προβλήματος. Εφ' όσον όμως οι πίνακες που δημιουργούνται είναι δύο διαστάσεων [σημεία x ημέρες], τότε προκύπτουν δύο βασικές (ακραίες) επιλογές μεταξύ των οποίων μπορεί να σχεδιαστεί η πειραματική διαδικασία:

1. Να επιχειρηθεί η πρόγνωση μιας χρονοσειράς (ή κάποιου χρονικού τμήματός της) έχοντας ως δεδομένα εισόδου άλλες χρονοσειρές (σημειακή πρόγνωση) (πίνακας 3.1)

	t = 0	**t = 1**	**t = 2**	**t = 3**	**t = 4**
X_1	train	train	train	train	test
X_2	train	train	train	train	test
Y	?	?	?	?	?

Πίνακας 3.1: *Περίπτωση πρόγνωσης της μεταβλητής Y με βάση τις μεταβλητές X_1 και X_2. Η εκπαίδευση πραγματοποιείται για τις χρονικές στιγμές t=0-3 και η αξιολόγηση για την χρονική στιγμή t=4.*

2. Να επιχειρηθεί η ταυτόχρονη πρόγνωση όλων των χρονοσειρών για κάθε ένα χρονικό βήμα (χωρική πρόγνωση) (πίνακες 3.2α & β).

	t = 0	**t = 1**	**t = 2**	**t = 3**	**t = 4**
X_1	train	train	train	train	Y_1
X_2	train	train	train	train	Y_2
X_3	train	train	train	train	Y_3

(α)

	t = 5	**t = 6**	**t = 7**	**t = 8**	**t = 9**
X_1	test	test	test	test	Y_1
X_2	test	test	test	test	Y_2
X_3	test	test	test	test	Y_3

(β)

Πίνακες 3.2α,β*: Περίπτωση πρόγνωσης των μεταβλητών Y_j με βάση τις μεταβλητές X_i. Οι Y αποτελούν την τελευταία τιμή κάθε χρονοσειράς τόσο για τα δεδομένα εκπαίδευσης (t=0-4s) όσο και για τα δεδομένα ελέγχου (t=5-9s).*

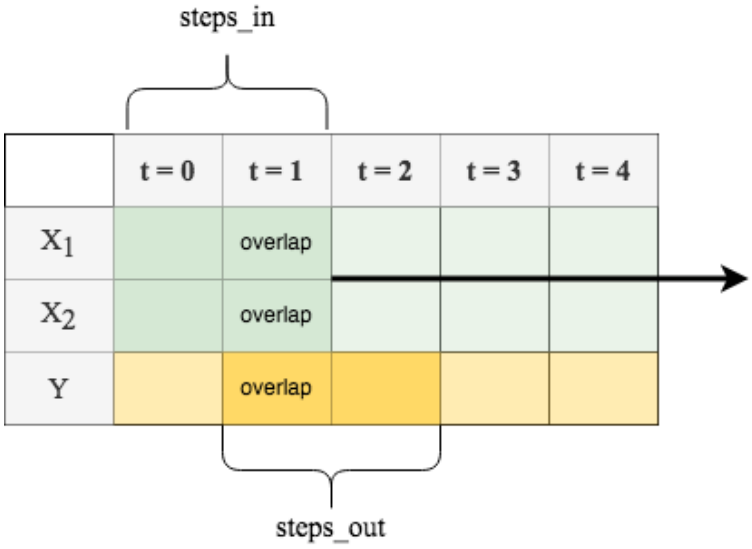
Σε κάθε περίπτωση πάντως, ένα τμήμα της χρονοσειράς δεσμεύεται ως "test", δηλαδή πρόκειται για δεδομένα στα οποία το δίκτυο δεν έχει εκτεθεί κατά τη διάρκεια της εκπαίδευσής του, ώστε στη συνέχεια να αξιολογηθεί με την αναλογία δεδομένων που περιγράφηκε σε προηγούμενη παράγραφο. Στο πρώτο παράδειγμα του πίνακα 3.1 πρόκειται για τη χρονική στιγμή t=4, ενώ στο δεύτερο παράδειγμα που απεικονίζεται στους πίνακες 3.2α και β, πρόκειται για τον ίδιο τον πίνακα 3.2β. Ωστόσο, με οποιονδήποτε τρόπο διαχωριστεί το dataset, θα ισχύει $\boldsymbol{j \leq i}$ για το πλήθος των Y και X.

3.2.1.1: Τεχνική "Sliding Window"

Οι παραπάνω δύο (ακραίες) περιπτώσεις που περιγράφηκαν εμπίπτουν στην κατηγορία **"*εκπαίδευσης χωρίς επίβλεψη*" (unsupervised learning)** καθώς το μοντέλο μαθαίνει σε συσχετίσεις μεταξύ X_i και Y_j μέσα από διαδοχικές χρονοσειρές που του παρουσιάζονται χωρίς κάποια άλλη παρέμβαση. Στην παρούσα εργασία, εφαρμόζεται μια μέθοδος που αποτελεί συνδυασμό των ανωτέρω παραδειγμάτων, καθώς η προγνωστική μεταβλητή είναι μεν μία (όπως στον πίνακα 3.1) αλλά μόνο για μία χρονική στιγμή (όπως στον πίνακα 3.2α&β) ή προκαθορισμένο διάστημα διαδοχικών χρονικών στιγμών. Άρα λοιπόν τα datasets της εκπαίδευσης και της αξιολόγησης θα πρέπει να αναδιοργανωθούν και να παρουσιαστούν με ανάλογο τρόπο στο μοντέλο. Η συγκεκριμένη αλλαγή μετατρέπει τελικά το πρόβλημα σε ***supervised learning,*** διότι πλέον "υποδεικνύουμε" μέσω κατάλληλου κώδικα στο AI μοντέλο ποια τμήματα των χρονοσειρών και με ποιον τρόπο θα λάβει υπόψιν.

Για τον σκοπό αυτό, η τεχνική του "**sliding window**" (μετακινούμενου παραθύρου), όπως και η παραλλαγή της, "**recurring sliding window**", είναι μία από τις βασικές που χρησιμοποιούνται (Dietterich, 2002). Για την εφαρμογή της sliding window τεχνικής, χρησιμοποιώντας το παράδειγμα των τριών μεταβλητών και των πέντε χρονικών στιγμών (πίνακες 3.1, 3.2α&β), ορίζουμε νέες συσχετίσεις μεταξύ των δεδομένων ως εξής: Θεωρούμε σταθερές χρονικές εισόδους **steps_in > 1** για τα X_i (features), οι οποίες παρουσιάζονται στο μοντέλο διαδοχικά (μετακινούμενο παράθυρο), ώστε να προβλέψει τη μεταβλητή Y. Η recurring εκδοχή του "μετακινούμενου παραθύρου" ορίζει πως η Y θα πρέπει να προβλεφθεί για τουλάχιστον 2 διαδοχικές στιγμές, όπου υπάρχει τουλάχιστον μία κοινή χρονική στιγμή με τις μεταβλητές X, η οποία μπορεί να οριστεί από μια σταθερά που για τις ανάγκες της εργασίας, ονομάζεται "**overlap"**. Η παραπάνω τεχνική βασίζεται στην υπόθεση πως εμφανίζονται

συσχετίσεις μεταξύ των παρελθοντικών διακυμάνσεων των μεταβλητών X και της μεταβλητής Y, οι οποίες είτε διερευνώνται είτε είναι ήδη γνωστές. Η απλοποιημένη εκδοχή της παραπάνω τεχνικής για ένα εικονικό dataset εκπαίδευσης X και Y μεταβλητών με ***i = 2***, ***j = 1***, πέντε χρονικών στιγμών ***t = 5 (0-4)***, ***steps_in = 2***, ***steps_out = 2*** και ***overlap = 1***, παρουσιάζεται στον πίνακα 3.2.



Πίνακας 3.3: *Υποθετικό παράδειγμα της recurring sliding window τεχνικής σε δεδομένα δύο ανεξάρτητων μεταβλητών και μίας εξαρτημένης. Διακρίνονται δύο βήματα εισόδου (steps_in) και δύο βήματα εξόδου (steps_out), τα οποία ταυτίζονται χρονικά σε ένα βήμα (overlap). Βάσει αυτών των σταθερών, η εκπαίδευση του μοντέλου προχωράει διαδοχικά κατά ένα βήμα κάθε φορά, στις επόμενες χρονικές στιγμές.*

Αξίζει να επισημανθεί ότι στην περίπτωση όπου steps_in = steps_out = overlap, τότε εμπίπτουμε στη γενική κατηγορία της σημειακής πρόγνωσης χρονοσειράς, όπως περιγράφηκε στην προηγούμενη παράγραφο. Αντιθέτως, εάν overlap = 0 και έχουμε μια εξαρτημένη μεταβλητή, τότε διερευνάται η συσχέτιση μεταξύ των X, Y αλλά σε διαφορετικό χρονικό σημείο ενώ, εάν επιχειρείται η ταυτόχρονη πρόγνωση όλων των χρονοσειρών, τότε σημαίνει πως λαμβάνονται υπόψιν οι τελευταίες ημέρες ταυτόχρονα σε όλες τις περιοχές, ώστε να προβλεφθούν οι επόμενες. Γενικότερα, οι παραλλαγές των παραπάνω τεχνικών είναι πάρα πολλές, εξαρτώμενες πάντα από τη φύση του προβλήματος και το είδος των δεδομένων.

Στην παρούσα έρευνα, εφ' όσον το ζητούμενο είναι η πρόγνωση κάποιας μελλοντικής χρονικής στιγμής της εξαρτημένης μεταβλητής Y, έχοντας ως **δεδομένες** τις τιμές των X (έχουν ήδη συμβεί), τότε θα πρέπει να ισχύει: **overlap < steps_in, steps_out** και **overlap > 0**.

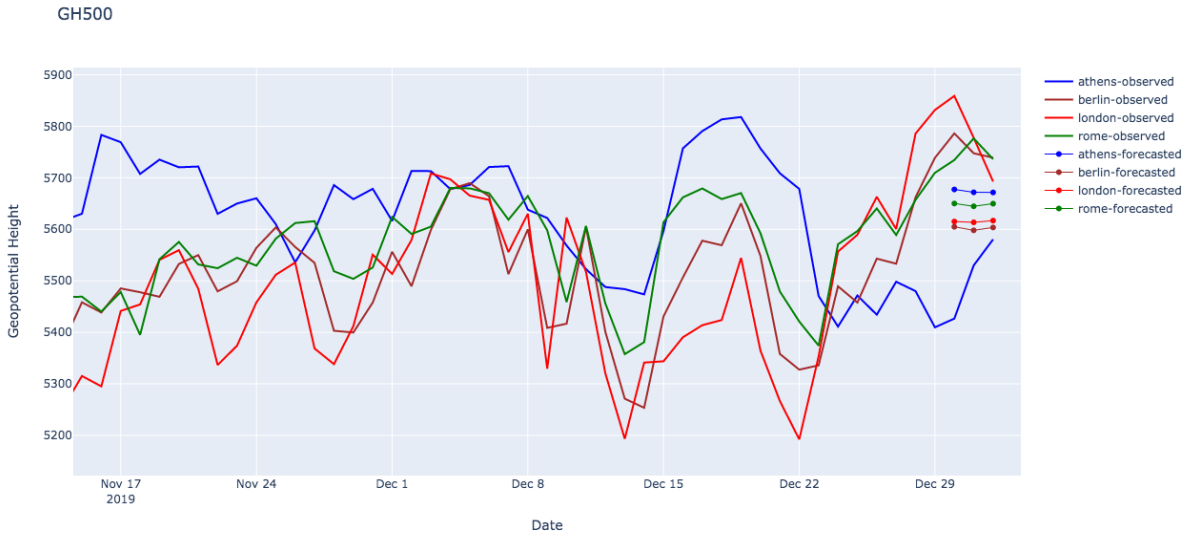
3.2.2 Μορφοποίηση και Εξαγωγή Χαρακτηριστικών

Τις περισσότερες φορές τα δεδομένα πάνω στα οποία καλείται να εκπαιδευτεί ένα νευρωνικό δίκτυο πρέπει πρώτα να υποστούν επεξεργασία (data pre-processing), η οποία εξαρτάται από το είδος και την αρχιτεκτονική του. Στην τελική τους μορφή οι τιμές καταλήγουν να μην έχουν κάποια φυσική σημασία και μπορεί να είναι αδιάστατες. Ωστόσο όποια μορφή κι αν έχουν λάβει, θα πρέπει οι αριθμοί να διατηρούν την μεταξύ τους σχέση και τα δεδομένα να έχουν την ίδια εσωτερική συνοχή και ερμηνεία. Για παράδειγμα, ο αριθμός -2 θα σημαίνει για το συγκεκριμένο δίκτυο πάντα το ίδιο, από όποιο dataset κι αν προέρχεται. Η συγκεκριμένη εργασία χρησιμοποιεί τα δεδομένα γεωδυναμικού ύψους (GH500), τα οποία, για να εισαχθούν ως γνωρίσματα στο μοντέλο, θα πρέπει οπωσδήποτε να υποστούν κάποια επεξεργασία. Καθώς πρόκειται για χρονοσειρές χρειάζεται: 1. Κάποιο είδος κανονικοποίησης (normalization), το οποίο συνήθως γίνεται αφού πρώτα προηγηθούν συγκεκριμένοι διαγνωστικοί έλεγχοι και 2. Αφαίρεση του στοιχείου της εποχικότητας που εμφανίζεται έντονο στα μετεωρολογικά δεδομένα.

3.2.2.1 Αφαίρεση τάσης και εποχικότητας

Οι χρονοσειρές συχνά εμφανίζουν κάποια ιδιαίτερα χαρακτηριστικά με κυρίαρχα 1) την **εποχικότητα / επαναληπτικότητα** (κυκλικές διακυμάνσεις) και 2) την **τάση** (Bontempi et al., 2013), οπότε τα δεδομένα τους χαρακτηρίζονται ως "**μή στατικά**" (non-stationary). Αντιθέτως στα "**στατικά**" stationary δεδομένα τα δύο παραπάνω χαρακτηριστικά απουσιάζουν και η διακύμανσή τους καθορίζεται από άλλους παράγοντες (τυχαίους ή προβλέψιμους). Σε κάθε περίπτωση για να χαρακτηριστεί μια χρονοσειρά ως "stationary", θα πρέπει τουλάχιστον να διατηρεί σταθερούς τους στατιστικούς της δείκτες, όπως η μέση τιμή και η διακύμανση. Τις περισσότερες φορές, είναι αναγκαία η μετατροπή των δεδομένων σε στατικά πριν εισαχθούν σε κάποιο νευρωνικό δίκτυο, διότι εάν κάποιο δίκτυο εκπαιδευτεί σε non-stationary χρονοσειρές, στην προσπάθειά του να ελαχιστοποιήσει το αποτέλεσμα της συνάρτησης κόστους, αυτό που πρωτίστως θα "μάθει" είναι την ενδεχόμενη κυκλικότητα ή την τάση τους. Όταν εκτεθεί σε νέα δεδομένα, το πιθανότερο αποτέλεσμα είναι να καταλήξει σε προβλέψεις που θα αντικατοπτρίζουν αυτά τα δύο χαρακτηριστικά, αγνοώντας όμως άλλες συσχετίσεις που ενδεχομένως υπάρχουν. Παράδειγμα τέτοιας πρόβλεψης απεικονίζεται στο σχήμα 3.1, όπου επιχειρείται η **παράλληλη** πρόγνωση τεσσάρων χρονοσειρών (i = j = 4, περισσότερα στο υποκεφάλαιο 3.2.1), στην ισοβαρική στάθμη των 500hpa για steps_in = steps_out = 3 και overlap = 0 (οι μεταβλητές έχουν οριστεί στο υποκεφάλαιο 3.2.1.1). Ως

δεδομένα εκπαίδευσης έχουν χρησιμοποιηθεί οι τελευταίες 1095 ημέρες (3 έτη) χωρίς κάποια προεργασία και επιχειρείται η πρόγνωση των επόμενων τριών ημερών. Όπως φαίνεται εκ του αποτελέσματος, το νευρωνικό δίκτυο έχει απλώς "μάθει" τον γενικό μέσο όρο και την τάση που εμφανίζεται προς το τέλος του dataset για κάθε μία από τις χρονοσειρές, χωρίς όμως να μπορεί να ανιχνεύσει μικρότερης κλίμακας αλλαγές.



Σχήμα 3.1: *Παράδειγμα πρόγνωσης παράλληλων χρονοσειρών της μεταβλητής GH500 ενός νευρωνικού δικτύου το οποίο έχει εκπαιδευτεί σε non-stationary δεδομένα 1095 ημερών.*

Ο πιο άμεσος τρόπος διάγνωσης της κυκλικότητας και της εποχικότητας στα δεδομένα είναι ο οπτικός. Όσον αφορά την ατμόσφαιρα, αυτά τα δύο χαρακτηριστικά είναι αναμενόμενα, όπως για παράδειγμα η πορεία της θερμοκρασίας σε ημερήσια και ετήσια βάση. Στην περίπτωση αυτή όμως, διαγνωστικοί έλεγχοι χρησιμοποιώντας δείκτες όπως ο μέσος όρος και η διακύμανση, αν και χρήσιμοι για την εξαγωγή συμπερασμάτων, σε πολλές περιπτώσεις αποδεικνύονται πολύ γενικοί και τελικά, αναποτελεσματικοί. Εάν για παράδειγμα στα non-stationary δεδομένα της θερμοκρασίας θεωρήσουμε ένα πολύ μεγάλο χρονικό διάστημα πέραν των 2 ετών, προκύπτει πως, τόσο ο μέσος όρος όσο και η διακύμανση της τιμής παραμένουν περίπου σταθερές, ενώ το ίδιο συμβαίνει και σε άλλες μετεωρολογικές παραμέτρους συμπεριλαμβανομένου και του γεωδυναμικού ύψους στα 500hpa (σχήμα 3.2). Ως εκ τούτου, διαπιστώνεται η ανάγκη για περισσότερο εξειδικευμένους στατιστικούς ελέγχους οι οποίοι εκτός από την οπτική παρατήρηση μπορούν να διαχωρίσουν μια χρονοσειρά σε στατική / μη στατική.

Ένας ευρέως χρησιμοποιούμενος στατιστικός δείκτης είναι ο Dickey-Fuller (1979). Κατά την $\boldsymbol{H_0}$ **αρχική υπόθεση** θεωρείται πως η χρονοσειρά είναι **"non-stationary"**, δηλαδή υπάρχει κάποιος παράγοντας εξαρτώμενος του χρόνου, ο οποίος καθορίζει τις τιμές (Unit-Root). Κατά τη διαδικασία του ελέγχου, εξετάζεται η εναλλακτική $\boldsymbol{H_1}$ **υπόθεση** πως δεν υπάρχει κάποια εξαρτώμενη από το χρόνο αιτία διακύμανσης της χρονοσειράς, άρα η χρονοσειρά είναι **"stationary"**. Ο συγκεκριμένος δείκτης βασίζεται στη θεωρία του Auto-Regression (AR), ως τμήμα του μοντέλου περιγραφής και πρόγνωσης των χρονοσειρών "ARMA" (Auto-Regression Moving Average), όπως ορίστηκε από τον Whittle (1951). Σύμφωνα με αυτή, κάθε χρονοσειρά δεδομένων μπορεί να περιγραφεί από την εξής σχέση[7], η οποία συνδέει προηγούμενες και επόμενες τιμές μιας μεταβλητής $\boldsymbol{y}$:

$$y_t = \rho y_{t-1} + u_t \qquad (3.1)$$

όπου,

ρ : η παράμετρος που περιγράφει το αίτιο διακύμανσης (Unit Root) και

u_t : η παράμετρος τυχαιότητας (θόρυβος / λευκός θόρυβος).

Για την ύπαρξη τυχαιότητας στα δεδομένα η κρίσιμη τιμή του ρ είναι η μονάδα: Εάν **ρ < 1**, τότε η χρονοσειρά καθορίζεται κυρίως από τον παράγοντα της τυχαιότητας, καθώς με την πάροδο του χρόνου η επιρροή του ρ τείνει να εκμηδενίζεται και άρα το $y_t \to u_t$, δηλαδή πρόκειται για "**stationary**" χρονοσειρά. Εάν ρ = 1, τότε η χρονοσειρά μετατρέπεται σε "non-stationary" ενώ εάν ρ > 1, τότε η τάση είναι τόσο ισχυρή που η τιμή y τείνει στο άπειρο.

Στην περίπτωση της 3.1 έχει γίνει η υπόθεση πως τα δεδομένα είναι τυχαία, όμως στην πραγματικότητα η H_0 αρχική υπόθεση είναι η ύπαρξη μιας Unit Root, δηλαδή η **μη τυχαιότητα** στα δεδομένα. Θεωρώντας ως $\delta = \rho - 1$ τότε για την πεπερασμένη μεταβολή **Δ** του y, η 3.1 μπορεί να γραφτεί και ως εξής:

$$\Delta y_t = \delta \cdot y_{t-1} + u_t \qquad (3.2)$$

---

[7] *Η σχέση εδώ παρουσιάζεται με απλοποιημένο τρόπο. Ο ορισμός στην πλήρη της μορφή περιγράφει ένα άθροισμα (Σ), παραγόντων (φ_i) για κάθε μεταβλητή y_i.*

οπότε εφ' όσον αναφερόμαστε σε μεταβολή, παρατηρούμε πως η κρίσιμη τιμή της δ είναι το μηδέν. Εάν δ < 0 τότε $\Delta y_t \rightarrow u_t$, δηλαδή η μεταβολή του y τείνει να προσδιοριστεί μόνο από το λευκό θόρυβο και άρα η χρονοσειρά είναι "stationary" απορρίπτοντας την αρχική υπόθεση (H_0) ότι η χρονοσειρά καθορίζεται από την ύπαρξη μιας Unit-Root.

