



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ  
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ  
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ & ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

---

**Πρόγραμμα Μεταπτυχιακών Σπουδών**  
**«ΔΙΑΔΙΚΤΥΟ ΤΩΝ ΠΡΑΓΜΑΤΩΝ ΚΑΙ ΕΥΦΥΗ ΠΕΡΙΒΑΛΛΟΝΤΑ»**

---

**ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

---

Πρόβλεψη των θανάτων από COVID-19 με χρήση αναδρομικών  
νευρωνικών δικτύων



Μεταπτυχιακός Φοιτητής: Φώτιος Θεμελής, AM msciot19001

Επιβλέπων: Αλεξανδρίδης Αλέξανδρος, Καθηγητής

**ΑΙΓΑΛΕΩ, Ιούλιος 2023**



UNIVERSITY OF WEST ATTICA  
FACULTY OF ENGINEERING  
DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING

---

**Master of Science in  
“INTERNET of THINGS AND INTELLIGENT ENVIRONMENTS”**

---

**MSc Thesis**

---

Predicting Covid-19 deaths using Recurrent Neural Networks



---

Student: Themelis Fotis, Registration Number msciot19001

MSc Thesis Supervisor: Alexandridis Alexandros, Professor

ATHENS-EGALEO, JULY 2023

Η Διπλωματική Εργασία έγινε αποδεκτή και βαθμολογήθηκε από την εξής τριμελή επιτροπή:

Αλεξανδρίδης Αλέξανδρος Καθηγητής (Επιβλέπων)	Ζώης Ηλίας Αναπληρωτής Καθηγητής (Μέλος)	Παπαδόπουλος Περικλής Καθηγητής (Μέλος)

**Copyright ©** Με επιφύλαξη παντός δικαιώματος. All rights reserved.

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ και Φώτιος Θεμελής,**

**Ιούλιος 2023**

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

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον/την συγγραφέα του και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις θέσεις του επιβλέποντος, της επιτροπής εξέτασης ή τις επίσημες θέσεις του Τμήματος και του Ιδρύματος.

#### **ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΟΥ**

Ο κάτωθι υπογεγραμμένος: Φώτιος Θεμελής του Χρήστου, με αριθμό μητρώου msciot19001, φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής ΜΗΧΑΝΙΚΩΝ του Τμήματος ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ,

#### **δηλώνω υπεύθυνα ότι:**

«Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου».

Ο Δηλών

Φώτιος Θεμελής

## ΠΕΡΙΛΗΨΗ

Η παρούσα πτυχιακή εργασία αποσκοπεί στην μελέτη αλγορίθμων νευρωνικών δικτύων με σκοπό την πρόβλεψη δεδομένων και εξαγωγή συμπερασμάτων που αφορούν στην πανδημία Covid-19. Το θέμα της πανδημίας απασχολεί με ποικίλους τρόπους την ανθρωπότητα παντού και σε πολλούς διαφορετικούς τομείς, οπότε το αυξημένο ενδιαφέρον για την πορεία της διασποράς της αποτελεί συχνή πυξίδα για τον προσανατολισμό των ερευνών.

Τα νευρωνικά δίκτυα υπάγονται στο ευρύτερο πεδίο της τεχνητής νοημοσύνης η οποία είναι ιδιαίτερα επίκαιρη και έχει ανασύρει στην επιφάνεια νέες προοπτικές, προκλήσεις και λύσεις στο χώρο της τεχνολογίας.

Στην παρούσα έρευνα το ερώτημα είναι, αν είναι εφικτό να αναπτυχθεί ένα μοντέλο πρόβλεψης το οποίο να έχει την δυνατότητα να τροφοδοτείται με δεδομένα σχετικά με την διασπορά του Covid-19 και να εκτιμάται με ικανοποιητική ακρίβεια το ύψος του πλήθους των θανάτων στο μέλλον. Για την ανάπτυξη του μοντέλου πρόβλεψης χρησιμοποιούνται αλγόριθμοι που στηρίζονται στα νευρωνικά δίκτυα. Η αρχιτεκτονική που εξετάστηκε ήταν αυτή των αναδρομικών νευρωνικών δικτύων (Recurrent Neural Networks) που στηρίζεται στην αναδρομική τροφοδοσία των εισόδων.

Στην διάρκεια της πειραματικής διαδικασίας, πραγματοποιήθηκε σύγκριση του νευρωνικού δικτύου με το δίκτυο ακτινικής συμμετρίας (Radial Basis Function) με σκοπό την εύρεση του καταλληλότερου μοντέλου προβλέψεων για τα συγκεκριμένα δεδομένα βάσει αποτελεσμάτων.

Τα αποτελέσματα επιβεβαίωσαν ότι τα αναδρομικά νευρωνικά δίκτυα αποδίδουν καλύτερα σε σχέση με τα δίκτυα RBF για τα συγκεκριμένα δεδομένα χρονοσειρών.

**Λέξεις-Κλειδιά:** Αναδρομικά Νευρωνικά Δίκτυα, Μηχανική Μάθηση, Covid-19, δίκτυα ακτινικής συμμετρίας RBF

## ABSTRACT

The aim of this Thesis is to study neural network algorithms in order to predict data and draw conclusions related to the Covid-19 pandemic. The topic of the pandemic concerns the entire humanity in several ways in many different fields, so the increased interest in the course of its spread often guides the orientation of research. Neural Networks is a subfield of artificial intelligence which has brought to the surface new perspectives, challenges and solutions in the field of technology.

In the present research, the question is whether it is possible to develop a prediction model that has the ability to be fed with data about the spread of Covid-19 and estimate with sufficient accuracy the number of deaths in the future. Algorithms based on neural networks are used to develop the prediction model.

The architecture that was examined was that of recurrent neural networks (RNNs) which is based on the recursive feeding of the inputs. During the experimental process, a comparison of the neural network with the Radial Basis Function (RBF) model was carried out in order to find the model which more efficiently handles that type of data.

The results confirmed that RNNs perform better than RBF networks for the particular time series data.

**Keywords:** Recurrent Neural Networks, Machine Learning, Covid-19, radial basis function neural networks.

## ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Προκλήσεις	10
Σχετικές έρευνες	12
Παρούσα Έρευνα	13
Δομή Εργασίας	14
1 Μηχανική Μάθηση	15
1.1 Αλγόριθμοι μηχανικής μάθησης με επίβλεψη	16
1.2 Αλγόριθμοι μηχανικής μάθησης χωρίς επίβλεψη	17
1.3 Αλγόριθμοι Ενισχυτικής Μάθησης	18
1.4 Υβριδικές προσεγγίσεις	19
1.4.1 Ημιεποπτευόμενη μάθηση	19
1.4.2 Αυτοεποπτευόμενη μάθηση	19
1.4.3 Αυτοδίδακτη μάθηση	19
1.4.4 Άλλες προσεγγίσεις	20
2. Νευρωνικά Δίκτυα	22
2.1 Βιολογικά και Τεχνητά Νευρωνικά δίκτυα	22
2.2 Τεχνητά Νευρωνικά Δίκτυα	23
2.3 Δομή και Λειτουργία των Τεχνητών Νευρωνικών Δικτύων	23
2.4 Συναρτήσεις ενεργοποίησης	24
2.4.1 Hard Limit	24
2.4.2 Linear Functions	25
2.4.3 Log-Sigmoid Functions	25
2.4.4 Νευρώνες πολλαπλών εισόδων	26
3. Αρχιτεκτονικές Νευρωνικών Δικτύων	27
3.1 Χαρακτηριστικά	27
3.1.1 Επίπεδο Νευρωνικού Δικτύου	27



3.1.2 Πολυεπίπεδα Νευρωνικά Δίκτυα	28
3.1.3 Επίπεδο Εισόδου - Επίπεδο Εξόδου	29
3.1.4 Εσωτερικά (κρυφά) επίπεδα	30
3.2 Deep Learning	30
3.3 Feed Forward Τεχνητά Νευρωνικά Δίκτυα	31
3.4 Backpropagation	32
3.5 Recursive Neural Networks	34
3.6 Convolutional Neural Networks	34
3.7 Radial Basis Function Networks	35
4. Recurrent Neural Networks	38
4.1 Χαρακτηριστικά	38
4.2 Δομή	39
4.3 Εκπαίδευση RNN	41
4.4 Long-Short Term Τεχνητά Νευρωνικά Δίκτυα (LSTMNN)	42
5. Υπολογιστικά Εργαλεία για Διαδικασίες Μηχανικής Μάθησης	45
5.1 Η γλώσσα προγραμματισμού Python	45
5.2 Η πλατφόρμα Tensorflow	48
5.3 Η βιβλιοθήκη Keras	48
6. Case study	50
6.1 Περιγραφή	50
6.2 Χαρακτηριστικά της διασποράς του Covid-19	50
6.3 Σύνολα Δεδομένων	51
6.4 Υλοποίηση	51
6.5 Αποτελέσματα	54

---

## ΕΙΣΑΓΩΓΗ

### Αντικείμενο, ερευνητικά ερωτήματα και διάρθρωση της εργασίας

---

#### **Σκοπός και σημασία**

Από το τέλος του 2019, ο πληθυσμός του πλανήτη ταλανίζεται από τις επιπτώσεις της ραγδαίας εξάπλωσης του ιού Covid-19. Η πανδημία – όπως αναφέρεται από μεγάλο μέρος της επιστημονικής κοινότητας – καθορίζει εδώ και κάποια χρόνια την καθημερινότητα αλλά και κρίσιμες δραστηριότητες σε κάθε επίπεδο της ανθρώπινης δραστηριότητας. Η εξέλιξη και οι μεταλλάξεις του ιού, επηρεάζουν την κατάσταση της υγείας μεγάλου μέρους του πληθυσμού καθημερινά. Το γεγονός αυτό, καθιστά την καθημερινή παρακολούθηση του πλήθους των ανθρώπων που προσβάλλονται από τον ιό, απαραίτητη, προκειμένου αυτό το πλήθος να παραμένει σε διαχειρίσιμα επίπεδα. Μέσα από την παρακολούθηση αυτή, μπορεί να προκύψουν πολύτιμα συμπεράσματα που μπορούν να χρησιμοποιηθούν στις διαδικασίες λήψης αποφάσεων για την διαχείριση της πανδημίας. Ένα σημαντικό εργαλείο που μπορεί να προσφέρει σημαντικές υπηρεσίες στην προσπάθεια αυτή είναι η μηχανική μάθηση[1]. Μέσω κατάλληλα σχεδιασμένων διαδικασιών είναι εφικτό να αναπτύσσονται ισχυρά μοντέλα πρόβλεψης του αριθμού των κρουσμάτων προσβολής ανθρώπων καθώς και την θνησιμότητα νοσούντων από τον ιό. Τα μοντέλα θα πρέπει να βασίζονται στις παραμέτρους που συνθέτουν τις αιτίες νόσησης και μετάδοσης της νόσου στον πληθυσμό. Αν οι υγειονομικές αρχές είναι σε θέση να προβλέπουν την καθημερινή εξέλιξη του αριθμού των κρουσμάτων, θα μπορούν να καθορίζουν τις αποτελεσματικότερες πολιτικές αντιμετώπισης της διασποράς αλλά και να προετοιμάζουν κατάλληλα το σύστημα υγείας για να αντέχει στην ενδεχόμενη πίεση που μπορεί να δεχθεί.