Σε περίπτωση ύπαρξης κάποιας σταθεράς απόκλισης (drift) a_0 η 3.2 μετατρέπεται σε:

$$\Delta y_t = a_0 + \delta \cdot y_{t-1} + u_t \qquad (3.2b)$$

ενώ εάν επίσης υφίσταται και κάποια αιτιοκρατική τάση (trend) συναρτήσει του χρόνου $a_1 \cdot t$ τότε:

$$\Delta y_t = a_0 + a_1 \cdot t + \delta \cdot y_{t-1} + u_t \qquad (3.2c)$$

Ο έλεγχος Dickey-Fuller πραγματοποιείται ως προς την τιμή της δ σε κάποια εκ των τριών περιπτώσεων (3.2, 3.2b, 3.2c). Ειδικότερα εάν αφορά την 3.2c τότε καλείται ως Augmented Dickey-Fuller test, το οποίο ορίζεται από την εξής σχέση:

$$ADF_\tau = \frac{\delta}{SE(\delta)} \qquad (3.3)$$

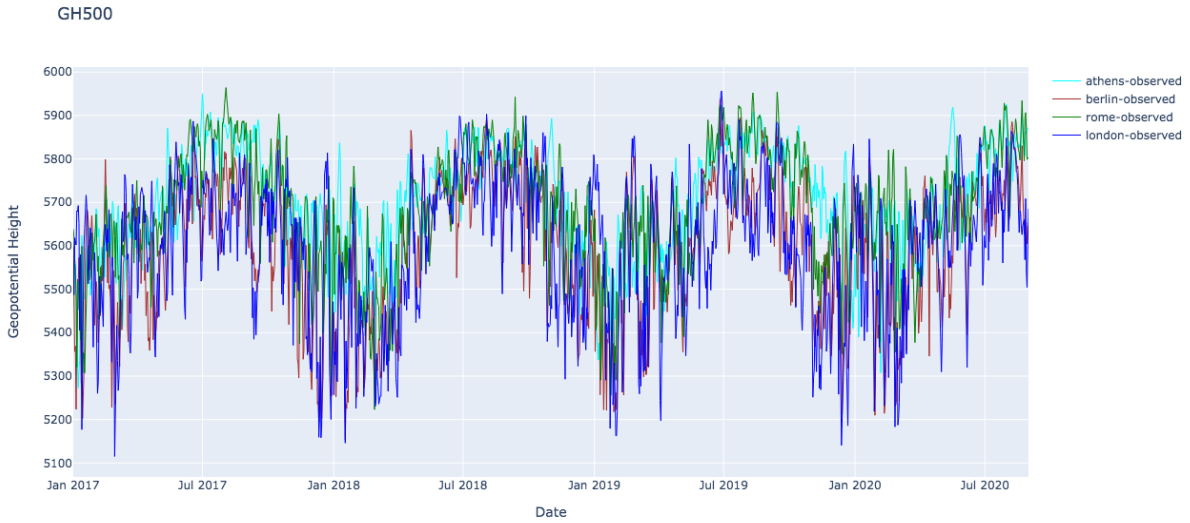
όπου SE: τυπικό σφάλμα.

Εφ' όσον για να απορριφθεί η H_0 υπόθεση πρέπει δ < 0, τότε συνακολούθως και ο ADF θα πρέπει να έχει αρνητικές τιμές. Το αποτέλεσμα του υπολογισμού συγκρίνεται με τον πίνακα τιμών (πίνακας 3.3) για επίπεδο σημαντικότητας $p < 0.01$ ή $p < 0.05$ και συγκεκριμένους αριθμούς δειγμάτων και εάν το αποτέλεσμα είναι μικρότερο της κρίσιμης τιμής, τότε η H_0 απορρίπτεται.

ADF	Χωρίς τάση		Με τάση	
Δείγμα (τ)	**p = 1%**	**p = 5%**	**p = 1%**	**p = 5%**
25	-3,75	-3,0	-4,38	-3,60
50	-3,58	-2,93	-4,15	-3,50
100	-3,51	-2,89	-4,04	-3,45
250	-3,46	-2,88	-3,99	-3,43
500	-3,44	-2,87	-3,98	-3,42
∞	-3,43	-2,86	-3,96	-3,41

Πίνακας 3.3: *Κρίσιμες τιμές της ADF συναρτήσει του αριθμό του δείγματος* τ *και για επίπεδα σημαντικότητας p = 1% και p = 5%, κάτω από τις οποίες η* H_0 *υπόθεσης ύπαρξης Unit Root στα δεδομένα μιας χρονοσειράς, απορρίπτεται. Πηγή: Fuller, 1976.*

Η χρονοσειρά των δεδομένων του γεωδυναμικού ύψους στα 500hpa (GH500) στην περιοχή των Αθηνών και άλλων τριών τυχαίων περιοχών (π.χ. της Ρώμης, του Λονδίνου και του Ρέϋκιαβικ) για δείγμα τ = 1338 ημερών (1/1/2017 - 31/8/2020) παρουσιάζεται στο σχήμα 3.2. Οπτικά διαπιστώνεται η εποχικότητα στις τιμές (υψηλότερες τιμές τους θερινούς μήνες και χαμηλότερες τους χειμερινούς), από την άλλη πλευρά όμως δεν υπάρχει κάποια σαφής τάση ανόδου / καθόδου κατά την περίοδο των 3.5 ετών.



Σχήμα 3.2: *Τμήμα της χρονοσειράς της μεταβλητής του γεωδυναμικού ύψους στα 500hpa από αρχές Ιανουαρίου 2017 μέχρι και τέλος Αυγούστου 2020 για την περιοχή των Αθηνών, του Βερολίνου, της Ρώμης και του Λονδίνου.*

Ταυτόχρονα όμως, παρατηρείται απόκλιση των τιμών της Αθήνας και της Ρώμης σε σχέση με αυτές του Λονδίνου και του Βερολίνου, με τις δύο πρώτες να εμφανίζονται συστηματικά υψηλότερες (γαλάζια και πράσινη γραμμή). Αυτό ερμηνεύεται ως αποτέλεσμα της κλιματικά αντίστροφης σχέσης του γεωγραφικού πλάτους με τις τιμές των γεωδυναμικών υψών στα 500hpa. Οι αποκλίσεις είναι εντονότερες κατά τις χειμερινές περιόδους.

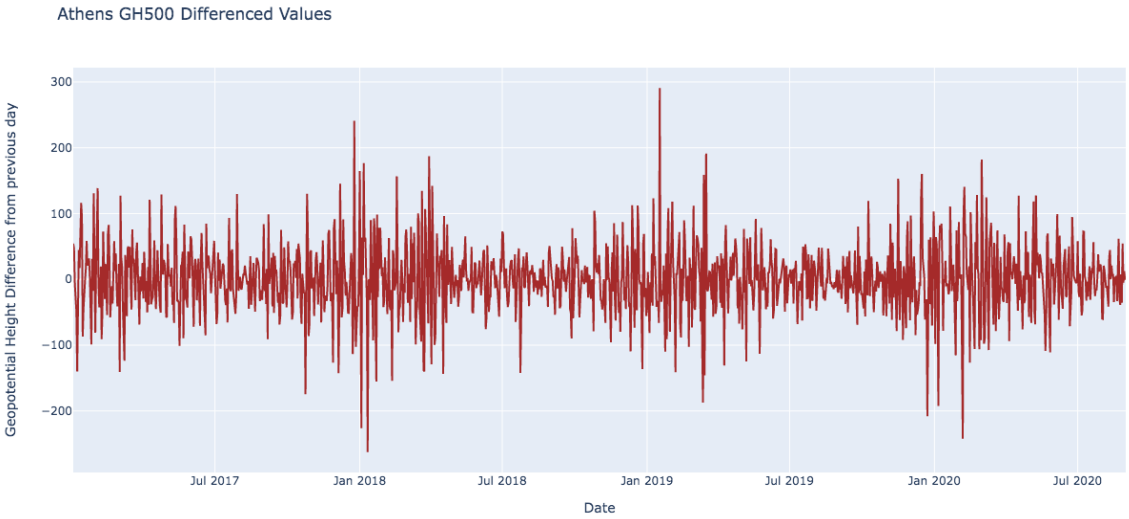
Θεωρώντας ότι 1. τα τέσσερα προαναφερθέντα σημεία είναι αντιπροσωπευτικά διαφόρων περιοχών και κλιματικών τύπων της Ευρώπης και 2. το δείγμα των 1338 ημερών είναι επαρκές, πραγματοποιήθηκε ο έλεγχος του δείκτη ADF για κάθε ένα από αυτά, τα αποτελέσματα του οποίου παρουσιάζονται στον πίνακα 3.4. Η σύγκριση με τις τιμές του πίνακα 3.3 οδηγεί στο συμπέρασμα ότι ο δείκτης βρίσκεται κάτω από τις κρίσιμες τιμές μόνο για τις βορειότερες περιοχές της ηπείρου (Λονδίνο, Βερολίνο) ενώ στις νοτιότερες (Αθήνα, Ρώμη) συμβαίνει το αντίθετο. Αυτό σημαίνει ότι στις μεν βορειότερες περιοχές η αρχική υπόθεση H_0 ύπαρξης μιας σχετιζόμενης με το χρόνο αιτίας διακύμανσης των τιμών (Unit Root) μπορεί να απορριφθεί, κάτι το οποίο δεν ισχύει για τις νοτιότερες, οι χρονοσειρές των οποίων μπορούν να χαρακτηριστούν ως "non-stationary", δηλαδή **παρουσιάζουν εποχικότητα**. Αυτή η διαφορά μπορεί να ερμηνευθεί από το γεγονός της σημαντικής εξασθένησης της κυκλωνικής δραστηριότητας που παρατηρείται κατά τους θερινούς μήνες στη Μεσόγειο, κάτι το οποίο δεν ισχύει σε τόσο έντονο βαθμό στις βορειότερες και δυτικότερες περιοχές της Ευρώπης.

Περιοχή	Γ. Πλάτος	Γ. Μήκος	ADF	p-value
Αθήνα	38.0	24.0	-3,38	0,011
Ρώμη	42.0	12.0	-3,11	0,025
Λονδίνο	52.0	0.0	-5,45	0,000
Βερολίνο	52.0	14.0	-4,31	0,000

Πίνακας 3.4: *Οι τιμές και τα επίπεδα σημαντικότητας του ελέγχου ADF για τέσσερις περιοχές της Ευρώπης σε δείγμα δεδομένων 1338 ημερών της τιμής GH500. Παρατηρείται ότι σε Αθήνα και Ρώμη, οι τιμές του δείκτη βρίσκονται πάνω από το κρίσιμο όριο (κόκκινο χρώμα - εμφάνιση εποχικότητας) ενώ το αντίθετο συμβαίνει σε Λονδίνο και Βερολίνο (πράσινο χρώμα - δεν εμφανίζεται εποχικότητα).*

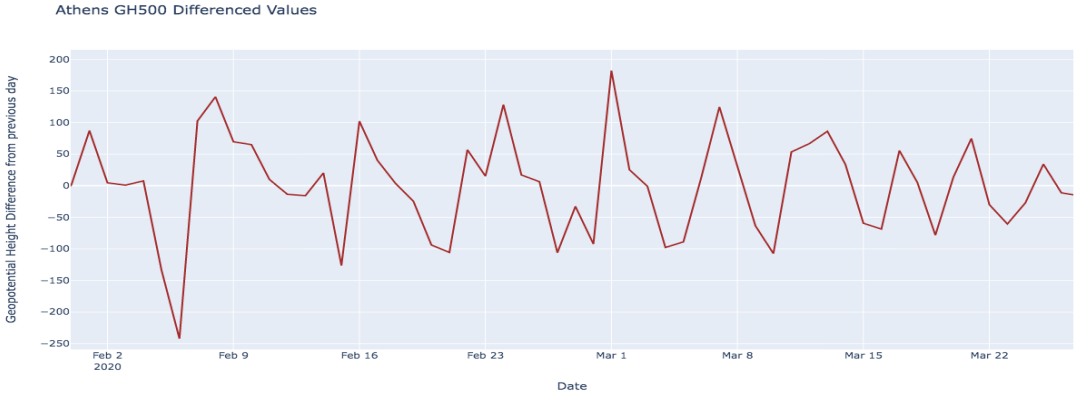
Ένας απλός τρόπος αφαίρεσης της εποχικότητας (ή και της παρατηρούμενης τάσης) στα δεδομένα είναι η δημιουργία μιας νέας χρονοσειράς y' (παράγωγος της y), η οποία

προκύπτει από τη διαφορές των διαδοχικών μετρήσεων, έτσι ώστε για κάθε χρονική στιγμή t να ισχύει $y_t = x_t - x_{t-1}$. Κατ'αυτόν τον τρόπο προκύπτει μια χρονοσειρά που εκφράζει την "κλίση" μεταξύ δύο διαδοχικών μετρήσεων. Όπως γίνεται αντιληπτό, ο αριθμός των τιμών στα καινούργια δεδομένα είναι κατά μία λιγότερη, τα οποία παρουσιάζουν διακύμανση γύρω από το μηδέν. Στο σχήμα 3.3 παρουσιάζεται η παράγωγος χρονοσειρά της Αθήνας y'.



Σχήμα 3.3: *Η παράγωγος χρονοσειρά της μεταβλητής GH500 για την περιοχή των Αθηνών, ως αποτέλεσμα των ημερήσιων διαφορών στην τιμή της για την περίοδο μεταξύ 2/1/2017 και 31/8/2020.*

Παρατηρείται ότι αν και πλέον δεν εμφανίζεται κάποια εποχικότητα, υπάρχουν δύο χαρακτηριστικά: 1. οι διακυμάνσεις είναι εντονότερες κατά τους χειμερινούς μήνες, κάτι που ερμηνεύεται ως αποτέλεσμα της διέλευσης ισχυρότερων ατμοσφαιρικών διαταραχών και 2. υπάρχει μια τάση περιοδικότητας 6-8 ημερών, χαρακτηριστικό παράδειγμα της οποίας μπορεί να φανεί στο σχήμα 3.4 το οποίο παρουσιάζει το χρονικό διάστημα μεταξύ 31 Ιανουαρίου και 27 Μαρτίου 2020.



Σχήμα 3.4: *Τμήμα της παραγώγου χρονοσειράς της μεταβλητής GH500 για την περιοχή των Αθηνών, ως αποτέλεσμα των ημερήσιων διαφορών στην τιμή της για την περίοδο μεταξύ 31/1/2020 και 27/3/2020.*

Για τη νέα χρονοσειρά δεδομένων y' πραγματοποιήθηκε εκ νέου ο υπολογισμός του δείκτη Dickey-Fuller για τις τέσσερις προαναφερθείσες περιοχές και για δείγμα 1337 μετρήσεων, τα αποτελέσματα του οποίου παρουσιάζονται στον πίνακα 3.5:

Περιοχή	**Γ. Πλάτος**	**Γ. Μήκος**	**ADF**	**p-value**
Αθήνα	38.0	24.0	-15,95	~0,000
Ρώμη	42.0	12.0	-12,99	~0,000
Λονδίνο	52.0	0.0	-12,49	~0,000
Βερολίνο	52.0	14.0	-11,61	~0,000

Πίνακας 3.5: *Οι τιμές και τα επίπεδα σημαντικότητας του ελέγχου ADF για τέσσερις περιοχές της Ευρώπης σε δείγμα 1337 ημερών, στα δεδομένα της ημερήσιας μεταβολής της τιμής GH500.*

Η σύγκριση των τιμών του πίνακα 3.5 με τις τιμές του πίνακα 3.4 οδηγεί στο συμπέρασμα ότι βάσει του δείγματος οι παραχθείσες χρονοσειρές σε όλα τα σημεία ελέγχου δεν έχουν κάποιο αίτιο διακύμανσης που εξαρτάται από το χρόνο (Unit Root) και ως εκ τούτου η αρχική H_0 υπόθεση μπορεί να απορριφθεί για ένα πολύ μεγάλο επίπεδο σημαντικότητας, όπως εκφράζεται από την τιμή της p-value.

3.2.2.2: Εξαγωγή Γνωρισμάτων

Σε κάθε εφαρμογή της Μηχανικής Εκμάθησης το μοντέλο εκπαιδεύεται πάνω σε ένα πλήθος γνωρισμάτων, τα οποία δύναται να επηρεάσουν το αποτέλεσμα. Όπως ήδη αναφέρθηκε και στο υποκεφάλαιο 1.3, εάν αυτά τα γνωρίσματα είναι εξ' αρχής γνωστά και τα δεδομένα έχουν ήδη κατηγοριοποιηθεί, τότε πρόκειται για μια περίπτωση "*εκπαίδευσης με επίβλεψη*", ενώ το αντίθετο συμβαίνει στην "*εκπαίδευση χωρίς επίβλεψη*". Στην παρούσα έρευνα μπορεί να θεωρηθεί πως κάθε ένα από τα σημεία του πλανήτη που εξετάζονται αποτελεί κι ένα γνώρισμα. Η εκπαίδευση του μοντέλου πραγματοποιήθηκε και με τους δύο τρόπους.

Όπως έχει ήδη αναφερθεί στο υποκεφάλαιο 3.2.1, για ανάλυση 1°, ο αρχικός αριθμός των γνωρισμάτων / σημείων είναι *61560*, πλήθος το οποίο είναι εξαιρετικά μεγάλο για να είναι

διαχειρίσιμο από τη μνήμη και τον επεξεργαστή ενός τυπικού υπολογιστή[8]. Για το λόγο αυτό, η αρχική ανάλυση μειώθηκε κατά 5 φορές (δηλαδή από 1° στις 5°), μειώνοντας και το πλήθος των σημείων περίπου κατά 25 φορές, με αποτέλεσμα ένα τελικό dataset 2664 σημείων (features). Η μείωση της ανάλυσης δεν αναμένεται να επηρεάσει σημαντικά την πειραματική διαδικασία, διότι η πιθανή ύπαρξη συσχετισμών διαφόρων περιοχών του πλανήτη αφορά μεγάλης κλίμακας ατμοσφαιρικούς κυματισμούς οι οποίοι μπορούν να εντοπιστούν με την συγκεκριμένη ανάλυση.

Σε πολλές περιπτώσεις μηχανικής εκμάθησης η εξαγωγή χαρακτηριστικών (feature extraction) αποτελεί ένα απαραίτητο τμήμα της διαδικασίας. Αυτό όμως εξαρτάται τόσο από το είδος του προβλήματος, από την αρχιτεκτονική του μετέπειτα νευρωνικού δικτύου που θα χρησιμοποιηθεί και γενικότερα, εάν θα ακολουθηθεί η προσέγγιση της εκπαίδευσης με επίβλεψη ή χωρίς επίβλεψη. Καθώς τα δεδομένα χρονοσειρών αποτελούν αντικείμενο της παρούσας μελέτης, κρίθηκε αναγκαίο η διερεύνηση συσχετίσεων μεταξύ τους, με την εφαρμογή στατιστικών δεικτών. Μερικοί ευρέως χρησιμοποιούμενοι δείκτες είναι οι Pearson και Spearman, οι οποίοι παρουσιάζονται παρακάτω:

3.2.2.2.1: Οι δείκτες Pearson και Spearman

Ο δείκτης **Pearson** (Pearson, 1895) εκφράζει τη γραμμική συσχέτιση ρ μεταξύ δύο μεταβλητών X, Y και ορίζεται ως εξής (Pearson, 1895)[9]:

$$\rho_{pearson} = \frac{\Sigma(x_i-\underline{x})(y_i-\underline{y})}{\sqrt{\Sigma(x_i-\underline{x})^2}\sqrt{\Sigma(y_i-\underline{y})^2}} \qquad (3.4)$$

όπου: x_i *,* y_i *οι μεταβλητές για τις οποίες εξετάζεται η συσχέτιση και* $\underline{x}$ *,* $\underline{y}$ *η μέση τιμή των* x_i *και* y_i *αντίστοιχα.*

[8] *Ο υπολογιστής στον οποίο εκτελέστηκε η πειραματική διαδικασία είναι Mac, με μνήμη 4GB και επεξεργαστή Intel i5 @ 2.5GHz.*

[9] *Στην πραγματικότητα ο ορισμός της συγκεκριμένης γραμμικής συσχέτισης υπάρχει και προγενέστερα, όπως για παράδειγμα στις εργασίες του Francis Galton (γύρω στο 1880) και του Auguste Bravais (1844).*

Από την άλλη πλευρά, ο δείκτης **Spearman** ορίζεται ως εξής (Spearman, 1904):

$$\rho_{spearman} = 1 - \frac{6\Sigma_i d_i^2}{n(n^2-1)} \qquad (3.5)$$

όπου: n ο αριθμός των παρατηρήσεων και για τις δύο μεταβλητές και d η διαφορά στην κατάταξή τους (ranking).

Η βασική διαφορά του $\rho_{pearson}$ σε σχέση με τον $\rho_{spearman}$ είναι ότι ο πρώτος υποθέτει κάποια γραμμική συσχέτιση μεταξύ των εξεταζόμενων μεταβλητών ενώ ο δεύτερος μόνο μονοτονικότητα, διότι στον υπολογισμό λαμβάνεται υπόψιν η κατάταξη (η σχετική θέση) κάθε μεταβλητής και όχι η απόλυτη τιμή της. Συνεπώς, μέσω του δείκτη spearman μπορούν να συμπεριληφθούν ομάδες ακραίων δεδομένων (outliers) ή συστάδων δεδομένων (clusters). Το εύρος και των δύο δεικτών κυμαίνεται από -1 έως +1, όπου -1 σημαίνει απόλυτα αρνητική συσχέτιση και +1 απόλυτα θετική συσχέτιση, ενώ αποτέλεσμα που ισούται με 0 σημαίνει πως δεν εμφανίζεται κάποια.

Κάνοντας χρήση κάποιου δείκτη για κάθε συνδυασμό γνωρισμάτων, μπορεί να δημιουργηθεί ένας τετραγωνικός πίνακας [i x j], όπου i = j = το πλήθος τους, κάθε κελί του οποίου έχει μια τιμή μεταξύ -1 και +1. Καθώς στην παρούσα έρευνα δεν πραγματοποιείται η παράλληλη πρόγνωση χρονοσειρών αλλά η πρόγνωση μόνο μίας λαμβάνοντας ως δεδομένα εισόδου τις διακυμάνσεις των υπολοίπων (διαδικασία που περιγράφεται στο υποκεφάλαιο 3.2.1), εξετάζεται η συσχέτιση μίας μόνο περιοχής (j = Αθήνα), με τα υπόλοιπα γνωρίσματα του dataset (i = 2664) και δοκιμάστηκαν διάφορες εναλλακτικές τιμές του δείκτη ρ. Η συγγραφή του κώδικα έγινε κατά τέτοιον τρόπο ώστε η επιλογή των σημείων (features) για το μοντέλο να πραγματοποιείται με αυτοματοποιημένο τρόπο θέτοντας απλώς τα επιθυμητά όρια του συγκεκριμένου δείκτη ρ.

3.3: Δημιουργία Νευρωνικού Δικτύου

Τα RNN (και ειδικότερα τα LSTM) δίκτυα θεωρούνται κατ' αρχήν ως τα πλέον κατάλληλα για την εκπαίδευσή τους σε χρονοσειρές. Στο πλαίσιο της παρούσας εργασίας, διερευνήθηκε η δυνατότητα εκμάθησης ενός LSTM δικτύου πάνω σε δεδομένα χρονοσειρών.

3.3.1: Ορισμός της Αρχιτεκτονικής του Δικτύου

Ιστορικά, τα RNN/LSTM δίκτυα αναπτύχθηκαν για να επιλύσουν προβλήματα όπου τα δεδομένα οργανώνονται σε χρονοσειρές, ενώ από την άλλη πλευρά τα CNN δίκτυα αναπτύχθηκαν αρχικά για επίλυση προβλημάτων κατηγοριοποίησης και αναγνώρισης εικόνων. Πιο σύγχρονες προσεγγίσεις μπορεί να περιλαμβάνουν όμως μικτά CNN-LSTM δίκτυα, δίκτυα ConvLSTM, νέες μορφές όπως τα GCN ή επεκτάσεις των ήδη υπάρχοντων. Στις Deep Learning προσεγγίσεις όμως, όποιο δίκτυο (ή παραλλαγή του) κι αν χρησιμοποιηθεί, τα τρία βασικά στοιχεία της αρχιτεκτονικής τους είναι παρόντα και περιλαμβάνουν πάντα 1) ένα επίπεδο εισόδου, 2) ένα ή περισσότερα ενδιάμεσα επίπεδα και 3) ένα επίπεδο εξόδου.

Η πολυμορφία των δικτύων οφείλεται στο γεγονός ότι με τη χρήση βασικών δομικών στοιχείων, όπως για παράδειγμα το είδος των νευρώνων και οι τρόποι διασύνδεσής τους, μπορούν να επιτευχθούν πολλοί διαφορετικοί συνδυασμοί. Ανεξαρτήτως όμως των επιμέρους στοιχείων, τα οποία πολλές φορές ανταποκρίνονται στις ανάγκες κάποιου προβλήματος, ο αριθμός των επιπέδων και των νευρώνων σε κάθε επίπεδο αποτελούν τις βασικότερες παραμέτρους. Αν και δεν υπάρχει κάποιος αριθμητικός κανόνας που να συσχετίζει τα παραπάνω μεγέθη με τις βασικές διαστάσεις των δεδομένων εκπαίδευσης (όπως ο αριθμός των γνωρισμάτων και το χρονικό διάστημα όσον αφορά τις χρονοσειρές), προκύπτει[10] πως:

1. Ο αριθμός των νευρώνων εισόδου θα πρέπει να είναι τουλάχιστον ίσος με τον αριθμό των γνωρισμάτων + ένας ακόμη.

2. Η χρήση ενός ενδιάμεσου επιπέδου είναι συνήθως αρκετή, ο αριθμός των νευρώνων του οποίου μπορεί να είναι ο μέσος όρος του αριθμού του πρώτου και του τελευταίου επιπέδου.

3. Ο αριθμός των νευρώνων εξόδου θα πρέπει να είναι ίσος με τον αριθμό των προβλεπόμενων τιμών, ή των προβλεπόμενων χρονικών βημάτων.

Ερμηνεύοντας λίγο περισσότερα τα παραπάνω σημεία (1,2,3), είναι χρήσιμο να αναφερθεί ότι:

---

[10] *Κατά την άποψη του συγγραφέα, το παρακάτω άρθρο - απάντηση σε σχετικό ερώτημα, συνοψίζει αρκετά καλά την υπάρχουσα γνώση, εμπειρία και πρακτική γύρω από αυτό το ζήτημα:* *https://stats.stackexchange.com/questions/181/how-to-choose-the-number-of-hidden-layers-and-nodes-in-a-feedforward-neural-netw/136542#136542*

1. Είναι προφανές πως ο αριθμός των νευρώνων εισόδου (το πλάτος του δικτύου) σχετίζεται απόλυτα με τον αριθμό των γνωρισμάτων (ο αριθμός των στηλών στα δεδομένα).
2. Το βάθος του δικτύου (ο αριθμός των επιπέδων - CAP) σχετίζεται τελικά με την εξαγωγή γενικότερων χαρακτηριστικών από τα δεδομένα, λόγω του αυξημένου αριθμού των διασυνδέσεων που θα προκύψουν. Ωστόσο, στην παρούσα έρευνα και με βάση τα όσα παρουσιάστηκαν στο υποκεφάλαιο 3.2.2.3, η συγκεκριμένη διαδικασία αποτέλεσε κομμάτι της προεπεξεργασίας των δεδομένων, κατά συνέπεια σε όλες τις περιπτώσεις εναλλακτικών υλοποιήσεων δεν προκύπτει η ανάγκη ορισμού περισσότερων του ενός ενδιάμεσων επιπέδων.
3. Οι νευρώνες εξόδου είναι όσες και οι ημέρες πρόγνωσης.

3.3.2: Κύκλοι εκπαίδευσης και αποφυγή overfitting

Κατά την εκπαίδευση κάθε δικτύου το ενδιαφέρον εστιάζεται σε συγκεκριμένες μετρικές όπως loss και accuracy, οι οποίες εκφράζουν τη μέση απόκλιση σε σχέση με τις πραγματικές τιμές και την ακρίβεια αντίστοιχα. Ιδίως όσον αφορά τη συνάρτηση loss, υπάρχει η δυνατότητα πολλών επιλογών με κύριες τις Mean Squared Error (MSE), Mean Absolute Error (MAE), Categorical Cross-Entropy κ.ά. Στη συγκεκριμένη εργασία έγινε η χρήση της MSE (2.2), η οποία στην περίπτωση αυτή ορίζεται ως:

$$MSE = \frac{1}{N}\sum_{(x,y)\in D} (y - prediction(x))^2 \qquad (3.6)$$

όπου D: το dataset, (x,y): τα ζεύγη των πραγματικών και προβλεπόμενων τιμών αντίστοιχα και N: ο αριθμός των παραδειγμάτων (ζευγών).

Η ακρίβεια πρόβλεψης (accuracy) ορίζεται ως εξής (ως ποσοστό %):

$$accuracy = (1 - \frac{|y-\hat{y}|}{y})\% \qquad (3.7)$$

όπου $\hat{y}$: η πρόγνωση, y: η πραγματική τιμή.

Κάθε κύκλος εκπαίδευσης ονομάζεται “epoch”, όπου υπολογίζονται οι loss και accuracy μετρικές. Κατά τον ορισμό του δικτύου προσδιορίζεται μέσω ειδικής μεταβλητής το ποσοστό του dataset εκπαίδευσης, το οποίο θα χρησιμοποιηθεί για testing (συνήθως το 20%), συνεπώς προκύπτουν οι μεταβλητές val_loss και val_accuracy αντίστοιχα οι οποίες αφορούν μόνο αυτό και ***δεν θα πρέπει να συγχέεται με τον αρχικό διαχωρισμό σε train / test dataset που προσδιορίστηκε στο υποκεφάλαιο 3.2.1***. Ο βασικός λόγος που τελικά ορίζονται δύο

ομάδες μετρικών (loss / val_loss και accuracy / val_accuracy) είναι για να ανιχνευθεί το φαινόμενο του “overfitting” (υπερ-εκπαίδευσης), κατά το οποίο ένα νευρωνικό δίκτυο μετά από πολλούς κύκλους εκπαίδευσης τείνει να ανταποκρίνεται πολύ καλά στα δεδομένα εκπαίδευσής του, αδυνατώντας όμως να πραγματοποιήσει σωστές προβλέψεις, όταν εκτεθεί σε νέα (άγνωστα) δεδομένα. Για την αποφυγή του “overfitting” ισχύει ο γενικός κανόνας: loss ≥ val_loss και accuracy ≤ val_accuracy ενώ η εκπαίδευση του μοντέλου θα πρέπει να σταματάει, όταν μετά από k αριθμό epochs ισχύει loss ≈ val_loss και accuracy ≈ val_accuracy, το οποίο θεωρείται ως το βέλτιστο σημείο. Όπως είναι προφανές, τόσο η ακρίβεια όσο και το σφάλμα δεν μπορούν ποτέ να λάβουν την μέγιστη και ελάχιστη (αντίστοιχα) τιμή τους χωρίς το μοντέλο να οδηγηθεί σε κατάσταση overfitting, συνεπώς αναζητείται η καλύτερη αρχιτεκτονική και οι επιμέρους παραμετροποιήσεις που μπορούν να οδηγήσουν σε βελτίωση των συγκεκριμένων μετρικών με την ταυτόχρονη μείωση των κύκλων εκπαίδευσης k αλλά και του συνολικού χρόνου για κάθε μία εξ αυτών.

3.4: Εργαλεία

Στα σύγχρονα προγραμματιστικά περιβάλλοντα, οι τρόποι επίλυσης ενός προβλήματος είναι συνήθως πολλοί, εξαρτώμενοι από την επιλογή της γλώσσας ή των γλωσσών που θα χρησιμοποιηθούν. Οι περισσότερες εξ αυτών, υποστηρίζονται από τα περισσότερα λειτουργικά συστήματα (ενδεικτικά, Unix / MacOs / Windows). Έτσι λοιπόν, μια σχετικά χαμηλού επιπέδου, συναρτησιακή και αντικειμενοστραφής γλώσσα όπως η C++ ή η Java θα μπορούσαν να χρησιμοποιηθούν για τη κατασκευή των νευρώνων, των σχέσεων μεταξύ τους, της διαδικασίας εκπαίδευσής τους αλλά και της ανάλυσης των αποτελεσμάτων. Μερικά κριτήρια επιλογής θα μπορούσαν να είναι:

1) το είδος του προβλήματος,
2) η παρεχόμενη υποστήριξη,
3) το εύρος των διαθέσιμων βιβλιοθηκών,
4) ο τρόπος προσέγγισης του προβλήματος/συγγραφής του κώδικα (high/low level),
5) η ταχύτητα εκτέλεσης,
6) η μεταφερσιμότητά του,
7) η επικοινωνία με άλλες.

Οι προαναφερθείσες γλώσσες πληρούν τα περισσότερα εξ αυτών, όμως ιδιαίτερα όσον αφορά το επίπεδο συγγραφής (σημείο 4), αποτελούν μια περισσότερο low-level επιλογή. Λαμβάνοντας υπόψη τις ανάγκες του πειράματος, η Python ως περισσότερο υψηλού επιπέδου

γλώσσα κρίθηκε καταλληλότερη, εξαιτίας των πολλών διαθέσιμων βιβλιοθηκών και στατιστικών πακέτων που τις υπερκαλύπτουν. Από την άλλη πλευρά, η Python σε σύγκριση με κάποια χαμηλότερου επιπέδου γλώσσα παρουσιάζει πιο μεγάλους χρόνους εκτέλεσης.

Με βάση τα παραπάνω, όλη η διαδικασία της επεξεργασίας των δεδομένων, της δημιουργίας των νευρωνικών δικτύων / μοντέλων, η στατιστική ανάλυση των αποτελεσμάτων και η οπτικοποίησή τους, πραγματοποιήθηκαν σχεδόν αποκλειστικά με την χρήση της γλώσσας python στην έκδοση 3.9.13 (17/5/2022, περισσότερες πληροφορίες εδώ: https://www.python.org/downloads/release/python-3913/), αλλά και με τη χρήση συγκεκριμένων εξειδικευμένων βιβλιοθηκών και πακέτων, για τα οποία θα γίνει αναφορά στο υποκεφάλαιο 3.4.1. Το λειτουργικό σύστημα στο οποίο εκτελέστηκαν τα πειράματα είναι το MacOS στην έκδοση 10.13.6 (High Sierra), με επεξεργαστή Intel i5 δύο πυρήνων, μνήμη 4 GB και με σκληρό δίσκο τύπου SSD, χωρητικότητας 1 TB.

Προκειμένου να ελαχιστοποιηθούν τα θέματα συμβατότητας σε περίπτωση μεταφοράς του κώδικα σε άλλο σύστημα, δημιουργήθηκε εξ' αρχής ένα εικονικό περιβάλλον (virtual environment), εντός του οποίου πραγματοποιήθηκε η εγκατάσταση των απαραίτητων βιβλιοθηκών. Το εικονικό περιβάλλον δημιουργείται με το module "venv", το οποίο εγκαθιστά την python σε φάκελο που ορίζει ο χρήστης (στην ίδια έκδοση με αυτήν της οποίας το module venv κλήθηκε). Σκοπός του εικονικού περιβάλλοντος είναι να μπορεί να έχει τις δικές του βιβλιοθήκες εγκατεστημένες, ανεξάρτητες από το υπόλοιπο σύστημα, κάτι το οποίο τελικά στην πλειονότητα των περιπτώσεων μπορεί να επιτευχθεί (περισσότερες πληροφορίες: https://docs.python.org/3.9/library/venv.html)

3.4.1 Βιβλιοθήκες και APIs

Η εγκατάσταση των βιβλιοθηκών πραγματοποιείται με τη χρήση της εντολής "*python -m pip install*" ως εξής:

```
python3 -m pip install <library>==<version>
```

Στον πίνακα 3.6 παρατίθενται όσες χρησιμοποιήθηκαν καθώς και η έκδοσή τους, χωρίς να γίνεται αναφορά σε όσες είναι απαραίτητες για την λειτουργία τους (dependencies).

Κατηγορία	Βιβλιοθήκη	Έκδοση	Πληροφορίες
Άντληση Δεδομένων	requests	2.28.1	https://pypi.org/project/requests/
Ανάλυση Δεδομένων	pandas	1.4.3	https://pandas.pydata.org/
	numpy	1.19.5	https://numpy.org/
Οπτικοποίηση	plotly	5.9.0	https://plotly.com/
	matplotlib	3.5.2	https://matplotlib.org/
Νευρωνικά Δίκτυα	keras &	2.6.0	https://keras.io/
	tensorflow	2.6.5	https://www.tensorflow.org/

Πίνακας 3.6: *Οι βιβλιοθήκες της python και οι εκδόσεις τους που χρησιμοποιήθηκαν για κάθε κατηγορία εργασιών.*

Σχετικά με την **Άντληση, Ανάλυση και Οπτικοποίηση των δεδομένων**, το κριτήριο της επιλογής των συγκεκριμένων βιβλιοθηκών που παρουσιάζονται στον πίνακα 3.6 , ήταν κυρίως η *ευκολία εκμάθησης και χρήσης* τους. Σε σχέση όμως με τη **ορισμό και την εκπαίδευση των νευρωνικών δικτύων**, επιπλέον παράγοντες για την επιλογή τους ήταν και η *δημοφιλία - παρεχόμενη υποστήριξη* και η *καλή συνεργασία με το υπόλοιπο λειτουργικό σύστημα*, ιδίως ως προς τη διεπαφή με την python. Για το συγκεκριμένο στάδιο της έρευνας, η βιβλιοθήκη tensorflow (η οποία συνδέεται στενά με το keras api) και το περιβάλλον της PyTorch θεωρούνται τα πληρέστερα, μεταξύ διαφόρων ακόμη διαθέσιμων επιλογών (MXNet, Theano, Caffe2, CNTK, κ.ά.).

Στα υποκεφάλαια 3.4.1.1 και 3.4.1.2 παρακάτω, παρουσιάζονται και συγκρίνονται αναλυτικά οι δυνατότητες αλλά και οι επιδόσεις των πιο γνωστών διαθέσιμων βιβλιοθηκών για τη δημιουργία και εκπαίδευση νευρωνικών δικτύων.

3.4.1.1: Keras / Tensorflow Vs Pytorch

Η βιβλιοθήκη **Tensorflow** (η οποία υπάρχει διαθέσιμη και για την Javascript), δημιουργήθηκε από την Google το 2015 και αποτελεί έργο ανοιχτού κώδικα (Abadi et al., 2016) βασισμένη στην προγενέστερη βιβλιοθήκη Theano. Είναι στενά συνδεδεμένη με το keras, το οποίο είναι ουσιαστικά ένα περιβάλλον διεπαφής (API) για την tensorflow (καθώς

και τη Theano και την CNTK). Μεταξύ των κύριων χαρακτηριστικών της είναι ότι δημιουργεί ένα υπολογιστικό γράφο (computational graph), τα στοιχεία του οποίου (κόμβοι), συνδέονται μεταξύ τους μέσω συναρτησιακών σχέσεων τύπου $z = f(x)$, όπου *f: sigmoid, relu, κ.ά.* ενώ *z, x*, αποτελούν πίνακες δεδομένων n διαστάσεων. (https://wiki.pathmind.com/comparison-frameworks-dl4j-tensorflow-pytorch, πρόσβαση: Αύγουστος 2022) . Οι τιμές αυτές ονομάζονται "tensors" διατρέχοντας (flow) τους κόμβους (nodes) του δικτύου. Ένα από τα βασικά πλεονεκτήματά της είναι ότι ο χρήστης έχει τη δυνατότητα επιλογής μεταξύ όλων των γνωστών συναρτήσεων κόστους και των συναρτήσεων ενεργοποίησης κατά τη διάρκεια του ορισμού των επιπέδων του δικτύου. Από την άλλη πλευρά, ο ορισμός των υπολογιστικών γράφων του δικτύου γίνεται στατικά, πριν τη διαδικασία της εκπαίδευσής του.

Το περιβάλλον **PyTorch** παρουσιάστηκε ως ανοιχτού κώδικα τον Οκτώβριο του 2016 από το Facebook (Paszke et al., 2017), ως η python έκδοση της ήδη υπάρχουσας Torch, ενώ πλέον έχει ενσωματώσει και την Caffe2 (https://caffe2.ai/, πρόσβαση, Αύγουστος 2022). Η μεγαλύτερη διαφορά της σε σχέση με τις υπόλοιπες βιβλιοθήκες είναι ότι ο ορισμός των υπολογιστικών γράφων γίνεται με δυναμικό τρόπο (dynamic computational graphs) (Elsawi et al., 2021), χαρακτηριστικό το οποίο επιτρέπει στον χρήστη την δημιουργία του δικτύου κατά τη διάρκεια της εκπαίδευσής του. Αυτό μπορεί να αποδειχθεί αρκετά χρήσιμο στην περίπτωση των RNN δικτύων (Awni Hanun, https://awni.github.io/pytorch-tensorflow/, πρόσβαση: Αύγουστος 2022), λόγω της μεταβλητής φύσης των δεδομένων εισαγωγής. Παρ' όλο που η δυνατότητα αυτή υπάρχει και στην Tensorflow (Control Flow Operations), όπως για παράδειγμα με τη χρήση της `tensorflow.while_loop` , η αποσφαλμάτωση της διαδικασίας είναι δυσκολότερη σε σχέση με την PyTorch. Από την άλλη πλευρά όμως, για τον ίδιο ακριβώς λόγο, αρκετά χαρακτηριστικά της αρχιτεκτονικής ενός νευρωνικού δικτύου, θα πρέπει να οριστούν από το χρήστη και δεν υπάρχουν έτοιμα προς χρήση.

Η PyTorch θα μπορούσε να χαρακτηριστεί ως περισσότερο "low-level" σε σχέση με την tensorflow και συνηθέστερα αποτελεί επιλογή σε ερευνητικό επίπεδο, όταν το ζητούμενο είναι η μελέτη / πειραματισμός πάνω σε θέματα της αρχιτεκτονικής των δικτύων. Η tensorflow υλοποιεί τα μοντέλα σε περισσότερο "high-level" επίπεδο, αν και, όπως αναφέρεται, η διαδικασία της εκμάθησής της σχετικά με τη λειτουργικότητά της μπορεί να είναι δυσκολότερη, όπως για παράδειγμα τα sessions, placeholders κ.ά. (Yashwardhan Jain, 2018).

Τα συγκεκριμένα χαρακτηριστικά καθώς και μερικά ακόμη παρουσιάζονται συνοπτικά στον συγκριτικό πίνακα 3.7:

Χαρακτηριστικά	**PyTorch**	**Tensorflow**
Έλεγχος κώδικα	Ευκολότερος	Δυσκολότερος
Ταχύτητα εκτέλεσης	Υψηλή	Υψηλή
Documentation	Λιγότερο αναλυτικό	**Πλήρες / Κατανοητό ✓**
Δημοφιλία / Κοινότητα	Μικρότερη	**Μεγαλύτερη ✓**
Όγκος Δεδομένων	Μεγάλος	Μεγάλος
Δυσκολία εκμάθησης	Μικρότερη	Μεγαλύτερη
Οπτικοποίηση Μοντέλων	Δυσκολότερη	Ευκολότερη

Πίνακας 3.7: *Σύγκριση των γενικών χαρακτηριστικών των βιβλιοθηκών PyTorch και Tensorflow*

Ενώ και οι δύο επιλογές είναι κατάλληλες για το ερευνητικό πλαίσιο της παρούσας εργασίας, η υλοποίηση των μοντέλων πραγματοποιήθηκε τελικά με τη χρήση της tensorflow, κυρίως λόγω του πληρέστερου documentation και της δημοφιλίας της, χαρακτηριστικά που οδηγούν σε μεγαλύτερο όγκο σχετικής βιβλιογραφίας αλλά και πλήθος σχετικών παραδειγμάτων και υλοποιήσεων.