#### *Προκλήσεις*

Κατά την διάρκεια μελέτης του φαινομένου της πανδημίας και των δεδομένων που εξάγονται καθημερινά απ' αυτό, δημιουργήθηκαν αρκετά εμπόδια όπου συνδυαστικά με την σοβαρότητα του ζητήματος ανασύρουν στην επιφάνεια πολλές προκλήσεις.

Η επιστημονική κοινότητα το τελευταίο διάστημα έκανε βήματα για την αποτελεσματική μέθοδο αντιμετώπισης του ιού. Η κατάσταση δείχνει πιο σταθερή και πολλές χώρες έχουν προχωρήσει σε ελάφρυνση των μέτρων παρ' όλα αυτά δεν πρέπει να

υπάρχει εφησυχασμός, γιατί παρατηρείτε ανά περιόδους έξαρση του φαινομένου και η ανάγκη για παρακολούθηση των κρουσμάτων και των συνεπειών τους είναι ακόμα υπαρκτή.

Σημαντική πρόκληση είναι η ανακάλυψη ικανού όγκου κατάλληλων δεδομένων που θα χρησιμοποιηθούν ως είσοδος στις σχεδιαζόμενες διαδικασίες. Πολλές φορές είναι αναγκαία και η ανάπτυξη ανάλογων διαδικασιών για την δημιουργία των δεδομένων αυτών. Το διαδίκτυο την σύγχρονη εποχή είναι μία πλούσια πηγή δεδομένων σχετικών για έρευνες προσανατολισμένων σε μία ποικιλία επιστημονικών τομέων. Το πρόβλημα πλέον έγκειται στην ανίχνευση εκείνου του μέρους των δεδομένων που είναι ικανά να οδηγήσουν στην αξιόπιστη εκτίμηση καταστάσεων.

Σε κάθε περίπτωση τα υπό παρατήρηση δεδομένα για τέτοιου είδους προβλήματα είναι χρονικές ακολουθίες που ονομάζονται χρονοσειρές (timeseries). Οι χρονοσειρές αναπαριστούν την εξέλιξη μιας μεταβλητής με την πάροδο του χρόνου, όπου οι τιμές είναι συνδεδεμένες μεταξύ τους με κάποιον τρόπο.

Το κύριο πρόβλημα στην ανάλυση χρονοσειρών είναι η πρόβλεψη των μελλοντικών τιμών μιας μεταβλητής. Αυτό απαιτεί την κατανόηση των παρελθοντικών παρατηρήσεων και των παραμέτρων που επηρεάζουν την εξέλιξη της χρονοσειράς. Οι μέθοδοι μηχανικής μάθησης, όπως οι αλγόριθμοι πρόβλεψης χρονοσειρών, μπορούν να χρησιμοποιηθούν για να εξαγάγουν μοτίβα και να κάνουν προβλέψεις βάσει αυτών.

Ως επί το πλείστον οι μελέτες που γίνονται σε χρονοσειρές αφορούν μη γραμμικές χρονοσειρές διότι τα γραμμικά δυναμικά συστήματα έχουν πιο απλές λύσεις. Το κεντρικό σημείο που διαφοροποιεί τα προβλήματα χρονοσειρών από τα περισσότερα άλλα στατιστικά προβλήματα είναι ότι σε μια χρονοσειρά, οι τιμές δεν είναι αμοιβαία ανεξάρτητες. Ένα τυχαίο συμβάν μπορεί να επηρεάσει όλα τα μεταγενέστερα σημεία δεδομένων. Αυτό κάνει την ανάλυση χρονοσειρών αρκετά διαφορετική από τους περισσότερους άλλους τομείς στατιστικών [2], [3].

Ένα πολύ σημαντικό στοιχείο στις χρονοσειρές είναι ο θόρυβος, σε κάποιες περιπτώσεις μπορεί να δημιουργεί ένα μοτίβο και να εξαχθούν στοιχεία της δομής της ακολουθίας όπως οι τάσεις και η εποχικότητα τα οποία μπορούν να αξιοποιηθούν για τη βελτίωση της ικανότητας του μοντέλου, σε άλλες περιπτώσεις ωστόσο το στοιχείο της τυχαιότητας είναι πιο έντονο και είναι αρκετά δύσκολο να αντιμετωπιστεί [4]

Τα νευρωνικά δίκτυα προσεγγίζουν μια συνάρτηση από μεταβλητές εισόδου σε μεταβλητές εξόδου. Αυτή η γενική ικανότητα είναι πολύτιμη για χρονοσειρές για διάφορους λόγους.

Τα νευρωνικά δίκτυα έχουν πολλά οφέλη όταν πρόκειται για την πρόβλεψη χρονοσειρών. Έχουν τη δυνατότητα να αντιμετωπίσουν το θόρυβο στα δεδομένα εισόδου, καθώς μπορούν να εκπαιδευτούν για να εντοπίζουν τα μοτίβα και τις τάσεις που υπάρχουν πέρα από τον θόρυβο. Επίσης μπορούν να αντιμετωπίσουν την απουσία τιμών (missing values) στις χρονοσειρές, καθώς μπορούν να εκμεταλλευτούν τις σχέσεις μεταξύ των διαθέσιμων τιμών και να προβλέψουν τις απουσιάζουσες τιμές.

Τα νευρωνικά δίκτυα είναι επίσης ευέλικτα όσον αφορά τον αριθμό των εισόδων και των εξόδων. Μπορούν να ρυθμιστούν για να υποστηρίζουν οποιονδήποτε αριθμό εισόδων και εξόδων στη συνάρτηση χαρτογράφησης, ανάλογα με τις απαιτήσεις του προβλήματος.

Τέλος, τα αναδρομικά νευρωνικά δίκτυα έχουν τη δυνατότητα να αναγνωρίζουν τη χρονική εξάρτηση των δεδομένων. Αυτό σημαίνει ότι μπορούν να αναλύσουν τα μοτίβα και τις τάσεις που εμφανίζονται σε διαφορετικά χρονικά σημεία και να τις χρησιμοποιήσουν για προβλέψεις στο μέλλον.

Όλα αυτά τα χαρακτηριστικά καθιστούν τα αναδρομικά νευρωνικά δίκτυα ιδιαίτερα ισχυρά εργαλεία για την ανάλυση και πρόβλεψη χρονοσειρών [4].

## Σχετικές έρευνες

Η μεγάλη σημασία που έχει η πρόβλεψη της υγειονομικής κατάστασης στο επίπεδο του αριθμού των κρουσμάτων προσβολής ανθρώπων από τον Covid-19, έχει ενεργοποιήσει την επιστημονική κοινότητα σε μια διαρκή προσπάθεια αναζήτησης αξιόπιστων μεθόδων και τεχνικών για τον σκοπό αυτό.

Μια από τις προτάσεις αφορούσε την εφαρμογή μιας Bayesian μεθόδου εκτίμησης για την δυναμική εκτίμηση των επιδημιολογικών παραμέτρων με τη χρήση των στατιστικών στοιχείων του πραγματικού αριθμού λοιμώξεων σε καθημερινή βάση. Η μελέτη αυτή κατέδειξε τη σημασία και την αποτελεσματικότητα της επιβολής αυστηρών ταξιδιωτικών περιορισμών (πέραν των βασικών) και των κρατικών παρεμβάσεων για τον αποτελεσματικό περιορισμό της διάδοσης του ιού στον πληθυσμό [5].

Μία άλλη μελέτη εστίασε στην επίδραση των διαφορετικών στρατηγικών ελέγχου ως παρεμβάσεις που εξαρτώνται από το χρόνο, χρησιμοποιώντας μαθηματική μοντελοποίηση και τεχνικές ελέγχου βελτιστοποίησης, για να εξακριβωθεί η συμβολή τους στη δυναμική μετάδοσης του COVID-19. Το μοντέλο αποδείχθηκε ότι οδηγούσε σε σύγκλιση των προβλέψεων με ικανοποιητική απόδοση. Τα αποτελέσματα χρησιμοποιήθηκαν για την ανάλυση της ευαισθησίας των παραμέτρων που εξετάστηκαν ως ύποπτες για την εξάπλωση του COVID-19. Τα ευρήματα της ανάλυσης και των αριθμητικών προσομοιώσεων αποκάλυψαν ότι οι εξαρτώμενες από το χρόνο παρεμβάσεις μείωσαν τον αριθμό των μολυσμένων ατόμων σε σύγκριση με παρεμβάσεις ανεξάρτητες από το χρόνο. Αξιολογήθηκαν επίσης και οι επιδράσεις μεμονωμένων ή συνδυασμένων επεμβάσεων. Τα αποτελέσματα οδήγησαν στο συμπέρασμα ότι ήταν εποικοδομητική η έγκαιρη εφαρμογή χρονικά εξαρτώμενων επεμβάσεων για τον περιορισμό της διασποράς του ιού [6].

Σε μελέτη που έλαβε χώρα στο Πακιστάν, εξετάστηκε η επίδραση των καιρικών συνθηκών (πχ θερμοκρασία, υγρασία) στην διάδοση και τη φονικότητα του ιού. Τα δεδομένα που συλλέχθηκαν υποβλήθηκαν σε επεξεργασία και χρησιμοποιήθηκαν για τη σύγκριση της απόδοσης διαφόρων μοντέλων πρόβλεψης χρονοσειρών ως προς στατιστικά μέτρα του σφάλματος μεταξύ προβλέψεων και πραγματικών παρατηρήσεων. Για τον σκοπό αυτό χρησιμοποιήθηκε το μοντέλο χρονοσειρών, για την εκτίμηση της σχέσης μεταξύ καιρικών συνθηκών και διασποράς του COVID-19, λαμβάνοντας υπόψη τη συσχέτιση μεταξύ των συνολικών μολυσμένων κρουσμάτων και του αριθμού των θανάτων και των καιρικών μεταβλητών σε μια συγκεκριμένη περιοχή. Τα αποτελέσματα κατέδειξαν ότι πράγματι οι καιρικές συνθήκες επηρεάζουν τον πλήθος των κρουσμάτων που παρουσιάζονται [7]

Σε μία ακόμα μελέτη που εξετάστηκαν δείγματα κρουσμάτων αναρρώσεων και θανάτων και η συσχέτιση τους με τον ρυθμό εξάπλωσης του COVID-19, χρησιμοποιήθηκαν πρακτικές βαθιάς μηχανικής μάθησης (deep learning). Εντοπίστηκαν προσεγγιστικές λύσεις για τις παραμέτρους του μοντέλου. Χρησιμοποιήθηκαν νευρωνικά δίκτυα και εξετάστηκε η απόδοση τους με διαφορετικές παραμέτρους και συναρτήσεις ενεργοποίησης. Το μοντέλο που δημιουργήθηκε, αξιολογήθηκε με δεδομένα προερχόμενα από διαφορετικές περιοχές του πλανήτη και είχε θετικά αποτελέσματα στην βραχυπρόθεσμη πρόβλεψη για την διασπορά του Covid19 [8].

Μία μελέτη η οποία στόχευε στην δημιουργία ενός μοντέλου πρόβλεψης του αριθμού των κρουσμάτων Covid-19 με την χρήση σχετικών δεδομένων από το παρελθόν, βασίστηκε σε πρώτο χρόνο στην αξιολόγηση διαφορετικών τεχνικών, προκειμένου να εντοπιστεί η τεχνική έρευνας που θα επέφερε τα ακριβέστερα και πιο αξιόπιστα αποτελέσματα. Το αντικείμενο της μελέτης ήταν διττό. Το ένα σκέλος αφορούσε την πρόβλεψη – εκτίμηση νόσησης με βάση τον συνδυασμό των εκδηλωμένων συμπτωμάτων και το δεύτερο αφορούσε την πρόβλεψη για τον αριθμό των επιβεβαιωμένων κρουσμάτων. Για το πρώτο σκέλος, καλύτερη ακρίβεια παρείχε το μοντέλο που αναπτύχθηκε με τεχνικές των δένδρων

απόφασης (η Extra Tree Classifier ήταν αποδοτικότερη). Για το δεύτερο σκέλος, αποδοτικότερη αποδείχθηκε η μέθοδος Autoregressive Integrated Moving Average Model (ARIMA). Τα δείγματα που χρησιμοποιήθηκαν προερχόταν από την Ινδία. Οι μετρήσεις για την ακρίβεια των μοντέλων ήταν ικανοποιητικές. Εκτιμάται ότι οι δοκιμές των μοντέλων θα πρέπει να επαναληφθούν με δεδομένα πιο αντιπροσωπευτικά (από διαφορετικές περιοχές) και περισσότερα. Σε κάθε περίπτωση αυτά τα εργαλεία μπορεί να σταθούν αρωγοί τόσο για την διάγνωση της νόσου (και επομένως τη εγκαίρως επέμβαση) είτε για την λήψη μέτρων εγκαίρως[9]

Σε έρευνα που στηρίχθηκε σε δεδομένα που προερχόταν από το Βέλγιο, την Ολλανδία και το Μαρόκο, τα ζητούμενα ήταν τα εξής:

- Μοντελοποίηση και πρόβλεψη εξάπλωσης του Covid-19
- Αξιολόγηση της αποτελεσματικότητας των μέτρων για την αντιμετώπιση της διασποράς του Covid-19
- Ανάπτυξη εκτιμήσεων για την χαλάρωση των παραπάνω αναφερομένων μέτρων.

Στόχος της ήταν να παράγει ένα μαθηματικό μοντέλο για προβλέψεις από βασικά δεδομένα που συλλέχθηκαν από χώρες που βρίσκονται σε βασικό στάδιο ως προς την διασπορά του ιού. Για τον έλεγχο της προόδου της ασθένειας σε έναν συγκεκριμένο πληθυσμό, χρησιμοποιήθηκε ένα μοντέλο που κατέτασσε του ανθρώπους σε τέσσερις διαφορετικές κατηγορίες:

- Ευπαθείς άνθρωποι (S): Ευάλωτοι σε περίπτωση που θα προσβληθούν από τον Covid-19
- Μολυσμένοι άνθρωποι (I): Έχουν διαγνωστεί θετικοί στον ιό.
- Διαγνωσμένοι άνθρωποι (D): Έχουν κάνει διαγνωστικό τεστ και έχουν βρεθεί θετικοί
- Υγιείς άνθρωποι (R): Αυτοί είτε έχουν εμβολιαστεί είτε έχουν νοσήσει.

Στην συνέχεια εκτελέστηκαν συγκριτικές προσομοιώσεις για την αξιολόγηση της αποτελεσματικότητας των μέτρων ελέγχου. Ελέγχθηκαν παράμετροι που καθόριζαν την επιβολή και την απουσία μέτρων. Η μελέτη απέδειξε ότι τα μέτρα ελέγχου βοήθησαν να περιοριστεί η πανδημία [10].

### *Παρούσα Έρευνα*

Η παρούσα έρευνα έχει ως στόχο την μελέτη των νευρωνικών δικτύων με έμφαση στα αναδρομικά νευρωνικά δίκτυα ώστε μέσω της μοντελοποίησης ενός αλγορίθμου να γίνουν προβλέψεις σε κάποια δεδομένα χρονοσειρών και να εξαχθούν συμπεράσματα όσον αφορά την καταλληλότητα της συγκεκριμένης αρχιτεκτονικής.

Συγκεκριμένα τα δεδομένα στα οποία γίνονται προβλέψεις είναι οι θάνατοι από Covid-19. Η μη γραμμικότητα των συγκεκριμένων δεδομένων καθώς και ο θόρυβος δημιουργούν ζητήματα στην απόδοση και προκλήσεις στην αντιμετώπιση τους.

## *Δομή Εργασίας*

Στα πρώτα κεφάλαια της παρούσας εργασίας καταγράφονται οι βασικές κατηγορίες Μηχανικής Μάθησης και αναλύονται τα τεχνητά νευρωνικά δίκτυα ως προς τα χαρακτηριστικά τους την δομή τους και την χρήση τους.

Στην συνέχεια γίνεται εκτενέστερη ανάλυση των αναδρομικών νευρωνικών δικτύων πάνω στα οποία βασίστηκε και η ανάπτυξη του αλγορίθμου του πειραματικού μέρους.

Στο 8ο κεφάλαιο καταγράφονται στοιχεία για την γλώσσα προγραμματισμού Python η οποία αποτέλεσε το εργαλείο για την ανάπτυξη του αλγορίθμου του μοντέλου.

Στο τελικό σκέλος η εργασία εστίασε στην πειραματική διαδικασία, στην χρησιμοποίηση του τεχνητού νευρωνικού δικτύου RNN (Recurrent Neural Network) στην σύγκριση με το μοντέλο RBF (Radial Basis Function) και εν τέλει στην ανάδειξη του καταλληλότερου μοντέλου προβλέψεων για τα συγκεκριμένα δεδομένα βάση αποτελεσμάτων.

Κλείνοντας γίνεται μια καταγραφή και αποτίμηση της όλης διαδικασίας και παρατίθενται τα συμπεράσματα που εξήχθησαν.

# ΚΕΦΑΛΑΙΟ 1

## Μηχανική Μάθηση

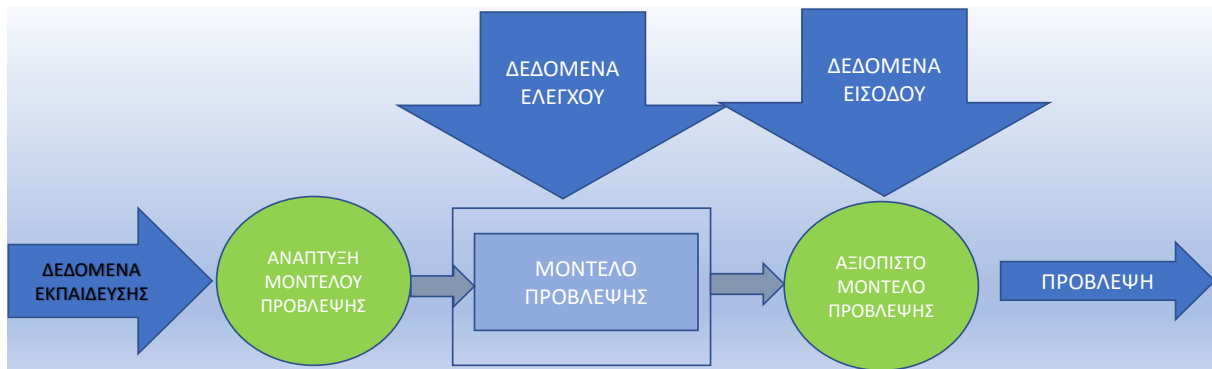
---

### 1 Μηχανική Μάθηση

Η μηχανική μάθηση ασχολείται με τη μελέτη και την ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν σε υπολογιστικά συστήματα να μαθαίνουν αυτόματα από δεδομένα, και να βελτιώνονται στην απόδοσή τους με τον χρόνο. [11] [10] Μαζί με την ασαφή λογική, τον εξελικτικό υπολογισμό [12] [13], και την ευφυΐα σμήνους [14]–[17] η μηχανική μάθηση θεωρείται γενικά ένας από τους βασικούς κλάδους της τεχνητής και υπολογιστικής νοημοσύνης. Οι αλγόριθμοι μηχανικής μάθησης μπορούν να ανακαλύψουν κρυμμένα πρότυπα και να εξάγουν σημαντικές πληροφορίες από μεγάλους όγκους πολύπλοκων δεδομένων. Αναγνωρίζουν πρότυπα, δομές και συσχετίσεις μεταξύ των δεδομένων που δεν είναι εύκολα αντιληπτά από τον ανθρώπινο εγκέφαλο. Με βάση την προσέγγιση εκμάθησης, τον τύπο των δεδομένων που δέχονται σαν παραμέτρους και που εξάγουν ως αποτέλεσμα και τον τύπο του προβλήματος που επιλύουν, διακρίνονται σε:

- Αλγορίθμους που λειτουργούν με επίβλεψη
- Αλγορίθμους που λειτουργούν χωρίς επίβλεψη
- Αλγορίθμους ενισχυτικής μάθησης
- Υβριδικοί αλγόριθμοι

Πέραν αυτών υπάρχουν υβριδικές προσεγγίσεις και άλλες κοινές μέθοδοι που προσφέρουν λύσεις σε ποικίλες μορφές προβλημάτων μηχανικής μάθησης. Ο κοινός παρονομαστής των αλγορίθμων αυτών είναι ότι βασίζονται στην επεξεργασία συνόλων δεδομένων για να κατασκευάσουν μοντέλα πρόβλεψης μέσα από διαδικασίες εκπαίδευσης. Στην συνέχεια τα μοντέλα που δημιουργούνται ελέγχονται για την αξιοπιστία τους και αν αυτή εκτιμάται ως ικανοποιητική, χρησιμοποιούνται σε διαδικασίες πρόβλεψης (των οποίων ο τύπος της εξόδου ποικίλει ανάλογα με την σκοπιμότητα του μοντέλου). Στο Σχήμα 1 φαίνεται αφαιρετικά η γενική τους λειτουργία.