3.4.1.2: Άλλα περιβάλλοντα / βιβλιοθήκες

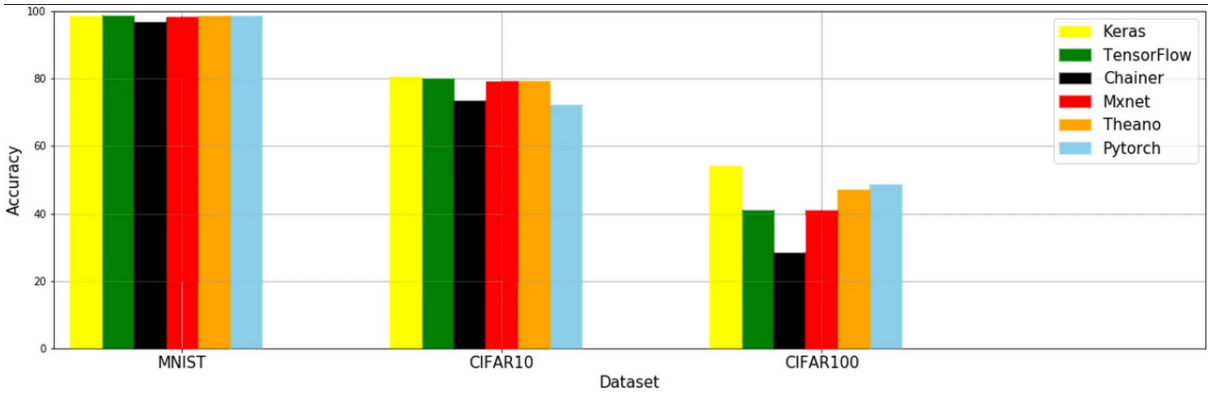
Μερικές ακόμη, λιγότερο δημοφιλείς, βιβλιοθήκες είναι οι Theano, Chainer, DyNET MXNet, Microsoft Cognitive Toolkit (CNTK) κ.ά. Πιο αναλυτικά:

- Όπως ήδη αναφέρθηκε, η **Theano** θεωρείται ο πρόγονος της Tensorflow και αναπτύχθηκε από το πανεπιστήμιο του Montreal το 2007 ως μια βιβλιοθήκη της Python εξειδικευμένη σε μαθηματικούς υπολογισμούς μεταξύ πινάκων, κάτι που την καθιστούσε κατάλληλη και για την επίλυση προβλημάτων μηχανικής εκμάθησης. Η εξέλιξή της σταμάτησε το 2017 (έκδοση 1.0.0), κυρίως λόγω της ύπαρξης περισσότερων σύγχρονων βιβλιοθηκών (πηγή: wikipedia), ενώ η τελευταία ανανέωσή της πραγματοποιήθηκε τον Ιούλιο του 2020 (έκδοση 1.0.5) (https://pypi.org/project/Theano/).

- Το **DyNET (Dynamic Neural Network Toolkit)** δημιουργήθηκε από ερευνητές του πανεπιστημίου Carnegie Mellon και άλλων ιδρυμάτων (Neubig et al., 2017). Πρόκειται για μια βιβλιοθήκη της C++, η οποία όμως υπάρχει διαθέσιμη και στην Python, εστιασμένη στη δημιουργία δυναμικών υπολογιστικών γράφων.
- Η **Chainer** είναι ένα ανοιχτού κώδικα περιβάλλον διεπαφής με την Python, το οποίο υλοποιήθηκε από τους προγραμματιστές της εταιρείας Preffered Networks με έδρα το Tokyo. Πρόκειται για το πρώτο με τη δυνατότητα υλοποίησης *δυναμικών υπολογιστικών γράφων* (πριν την εμφάνιση των DyNET και PyTorch), ενώ οι δημιουργοί του αναφέρουν πως είναι ταυτόχρονα και το γρηγορότερο υπολογιστικά (https://github.com/clab/dynet) .
- Το **MXNet** αποτελεί συνεργασία των ερευνητών του πανεπιστημίου Carnegie Mellon, Washington University και της Microsoft και αναφέρεται ως "scalable framework" ανοιχτού κώδικα, το οποίο επιτρέπει την δημιουργία δικτύων βαθιάς μηχανικής εκμάθησης σε διάφορες γλώσσες συμπεριλαμβανομένης της C++, Python, MATLAB, Javascript, R, Julia και Scala. Υποστηρίζει τη δυνατότητα παράλληλου προγραμματισμού δεδομένων και μοντέλων σε CPU και GPU καθώς και την σύγχρονη / ασύγχρονη εκπαίδευση των μοντέλων (Chen et al., 2015).
- Το **Microsoft Cognitive Toolkit** (πρώην CNTK - Computational Network Toolkit) είναι ένα περιβάλλον ανοιχτού κώδικα της Microsoft υλοποιημένο σε C++, το οποίο προσφέρει τη δυνατότητα διεπαφής με την Python. Περιέχει βιβλιοθήκες για υλοποίηση δικτύων τύπου MLP, CNN και RNN, ωστόσο η άδειά του δεν επιτρέπει τη χρήση του για εμπορικούς σκοπούς.

Τα συγκεκριμένα περιβάλλοντα / βιβλιοθήκες, αν και λιγότερο δημοφιλή, εμφανίζουν κατά περίπτωση και αναλόγως της φύσης και τους σκοπούς κάθε προβλήματος, αξιόλογες δυνατότητες. Ένας τρόπος σύγκρισης της ικανότητας εκμάθησής τους είναι η επίλυση γνωστών ML προβλημάτων χρησιμοποιώντας τις αρχικές προεπιλεγμένες τους ρυθμίσεις για CNN ή RNN / LSTM δίκτυα. Για παράδειγμα στο διάγραμμα του σχήματος 3.5 (Elshawi et al, 2021), συγκρίνεται η ακρίβεια των Keras, Tensorflow, Chainer, MXNet, Theano και Pytorch για προβλήματα αναγνώρισης εικόνων από τις βάσεις δεδομένων MNIST, CIFAR10 και CIFAR100, όπου με εξαίρεση την περίπτωση της CIFAR100, δεν διακρίνεται κάποια στατιστικά σημαντική διαφορά μεταξύ τους. Ανάλογη είναι και η εικόνα για LSTM

αρχιτεκτονικές όπου επιλύονται προβλήματα όπως η κατηγοριοποίηση σχολίων (Sentiment Analysis) στη βάση δεδομένων imdb κ.ά.



Σχήμα 3.5: *Ακρίβεια στην αναγνώριση εικόνων των βάσεων δεδομένων MNIST, CIFAR10 και CIFAR100 μερικών γνωστών ML Frameworks, χρησιμοποιώντας τις αρχικές τους ρυθμίσεις για τη δημιουργία δικτύων τύπου CNN. Πηγή: (Elshawi et al, 2021).*

3.5 Περιγραφή του κώδικα υλοποίησης

. Ο πυρήνας όλης της υλοποίησης είναι ο ορισμός των χαρακτηριστικών και της δομής του μοντέλου τεχνητής νοημοσύνης με τη χρήση των βιβλιοθηκών keras/tensorflow (υποκεφάλαιο 3.4.1) και η τροφοδοσία του με τα μετεωρολογικά δεδομένα της εκπαίδευσης και της αξιολόγησής του. Αυτά όμως, θα πρέπει πρώτα να έχουν υποστεί την κατάλληλη επεξεργασία (διαδικασία που περιγράφεται αναλυτικά στο υποκεφάλαιο 3.2). Με την ολοκλήρωση της εκπαίδευσής του, καλείται να προβλέψει τις νέες τιμές έχοντας ως είσοδο άλλα δεδομένα τα οποία δεν του έχουν παρουσιαστεί (η διαδικασία του διαχωρισμού τους περιγράφεται αναλυτικά στο υποκεφάλαιο 3.2.1.). Τα αποτελέσματα των προγνώσεων αποθηκεύονται και αφού μετατραπούν στην αρχική τους μορφή αναλύονται στατιστικά ενώ ένα άλλο τμήμα του κώδικα αναλαμβάνει την γραφική απεικόνιση των τιμών για το χρονικό διάστημα ενδιαφέροντος. Βάσει των παραπάνω, η γενική δομή της υλοποίησης του κώδικα περιλαμβάνει διάφορα αρχεία γύρω από το κεντρικό που αφορά στην εκπαίδευση και εκτέλεση (τρέξιμο) του μοντέλου.

Το σύνολο του κώδικα εκτελεί τέσσερις διαφορετικές λειτουργίες:

- Συλλογή, επεξεργασία και διαχωρισμός των δεδομένων που θα εκπαιδεύσουν και θα τροφοδοτήσουν το μοντέλο.
- Διαγνωστικοί έλεγχοι των χρονοσειρών (ADF tests), οι οποίοι αποτελούν κριτήριο καταλληλότητας για την τροφοδοσία των δεδομένων στο μοντέλο.

- Ορισμός των χαρακτηριστικών και της δομής του μοντέλου και η εκπαίδευσή του.
- Συλλογή των αποτελεσμάτων της πρόγνωσης, γραφική απεικόνιση και στατιστική ανάλυσή των αποτελεσμάτων.

Στην εικόνα του σχήματος 3.7 παρουσιάζεται το διάγραμμα της ροής του κώδικα της εργασίας.

Εκτός της γενικής αρχιτεκτονικής του κώδικα, αξίζει να αναφερθεί ότι τμήματα του έχουν γραφτεί κατά τέτοιον τρόπο ώστε να είναι εφικτή η επαναχρησιμοποίησή τους (functional programming), όπου αυτό ήταν εφικτό. Στην εικόνα του σχήματος 3.7 παρουσιάζεται το διάγραμμα ροής του κώδικα της εργασίας, ενώ στα επόμενα υποκεφάλαια αναλύονται τα προαναφερθέντα τμήματα του κώδικα.

3.5.1: Κώδικας συλλογής, επεξεργασίας και διαχωρισμού των δεδομένων

Προκειμένου να καταστεί εφικτή η ανάλυση / επεξεργασία των δεδομένων και η εισαγωγή τους στο νευρωνικό δίκτυο, είναι απαραίτητο να έχουν συγκεκριμένη μορφή, η οποία είναι κάποιος πίνακας $n \in N^*$ διαστάσεων ή συνηθέστερα διαδοχικοί πίνακες ίδιων διαστάσεων, οι οποίοι αντιπροσωπεύουν τα δεδομένα καθεαυτά ή κάποια αναπαράστασή τους. Αναλόγως, η έξοδος του δικτύου είναι και πάλι ένας ή διαδοχικοί πίνακες, οι οποίοι επίσης αντιπροσωπεύουν κάποια αναπαράστασή των ή της λύσης του προβλήματος (εάν λ.χ. πρόκειται για classification θα μπορούσε να είναι πίνακας-αριθμός, δυαδικής μορφής).

Ως εκ τούτου, εφ' όσον η κύρια πηγή των δεδομένων είναι σε grib1 / grib2 μορφή, θα πρέπει πρώτα να διαβαστούν, στη συνέχεια να επιλεχθούν ως προς τις παραμέτρους και την περιοχή ενδιαφέροντος και έπειτα να μετατραπούν στην κατάλληλη μορφή. Διάφορα εξειδικευμένα εργαλεία είναι το **wgrib / wgrib2** (https://www.cpc.ncep.noaa.gov/products/wesley/wgrib2/) που παρέχεται από το NOAA (και άλλους), το **ecCodes** (https://confluence.ecmwf.int/display/ECC) του Ευρωπαϊκού Κέντρου Μεσοπρόθεσμων Προγνώσεων (ECMWF), το **Climate Data Operators (CDO)** του Ινστιτούτου Max Planck (https://code.mpimet.mpg.de/projects/cdo/wiki/Tutorial) κ.ά. Για τις ανάγκες της έρευνας χρησιμοποιήθηκε το ecCodes και πιο συγκεκριμένα οι εντολές `grib_copy` και `grib_get_data` για τον διαχωρισμό της παραμέτρου του γεωδυναμικού ύψους στα 500hpa από το υπόλοιπο grib αρχείο και για την εξαγωγή των τιμών της αντίστοιχα,

οι οποίες στη συνέχεια αποθηκεύονται σε ημερήσια csv αρχεία. Τα αρχεία αυτά έχουν τη μορφή: Date, Latitude, Longitude, Value. Όπως αποδείχθηκε, τα συγκεκριμένα δεδομένα ήταν πλήρη χωρίς null τιμές. Στο σχήμα 3.8 παρουσιάζεται ο έλεγχος της πληρότητας των τιμών μέσω της συνάρτησης notnull της βιβλιοθήκης pandas.

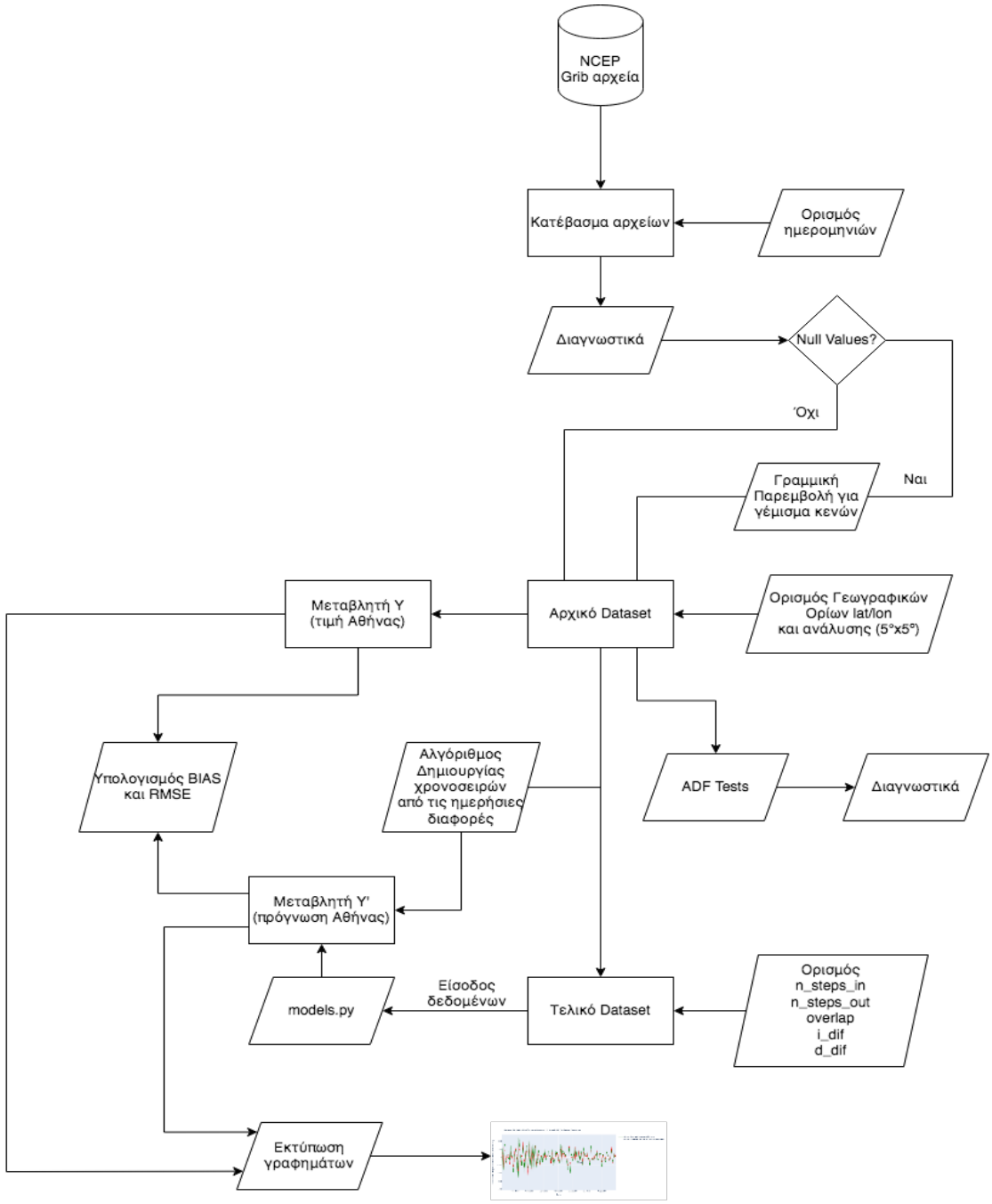
```
[Clang 10.0.0 (clang-1000.11.45.5)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> import pandas as pd
>>> dataframe = pd.read_csv('dataframe_big.csv')
>>> print(dataframe.shape)
(1338, 2667)
>>> not_null_values = pd.notnull(dataframe)
>>> print(not_null_values.shape)
(1338, 2667)
>>>
```

Σχήμα 3.6: *Έλεγχος για null τιμές στο αρχικό dataset 1338 ημερών x 2667 features.*

Σχετικά με την εξαγωγή δεδομένων από τα grib αρχεία δοκιμάστηκε επίσης η χρήση της βιβλιοθήκης **xarray** (https://docs.xarray.dev) της python, η οποία εκτός των άλλων έχει τη δυνατότητα ανάγνωσης αρχείων τύπου grib (1ή2) (https://docs.xarray.dev/en/stable/examples/ERA5-GRIB-example.html) μέσω της χρήσης της cfgrib (https://github.com/ecmwf/cfgrib). Όμως στις δοκιμές εγκατάστασης που πραγματοποιήθηκαν, διαπιστώθηκε η **έλλειψη συμβατότητας της cfgrib με την έκδοση του λειτουργικού συστήματος** όπου εκτελείται το πείραμα, *ανεξαρτήτως της έκδοσης της python μέσω της οποίας δημιουργήθηκε το εικονικό περιβάλλον*, γεγονός που καθιστά το συγκεκριμένο κομμάτι του κώδικα μη μεταφέρσιμο για όλα τα λειτουργικά συστήματα. Συνεπώς, ενώ τα αρχεία τύπου grib μπορούν να παίξουν το ρόλο της βάσης μετεωρολογικών δεδομένων λόγω της φύσης τους (υπερσυμπιεσμένα binary μορφής) χωρίς να υπάρχει η ανάγκη μετατροπής τους σε csv, αυτό ήρθε σε αντίθεση με τη δυνατότητα τρεξίματος του κώδικα σε άλλα λειτουργικά συστήματα, κάτι που όμως αποτελεί σημαντικό ζητούμενο σε ερευνητικές εργασίες.

Εφ’ όσον είναι δυνατή η μετατροπή των μετεωρολογικών αρχείων σε μορφή csv κατά τέτοιον τρόπο ώστε να είναι πλήρη, στη συνέχεια εισάγονται σε κάποιο ANN μέσα από επιπρόσθετους μετασχηματισμούς, τους οποίους αναλαμβάνει ο κώδικας σε επόμενο στάδιο. Τμήματά του έχουν γραφτεί σε όσο το δυνατόν πιο γενικό επίπεδο (functional programming) με σκοπό την επαναχρησιμοποίησή τους, ώστε να ανταποκρίνονται στο δίκτυο που πρόκειται να δημιουργηθεί. Η δημιουργία κάθε νευρωνικού δικτύου ορίζεται σε ξεχωριστό αρχείο, το

οποίο μπορεί να κληθεί ως συνάρτηση από το κυρίως σώμα του κώδικα. Κατά τον ίδιο τρόπο, ένα άλλο τμήμα του αναλαμβάνει την εκτύπωση διαγνωστικών στοιχείων σε σχέση με τις υπό μελέτη χρονοσειρές, την εκπαίδευση και επίδοση του δικτύου, την εκτύπωση των προγνώσεων και τις στατιστικές συγκρίσεις σε σχέση με τα πραγματικά (ground truth) δεδομένα.



Σχήμα 3.7: *Η ροή του κώδικα που αναπτύχθηκε στο πλαίσιο της παρούσας εργασίας*

Όλος ο κώδικας που αναπτύχθηκε στο πλαίσιο της παρούσας εργασίας βρίσκεται στη διεύθυνση: ***https://github.com/zarakoinos/ai-weather-forecast***

Κεφάλαιο 4° : Πειραματική Διαδικασία

Για την εκπαίδευση των νευρωνικών δικτύων που θα παρουσιαστούν στη συνέχεια χρησιμοποιήθηκαν δεδομένα 1338 ημερών (1/1/2017 έως 31/8/2020), οι οποίες χωρίστηκαν σε 1095 εκπαιδευτικές ημέρες (81,8%) και 243 ημέρες ελέγχου (18,2%), από 1/1/2017 έως 31/12/2019 και από 1/1/2020 εώς 31/8/2020 αντίστοιχα. Το συγκεκριμένο dataset θεωρείται πως καλύπτει επαρκώς το ζητούμενο της έρευνας, καθώς περιλαμβάνει τρία έτη εκπαίδευσης και 9 μήνες αξιολόγησης, συμπεριλαμβάνοντας το μεγαλύτερο φάσμα της εποχικής μεταβλητότητας.

Μεταξύ δυνητικά πολλών εναλλακτικών υλοποιήσεων που θα ήταν εφικτό να ελεγχθούν, το ζητούμενο περιορίστηκε στον αρχικό έλεγχο τριών παραμέτρων, οι οποίες είναι: *ο αριθμός των γνωρισμάτων*, *ο αριθμός των νευρώνων / συνάψεων και οι κύκλοι εκπαίδευσης*. Η βέλτιστη τιμή της πρώτης παραμέτρου (αριθμός γνωρισμάτων) υπήρξε το αντικείμενο διερεύνησης όλης της πειραματικής διαδικασίας, όπως παρουσιάζεται στα υποκεφάλαια 4.4.1, 4.4.2 και 4.4.3, ενώ οι τιμές των υπολοίπων δύο (αριθμός συνάψεων και κύκλοι εκπαίδευσης) καθορίστηκαν εξ αρχής, διότι εξαρτώνται άμεσα από την πρώτη. Η αλληλεξάρτηση των τριών παραμέτρων μελετήθηκε ξεχωριστά και παρουσιάζεται αναλυτικότερα στο υποκεφάλαιο 4.5. Όπως προκύπτει από τα Κεφάλαια 2 και 3, οι παράμετροι ορισμού και εκπαίδευσης ενός νευρωνικού δικτύου είναι πολλές και ακόμη περισσότεροι οι πιθανοί συνδυασμοί τους. Για τον σκοπό της παρούσας έρευνας, κάποιες παράμετροι θα πρέπει να θεωρηθούν σταθερές καθώς επιλέχθηκαν με κριτήριο την καταλληλότητά τους (πίνακας 4.1).