Σχήμα 1: Γενική λειτουργία των αλγορίθμων μηχανικής μάθησης

### 1.1 Αλγόριθμοι μηχανικής μάθησης με επίβλεψη

Οι αλγόριθμοι μηχανικής μάθησης με επίβλεψη λειτουργούν χρησιμοποιώντας ένα σύνολο δεδομένων εκπαίδευσης που περιλαμβάνει τόσο τις παραμέτρους εισόδου όσο και τις αντίστοιχες ετικέτες ή τις επιθυμητές εξόδους. Οι παράμετροι εισόδου αντιπροσωπεύουν τα χαρακτηριστικά ή τις παρατηρήσεις για κάθε παράδειγμα στο σύνολο δεδομένων, ενώ οι ετικέτες ή οι επιθυμητές έξοδοι αντιπροσωπεύουν την επιθυμητή απόκριση ή το αναμενόμενο αποτέλεσμα για κάθε παράδειγμα. Αφού δημιουργηθεί το μοντέλο, χρησιμοποιείται ένα σύνολο, ίδιας δομής με αυτό που χρησιμοποιήθηκε για την εκπαίδευση του, προκειμένου να αξιολογηθεί η αξιοπιστία του. Αυτό γίνεται με εφαρμογή των παραμέτρων των δεδομένων στο νέο δημιουργηθέν μοντέλο και σύγκριση της πραγματικής εξόδου με την αναμενόμενη. Η διαδικασία αυτή επαναλαμβάνεται μέχρι ο βαθμός αξιοπιστίας να ξεπεράσει ένα κατώφλι το οποίο κρίνεται ως ικανοποιητικό. Από το σημείο αυτό και μετά, το αξιόπιστο μοντέλο μπορεί να χρησιμοποιηθεί σε διαδικασίες πρόβλεψης με νέα δεδομένα με τιμές παραμέτρων που έχουν δομή ίδια με την δομή αυτών των συνόλων εκπαίδευσης και ελέγχου. Η διαθεσιμότητα δειγμάτων δεδομένων μεγάλης κλίμακας με ετικέτα το καθιστά αυτό μια δαπανηρή προσέγγιση για εργασίες όπου τα δεδομένα είναι σπάνια.

Οι βασικότερες κατηγορίες αλγορίθμων εποπτευόμενης μηχανικής μάθησης είναι:

- Η κατηγοριοποίηση (Classification): Οι αλγόριθμοι αυτής της κατηγορίας στοχεύουν στην δημιουργία μοντέλων πρόβλεψης τα οποία κατατάσσουν τα δείγματα σε κλάσεις από ένα σύνολο που περιλαμβάνει πεπερασμένο πλήθος. Στην κατηγορία αυτή των αλγορίθμων περιλαμβάνονται:
  - Δένδρα απόφασης: Είναι γραφικές αναπαραστάσεις της ανάλυσης προβλημάτων και της διαδικασίας επίλυσής τους. Η βασική ιδέα στην οποία βασίζονται είναι ότι η επίλυση προβλημάτων περνά από διαδοχικές αποφάσεις. Τα δένδρα απόφασης έχουν τη μορφή ενός διακλαδωμένου δέντρου, όπου κάθε κόμβος του δέντρου αντιπροσωπεύει μια επιλογή ή ένα κριτήριο που εφαρμόζεται στις παρατηρήσεις. Κάθε κλάδος αντιπροσωπεύει μια πιθανή πρόβλεψη ή απόφαση. Ο τρόπος που λειτουργούν τα δένδρα απόφασης είναι ο εξής: ξεκινώντας από τη ρίζα του δέντρου, κάθε κόμβος ελέγχει μια παράμετρο εισόδου και ανακατευθύνει τη ροή της εκτέλεσης σε έναν από τους υποκόμβους, ανάλογα με την τιμή της παρατήρησης. Η διαδικασία αυτή επαναλαμβάνεται για κάθε κόμβο μέχρι να φτάσουμε



σε ένα φύλλο του δέντρου, που αντιπροσωπεύει την τελική απόφαση ή πρόβλεψη.

- Naïve Bayes: Ο ταξινομητής Naive Bayesian βασίζεται στο θεώρημα του Bayes και κάνει την αφελή υπόθεση της ανεξαρτησίας μεταξύ των προβλέψεων. Αυτό σημαίνει ότι θεωρεί ότι οι χαρακτηριστικά των δεδομένου είναι ανεξάρτητα μεταξύ τους, δεδομένης της κλάσης πρόβλεψης. Ο ταξινομητής Naive Bayesian χρησιμοποιείται συχνά για προβλήματα ταξινόμησης, όπου ο στόχος είναι να αναθέσουμε μια κλάση σε ένα δείγμα βάσει των χαρακτηριστικών του. Με τη χρήση του θεωρήματος του Bayes, μπορούμε να υπολογίσουμε την πιθανότητα μιας κλάσης, δεδομένων των χαρακτηριστικών του δείγματος. Η σχέση που εκφράζει το θεώρημα και κατ' επέκταση την γενική συμπεριφορά των ταξινομητών αυτού του είδους είναι<sup>1</sup>:

$$P(c|x) = \frac{P(c)P(x|c)}{P(x)} \quad (1-1)$$

- Νευρωνικά Δίκτυα: Πρόκειται για αλγόριθμους που η επεξεργασία των δεδομένων περνάει μέσα από μία δομή που μοιάζει με το νευρικό σύστημα των θηλαστικών.
- Η παλινδρόμηση (Regression): Οι αλγόριθμοι αυτοί χειρίζονται περιπτώσεις όπου η έξοδος είναι συνεχής τιμή. Δημιουργούν μοντέλα τα οποία στοχεύουν την πρόβλεψη της τιμής της εξόδου από τις τιμές των παραμέτρων της εισόδου. Οι βασικότεροι αλγόριθμοι της κατηγορίας αυτής είναι:
  - Γραμμική παλινδρόμηση: Η γραμμική παλινδρόμηση επιχειρεί να μοντελοποιήσει τη σχέση μεταξύ δύο μεταβλητών προσαρμόζοντας μια γραμμική εξίσωση στα παρατηρούμενα δεδομένα. Μια γραμμή γραμμικής παλινδρόμησης περιγράφεται από μια εξίσωση της μορφής  $Y = a + b \cdot X$ , όπου  $X$  είναι η ανεξάρτητη μεταβλητή και  $Y$  είναι η εξαρτημένη μεταβλητή.
  - Λογιστική παλινδρόμηση: Η λογιστική παλινδρόμηση είναι ένα στατιστικό μοντέλο που στη βασική του μορφή χρησιμοποιεί μια λογιστική συνάρτηση για να μοντελοποιήσει μια δυαδική εξαρτημένη μεταβλητή.

## 1.2 Αλγόριθμοι μηχανικής μάθησης χωρίς επίβλεψη

Οι αλγόριθμοι μηχανικής μάθησης χωρίς επίβλεψη είναι αλγόριθμοι που λειτουργούν χωρίς τη χρήση ετικετών ή προσδιορισμένων εξόδων για τα δεδομένα εκπαίδευσης.

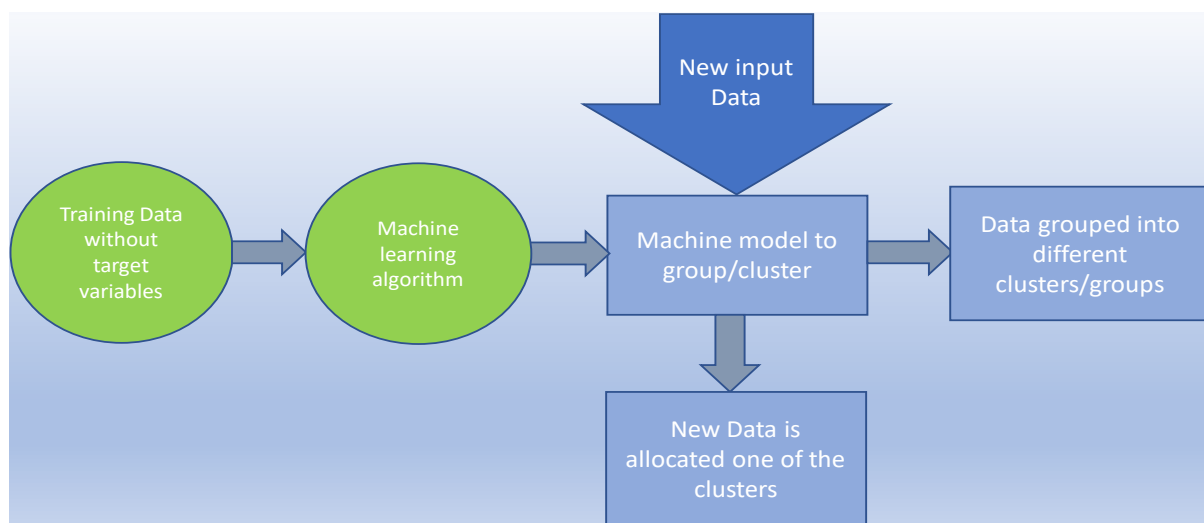
---

<sup>1</sup> Όπου  $P(c|x)$  ένταξης στην κλάση  $c$ ,  $P(x|c)$  είναι η πιθανότητα της παρατήρησης μέσα στην κλάση,  $P(x)$  η πιθανότητα του  $x$  γενικά και  $P(c)$  η πιθανότητα της κλάσης.

Αντίθετα, αυτοί οι αλγόριθμοι εξάγουν πληροφορίες και δομές από τα ίδια τα δεδομένα, προσπαθώντας να ανακαλύψουν κρυμμένα μοτίβα ή ομαδοποιήσεις μεταξύ τους.

Οι αλγόριθμοι μηχανικής μάθησης χωρίς επίβλεψη είναι χρήσιμοι όταν δεν έχουμε προκαθορισμένες ετικέτες για τα δεδομένα μας ή όταν θέλουμε να εξερευνήσουμε τη δομή των δεδομένων και να ανακαλύψουμε ενδιαφέρουσες πληροφορίες χωρίς προκαθορισμένες υποθέσεις.

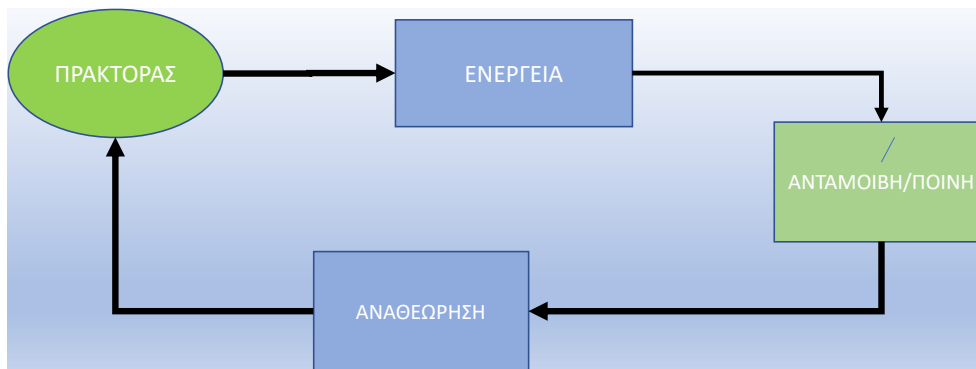
Ένας από τους κύριους τύπους αλγορίθμων χωρίς επίβλεψη είναι η ομαδοποίηση (clustering). Οι αλγόριθμοι αυτοί περνάνε σε επεξεργασία των δεδομένων του συνόλου εκπαίδευσης ώστε να καταφέρουν να ανακαλύψουν εγγενείς ομαδοποιήσεις και να εκτιμήσουν πως οι τιμές των παραμέτρων τους επηρεάζουν την κατάταξη τους στις ομάδες αυτές. Και στην περίπτωση της ομαδοποίησης, χρησιμοποιούνται κατάλληλες μετρικές για την αξιολόγηση της αξιοπιστίας των μοντέλων πρόβλεψης που παράγονται, πριν χρησιμοποιηθούν. Στο Σχήμα 2 περιγράφεται η γενική λειτουργία των αλγορίθμων μη εποπτευόμενης μηχανικής μάθησης.



Σχήμα 2: Γενική λειτουργία αλγορίθμων μη εποπτευόμενης μηχανικής μάθησης

### 1.3 Αλγόριθμοι Ενισχυτικής Μάθησης

Οι αλγόριθμοι ενισχυτικής μάθησης χρησιμοποιούνται στην επίλυση προβλημάτων που το ζητούμενο είναι η λήψη μιας σειράς αποφάσεων με στόχο την μεγιστοποίηση ενός τελικού οφέλους (ανταμοιβή). Κατά τη διάρκεια της μαθησιακής διαδικασίας, ένας τεχνητός πράκτορας λαμβάνει είτε ανταμοιβές είτε ποινές για τις ενέργειες που εκτελεί. Η δράση του είναι προσανατολισμένη στη μεγιστοποίηση της συνολικής ανταμοιβής. Οι αλγόριθμοι αυτοί χρησιμοποιούνται στην ανάπτυξη ψυχαγωγικών εφαρμογών – παιχνιδιών και την ρομποτική. Στην Σχήμα 3 περιγράφεται σχηματική η γενική λειτουργία των αλγορίθμων ενισχυμένης μάθησης.



Σχήμα 3: Γενική λειτουργία των αλγορίθμων ενισχυτικής μάθησης

## 1.4 Υβριδικές προσεγγίσεις

### 1.4.1 Ημιοποπτευόμενη μάθηση

Πρόκειται για μία ενδιάμεση προσέγγιση μεταξύ των τεχνικών εκμάθησης με επίβλεψη και χωρίς επίβλεψη. Χαρακτηριστικό τους γνώρισμα είναι ότι η εκπαίδευση πραγματοποιείται χρησιμοποιώντας έναν συνδυασμό δεδομένων με ετικέτες και χωρίς ετικέτες (συνήθως χρησιμοποιείται ένας μικρός αριθμός δεδομένων με ετικέτα και ένας αρκετά μεγαλύτερος αριθμός δεδομένων χωρίς ετικέτα). Η βασική λειτουργία των αλγορίθμων αυτού του είδους είναι να ομαδοποιήσουν τα δεδομένα με βάση το σύνολο χωρίς ετικέτες, χρησιμοποιώντας έναν αλγόριθμο μάθησης χωρίς επίβλεψη. Το επόμενο στάδιο αφορά την χρήση του συνόλου με ετικέτα για την επισήμανση των υπόλοιπων στις ομάδες με ετικέτα.

### 1.4.2 Αυτοεποπτευόμενη μάθηση

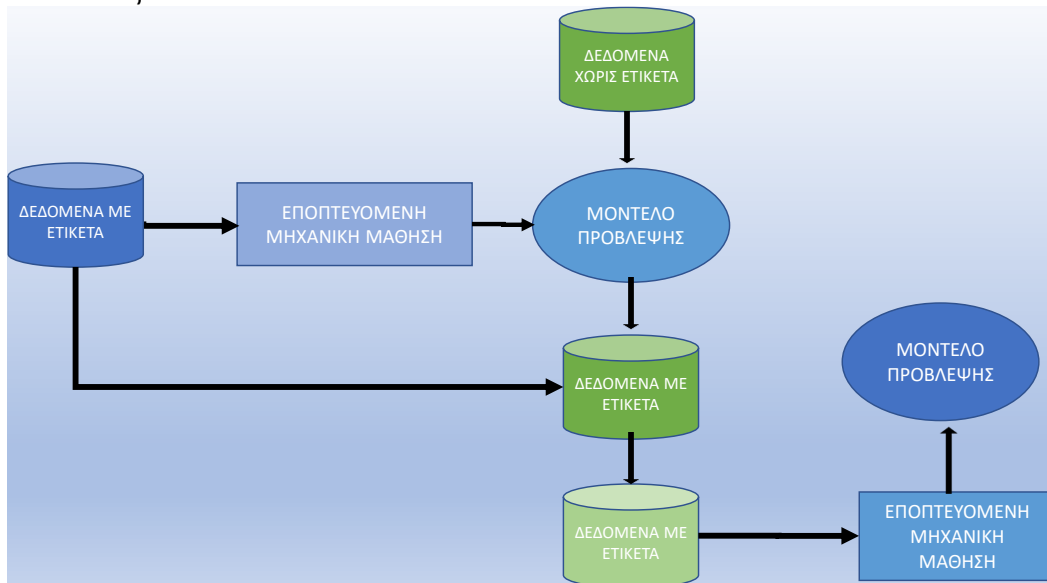
Η αυτοεποπτευόμενη μάθηση είναι μια μορφή μηχανικής μάθησης όπου το μοντέλο εκπαιδεύεται να αντλεί γνώση από τα δεδομένα εισόδου χωρίς την ανάγκη ετικετών ή επιβλέψεων από τον άνθρωπο. Αντ' αυτού, το μοντέλο επιδιώκει να εντοπίσει δομικά μοτίβα, συσχετίσεις ή κρυμμένες δομές στα δεδομένα εισόδου χωρίς προηγούμενη γνώση. Η γνώση που έχει αναπτυχθεί στο στάδιο αυτό, μεταφέρεται στο μοντέλο για υλοποίηση της κύριας εργασίας. Στην πραγματικότητα, η διαφορά των αλγορίθμων αυτών από τους αντίστοιχους της εποπτευόμενης μάθησης είναι ότι η επισήμανση των δεδομένων εισόδου δεν γίνεται από τον άνθρωπο αλλά εντάσσεται και αυτή στα αρχικά βήματα του αλγορίθμου.

### 1.4.3 Αυτοδίδακτη μάθηση

Οι αλγόριθμοι αυτοί δέχονται ως είσοδο δεδομένα και με ετικέτα και χωρίς ετικέτα. Η κεντρική ιδέα είναι η μεταφορά της γνώσης από τα δεδομένα με ετικέτα προς τα δεδομένα χωρίς ετικέτα. Οι αλγόριθμοι αυτοί έχουν τρία στάδια. Στο πρώτο στάδιο χρησιμοποιείται ένας αλγόριθμος εποπτευόμενης μηχανικής μάθησης, με δεδομένα εκπαίδευσης και ελέγχου με ετικέτα. Με την ολοκλήρωση του πρώτου σταδίου, παράγεται ένα μοντέλο πρόβλεψης. Στο δεύτερο στάδιο, χρησιμοποιείται το μοντέλο αυτό προκειμένου να αποδοθούν ετικέτες στα δεδομένα του συνόλου εκείνων που δεν έχουν. Μετά και από το στάδιο αυτό, τα

δεδομένα χωρίς ετικέτα, με την νέα ετικέτα που απέκτησαν στο δεύτερο στάδιο, συνενώνονται με τα δεδομένα με ετικέτα και το ενοποιημένο σύνολο αποτελεί πλέον το σύνολο δεδομένων εισόδου και ελέγχου σε μία επανάληψη της διαδικασίας του πρώτου σταδίου.

Στο Σχήμα 4 φαίνεται μια σχηματική αναπαράσταση της λειτουργίας των αλγορίθμων αυτού του είδους.



Σχήμα 4: Αυτοδίδακτη μηχανική μάθηση

#### 1.4.4 Άλλες προσεγγίσεις

Κάποιες από τις λοιπές προσεγγίσεις στον τομέα της μηχανικής μάθησης είναι οι εξής:

- Εκμάθηση πολλαπλών εργασιών: Η προσέγγιση εκμεταλλεύεται το γεγονός ότι πολλές εργασίες έχουν κοινά χαρακτηριστικά και επομένως μπορούν να χρησιμοποιούν κοινά μοντέλα πρόβλεψης. Μια παραλλαγή της είναι η μάθηση με μεταφορά, όπου ένα ήδη δημιουργηθέν μοντέλο (που εξυπηρετεί έναν άλλο συναφή σκοπό), εκπαιδεύεται περαιτέρω για να εξυπηρετήσει μία νέα σκοπιμότητα.
- Διαδραστική μάθηση: Με την προσέγγιση αυτή, επιλέγεται αρχικά ένα υποσύνολο των δειγμάτων δεδομένων για να χρησιμοποιηθεί στην διαδικασία της μάθησης χωρίς επίβλεψη. Με την ολοκλήρωση του αλγορίθμου, αποδίδονται στα δεδομένα ετικέτες που αντιστοιχούν στις ομάδες που δημιουργήθηκαν. Οι έξοδοι οπτικοποιούνται και εκεί επεμβαίνει ο ανθρώπινος παράγοντας, διορθώνοντας τις παραμέτρους που επηρεάζουν την πρόβλεψη. Με τον τρόπο αυτό συνδυάζονται η ταχύτητα και η ακρίβεια των υπολογισμών των αυτοματοποιημένων διεργασιών με την αυξημένη ανθρώπινη αντίληψη.
- Διαδικτυακή μάθηση: Αυτή η προσέγγιση χρησιμοποιεί σαν είσοδο ροές δεδομένων (το σύνολο των δεδομένων δεν είναι ολόκληρο εξ αρχής διαθέσιμο). Χρησιμοποιείται κυρίως σε περιπτώσεις όπου απαιτείται δυναμική προσαρμογή των μοντέλων που αναπτύσσονται σε νέα μοτίβα δεδομένων.