Εκπαίδευση Δικτύου	
Αλγόριθμος Βελτιστοποίησης	**Adam**
Συνάρτηση Ενεργοποίησης	**ReLU**
Συνάρτηση Κόστους	**MSE**
Επιλογή Γνωρισμάτων	
Αλγόριθμος Συσχετισμού	**Spearman**
Διαχωρισμός Dataset	
Steps in	**2**
Steps out	**2**
Overlap	**1**

Πίνακας 4.1: *Οι παράμετροι που θεωρήθηκαν σταθερές κατά τη διάρκεια της πειραματικής διαδικασίας.*

4.1: Μεταβλητές Παράμετροι Εκπαίδευσης

Στα υποκεφάλαια 4.1.1-4.1.3 περιγράφονται οι παράμετροι που ελέγχθηκαν κατά τη διάρκεια της πειραματικής διαδικασίας και ως εκ τούτου θα πρέπει να θεωρηθούν ως μεταβλητές.

4.1.1: Αριθμός Γνωρισμάτων

Με βάση τα όσα παρουσιάστηκαν στο κεφάλαιο 3, η πρόγνωση της τιμής του Γεωδυναμικού Ύψους στα 500hpa στην περιοχή των Αθηνών προσεγγίστηκε ως ένα ζητούμενο εκμάθησης με επίβλεψη (supervised learning), που σημαίνει ότι η εξαγωγή γνωρισμάτων (feature extraction) αποτελεί μια καθοδηγούμενη διαδικασία. Στο πλαίσιο της συγκεκριμένης εργασίας, γνωρίσματα θεωρούνται οι γεωγραφικές συντεταγμένες, το πλέγμα των οποίων καλύπτει όλον τον πλανήτη σε αρχική ανάλυση 1° x 1º. Το πρώτο βήμα μείωσης του όγκου των δεδομένων ήταν η δημιουργία ενός αραιότερου πλέγματος ανάλυσης 5° x 5º, όπως περιγράφεται στο υποκεφάλαιο 3.2.2.2, το οποίο είχε ως αποτέλεσμα την μείωση του όγκου των γνωρισμάτων από 61560 σε 2664 σημεία (περίπου 23 φορές). Όπως αναφέρεται, η αραίωση του πλέγματος δεν αναμένεται να έχει σαν συνέπεια την απώλεια σημαντικής πληροφορίας, διότι μεγάλης κλίμακας φαινόμενα της πλανητικής κυκλοφορίας όπως τηλεσυνδέσεις, κύματα Rossby, κ.ά., μπορούν να αποδοθούν ικανοποιητικά και σε μικρές αναλύσεις. Κατά συνέπεια, το αρχικό dataset που αποτελεί τη βάση του πειράματος αποτελείται τελικά από 1338 ημέρες x 2664 περιοχές (γνωρίσματα). Ως επιπρόσθετο κριτήριο επιλογής γνωρισμάτων χρησιμοποιήθηκε ο δείκτης συσχέτισης Spearman (υποκεφάλαιο 3.2.2.3.1), διότι δεν ανιχνεύει απλώς γραμμικές συσχετίσεις αλλά είναι σε θέση να διακρίνει και συστάδες δεδομένων ακόμη κι αν πρόκειται για ακραίες τιμές. Βάσει αυτού, μια παράμετρος διαφοροποίησης των δεδομένων ήταν η εισαγωγή διαφορετικών ορίων στο δείκτη $\rho_{spearman}$ κάθε φορά.

4.1.2: Αρχιτεκτονική Δικτύου

Λαμβάνοντας υπόψιν τα συμπεράσματα που αναφέρονται στο υποκεφάλαιο 3.3.1, προκύπτει ότι για αριθμό γνωρισμάτων έως ~100 σημεία το πλάτος ενός LSTM δικτύου θα πρέπει να είναι τουλάχιστον 100 νευρώνες. Από την άλλη πλευρά, το βάθος του δικτύου (αριθμός επιπέδων), σχετίζεται με τον αριθμό των ημερών αλλά και το βαθμό γενίκευσης των χαρακτηριστικών που τείνει να "μαθαίνει". Σε γενικές γραμμές και για τον συγκεκριμένο αριθμό ημερών, περισσότερα των 2 επιπέδων μπορούν να θεωρηθούν ως ένα "βαθύ δίκτυο".