- Προοδευτική μάθηση: Πρόκειται για μία προσέγγιση η οποία ομοιάζει αρκετά με τη διαδικτυακή μάθηση, με την διαφορά ότι κάθε δείγμα μπορεί να χρησιμοποιηθεί περισσότερες από μία φορές κατά τη διάρκεια της επεξεργασίας.
- Κατανεμημένη Μάθηση: Οι τεχνικές αυτές επιτρέπουν την κατανομή της εκπαίδευσης σε διαφορετικές διαδικασίες και με χρήση διαφορετικών συνόλων δεδομένων.
- Συνολική Μάθηση: Είναι τεχνικές που συνδυάζουν τα αποτελέσματα της διαδικασίας εκπαίδευσης διαφορετικών επιμέρους τεχνικών.
- Μάθηση Αντιπαράθεσης: Πρόκειται για μεθοδολογία που συνδυάζει την εκπαίδευση των μοντέλων χρησιμοποιώντας σύνολα από δεδομένα με έντονες διαφοροποιήσεις.

# ΚΕΦΑΛΑΙΟ 2

## Νευρωνικά Δίκτυα

---

### 2. Νευρωνικά Δίκτυα

#### 2.1 Βιολογικά και Τεχνητά Νευρωνικά δίκτυα

Ο όρος νευρωνικά δίκτυα παραδοσιακά περιγράφει δίκτυα νευρώνων στον εγκέφαλο των θηλαστικών. Οι νευρώνες είναι οι θεμελιώδεις μονάδες που διενεργούν επεξεργασία των ερεθισμάτων του περιβάλλοντος και συνδέονται μεταξύ τους σε δίκτυα για την μεταφορά τους από το ένα στον άλλο. Υπό το πρίσμα αυτό, η επεξεργασία αυτή των εξωτερικών ερεθισμάτων που έχει ως αποτέλεσμα την παραγωγή απόκρισης, μπορεί να παρουσιάζει υψηλή πολυπλοκότητα. Οι είσοδοι και οι έξοδοι κάθε νευρώνα ποικίλλουν σε συνάρτηση με τον χρόνο όπως επίσης και το ίδιο το δίκτυο υφίσταται μεταβολές με την πάροδο του χρόνου. Αυτή το χαρακτηριστικό εξηγεί και την διαδικασία της μάθησης και της βελτίωσης των δεξιοτήτων μέσα από αναπροσαρμογές των συνδέσεων των νευρώνων ή ανάπτυξη νέων. Η δομή του νευρώνα του εγκεφάλου των θηλαστικών έχει την εξής δομή:

- Μία διακλαδισμένη δομή εισόδου, τους δενδρίτες. Οι δενδρίτες είναι δίκτυα νευρικών ινών, που μοιάζουν με δέντρα και μεταφέρουν ηλεκτρικά σήματα στο κυτταρικό σώμα.
- Ένα βασικό κυτταρικό σώμα. Το κυτταρικό σώμα αθροίζει και επεξεργάζεται τα εισερχόμενα σήματα.
- Μία δομή εξόδου, τους νευροάξονες. Ο άξονας είναι μια ενιαία μακριά ίνα που μεταφέρει το σήμα από το κυτταρικό σώμα προς άλλους νευρώνες. Το σημείο επαφής μεταξύ ενός άξονα ενός κυττάρου και ενός δενδρίτη ενός άλλου κυττάρου ονομάζεται σύναψη. Η διάταξη των νευρώνων και οι δυνάμεις των επιμέρους συνάψεων, που καθορίζονται από μια πολύπλοκη χημική διαδικασία, καθορίζει τη λειτουργία του νευρωνικού δικτύου.

Ένα ποσοστό της νευρικής δομής των θηλαστικών καθορίζεται κατά τη γέννηση τους. Ένα άλλο μέρος αναπτύσσεται μέσω της μάθησης, καθώς δημιουργούνται νέες συνδέσεις, ενώ ένα άλλο μέρος χάνεται με τον καιρό. Αυτή η εξέλιξη είναι πιο αισθητή στα πρώτα στάδια της ζωής. Οι νευρικές δομές συνεχίζουν να αλλάζουν καθ' όλη τη διάρκεια της ζωής. Οι μεταγενέστερες αλλαγές τείνουν να εκδηλώνονται ως ενίσχυση ή αποδυνάμωση των συναπτικών συνδέσεων.

Με την ενεργοποίηση των νευρώνων, κάθε ένας πυροδοτεί ένα ηλεκτροχημικό σήμα κατά μήκος του νευροάξονα που διασχίζει την σύναψη προς το επόμενο κύτταρο, το οποίο

μπορεί επίσης να ενεργοποιηθεί. Ένας νευρώνας μπορεί να πυροδοτηθεί μόνο όταν το σήμα που φτάσει από τους δενδρίτες του, υπερβαίνει μία συγκεκριμένη τιμή που αναφέρεται ως κατώφλι πυροδότησης. Σαν είσοδο ένας νευρώνας μπορεί να δέχεται πολλά ερεθίσματα, κάποια εκ των οποίων συμβάλλουν στην ενεργοποίησή του. Αποτέλεσμα αυτών είναι μία και μόνη έξοδος. Η εκπαίδευση του δικτύου στην αποδοχή των ερεθισμάτων γίνεται με την ρύθμιση των σημάτων που εφαρμόζονται στις συνάψεις.

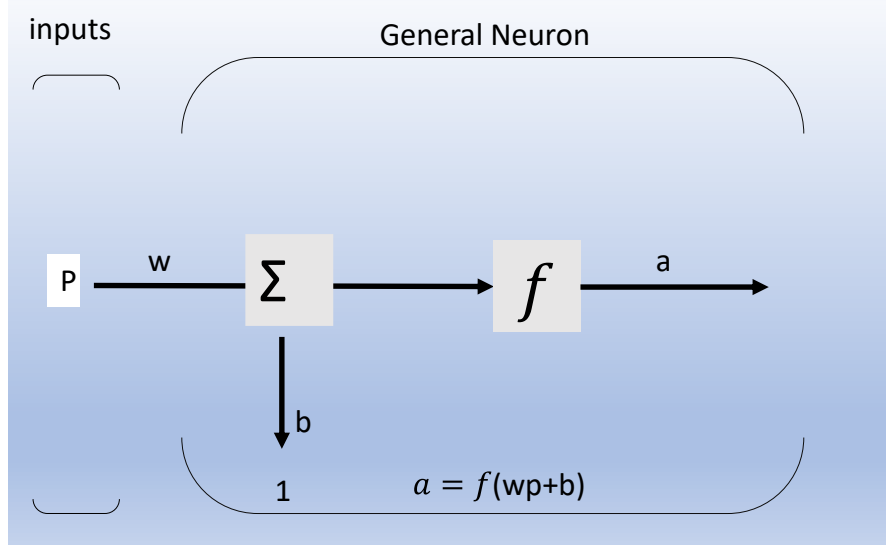
## 2.2 Τεχνητά Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα αναπτύσσονται σε αντιστοιχία με τα βιολογικά χωρίς να προσεγγίζουν την πολυπλοκότητα και την τελειότητα του εγκεφάλου. Υπάρχουν, ωστόσο, δύο βασικές ομοιότητες μεταξύ βιολογικών και τεχνητών νευρωνικών δικτύων. Η πρώτη αφορά τα δομικά στοιχεία τους στοιχεία τα οποία συνιστούν απλές υπολογιστικές συσκευές διασυνδεδεμένες. Η δεύτερη αφορά το γεγονός ότι οι συνδέσεις μεταξύ των νευρώνων καθορίζουν τη λειτουργία του δικτύου. Παρά το γεγονός ότι οι βιολογικοί νευρώνες είναι πολύ αργοί σε σύγκριση με τα ηλεκτρικά κυκλώματα, ο ανθρώπινος εγκέφαλος είναι σε θέση να εκτελεί πολλές εργασίες πολύ πιο γρήγορα από οποιονδήποτε συμβατικό υπολογιστή. Αυτό οφείλεται κυρίως στην παράλληλη δομή των βιολογικών νευρωνικών δικτύων που επιτρέπει στους νευρώνες λειτουργούν ταυτόχρονα. Τα τεχνητά νευρωνικά δίκτυα υλοποιούν την δομή αυτή και δύνανται να εκμεταλλευτούν τις τεχνολογίες παράλληλης επεξεργασίας. Οι αλγόριθμοι χρησιμοποιούν εξαιρετικά εξιδανικευμένα μοντέλα νευρώνων. Ωστόσο, η θεμελιώδης αρχή είναι η ίδια: τα τεχνητά νευρωνικά δίκτυα μαθαίνουν αλλάζοντας τις συνδέσεις μεταξύ των νευρώνων τους. Τέτοια δίκτυα μπορούν να εκτελέσουν ένα πλήθος εργασιών επεξεργασίας πληροφοριών ώστε να αναγνωρίζουν δομές και μοτίβα σε ένα σύνολο δεδομένων και, σε κάποιο βαθμό, να γενικεύουν αυτά που έχουν μάθει. Ένα σύνολο εκπαίδευσης περιέχει μια λίστα μοτίβων εισόδου μαζί με μια λίστα αντίστοιχων ετικετών, που κωδικοποιούν τις ιδιότητες των μοτίβων εισόδου που υποτίθεται ότι θα μάθει το δίκτυο. Τα τεχνητά νευρωνικά δίκτυα μπορούν να εκπαιδευτούν ώστε να ταξινομήν τέτοια δεδομένα με μεγάλη ακρίβεια προσαρμόζοντας τις ισχύς σύνδεσης μεταξύ των νευρώνων τους και μπορούν να μάθουν να γενικεύουν το αποτέλεσμα σε άλλα σύνολα δεδομένων - υπό την προϋπόθεση ότι τα νέα δεδομένα δεν διαφέρουν πολύ από τα δεδομένα εκπαίδευσης. Έχουν εφαρμοστεί με επιτυχία σε ένα ευρύ φάσμα τομέων όπως η αναγνώριση αντικειμένων σε εικόνες, η αυτόματη μετάφραση ή η ανίχνευση συναισθήματος σε κείμενο.