Με βάση τα παραπάνω, ένα δίκτυο 100 x 100 νευρώνων είναι το ελάχιστο απαιτούμενο ώστε να προσεγγιστεί με τη μέγιστη δυνατή ακρίβεια η πρόγνωση της συγκεκριμένης μεταβλητής. Τελικά ο αριθμός των νευρώνων διπλασιάστηκε σε κάθε επίπεδο καθώς τα χαρακτηριστικά του συστήματος σε σχέση με το χρόνο εκπαίδευσης ήταν επαρκή και τελικά, η βάση μελέτης αποτέλεσε ένα δίκτυο 200 x 200 cells. Ο αριθμός των νευρώνων εξόδου είναι δύο, δηλαδή ίδιος με τον αριθμό των προγνωστικών ημερών.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 2, 200)            249600
_________________________________________________________________
dropout (Dropout)            (None, 2, 200)            0
_________________________________________________________________
lstm_1 (LSTM)                (None, 200)               320800
_________________________________________________________________
dropout_1 (Dropout)          (None, 200)               0
_________________________________________________________________
dense (Dense)                (None, 2)                 402
=================================================================
Total params: 570,802
Trainable params: 570,802
Non-trainable params: 0
```

Σχήμα 4.1: *Εκτύπωση της αρχιτεκτονικής του δικτύου στη γραμμή εντολών. Στην αριστερή στήλη αναγράφεται το όνομα του επιπέδου, στη μεσαία στήλη αναγράφεται το σχήμα (αριθμός νευρώνων) και στη δεξιά ο αριθμός των παραμέτρων που εκπαιδεύτηκαν σε κάθε επίπεδο.*

4.1.3: Κύκλοι Εκπαίδευσης

Η ανίχνευση του βέλτιστου αριθμού κύκλων εκπαίδευσης (epochs) είναι ζητούμενο κρίσιμης σημασίας για την επιτυχή εκπαίδευση ενός νευρωνικού δικτύου καθώς, λιγότεροι κύκλοι εκπαίδευσης οδηγούν σε ελλιπή εκμάθηση του δικτύου, ενώ περισσότεροι μπορούν να οδηγήσουν σε φαινόμενα υπερεκπαίδευσης (overfitting). Το δεύτερο πρακτικά σημαίνει ότι το νευρωνικό δίκτυο προσαρμόζεται υπερβολικά στα δεδομένα εκπαίδευσης, αδυνατώντας να αντιληφθεί τους συσχετισμούς νέων δεδομένων που του παρουσιάζονται. Κατά τη διαδικασία της εκπαίδευσης των νευρωνικών δικτύων υπάρχουν δύο βασικές επιλογές:

1. Ο προγραμματιστής ορίζει εξ αρχής τον αριθμό των κύκλων εκπαίδευσης,
2. Ορίζεται μια συνάρτηση - σχήμα διακοπής της εκπαίδευσης, βάση κάποιων μετρικών.

Η βιβλιοθήκη tensorflow.keras παρέχει την συνάρτηση EarlyStopping (https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/EarlyStopping), η οποία ορίζει μια συνθήκη τερματισμού ως εξής: Ελέγχεται η βελτίωση μιας μετρικής (συνηθέστερα loss ή accuracy) σε κάθε κύκλο εκπαίδευσης σε σχέση με την τιμή της σε προγενέστερο κύκλο

και σε περίπτωση που είναι κάτω από κάποιο όριο τότε η εκπαίδευση σταματάει. Στη συγκεκριμένα έρευνα, για κάθε ένα από τα μοντέλα που δημιουργήθηκαν, εφαρμόστηκαν αρχικά πολλοί κύκλοι εκπαίδευσης και συγκρίθηκαν οι τιμές loss και accuracy για τις εσωτερικές μεταβλητές train/test του μοντέλου. Λόγω του στοχαστικού χαρακτήρα της διαδικασίας, η "εικονική" εκπαίδευση πραγματοποιήθηκε τρεις φορές, ώστε να διαπιστωθεί το σημείο έναρξης της υπερεκπαίδευσης. Ο αριθμός των "epochs", όπως προκύπτει από τις τρεις εικονικές εκπαιδεύσεις, είναι τελικά αυτός που χρησιμοποιείται κατά την πειραματική διαδικασία. Η ανίχνευση αυτού του αριθμού αποτελεί και το πρώτο στάδιο κάθε πειραματικής διαδικασίας.

4.2: Model Ensembling και Μετρικές Αξιολόγησης (Scores)

Για την αξιολόγηση των μοντέλων που προέκυψαν από τον συνδυασμό διαφορετικών παραμετροποιήσεων, έγινε η χρήση δύο ευρέως χρησιμοποιούμενων στατιστικών δεικτών (Wilks, 2011): Η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (Root Mean Squared Error - RMSE) και η Μέση Απόκλιση (Bias).

Το RMSE, ορίζεται ως εξής:

$$RMSE = \sqrt{\frac{1}{k}\sum_{i=1}^{k} (for(i) - obs(i))^2} \quad (4.1)$$

Ενώ το BIAS ως εξής:

$$BIAS = \frac{1}{k}\sum_{i=1}^{k} (for(i) - obs(i)) \quad (4.2)$$

όπου:
k: ο αριθμός των ζευγών προγνώσεων (for) και παρατηρήσεων (obs).

Όπως είναι φανερό από τον ορισμό των νευρωνικών δικτύων αλλά και του τρόπου με τον οποίο πραγματοποιείται η εκπαίδευσή τους, η διαδικασία είναι στοχαστική, που σημαίνει ότι η ακριβής δομή κάθε νευρωνικού δικτύου (όπως πχ ο καθορισμός των βαρών σε κάθε ξεχωριστό νευρώνα), θα είναι διαφορετική κάθε φορά και άρα αναμένονται ελαφρώς

διαφορετικά αποτελέσματα. Έχοντας εξ αρχής ανιχνεύσει τον βέλτιστο αριθμό κύκλων εκπαίδευσης για κάθε μοντέλο υπό δοκιμή, εκπαιδεύονται και ελέγχονται πέντε διαφορετικά μοντέλα καταγράφοντας σε όλες τις περιπτώσεις τους δείκτες σφάλματος. Προκειμένου η διαδικασία να είναι αξιόπιστη, θα πρέπει η απόκλιση των δεικτών να μην είναι μεγάλη (π.χ. +/- 2,5%). Η παραπάνω διαδικασία μπορεί να ονομαστεί και "model ensembling".

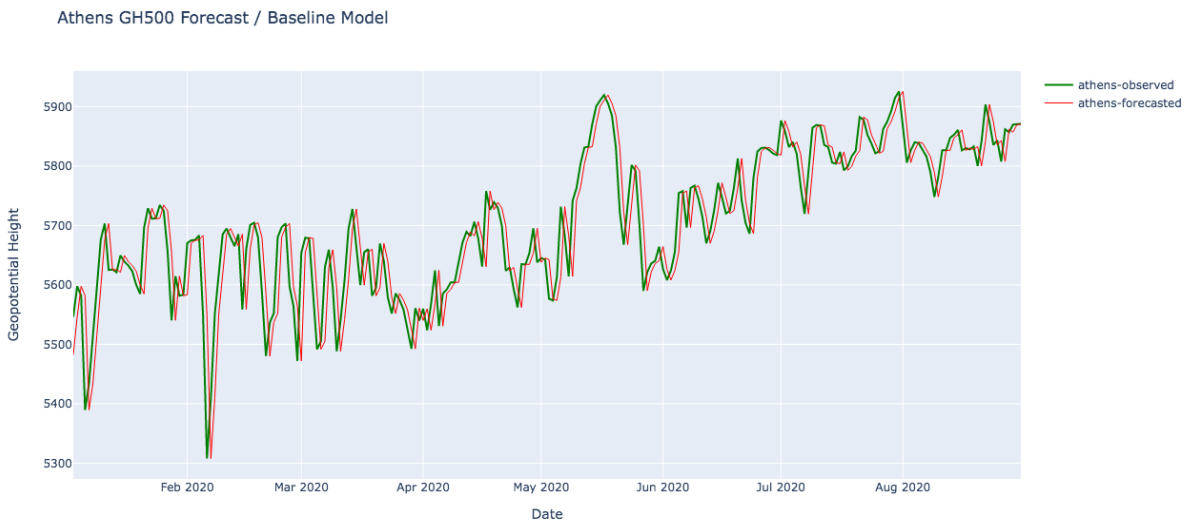
4.3: Μοντέλο Βάσης (Baseline Model)

Αναγκαία προϋπόθεση αξιολόγησης των προβλέψεων κάθε εναλλακτικού μοντέλου που δημιουργείται είναι η δημιουργία ενός αρχικού μοντέλου βάσης (Baseline Model - BM), προκειμένου να συγκριθούν τα αποτελέσματά του με αυτά των εναλλακτικών. Όπως αναφέρει ο Jason Brownlee σε άρθρο του στην προσωπική του ιστοσελίδα (2019)[11], υπάρχουν τρείς προϋποθέσεις καλής τεχνικής που πρέπει να τηρηθούν για τη δημιουργία ενός αξιόπιστου BM:

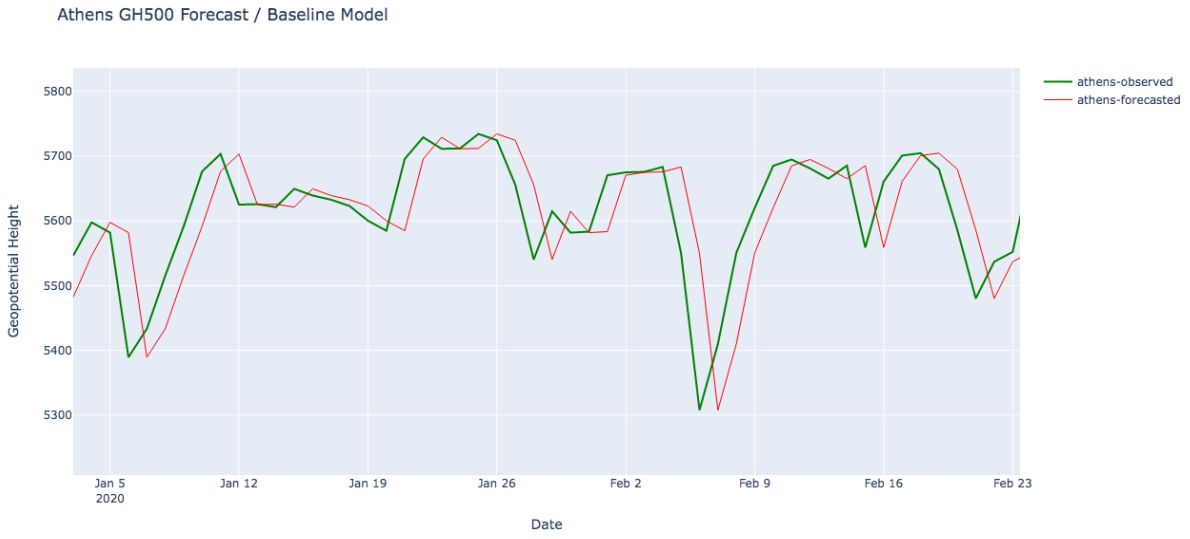
1. **Απλότητα:** Η μέθοδος που χρησιμοποιήθηκε για τη δημιουργία του να μην προϋποθέτει κάποια εξελιγμένη νοημοσύνη.
2. **Ταχύτητα:** Η μέθοδος θα πρέπει να είναι πολύ γρήγορα υλοποιήσιμη απαιτώντας ελάχιστους υπολογιστικούς πόρους.
3. **Επαναληψιμότητα:** Τα αποτελέσματα της μεθόδου να είναι τα ίδια κάθε φορά που εφαρμόζεται - με άλλα λόγια να είναι αιτιοκρατική.

Κοινή πρακτική δημιουργίας BM για τις ανάγκες των προγνώσεων χρονοσειρών είναι η δημιουργία μιας "persistent" Y', η οποία πρόκειται για την μετατόπιση της αρχικής Y, κατά ένα συγκεκριμένο αριθμό χρονικών βημάτων και η σύγκριση αυτής της "πρόβλεψης" με τις πραγματικές τιμές. Στη συγκεκριμένη εργασία, εφ' όσον επιχειρείται η πρόγνωση ενός 24ώρου, η μετατόπιση της Y' θα είναι κατά ένα χρονικό βήμα (δηλαδή μία ημέρα) μπροστά. Στο σχήμα 4.2 παρουσιάζεται η πρόγνωση για όλο το dataset ελέγχου με βάση την παραπάνω παραδοχή, ενώ στο σχήμα 4.3 απεικονίζεται μεγέθυνση του πρώτου αλλά για μικρότερο χρονικό διάστημα.

[11] https://machinelearningmastery.com/persistence-time-series-forecasting-with-python/



Σχήμα 4.2: *Πρόγνωση του "μοντέλου βάσης" (baseline model), για τη χρονοσειρά του γεωδυναμικού ύψους στα 500hpa της Αθήνας, η οποία απεικονίζεται με κόκκινο χρώμα. Με πράσινο χρώμα απεικονίζονται οι παρατηρούμενες τιμές. Σε αυτήν την περίπτωση μοντέλου, ορίζεται ως η παρατηρούμενη τιμή μετατοπισμένη κατά μία ημέρα.*



Σχήμα 4.3: *Όπως στο σχήμα 4.1 αλλά εστιασμένη στο διάστημα 4 Ιανουαρίου έως 23 Φεβρουαρίου 2020.*

Η αξιολόγηση της παραπάνω πρόγνωσης με τη χρήση του RMSE και BIAS, έδωσε **55.13** και **-1.60** μονάδες αντίστοιχα. Στο σχήμα 4.4 παρουσιάζεται το αποτέλεσμα όπως εκτυπώνεται κατευθείαν στη γραμμή εντολών (terminal):

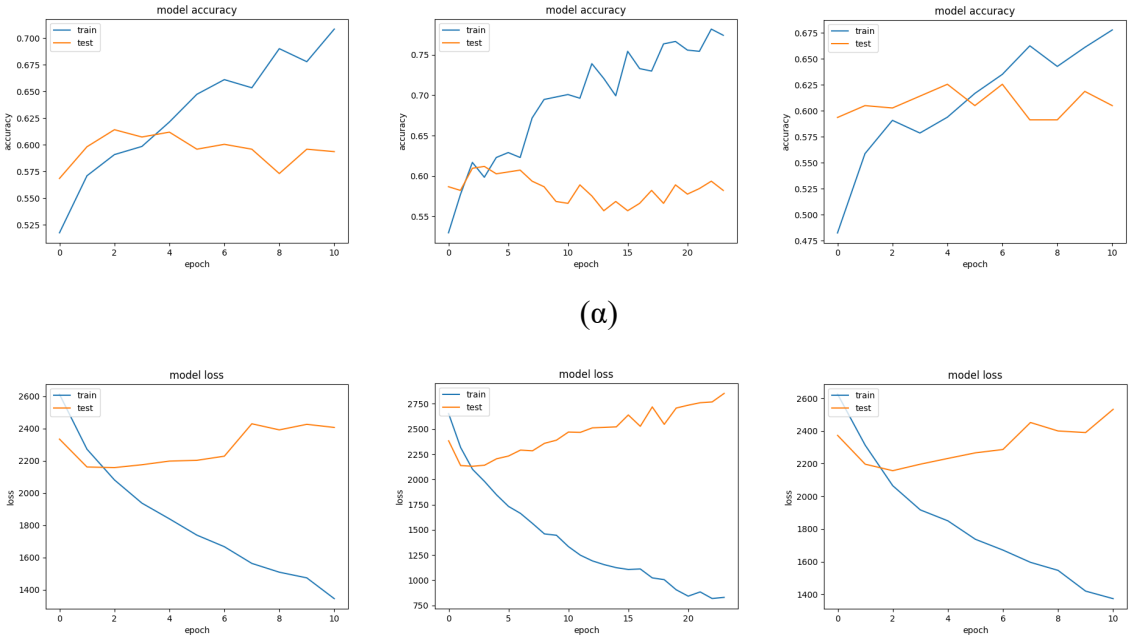
```
(1338, 2667)
initial number of points 2665
baseline_rmse: 55.134729728062794
baseline_bias: -1.6066967975206623
```

Σχήμα 4.4*: Η εκτύπωση των υπολογισμών για RMSE και BIAS του Baseline μοντέλου στη γραμμή εντολών.*

4.4: Εναλλακτικά Μοντέλα

4.4.1: Μοντέλο τριών τυχαίων γνωρισμάτων

Μέσω διαδοχικών δοκιμών στις οποίες έγινε η σύγκριση των loss και accuracy για τα δεδομένα εκπαίδευσης και ελέγχου, παρατηρείται το φαινόμενο της υπερεκπαίδευσης για αριθμό epochs > 4. Στο σχήμα 4.5 παρουσιάζεται η σύγκριση των παραπάνω μετρικών και για τις τρεις δοκιμές:

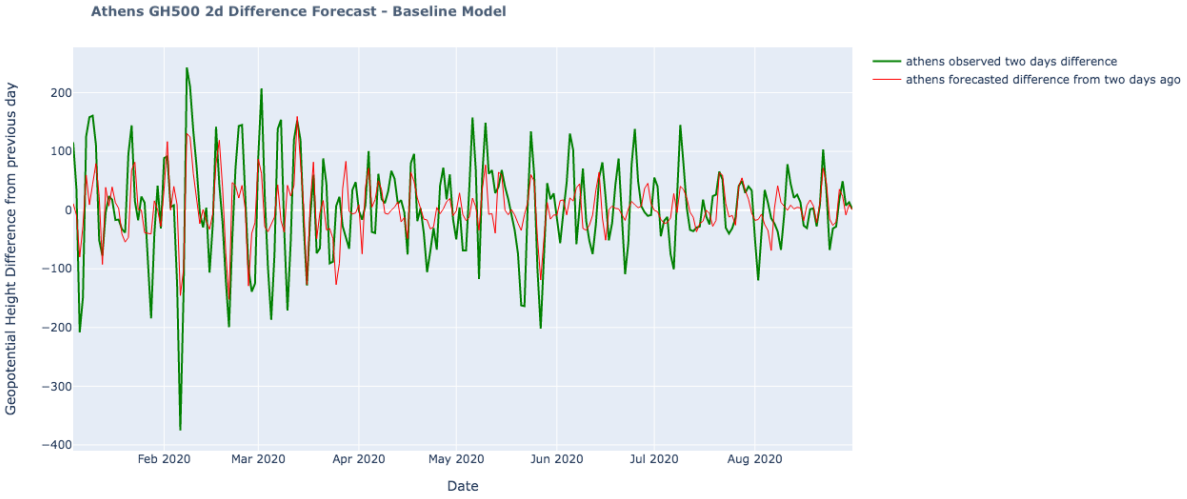


(α)

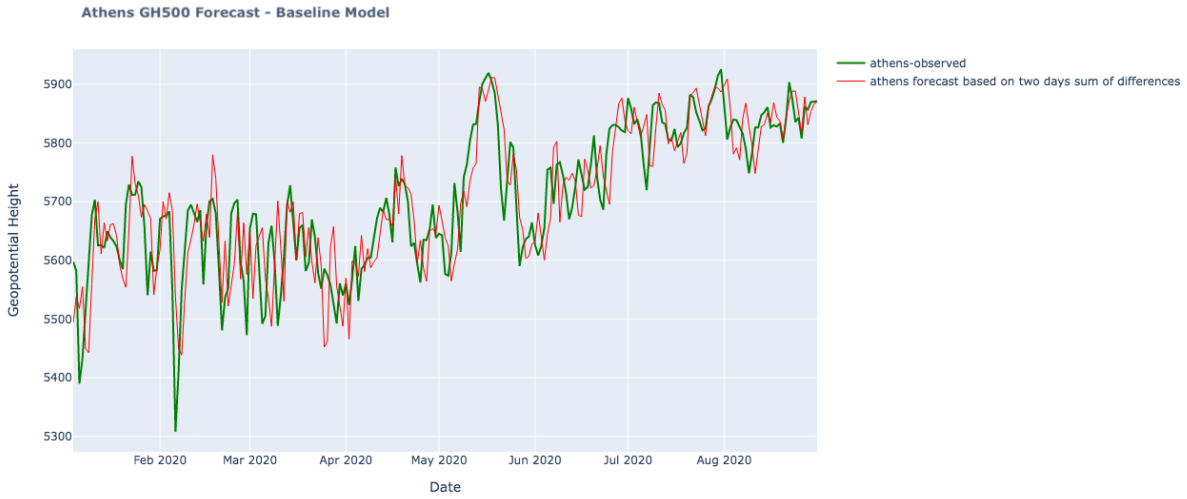
(β)

Σχήμα 4.5: *Οι μετρικές ελέγχου εκπαίδευσης accuracy (ομάδα α) και loss (ομάδα β) του μοντέλου τριών τυχαίων γνωρισμάτων σε συνάρτηση με τον αριθμό επαναλήψεων για τρεις διαδοχικές δοκιμές. Με μπλε χρώμα απεικονίζεται η τιμή που αφορά στα δεδομένα εκπαίδευσης (train) και με κόκκινο η τιμή που αφορά στα δεδομένα ελέγχου (test). Παρατηρείται ότι για αριθμό επαναλήψεων άνω των 4, η τιμή loss των δεδομένων εκπαίδευσης είναι χαμηλότερη της αντίστοιχης των δεδομένων ελέγχου ενώ η τιμή accuracy των δεδομένων εκπαίδευσης είναι υψηλότερη των δεδομένων ελέγχου (overfitting).*

Με βάση τα παραπάνω, οι εκπαιδευτικές επαναλήψεις του μοντέλου τριών τυχαίων γνωρισμάτων ορίστηκαν στις 4. Στο σχήμα 4.6 παρουσιάζεται η πρόγνωση της διήμερης διαφοράς της τιμής του γεωδυναμικού ύψους στα 500hpa (GH500) της Αθήνας, συναρτήσει της πραγματικής διαφοράς που παρατηρήθηκε. Στο σχήμα 4.7 παρουσιάζεται η τελική πρόγνωση του GH500 σε συνάρτηση με τις παρατηρούμενες τιμές.



Σχήμα 4.6: *Η διαφορά στην τιμή GH500 της Αθήνας σε σχέση με δύο ημέρες πριν για το διάστημα ελέγχου. Με κόκκινη γραμμή παρουσιάζεται η πρόγνωση σύμφωνα με το μοντέλο βάσης (baseline), ενώ με πράσινη η πραγματική τιμή.*



Σχήμα 4.7: *Η τιμή GH500 της Αθηνας για το διάστημα ελέγχου. Με κόκκινη γραμμή παρουσιάζεται η πρόγνωση σύμφωνα με το μοντέλο βάσης (baseline), ενώ με πράσινη η πραγματική τιμή.*

Από τα σχήματα 4.6 και 4.7 μπορεί να διαπιστωθεί ότι το συγκεκριμένο μοντέλο καταφέρνει μεν να διακρίνει τις περισσότερες από τις μεταβολές που σημειώθηκαν ως γεγονότα, αδυνατώντας όμως σε αρκετές περιπτώσεις να προσομοιώσει σωστά το μέγεθος της μεταβολής. Αυτό έχει σαν αποτέλεσμα η πρόγνωση της μεταβλητής GH500 να ακολουθεί τη γενική πορεία της πραγματικής τιμής στο χρόνο, χωρίς όμως να προσομοιώσει σωστά μεμονωμένα γεγονότα. Επιπλέον, σε αρκετές περιπτώσεις οι προγνωστικές τιμές

αποτυπώνονται με διαφορά φάσης 1-2 ημερών σε σχέση με την πραγματικότητα κάτι που μπορεί να αποδοθεί στα ελλιπή δεδομένα εκπαίδευσης του μοντέλου.

Στον πίνακα 4.2 σημειώνονται οι τιμές του μέσου τετραγωνικού σφάλματος (RMSE) και της απόκλισης (BIAS) για 5 διαδοχικά "μοντέλα τριών γνωρισμάτων" που δημιουργήθηκαν με τα ίδια δεδομένα εκπαίδευσης και στη συνέχεια εκτέθηκαν στα ίδια δεδομένα ελέγχου. Με έντονη γραφή τονίζονται οι τιμές του μοντέλου, η προσομοίωση του οποίου παρουσιάζεται στα σχήματα 4.6 και 4.7. Διαπιστώνεται ότι τα επίπεδα του RMSE είναι ακόμη υψηλότερα σε σχέση με μία πρόγνωσης η οποία απλώς θεωρεί την τιμή της προηγούμενης ημέρας ως προγνωστική για την επόμενη (baseline model).

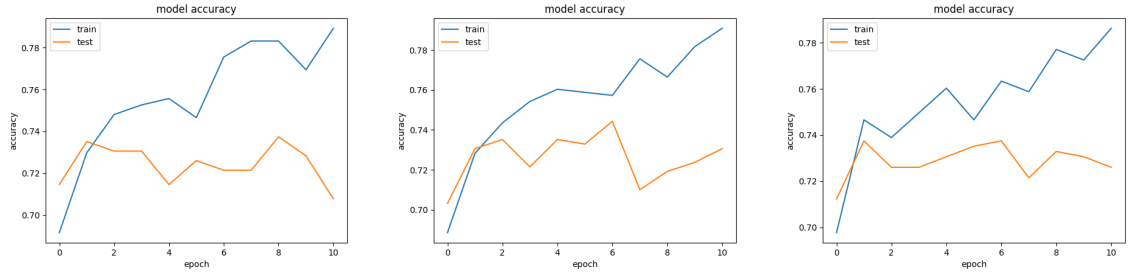
Baseline Model	**RMSE**	**BIAS**
# 1	64,44	+1,42
# 2	64,16	+0,15
# 3	64,69	-1,00
# 4	64,09	-1,04
# 5	64,80	-8,28

Πίνακας 4.2: *Οι τιμές των στατιστικών ελέγχων του μέσου τετραγωνικού σφάλματος (RMSE) και απόκλισης (BIAS) για την πρόγνωση του ίδιου δείγματος, για 5 διαδοχικά "μοντέλα τριών γνωρισμάτων" που δημιουργήθηκαν με τα ίδια δεδομένα εκπαίδευσης.*

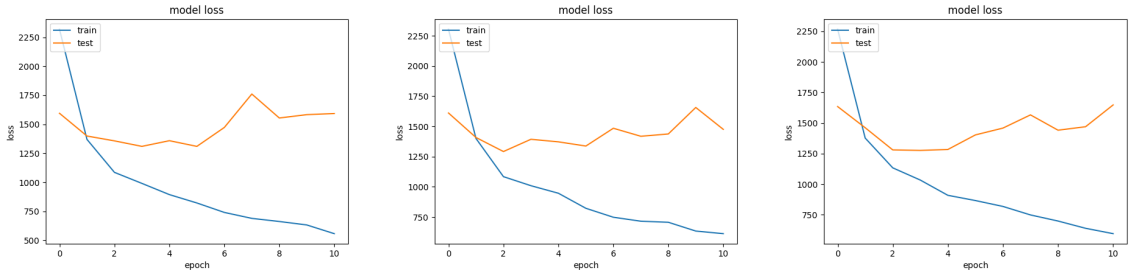
4.4.2: Μοντέλο τεσσάρων γνωρισμάτων με υψηλή συσχέτιση

Στο συγκεκριμένο μοντέλο ελήφθησαν υπόψιν μόνο υψηλώς συσχετιζόμενα (κατά Spearman) γνωρίσματα, δηλαδή latitude / longitude σημεία με βαθμό συσχέτισης >0.7 με το θεωρούμενο ως σημείο της Αθήνας (38.0N, 24.0E). Τα σημεία αυτά είναι τα [40.0, 25.0 - 35.0, 25.0 - 35.0, 20.0 - 40.0, 20.0]. Παρατηρείται ότι πρόκειται για σημεία που παρουσιάζουν γεωγραφική εγγύτητα με την Αθήνα.

Εφαρμόζοντας τη μέθοδο των διαδοχικών εκπαιδεύσεων του μοντέλου με μεγάλο αριθμό επαναλήψεων (epochs) σε κάθε μία εξ αυτών προκύπτουν τα γραφήματα του σχήματος 4.8 για τρεις επαναλήψεις. Είναι εμφανές ότι για αριθμό επαναλήψεων >2, το μοντέλο θα υπερεκπαιδευτεί στα δεδομένα εκμάθησης, έχοντας ως αποτέλεσμα την μείωση της ακρίβειάς του.



(α)



(β)

Σχήμα 4.8: *Οι μετρικές ελέγχου εκπαίδευσης accuracy (ομάδα α) και loss (ομάδα β) του μοντέλου υψηλών συσχετίσεων σε συνάρτηση με τον αριθμό επαναλήψεων για τρεις διαδοχικές δοκιμές. Με μπλε χρώμα απεικονίζεται η τιμή που αφορά στα δεδομένα εκπαίδευσης (train) και με κόκκινο η τιμή που αφορά στα δεδομένα ελέγχου (test). Παρατηρείται ότι για περισσότερες των τριών επαναλήψεων, η τιμή loss των δεδομένων εκπαίδευσης είναι χαμηλότερη της αντίστοιχης των δεδομένων ελέγχου ενώ η τιμή accuracy των δεδομένων εκπαίδευσης είναι υψηλότερη των δεδομένων ελέγχου.*

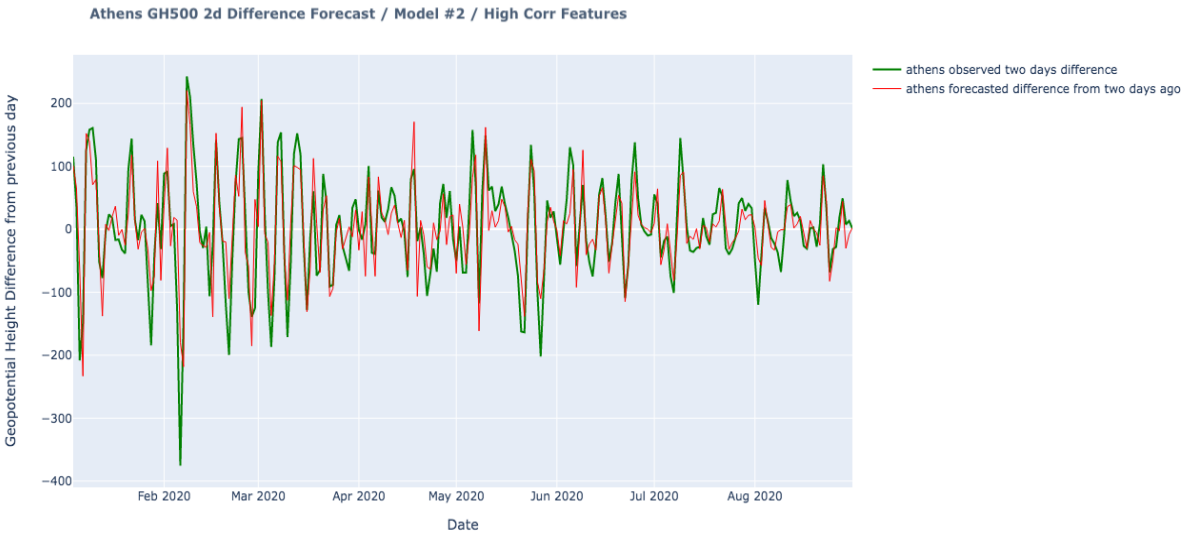
Έχοντας ανιχνεύσει το βέλτιστο αριθμό των epochs (3), εκπαιδεύτηκαν διαδοχικά πέντε διαφορετικά μοντέλα ώστε να υπολογιστούν οι τιμές σφάλματος RMSE και BIAS για κάθε ένα από αυτά, οι οποίες παρουσιάζονται στον πίνακα 4.3.

High Corr. Feature Model	RMSE	BIAS
# 1	47,29	4,40
# 2	45,86	-5,37
# 3	45,77	-3,00
# 4	**44,66**	**-0,10**
# 5	46,05	-2,47

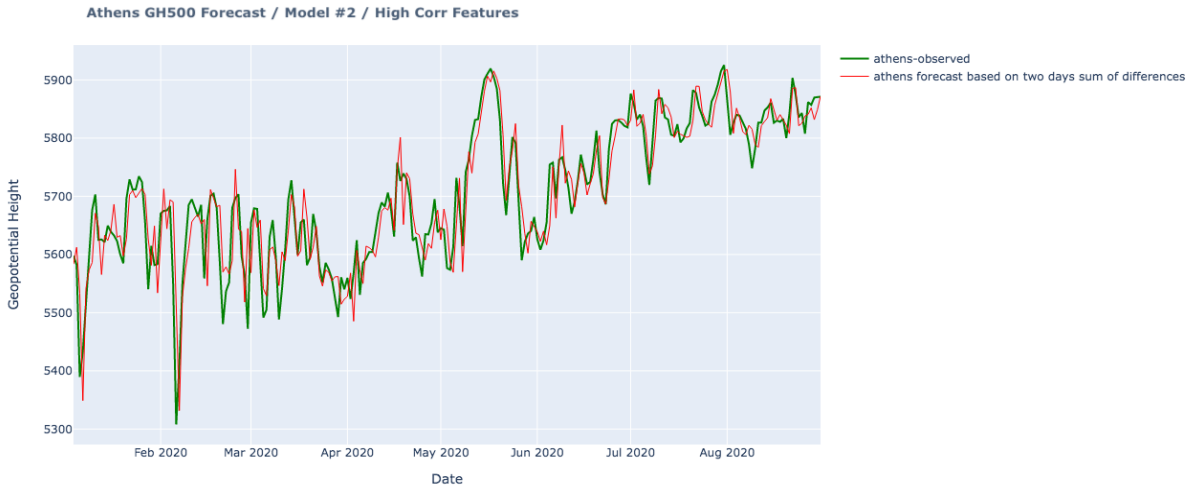
Πίνακας 4.3: *Οι τιμές των στατιστικών ελέγχων του μέσου τετραγωνικού σφάλματος (RMSE) και απόκλισης (BIAS) για την πρόγνωση του ίδιου δείγματος, για 5 διαδοχικά μοντέλα που δημιουργήθηκαν με τα ίδια δεδομένα εκπαίδευσης.*

Αυτό που διαπιστώνεται είναι το γεγονός ότι η εκπαίδευση του μοντέλου έχοντας ως δεδομένα υψηλώς συσχετιζόμενες περιοχές (features) μειώνει αισθητά το δείκτη RMSE σε σχέση με το μοντέλο βάσης αλλά όσον αφορά το δείκτη BIAS, δεν παρατηρείται κάποια βελτίωση. Στην πλειοψηφία των περιπτώσεων η απόκλιση φαίνεται πως είναι ελαφρώς αρνητική, δηλαδή το μοντέλο στην πρόβλεψή του τείνει να υποεκτιμά την τιμή του γεωδυναμικού ύψους.

Στα σχήματα 4.9 και 4.10 παρουσιάζονται η πρόγνωση της διήμερης διαφοράς της τιμής του γεωδυναμικού ύψους στα 500hpa (GH500) της Αθήνας συναρτήσει της πραγματικής διαφοράς που παρατηρήθηκε και η τελική πρόγνωση του GH500 σε συνάρτηση με τις παρατηρούμενες τιμές αντίστοιχα. Για την απεικόνιση των προγνώσεων χρησιμοποιήθηκε το μοντέλο με τα καλύτερα scores, όπως παρουσιάζονται στον πίνακα 4.3 (έχουν τονιστεί με έντονη γραφή).



Σχήμα 4.9: *Η διαφορά στην τιμή GH500 της Αθήνας σε σχέση με δύο ημέρες πριν για το διάστημα ελέγχου. Με κόκκινη γραμμή παρουσιάζεται η πρόγνωση σύμφωνα με το μοντέλο υψηλών συσχετιζόμενων γνωρισμάτων, ενώ με πράσινη η πραγματική τιμή.*



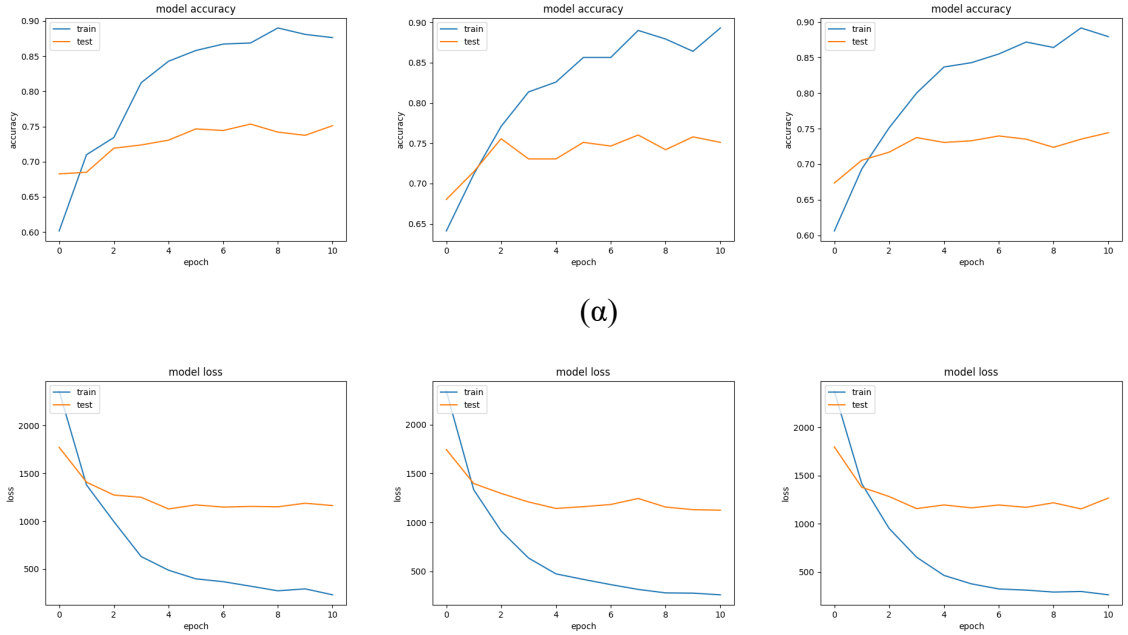
Σχήμα 4.10: *Η τιμή GH500 της Αθηνας για το διάστημα ελέγχου. Η κόκκινη γραμμή απεικονίζει την πρόγνωση σύμφωνα με το μοντέλο υψηλών συσχετιζόμενων γνωρισμάτων, ενώ η πράσινη την πραγματική τιμή.*