## 2.3 Δομή και Λειτουργία των Τεχνητών Νευρωνικών Δικτύων

Η πιο απλή μορφή νευρώνα είναι αυτή της μονής εισόδου. Η σταθμισμένη είσοδος, έστω  $p$ , πολλαπλασιάζεται με το αντίστοιχο βάρος, έστω  $w$ , για να σχηματιστεί το γινόμενο  $w_p$ , ένας από τους όρους που αποστέλλεται στον αθροιστή. Η άλλη είσοδος, έστω  $I$ , πολλαπλασιάζεται έναν παράγοντα που ονομάζεται πόλωση, έστω  $b$ , και μετά περνά και αυτή στον αθροιστή. Η αθροισμένη έξοδος, έστω  $n$ , περνάει από μια συνάρτηση ενεργοποίησης, έστω  $f$ , η οποία παράγει την έξοδο του σταθμισμένου νευρώνα, έστω  $a$ . Εάν συγκριθεί αυτό το απλό μοντέλο με τον βιολογικό νευρώνα, το βάρος  $w$ , αντιστοιχεί στη δύναμη μιας σύναψης, το κυτταρικό σώμα αντιπροσωπεύεται από την λειτουργία του

αθροιστή και της εφαρμογής της συνάρτησης ενεργοποίησης και η έξοδος του νευρώνα  $a$ , αντιπροσωπεύει το σήμα στον άξονα. Η δομή και λειτουργία αυτή φαίνεται στο Σχήμα 5.



Σχήμα 5: Υπόδειγμα δομής και λειτουργία νευρώνα απλής εισόδου

Η πραγματική έξοδος εξαρτάται από τη συνάρτηση ενεργοποίησης που θα επιλεγεί. Η πόλωση μπορεί και μην τεθεί στον υπολογισμό της εξόδου αν δεν παίζει κάποιο ρόλο στην εκτίμηση της. Σε κάθε περίπτωση, τόσο το βάρος όσο και η πόλωση είναι ρυθμιζόμενες. Η συνάρτηση ενεργοποίησης που θα εφαρμοστεί επιλέγεται από τον αναλυτή της εφαρμογής, όπως επίσης και οι παράμετροι που θα πρέπει να προσαρμόζονται από κάποιον κανόνα εκμάθησης έτσι ώστε η σχέση εισόδου/εξόδου νευρώνων να ανταποκρίνεται στον στόχο του δικτύου.

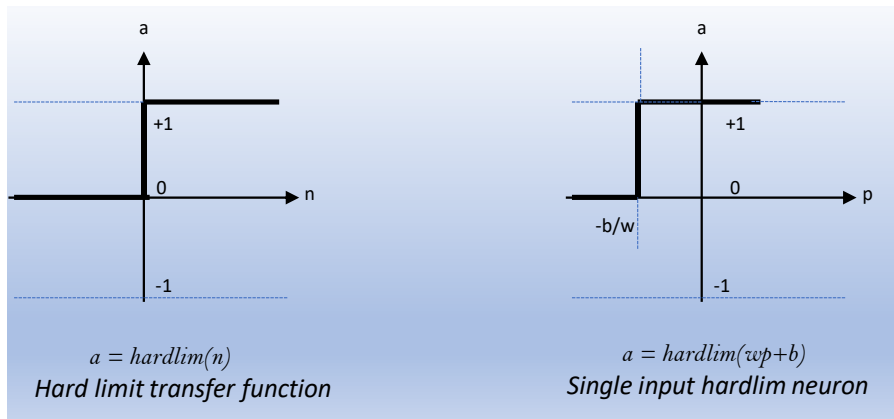
## 2.4 Συναρτήσεις ενεργοποίησης

Η συνάρτηση ενεργοποίησης μπορεί να είναι γραμμική ή μη γραμμική, ανάλογα με τον σκοπό του νευρωνικού δικτύου. Κάθε συνάρτηση ενεργοποίησης επιλέγεται με ορισμένες προδιαγραφές του προβλήματος που προσπαθεί να λύσει ο νευρώνας. Οι πιο συχνά χρησιμοποιούμενες συναρτήσεις ενεργοποίησης αναφέρονται στις ακόλουθες παραγράφους.

### 2.4.1 Hard Limit

Η συνάρτηση Hard Limit, καθορίζει ότι η έξοδος του νευρώνα θα είναι 0, αν η παράμετρος της είναι μικρότερη του 0, και θα είναι 1 αν η παράμετρος είναι μεγαλύτερη ή ίση με το 0. Χρησιμοποιείται σε περιπτώσεις όπου οι εισοδοί θα πρέπει να τοποθετηθούν σε δύο διακριτές κατηγορίες.

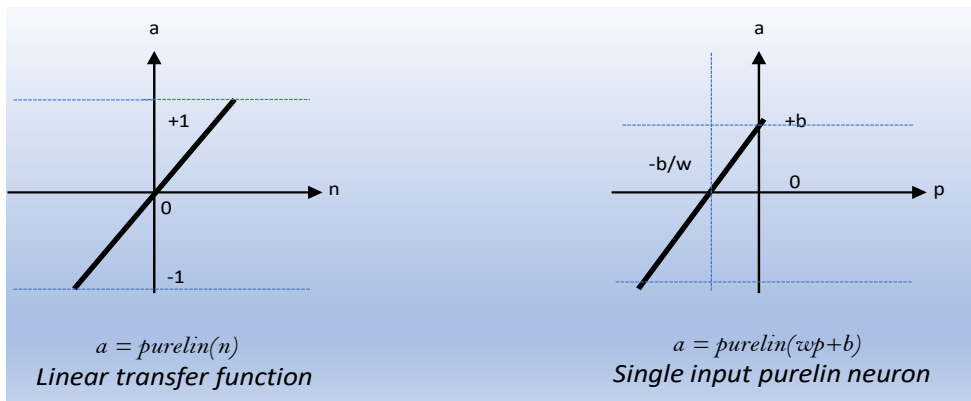




Σχήμα 6: Η συμπεριφορά των Hard Limit συναρτήσεων ενεργοποίησης.

### 2.4.2 Linear Functions

Η συνάρτηση ενός συστήματος είναι γραμμική όταν η έξοδος είναι ευθέως ανάλογη με την είσοδο, δηλαδή οποιαδήποτε αλλαγή στην είσοδο παράγει μια αναλογική αλλαγή στην έξοδο.



Σχήμα 7: Γραμμικές Συναρτήσεις Ενεργοποίησης

### 2.4.3 Log-Sigmoid Functions

Οι συναρτήσεις αυτού του είδους έχουν το χαρακτηριστικό να λαμβάνουν είσοδο με τιμή μεταξύ συν και πλην άπειρο και συμπιέζουν την έξοδο στο εύρος από 0 έως 1, σύμφωνα με τη σχέση:

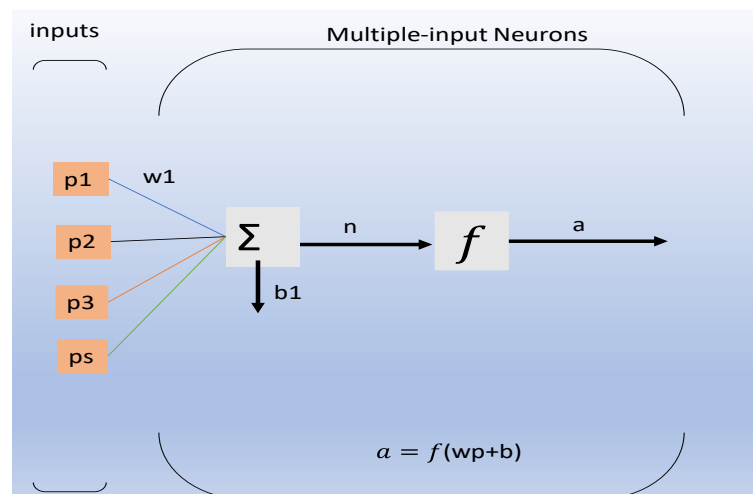
$$a = \frac{1}{1+e^{-n}} \tag{2-1}$$

Οι συναρτήσεις αυτού του είδους χρησιμοποιούνται συνήθως σε δίκτυα πολλαπλών επιπέδων που εκπαιδεύονται χρησιμοποιώντας τον αλγόριθμο backpropagation.

#### 2.4.4 Νευρώνες πολλαπλών εισόδων

Στην πράξη, στα περισσότερα νευρωνικά δίκτυα, χρησιμοποιούνται νευρώνες που λαμβάνουν πολλαπλές εισόδους  $p_1, p_2, p_3, \dots, p_n$ , στις οποίες εφαρμόζονται διαφορετικά βάρη  $w_1, w_2, w_3, \dots, w_n$ . Και στην περίπτωση αυτή μπορεί να εφαρμόζεται μία πόλωση η οποία επηρεάζει την έξοδο.

Σχηματικά οι νευρώνες πολλαπλών εισόδων περιγράφονται στο Σχήμα 8.



Σχήμα 8: Νευρώνας πολλαπλών εισόδων

## Αρχιτεκτονικές Νευρωνικών Δικτύων

---

### 3. Αρχιτεκτονικές Νευρωνικών Δικτύων

Η ανάγκη για διαφορετική αντιμετώπιση των διαφορετικών τύπων εισόδου στα νευρωνικά δίκτυα, δημιούργησε την ανάγκη για ανάπτυξη διαφορετικών αρχιτεκτονικών. Κάθε μία από αυτές έχει την δική της δομή και λειτουργία, που προσαρμόζεται στις απαιτήσεις του προβλήματος που αντιμετωπίζουμε. Ανάλογα με τον τύπο της εισόδου και τον τύπο της εξόδου που επιθυμούμε να προβλέψουμε, μπορούμε να επιλέξουμε την κατάλληλη αρχιτεκτονική.

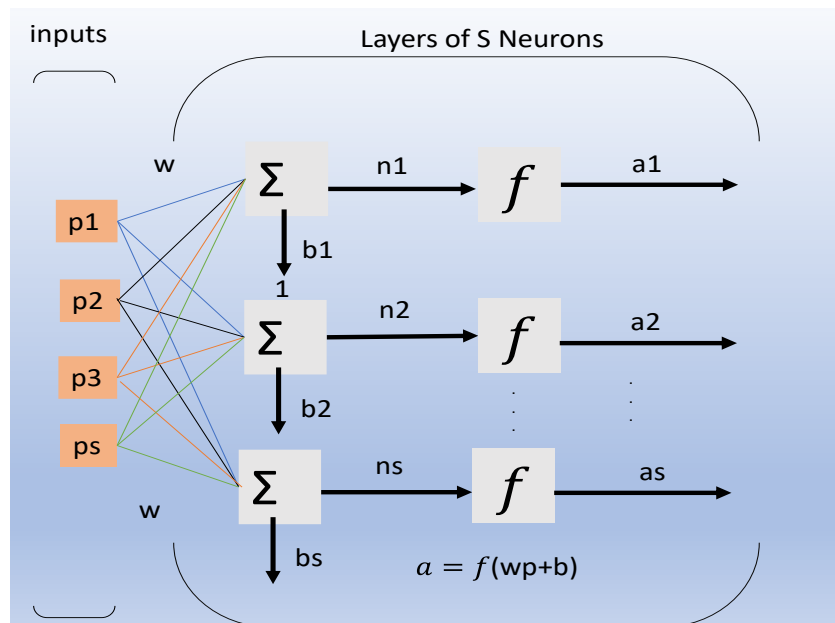
Στις επόμενες παραγράφους περιγράφονται χαρακτηριστικά γνωρίσματα που περιγράφουν σε γενικές γραμμές τις αρχιτεκτονικές και στην συνέχεια οι βασικότερες από αυτές.