Με βάση τις προγνώσεις που παρουσιάζονται στα σχήματα 4.9 και 4.10 διαπιστώνεται η μεγαλύτερη ακρίβεια του συγκεκριμένου μοντέλου σε σχέση με το μοντέλο τριών τυχαίων γνωρισμάτων. Ειδικότερα, είναι σε θέση να ανιχνεύσει τις περισσότερες φορές το μέγεθος των διακυμάνσεων που παρατηρούνται και όχι μόνο τα γεγονότα, κάτι που οδηγεί σε βελτιωμένη πρόγνωση της τιμής του γεωδυναμικού ύψους στα 500hpa. Επιπροσθέτως, το φαινόμενο της καθυστέρησης στην απεικόνιση των μεταβολών έχει μετριαστεί, αν και σε κάποιες μεμονωμένες περιπτώσεις επιμένει, ιδίως όταν λαμβάνουν χώρα έντονες διακυμάνσεις (πχ αρχές Ιανουαρίου 2020 και αρχές Φεβρουαρίου 2020) αλλά και προς το τέλος του διαστήματος ελέγχου.

4.4.3: Μοντέλο πολλών γνωρισμάτων

Καθώς η εισαγωγή γνωρισμάτων με υψηλή συσχέτιση αύξησε την ακρίβεια του μοντέλου, κρίθηκε σκόπιμη η δοκιμή του ίδιου μοντέλου με την εισαγωγή ακόμη περισσότερων γνωρισμάτων που να περιλαμβάνουν κι εκείνα που παρουσιάζουν πολύ ασθενή βαθμό συσχέτισης με την Αθήνα. Πιο συγκεκριμένα επιλέχθηκαν σημεία τα οποία η τιμή τους ήταν είτε μικρότερη του -0.1 είτε μεγαλύτερη του +0.1 σύμφωνα με τον έλεγχο κατά spearman. Κατ' αυτόν τον τρόπο ο αριθμός των σημείων (features) αυξήθηκε από 4 σε 112, όπου τα περισσότερα εξ αυτών δεν εμφανίζουν καμία εγγύτητα με τις γεωγραφικές συντεταγμένες της

Αθήνας. Ο λόγος της εισαγωγής αυτών των γνωρισμάτων προέρχεται από το γεγονός ότι το συγκεκριμένο μοντέλο μπορεί εκ φύσεως να ανιχνεύσει μη γραμμικές σχέσεις και ενδεχόμενα μοτίβα. Κατά συνέπεια, η εισαγωγή διαφόρων ακόμη σημείων δεν αναμένεται να μειώσει την ακρίβεια του μοντέλου και σε αυτό το πείραμα εξετάζεται το ενδεχόμενο περαιτέρω βελτίωσής της. Όπως προκύπτει από το σχήμα 4.11, ο βέλτιστος αριθμός εκπαιδεύσεων είναι και πάλι οι τρεις, πριν το μοντέλο περάσει σε κατάσταση “overfitting”.



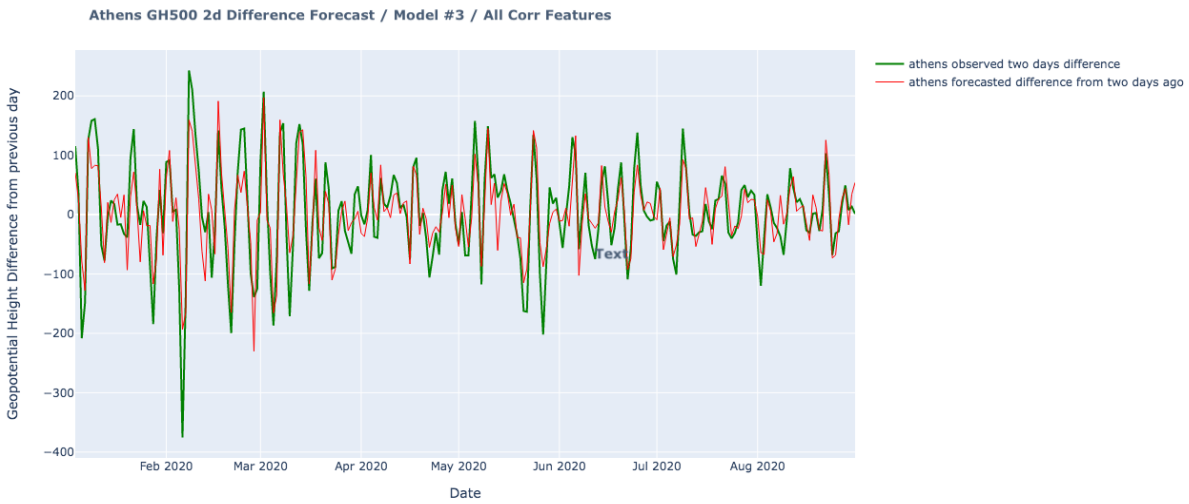
Σχήμα 4.11: *Οι τιμές accuracy (ομάδα α) και loss (ομάδα β) του μοντέλου πολλών συσχετίσεων σε συνάρτηση με τον αριθμό επαναλήψεων για τρεις διαδοχικές δοκιμές. Με μπλε χρώμα απεικονίζεται η τιμή των δεδομένων εκπαίδευσης (train) και με κόκκινο η τιμή των δεδομένων ελέγχου (test). Προκύπτει ότι ο μέγιστος αριθμός επαναλήψεων είναι τρεις, προτού το μοντέλο περάσει σε κατάσταση “overfitting”.*

Ακολουθώντας την ίδια διαδικασία με τα προηγούμενα πειράματα, για αριθμό epochs=3, εκπαιδεύτηκαν πέντε διαφορετικά μοντέλα ώστε να πραγματοποιηθούν οι υπολογισμοί του Μέσου Τετραγωνικού Σφάλματος (RMSE) και της Απόκλισης (BIAS). Τα αποτελέσματα παρουσιάζονται στον πίνακα 4.4. Η σύγκριση των τιμών με τις αντίστοιχες του μοντέλου υψηλών συσχετίσεων (υποκεφάλαιο 4.4.2 , πινακας 4.3) δείχνει μια μικρή βελτίωση του δείκτη RMSE, η οποία όμως λόγω του μικρού αριθμού των επαναλήψεων είναι στατιστικώς μη σημαντική, ενώ η καλύτερη τιμή επιτυγχάνεται από την πρώτη δοκιμή (μοντέλο #1).

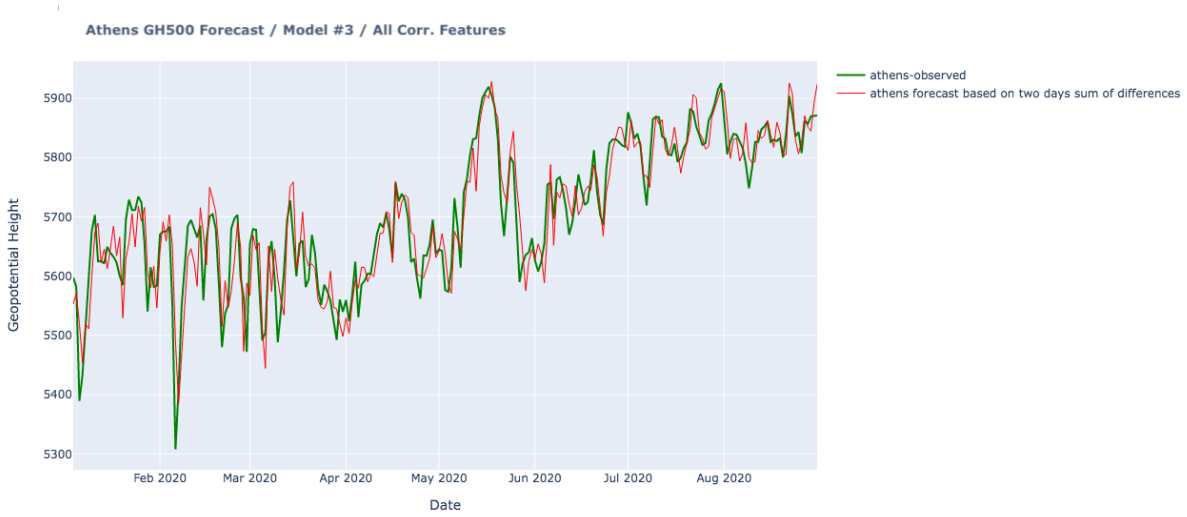
Στα σχήματα 4.12 και 4.13 παρουσιάζονται οι προγνώσεις για τη διαφορά της τιμής του GH500 σε σχέση με 48 ώρες πριν και η πρόγνωση της απόλυτης τιμής του δείκτη αντίστοιχα για το μοντέλο που σημείωσε τα καλύτερα scores. Σε σχέση με την πρόγνωση της προηγούμενης πειραματικής διαδικασίας (σχήματα 4.9 και 4.10), παρατηρείται ελαφρώς καλύτερη απόκριση του μοντέλου στις μεταβολές του γεωδυναμικού ύψους. Πιο συγκεκριμένα, το φαινόμενο της χρονικής καθυστέρησης στην πρόγνωση της τιμής που εμφανίζεται προς το τέλος του dataset, δείχνει να μετριάζεται, όχι όμως στις περιπτώσεις πολύ έντονων μεταβολών.

112 Feature Model	**RMSE**	**BIAS**
# 1	**43,53**	**-0,07**
# 2	45,77	-6,53
# 3	45,54	5,53
# 4	46,66	-1,51
# 5	46,26	6,26

Πίνακας 4.4: *Οι τιμές των στατιστικών ελέγχων του μέσου τετραγωνικού σφάλματος (RMSE) και απόκλισης (BIAS) για την πρόγνωση του ίδιου δείγματος, για 5 διαδοχικά μοντέλα που δημιουργήθηκαν με τα ίδια δεδομένα εκπαίδευσης.*



Σχήμα 4.12: *Η διαφορά στην τιμή GH500 της Αθήνας σε σχέση με δύο ημέρες πριν για το διάστημα ελέγχου. Με κόκκινη γραμμή παρουσιάζεται η πρόγνωση σύμφωνα με το μοντέλο πολλών συσχετιζόμενων γνωρισμάτων, ενώ με πράσινη η πραγματική τιμή.*



Σχήμα 4.13: *Η τιμή GH500 της Αθήνας για το διάστημα ελέγχου. Η κόκκινη γραμμή απεικονίζει την πρόγνωση σύμφωνα με το μοντέλο υψηλών συσχετιζόμενων γνωρισμάτων, ενώ η πράσινη την πραγματική τιμή.*

4.5: Μελέτη Αλληλεξάρτησης Μεταβλητών Εκπαίδευσης

Καθώς ένα νευρωνικό δίκτυο αποτελεί μια ενιαία οντότητα, οι μεταβλητές εκπαίδευσης που παρουσιάζονται και αναλύονται στα υποκεφάλαια 4.1.1 - 4.1.3, θα πρέπει να βρίσκονται σε αλληλεξάρτηση. Σε αυτήν την ενότητα επιχειρείται η τεκμηρίωση των μεταξύ τους σχέσεων ανά ζεύγη διατηρώντας τις υπόλοιπες μεταβλητές σταθερές. Ειδικότερα εξετάζονται:

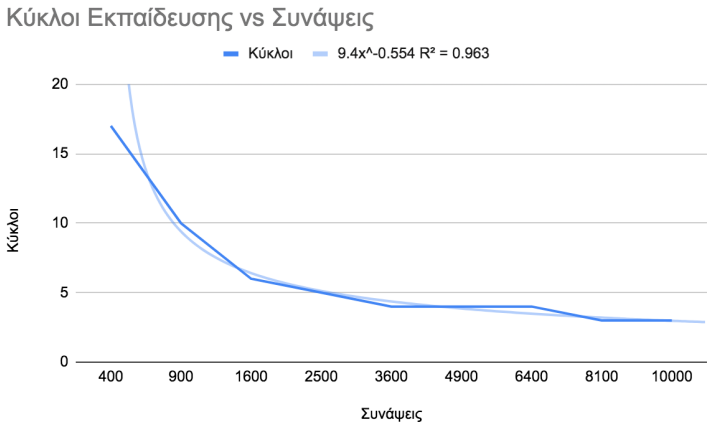
1. Η σχέση αριθμού νευρώνων και βέλτιστου αριθμού κύκλων εκπαίδευσης διατηρώντας ίδιο το μέγεθος του dataset.

2. Η σχέση μεταξύ αριθμού γνωρισμάτων και κύκλων εκπαίδευσης έχοντας ως βάση ένα συγκεκριμένο νευρωνικό δίκτυο.

Η σχέση μεταξύ αριθμού γνωρισμάτων και συνάψεων διατηρώντας σταθερό τον αριθμό των κύκλων εκπαίδευσης (epochs) στο βέλτιστο σημείο δεν έχει κάποια πειραματική αξία. Όπως έχει ήδη αναλυθεί στο υποκεφάλαιο 3.3.1 Από τον τρόπο με τον οποίο τα δεδομένα εισέρχονται σε ένα νευρωνικό δίκτυο, η μεταξύ τους σχέση είναι απολύτως προβλέψιμη, καθώς ο αριθμός των νευρώνων του πρώτου επιπέδου θα πρέπει να είναι τουλάχιστον ίσος με τον αριθμό των γνωρισμάτων.

4.5.1: Αριθμός Συνάψεων Vs. Epochs

Για την πραγματοποίηση των συγκεκριμένων πειραμάτων χρησιμοποιήθηκε το μοντέλο υψηλών συσχετίσεων (υποκεφάλαιο 4.4.2), το οποίο περιλαμβάνει 3 features. Ο ελάχιστος αριθμός νευρώνων κάθε επιπέδου είναι οι 20 και ο μέγιστος οι 100, δηλαδή συνολικός αριθμός νευρώνων από 400 έως 10.000. Εμπειρικά αναμένεται η ύπαρξη μιας αντίστροφης σχέσης, όπου για μικρότερο αριθμό νευρώνων ο βέλτιστος αριθμός epochs θα πρέπει να είναι μεγαλύτερος. Στο σχήμα 4.14 παρουσιάζεται γραφικά η συγκεκριμένη συσχέτιση, όπως προέκυψε μέσω διαδοχικών πειραματισμών. Παρατηρήθηκε ότι για συνολικό αριθμό νευρώνων μικρότερο των 400, η τιμή του RMSE υπολογιζόταν υψηλότερη (>49), ανεξαρτήτως του αριθμού των epochs, οδηγώντας στο συμπέρασμα ότι δίκτυο 20 x 20 cells είναι το ελάχιστο δυνατό που θα μπορούσε να χρησιμοποιηθεί.



Σχήμα 4.14: *Σχέση μεταξύ κύκλων εκπαίδευσης και συνολικού αριθμού συνάψεων για την εκπαίδευση ενός νευρωνικού δικτύου σε dataset τριών γνωρισμάτων.*

Με μια καλή προσέγγιση (R^2=0.963), διαπιστώνεται ότι η σχέση μεταξύ αριθμού συνάψεων και κύκλων εκπαίδευσης είναι η

$$y = 9{,}4 \cdot x^{-0{,}554}$$

όπου:

y: Οι κύκλοι εκπαίδευσης

x: Ο συνολικός αριθμός των συνάψεων μεταξύ των επιπέδων του δικτύου

4.5.2: Αριθμός Γνωρισμάτων Vs Κύκλοι εκπαίδευσης

Για την πραγματοποίηση των συγκεκριμένων πειραμάτων χρησιμοποιήθηκε νευρωνικό δίκτυο 100 x 100 νευρώνων (=10.000 συνάψεων), ενώ ο αριθμός των features καθορίστηκε αλλάζοντας τις τιμές του δείκτη $\rho_{spearman,}$ ώστε να διαφοροποιείται αντίστοιχα και ο αριθμός των γνωρισμάτων. Ο μέγιστος αριθμός των γνωρισμάτων με όρια τα $\rho < -0,1$ και $\rho > +0,1$ (τουλάχιστον ασθενής θετική ή αρνητική συσχέτιση) υπολογίστηκε στα 112 ενώ με όρια τα $\rho < -0,7$ και $\rho > +0,7$ (ισχυρή θετική ή αρνητική συσχέτιση), υπολογίστηκε στα 3. Παρατηρείται ότι για το συγκεκριμένο μέγεθος του δικτύου δεν υπάρχει καμία σχέση μεταξύ αριθμού γνωρισμάτων και κύκλων εκπαίδευσης διατηρώντας την τιμή του RMSE σε κάθε πείραμα στο 46 (+/ 2,5%).

Κεφάλαιο 5° : Σύνοψη και Συμπεράσματα

5.1: Σύνοψη

Στην παρούσα εργασία, εξετάστηκαν τρεις διαφορετικές υλοποιήσεις LSTM δικτύων με στόχο την πρόγνωση της τιμής του γεωδυναμικού ύψους στα 500hpa της περιοχής των Αθηνών, έχοντας ως δεδομένα εισόδου τις τιμές των υπολοίπων χρονοσειρών του πλέγματος. Η επιλογή του συγκεκριμένου τύπου δικτύου έγινε λόγω της ικανότητας "μνήμης" που διαθέτει, η οποία οφείλεται στην εσωτερική δομή κάθε νευρώνα. Επιπροσθέτως όμως και ανεξαρτήτως αρχιτεκτονικής, τα σύγχρονα νευρωνικά δίκτυα λόγω του τρόπου με τον οποίο "εκπαιδεύονται" είναι σε θέση να ανιχνεύσουν **μη γραμμικές** συσχετίσεις μεταξύ των δεδομένων εισόδου αλλά και της μεταβλητής που ζητείται. Η παρούσα έρευνα στηρίχθηκε στην υπόθεση ότι τέτοιου είδους συσχετίσεις θα πρέπει να υπάρχουν στα ατμοσφαιρικά δεδομένα, σε διαφορετική περίπτωση, η εκπαίδευση του μοντέλου δεν θα μπορούσε να πραγματοποιηθεί με επιτυχία. Διερευνήθηκε επιπλέον ο αριθμός των γνωρισμάτων που απαιτούνται, το φαινόμενο της υπερεκπαίδευσης (overfitting), η βέλτιστη αρχιτεκτονική του δικτύου κυρίως όσον αφορά τον αριθμό των νευρώνων και η μείωση της αβεβαιότητας της εκπαίδευσης που προκύπτει από τον στοχαστικό χαρακτήρα της διαδικασίας, μέσω model ensembling. Τέλος, ο αριθμός των δεδομένων που χρησιμοποιήθηκαν κρίθηκε επαρκής για να ικανοποιήσει το παραπάνω ζητούμενο καλύπτοντας ένα σύνολο 3 ετών και 9 μηνών.

Στον παρακάτω πίνακα (5.1), παρουσιάζονται συνοπτικά τα χαρακτηριστικά των μοντέλων που δημιουργήθηκαν και η καλύτερη τιμή RMSE και BIAS που μετρήθηκε για την περίοδο ελέγχου:

Μοντέλο	Νευρώνες	Γνωρίσματα	RMSE	BIAS	Σχόλια
Baseline	200 x 200	1	55,13	-1,60	Σε περίπτωση που θεωρήσουμε δύο προγνωστικές ημέρες το **RMSE = 88,17**
3 Features	200 x 200	3	64,16	+0,15	-
High Corr	200 x 200	4	44,66	-0,10	-
All Corr	200 x 200	112	43,53	-0,07	-

Πίνακας 5.1: *Συνοπτική παρουσίαση της αξιολόγησης και των χαρακτηριστικών των τεσσάρων μοντέλων που δημιουργήθηκαν.*

5.2 Συμπεράσματα

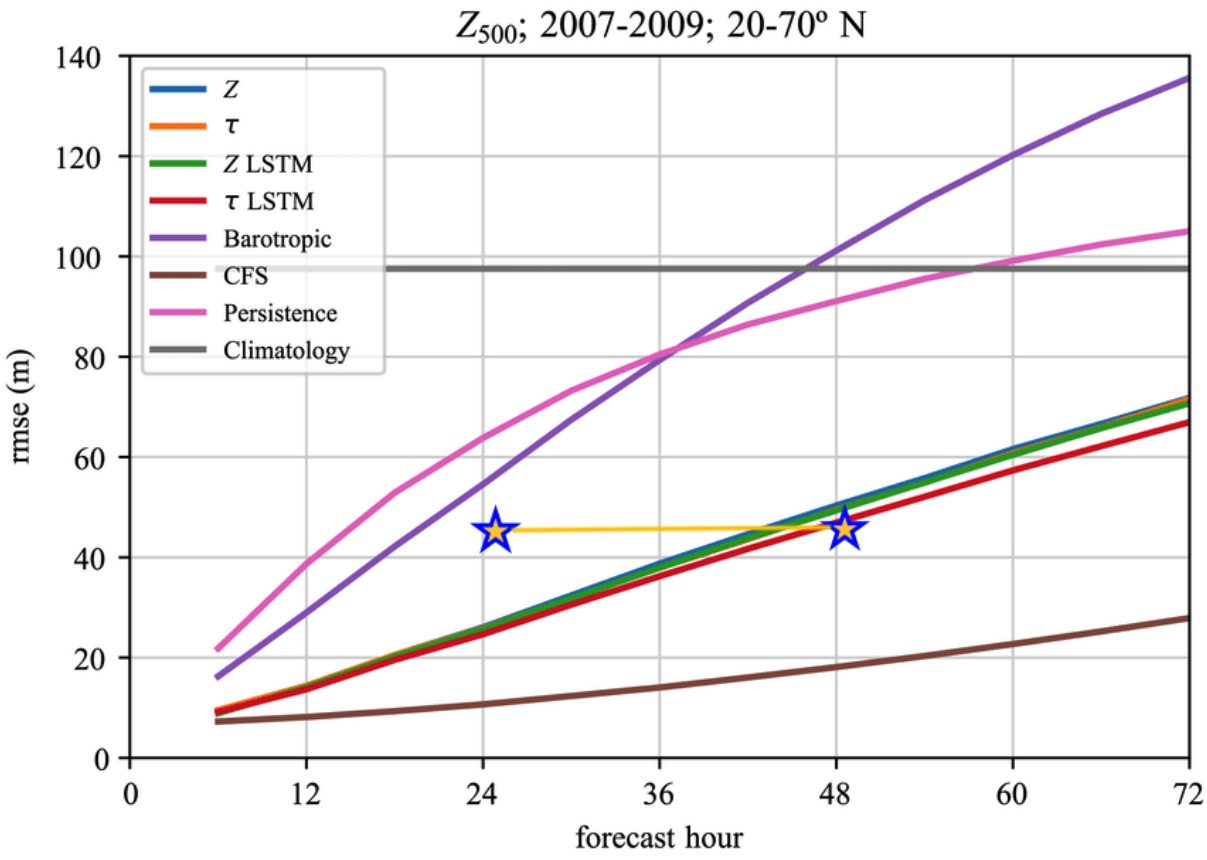
Όπως είναι αναμενόμενο, η αύξηση του αριθμού των γνωρισμάτων οδηγεί τελικά στη βελτίωση κυρίως του στατιστικού δείκτη RMSE, αλλά μέχρι κάποιο όριο, καθώς από εκεί και μετά φαίνεται πως δεν είναι σε θέση να παρέχει περισσότερη πληροφορία στο συγκεκριμένο μοντέλο. Ωθώντας σε ακραίες τιμές τις συνθήκες εκπαίδευσης, πραγματοποιήθηκαν έλεγχοι με δίκτυα 300 x 300 αλλά και 400 x 400 νευρώνων, με ακόμη μεγαλύτερο αριθμό γνωρισμάτων (έως και 1112), όπου τελικά παρατηρήθηκε **αύξηση** του δείκτη RMSE και ταυτόχρονα μείωση του βέλτιστου αριθμού των κύκλων εκπαίδευσης σε μόλις έναν, γεγονός που εγείρει δύο υποθέσεις:

1. Είτε απαιτείται μεγαλύτερο βάθος δικτύου (περισσότερα επίπεδα) για την εκμάθηση σε περισσότερο γενικές συσχετίσεις,

2. Είτε απαιτείται μεγαλύτερος όγκος δεδομένων.

Οι συγκεκριμένες υποθέσεις φαίνεται πως δεν αποκλείουν η μία την άλλη, αλλά λειτουργούν συμπληρωματικά, όπως επιβεβαιώνεται από την κοινή πρακτική (το βάθος δικτύου είναι ανάλογο του όγκου των δεδομένων). Συνεπάγεται πως, για τον συγκεκριμένο όγκο των δεδομένων, η συγκεκριμένη αρχιτεκτονική του μοντέλου που επιλέχθηκε ήταν η βέλτιστη. Το παραπάνω συμπέρασμα ενισχύει και η μελέτη αλληλεξάρτησης των μεταβλητών εκπαίδευσης και πιο συγκεκριμένα η σχέση μεταξύ αριθμού συνάψεων και βέλτιστου αριθμού κύκλων εκπαίδευσης, όπου για αριθμό συνάψεων > 10.000, η τιμή των epochs τείνει σε κάποια σταθερά - πολύ χαμηλός αριθμός.

Ενδιαφέρον παρουσιάζει το γεγονός ότι τα αποτελέσματα του δείκτη RMSE του μοντέλου 200 x 200 νευρώνων της παρούσας εργασίας, αν και υψηλότερα, παραμένουν συγκρίσιμα με άλλες αντίστοιχες. Ειδικότερα στην εργασία των Weyn et al. (2019), ένα DLWP (Deep Learning Weather Prediction) μοντέλο εκπαιδεύτηκε σε δεδομένα 30 ετών (3 εκ των οποίων χρησιμοποιήθηκαν για έλεγχο), για περιοχές μεταξύ 20-70°N. Καθώς πρόκειται για χωρική πρόγνωση, η αρχιτεκτονική του ήταν υβριδική περιλαμβάνοντας CNN επίπεδα. Στο σχήμα 5.1, παρουσιάζεται η σύγκριση των τιμών της ρίζας του μέσου τετραγωνικού σφάλματος (RMSE) ως μέσος όρος για όλο το πεδίο πρόγνωσης, για διάφορα μοντέλα συμπεριλαμβανομένου και του LSTM της έρευνας, συγκρινόμενες με την αριθμητική πρόγνωση (μοντέλο CFS).



Σχήμα 5.1: *Η εξέλιξη του δείκτη σφάλματος RMSE διαφόρων μοντέλων πρόγνωσης συμπεριλαμβανομένης και του κλιματικού μέσου όρου, ενός baseline μοντέλου (Persistence), αλλά και της αριθμητικής πρόγνωσης (CFS). Για λόγους σύγκρισης, σημειώνονται επίσης με κίτρινο αστέρι η θέση της αξιολόγησης του καλύτερου μοντέλου που αναπτύχθηκε στο πλαίσιο της παρούσας εργασίας. (Το σχήμα αποτελεί τροποποίηση του σχήματος των Weyn et al, 2019).*

Παρ’ όλο που, σε αντίθεση με το χωρικό μοντέλο των Weyn et al, η πρόγνωση στο πλαίσιο της παρούσας εργασίας πραγματοποιείται σημειακά, μια ποιοτικού τύπου σύγκριση μπορεί να οδηγήσει σε χρήσιμα συμπεράσματα. Από τον τρόπο μέσω του οποίου πραγματοποιείται ο υπολογισμός των επόμενων ημερών (τεχνική “sliding window” όπως περιγράφεται στο υποκεφάλαιο 3.2.1.1) και πιο συγκεκριμένα εφ’όσον λαμβάνονται υπόψιν οι δύο προηγούμενες ημέρες και επιχειρείται η πρόγνωση για τις δύο επόμενες έχοντας μία ημέρα κοινή (διαδικασία που περιγράφεται στο υποκεφάλαιο), τότε υπάρχουν δύο τρόποι να γίνει η σύγκριση των αποτελεσμάτων:

1.Εάν το ζητούμενο είναι η επιχειρησιακή εφαρμογή μιας τέτοιας μεθόδου, τότε, η σύγκριση με άλλες προγνώσεις πρέπει να πραγματοποιηθεί για χρονικό ορίζοντα 24 ωρών εφ’όσον η τιμή της πρώτης ημέρας που αποτελεί κοινή ημέρα με τα δεδομένα εισόδου, είναι ήδη γνωστή.

2. Ωστόσο, αυτό που στην πραγματικότητα υπολογίζεται είναι η διαφορά της τιμής σε σχέση με την προηγούμενη ημέρα για δύο συνεχόμενες ημέρες. Σε αυτήν την περίπτωση, η σύγκριση της αξιολόγησης του συγκεκριμένου μοντέλου θα πρέπει να γίνει με βάση τα scores των 48 ωρών

Στην πρώτη περίπτωση, με βάση το σχήμα 5.1 της εργασίας των Weyn et al, οι τιμές του RMSE είναι πλέον σε καλύτερα επίπεδα σε σχέση με ένα βαροτροπικό μοντέλο αριθμητικής πρόγνωσης αλλά υψηλότερες τόσο σε σχέση με το z LSTM μοντέλο των ερευνητών, όσο και σε σχέση με ένα σύγχρονου μοντέλο αριθμητικής πρόγνωσης (CFS). Στην δεύτερη περίπτωση όμως, η τιμή της ρίζας του μέσου τετραγωνικού σφάλματος είναι τελικά συγκρίσιμη με τα αποτελέσματα της συγκεκριμένης έρευνας.

Αν και ο ερευνητικός τομέας της μηχανικής εκμάθησης και της τεχνητής νοημοσύνης εξελίσσεται ραγδαία την τελευταία δεκαετία με εφαρμογές σε πολλούς και διαφορετικούς τομείς, φαίνεται πως οι έρευνα γύρω από ζητήματα πρόγνωσης της ατμόσφαιρας είναι πολύ πρόσφατη και περιορισμένη. Η πεποίθηση του συγγραφέα είναι ότι εφ’ όσον το ζητούμενο προσεγγιστεί και από τη σκοπιά των δεδομένων, η μηχανική εκμάθηση μπορεί να συμβάλλει στη μελέτη των πλανητικών κυματισμών και των τηλεσυνδέσεων, προσφέροντας μια επιπρόσθετη πληροφορία στις ήδη υπάρχουσες μεθόδους πρόγνωσης. Εξάλλου, ακόμη και στο περιορισμένο πλαίσιο της παρούσας μελέτης, προκύπτει το λογικό συμπέρασμα ότι υπάρχουν “κρυφές” συσχετίσεις στις χρονοσειρές των δεδομένων, τις οποίες ακόμη κι ένα περιορισμένης ικανότητας μοντέλο ήταν σε θέση να μάθει και να χρησιμοποιήσει σε καινούργια δεδομένα.

Αν και ο χαρακτήρας της ατμόσφαιρας είναι χαοτικός, θέτοντας όρια στην αιτιοκρατική προσέγγιση της πρόγνωσης (όπως απέδειξε και ο Edward N. Lorenz το 1963), εντούτοις είναι δυνατόν να προκύψουν μοτίβα, όπως συμβαίνει σε κάθε χαοτικό σύστημα. Πρώτα η αναγνώρισή τους σε χωρικό επίπεδο και στη συνέχεια η πρόγνωσή τους, κατά την άποψη του συγγραφέα, είναι ενδεχομένως και ο τρόπος με τον οποίο θα μπορούσε να η επιστήμη των δεδομένων να συμβάλλει αποφασιστικά στην επιχειρησιακή πρόγνωση του καιρού αλλά και του κλίματος.

Βιβλιογραφία

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. and Kudlur, M., 2016. ***{TensorFlow}: a system for {Large-Scale} machine learning***. In 12th USENIX symposium on operating systems design and implementation (OSDI 16) (pp. 265-283).

Aurélien Géron, 2017, ***Hands-On Machine Learning with Scikit-Learn and TensorFlow***, O'Reilly Media, Inc., ISBN: 9781491962299

Bjerknes, J., & Solberg H., 1922, ***Life Cycle of Cyclons and the Polar Front Theory of Atmospheric Circulation***. Monthly Weather Review, 50(9), 468-473

Bontempi G., Ben Taieb S., Le Borgne YA., 2013, ***Machine Learning Strategies for Time Series Forecasting***. In: Aufaure, MA., Zimányi, E. (eds) Business Intelligence. eBISS 2012. Lecture Notes in Business Information Processing, vol 138. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-36318-4_3

Bryson A., & Ho Y.-C., 1969, ***Applied Optimal Control***, New York: Blaisdell.

Charney J. G., R. FjÖrtoft & J. Von Neumann, 1950, ***Numerical Integration of the Barotropic Vorticity Equation***, Tellus, 2:4, 237-254

Chapman W.E., Subramanian A.C., Monache L.D., Xie S.P., Ralph F.M., 2019, ***Improving Atmospheric River Forecasts With Machine Learning,*** Geophysical Research Letters, V. 46, Issue 17-18, 10,627–10,635

Chen T., Li M., Li Y., Lin M., Wang N., Wang M., Xiao T., Xu B., Zhang C., Zhang Z., 2015, ***Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems***. arXiv preprint arXiv:1512.01274.

Dickey D. A., Fuller, W. A, 1979, "***Distribution of the Estimators for Autoregressive Time Series with a Unit Root***". Journal of the American Statistical Association. 74 (366): 427–431.

Dietterich, T.G., 2002, ***Machine Learning for Sequential Data: A Review***. In: Caelli, T., Amin, A., Duin, R.P.W., de Ridder, D., Kamel, M. (eds) Structural, Syntactic, and Statistical Pattern Recognition. SSPR /SPR 2002. Lecture Notes in Computer Science, vol 2396. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-70659-3_2

Dozat T., 2015, ***Incorporating Nesterov Momentum into Adam***, 4th International Conference for Learning Representations, San Juan, Puerto Rico, 2016

Duchi J., Hazan E., Singer Y., 2011, ***Adaptive subgradient methods for online learning and stochastic optimization***, Journal of Machine Learning Research, V.12, 2121-2159

Elshawi, R., Wahab, A., Barnawi, A. & Shakr S., 2021, ***DLBench: a comprehensive experimental evaluation of deep learning frameworks***, Cluster Computing 24, 2017–2038

Frankel, D. S., J. S. Draper, J. E. Peak, and J. C. McLeod, 1995, ***Artificial Intelligence Needs Workshop***, 4–5 November 1993, Boston, Massachusetts.Bull. Amer. Meteor. Soc.,76,728–738.

Fukushima K., 1980, ***Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position***, Biological Cybernetics. 36 (4): 193–202

Galanis G., Louka P., Katsafados P., Kallos G., Pytharoulis I., 2006, ***Applications of Kalman filters based on non-linear functions to numerical weather predictions***, Annales Geophysicae, 24, 2451-2460

Geron A., 2017, ***Hands-on Machine Learning with Sci-Kit Learn and Tensorflow***, O'Reilly Media Inc., ISBN: 9781491962299

Han, Kamber, Pei, Jaiwei, Micheline, Jian, 2011, ***Data Mining: Concepts and Techniques* (3rd ed.)**. Morgan Kaufmann. ISBN 9780123814791

Harper K, Uccelini L.W., Kalnay E., Carey K., Morone L., 2007, ***50th Anniversary of Operational Numerical Weather Prediction***, Bulletin of the American Meteorological Society, V.88, Issue 5, 639

Hoang D. T., Yang Pr. L., Cuong L.D., Trung Dr. P. D., Tu Dr. N. H., Truong L. V. , Hien T. T, Nha V. T., 2020, ***Weather prediction based on LSTM model implemented AWS Machine Learning Platform***, International Journal for Research in Applied Science & Engineering Technology (IJRASET), V.8, Issue V, 283-290

Hinton G.E., Osindero S., Teh Yee-Whye, 2006, ***A Fast Learning Algorithm for Deep Belief Nets,*** Neural Computation, 18 (7), 1527-1554.

Hochreiter S., & Schmidhuber J., 1997, ***Long Short-Term Memory***, Neural Computation, 9, 1735-1780.

Homma T., Atlas L. E., Marks II R.J., 1988, ***An artificial Neural Network for Bipolar Patterns: Application to Phoneme Classification***, Advances in Neural Information Processing Systems. 1: 31–40.

Izmailov P., Podoprikhin D., Garipov T., Vetrov D., Wilson A.G., 2019, ***Averaging Weights Leads to Wider Optima and Better Generalization***, Conference on Uncertainty in Artificial Intelligence (UAI), 2018

Kalnay E., 2002, ***Atmospheric Modeling, Data Assimilation and Predictability***, Cambridge University Press, ISBN: 9780521796293

Kingma D.P., Ba J., 2014, ***Adam: A Method for Stochastic Optimization***, 3rd International Conference for Learning Representations, San Diego, 2015

Kirkpatrick S., Gelatt Jr, C. D., Vecchi, M. P., 1983, ***Optimization by Simulated Annealing***,. Science. 220 (4598): 671–680

Krizhevsky A., Sutskever I., Hinton G.E., 2012, ***ImageNet Classification with Deep Convolutional Neural Networks***, Advances in Neural Information Processing Systems 25 (NIPS 2012)

Kuligowski, R. J., & Barros, A. P., 1998, ***Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks***, Weather and Forecasting, 13(4), 1194–1204.

LeCun Y., Boser B, Denker J. S., Henderson D., Howard R. E, Hubbard W., Jackel L. D., 1989, ***Backpropagation Applied to Handwritten Zip Code Recognition***, Neural Computation. 1 (4): 541–551.

LeCun Y., Bottou L., Bengio Y., Haffner P., 1998, ***Gradient-Based Learning Applied to Document Recognition,*** Proceedings of the IEEE. 86 (11): 2278–2324

Lorenz E., 1963, ***Deterministic Nonperiodic Flow***, Journal of Atmospheric Sciences, V.20, Issue 2, 130-141

Loshchilov I., Hutter F., 2016, ***SGDR: Stochastic Gradient Descent with Warm Restarts***, arXiv preprint arXiv:1608.03983

Lynch, P., 2008, ***The Origins of Computer Weather Prediction and Climate Modeling,*** Journal of Computational Physics, 227, 3441-3444

Lynch, P., 2008, ***The ENIAC Forecasts: A Re-creation*****,** Bulletin of the American Meteorological Society, V.89, Issue 1, 45-56

Lynch, P., 2008, ***The origins of computer weather prediction and climate modeling***, Journal of Computational Physics. 227 (7): 3431–44

Marzban, C., and G. J. Stumpf, 1996, ***A neural network for tornado prediction based on Doppler radar–derived attributes***, J. Appl.Meteor., V.35, 617–626

McCarthy, 2004, ***What is Artificial Intelligence?***, Stanford University Article

McCulloch W. and Pitts W., 1943, ***A logical calculus of the ideas immanent in nervous activity***, Bulletin of Mathematical Biophysics, 7:115–133.

McMahan H.B., Holt G., Sculley D., Young M., Ebner D., Grady J., Nie L., Phillips T., Davydov E., Golovin D., Chikkerur S., Liu D., Wattenberg M., Hrafnkelsson A.M., Boulos T., Kubica J., ***Ad Click Prediction: a View from the Trenches,*** Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (2013): n. pag.

Minsky, M. & Papert, S., 1969, ***Perceptrons: An Introduction to Computational Geometry*** , MIT Press , Cambridge, MA, USA .

Neubig, G., Dyer, C., Goldberg, Y., Matthews, A., Ammar, W., Anastasopoulos, A., Ballesteros, M., Chiang, D., Clothiaux, D., Cohn, T. and Duh, K., 2017. ***Dynet: The dynamic neural network toolkit.*** arXiv preprint arXiv:1701.03980.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L. and Lerer, A., 2017, ***Automatic Differentiation in Pytorch***, 31st Conference on Neural Information Processing Systems (NIPS 2017), 4-9 December 2017, Long Beach, CA, USA, 1-4.

Patterson, J. and Gibson, A., 2017. ***Deep learning: A practitioner's approach.*** "O'Reilly Media, Inc.", 6

Pearson K., 20 June 1895, ***Notes on regression and inheritance in the case of two parents***, Proceedings of the Royal Society of London. 58: 240–242.

Rasp S. & Lerch S., 2018, ***Neural Networks for Post Processing Ensemble Weather Forecasts***, Monthly Weather Review, Vol. 146, Issue 11, 3885-3900

Richardson, L., & Lynch, P., 2007, ***Weather Prediction by Numerical Process*** (2nd ed., Cambridge Mathematical Library), Cambridge: Cambridge University Press.

Robbins H & Monro S., 1951, ***A Stochastic Approximation Method,*** The Annals of Mathematical Statistics, Vol. 22, No. 3., 400-407

Rosenblatt F., 1958, ***The perceptron: A probabilistic model for information storage and organization in the brain***, Psychological Review, 65(6), 386–408

Rumelhart, David E.; Hinton, Geoffrey E.; Williams, Ronald J., 1986a, ***"Learning representations by back-propagating errors"***. Nature. 323 (6088): 533–536.

Russell S., and Norvig P., 1995, ***Artificial Intelligence: A Modern Approach***, Englewood Cliffs, NJ: Prentice Hall.

Sarker I.H., 2021, ***Deep Learning: A Comprehensive Overview on Techniques***, Taxonomy, Applications and Research Directions, SN Computer Science volume 2, Article number: 420

Schmidhuber J., 2014, ***Deep Learning in Neural Networks: An Overview,*** Technical Report IDSIA-03-14 / arXiv:1404.7828 v4

Spearman C., 1904, ***The Proof and Measurement of Association Between Two Things,*** American Journal of Psychology. 15 (1): 72–101

Staff Members of the Institute of Meteorology, University of Stockholm, 1954, ***Results of Forecasting with the Barotropic Model on an Electronic Computer (BESK)***, Tellus 6 (1954): 139-149.

Werbos P. J., 1981, ***Applications of advances in nonlinear sensitivity analysis***. In Proceedings of the 10th IFIP Conference, 31.8 - 4.9, NYC, pages 762–770.

Weyn J. A., Durran D. R. D. Caruana R.., 2019, ***Can Machines Learn to Predict Weather? Using Deep Learning to Predict Gridded 500-hPa Geopotential Height From Historical Weather Data***, Journal of Advances in Modeling Earth Systems, 11. 10.1029/2019MS001705.

Weyn J. A., Durran D., Caruana R., 2020, ***Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere***, Journal of Advances in Modeling Earth Systems, Volume 12, Issue 9.

Whittle P., 1951, ***Hypothesis Testing in Time Series Analysis***, Volume 4 of Statistics Upsala.

Environmental Modeling Center, EMC, College Park, MD; and K. Campana, P. Caplan, D. J. Halperin, B. Lapenta, S. Lilly, Y. Lin, A. B. Penny, S. Saha, V. Tallapragada, **G. H. White**, J. S. Woollen, and F. Yang, 2019, ***The Development and Success of NCEP's Global Forecast System,*** 99th Annual Meeting of the American Meteorological Society, 6-10th January 2019.

Wilks Daniel S., 2011, ***Statistical Methods in the Atmospheric Sciences***, 3rd edition, Academic Press, pp.306-315

World Meteorological Organization, 2019, Technical Regulations, Basic Documents No.2 , ***Volume I - General Meteorological Standards and Recommended Practices***, WMO-No.49

Zeiler M.D., 2012, Adadelta: ***An adaptive learning rate method***, arXiv preprint arXiv:1212.5701.

Zhang W., Tanida J., Itoh K. and Ichioka Y., 1988, ***Shift-invariant pattern recognition neural network and its optical architecture***, Proceedings of Annual Conference of the Japan Society of Applied Physics.

Κατσαφάδος Π., Μαυροματίδης Η., 2015, ***Εισαγωγή στη Φυσική της Ατμόσφαιρας και την Κλιματική Αλλαγή,*** Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράματα και Βοηθήματα, ISBN: 978-960-603-053-6

Διαδίκτυο

Brownlee J., 2019, ***How to Fix the Vanishing Gradients Problem Using the ReLU***, https://machinelearningmastery.com/how-to-fix-vanishing-gradients-using-the-rectified-linear-activation-function/, Ανακτήθηκε: Αύγουστος 2022

Brownlee, J., 2020, ***4 Types of Classification Tasks in Machine Learning,*** https://machinelearningmastery.com/types-of-classification-in-machine-learning/**,** Ανακτήθηκε: Απρίλιος 2022

Nielsen M., 2019, ***Neural Networks and Deep Learning***, Online book: http://neuralnetworksanddeeplearning.com/

Olah C., 2015, ***Understanding LSTM Networks,*** https://colah.github.io/posts/2015-08-Understanding-LSTMs/, Blog Post, Ανακτήθηκε: Αύγουστος 2022

Yashwardhan Jain, 2018. ***Tensorflow or PyTorch : The force is strong with which one?***, ιστότοπος medium.com, Ανακτήθηκε: Αύγουστος 2022

Ιστότοπος Εθνικής Μετεωρολογικής Υπηρεσίας, https://www.emy.gr, Φεβρουάριος 2021

Ιστότοπος προγράμματος COMET, https://www.comet.ucar.edu/, Φεβρουάριος 2023

Ιστότοπος Wikipedia, https://en.wikipedia.org/, Δεκέμβριος 2020

Ιστότοπος UK Met. Office, https://www.metoffice.gov.uk/weather/learn-about/how-forecasts-are-made/computer-models/history-of-numerical-weather-prediction, Φεβρουάριος 2023

Ιστότοπος IBM, Artificial Intelligence (AI), https://www.ibm.com/cloud/learn/what-is-artificial-intelligence, Μάιος 2022

Ιστοσελίδα Εθνικού Κέντρου Περιβαλλοντικών Προγνώσεων (National Centers for Environmental Prediction - NCEP) των ΗΠΑ, https://www.weather.gov/ncep/, Μάιος 2022

Ιστοσελίδα Παγκόσμιου Προγνωστικού Συστήματος του NCEP, (Global Forecasting System - GFS), https://www.nco.ncep.noaa.gov/pmb/products/gfs/, Μάιος 2022