#### 3.1 Χαρακτηριστικά

Τα χαρακτηριστικά που χρησιμοποιούνται για την γενική περιγραφή των νευρωνικών δικτύων είναι τα επίπεδα του, ο χειρισμός των εισόδων για την επεξεργασία τους και η διαχείριση των εξόδων των νευρώνων. Αποτελούν τα δομικά στοιχεία ενός νευρωνικού δικτύου και η διάταξη τους σε μεγάλο βαθμό καθορίζει την κάθε μία από τις αρχιτεκτονικές τους.

##### 3.1.1 Επίπεδο Νευρωνικού Δικτύου

Ένα επίπεδο νευρωνικού δικτύου, αποτελείται από ένα σύνολο  $n$  νευρώνων. Οι εισοδοί του επιπέδου, έστω  $R_1, R_2, R_3, \dots, R_n$ , αποτελούν μία από τις εισόδους σε κάθε ένα από τους νευρώνες που εντάσσονται στο επίπεδο. Στο Σχήμα 9 φαίνεται η τυπική μορφή ενός επιπέδου ενός νευρωνικού δικτύου.



Σχήμα 9: Τυπική μορφή επιπέδου νευρωνικού δικτύου

Κάθε επίπεδο περιλαμβάνει τους ακόλουθους βασικούς χαρακτηριστικούς τομείς:

1. Είσοδοι (Inputs): Το επίπεδο δέχεται είσοδο από άλλα επίπεδα ή από την είσοδο του ίδιου του δικτύου. Ο αριθμός των εισόδων μπορεί να είναι διαφορετικός από τον αριθμό των νευρώνων στο επίπεδο, και κάθε είσοδος μπορεί να συνδέεται με διαφορετικούς νευρώνες με οποιοδήποτε συνδυασμό.
2. Βάρη (Weights): Το επίπεδο περιλαμβάνει ένα σύνολο βαρών που εφαρμόζονται στις εισόδους των νευρώνων. Τα βάρη αντιπροσωπεύουν τη σημασία ή τη συνεισφορά της κάθε εισόδου στην έξοδο του νευρώνα.
3. Πολώσεις (Biases): Κάθε νευρώνας στο επίπεδο έχει μια πόλωση που προστίθεται στη συνάρτηση ενεργοποίησής του. Οι πολώσεις επιτρέπουν την προσαρμογή του συνολικού επιπέδου του νευρώνα και επηρεάζουν την επεξεργασία της εισόδου.
4. Συνάρτηση Ενεργοποίησης (Activation Function): Οι νευρώνες στο επίπεδο χρησιμοποιούν μια συνάρτηση ενεργοποίησης για να μετατρέψουν τη συνολική είσοδο σε μια έξοδο. Η συνάρτηση ενεργοποίησης καθορίζει την μη γραμμική συμπεριφορά του νευρώνα και επηρεάζει τον τρόπο με τον οποίο θα ανταποκρίνεται στις εισόδους.

Με βάση αυτά τα χαρακτηριστικά, το επίπεδο εκτελεί την επεξεργασία της εισόδου που δέχεται και παράγει μια έξοδο. Η συνδυασμένη διάταξη των νευρώνων, των βαρών, των πολώσεων και των συναρτήσεων ενεργοποίησης σχηματίζει την αρχιτεκτονική του επιπέδου και καθορίζει τον τρόπο λειτουργίας του.

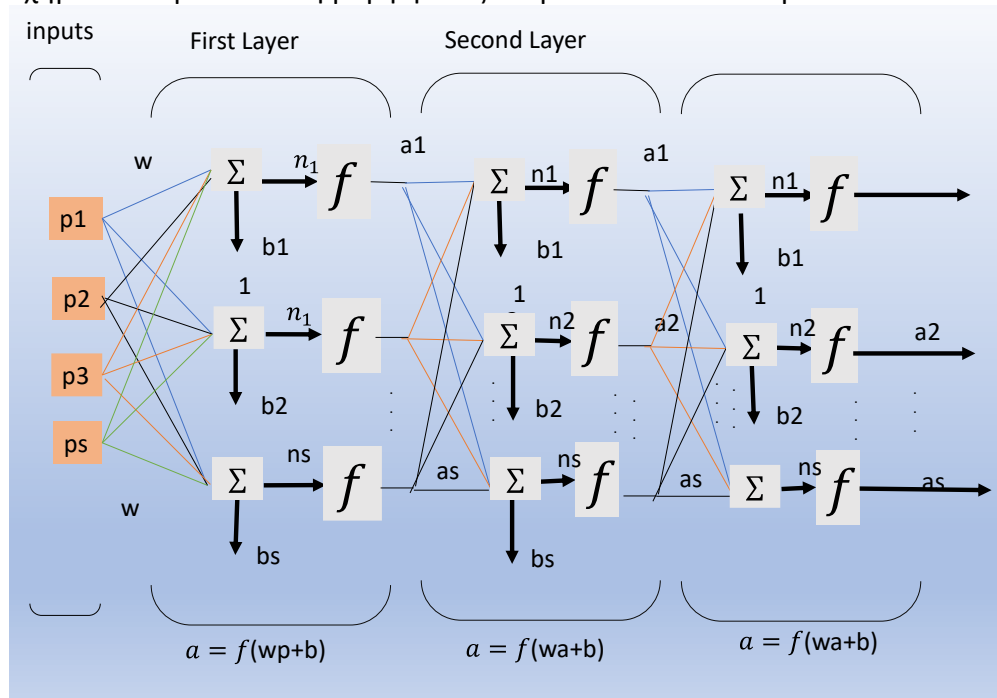
### 3.1.2 Πολυεπίπεδα Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα αποτελούνται από πολλά επίπεδα, και καθένα από αυτά περιλαμβάνει το δικό του σύνολο νευρώνων. Κάθε επίπεδο έχει τις δικές του εισόδους και εξόδους. Οι εξόδοι ενός επιπέδου μπορεί να αποτελούν είσοδο για άλλα επίπεδα, ή ακόμα να υπάρχουν συνδέσεις που ανατροφοδοτούν την ίδια την είσοδο του επιπέδου, δημιουργώντας ένα αναδρομικό σήμα.

Η διάταξη, το πλήθος και ο τρόπος με τον οποίο συνδέονται οι νευρώνες των διαφορετικών επιπέδων επηρεάζουν σημαντικά την απόδοση και την ικανότητα του δικτύου να επιλύει συγκεκριμένα προβλήματα. Η δομή των συνδέσεων και η σειρά των επιπέδων που ακολουθούνται καθορίζουν την αρχιτεκτονική του δικτύου, όπως τον αριθμό των επιπέδων, το μέγεθος κάθε επιπέδου και τον τρόπο με τον οποίο συνδέονται μεταξύ τους.

Ο σχεδιασμός και η επιλογή της κατάλληλης αρχιτεκτονικής για ένα συγκεκριμένο πρόβλημα είναι σημαντικοί παράγοντες που επηρεάζουν την απόδοση και την ικανότητα του νευρωνικού δικτύου να εκπαιδευτεί και να παράγει επιθυμητά αποτελέσματα.

Στο Σχήμα 10 παρατίθεται η μορφή ενός νευρωνικού δικτύου τριών επιπέδων.



Σχήμα 10: Μορφή νευρωνικού δικτύου τριών επιπέδων

### 3.1.3 Επίπεδο Εισόδου - Επίπεδο Εξόδου

Σε κάθε νευρωνικό δίκτυο υπάρχει ένα επίπεδο εισόδου και ένα επίπεδο εξόδου. Το επίπεδο εισόδου λαμβάνει το διάνυσμα των παραμέτρων ή των χαρακτηριστικών από το περιβάλλον και αυτές οι τιμές υπόκεινται σε επεξεργασία από τους νευρώνες του δικτύου. Το επίπεδο εξόδου αποδίδει το αποτέλεσμα της επεξεργασίας αυτής στο περιβάλλον.

Η είσοδος του νευρωνικού δικτύου συνήθως αποτελείται από ένα διάνυσμα πολλών παραμέτρων ή χαρακτηριστικών. Αυτές οι παράμετροι μπορεί να αντιπροσωπεύουν δεδομένα εισόδου όπως εικόνες, κείμενο, αριθμητικά χαρακτηριστικά κλπ.

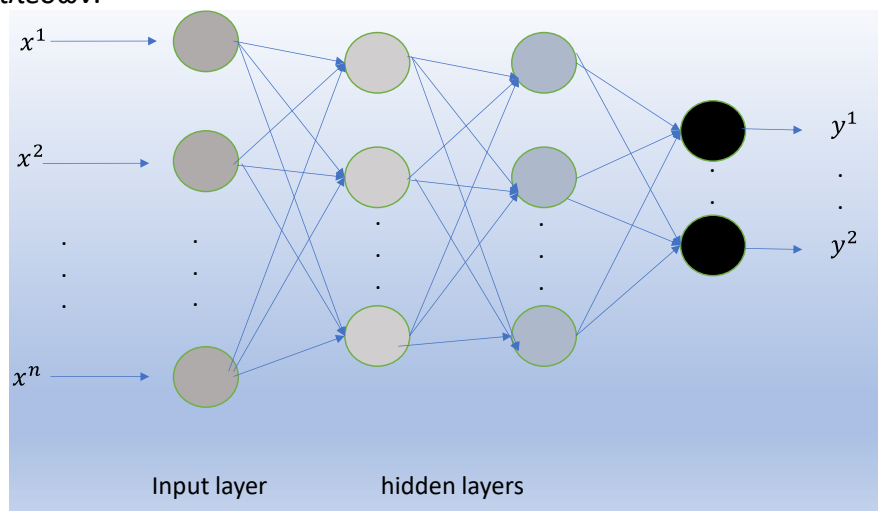
Η έξοδος του δικτύου μπορεί να αποτελείται είτε από μία μοναδική τιμή, είτε από ένα διάνυσμα τιμών, ανάλογα με την εφαρμογή και τον στόχο του δικτύου.

### 3.1.4 Εσωτερικά (κρυφά) επίπεδα

Τα εσωτερικά επίπεδα των νευρωνικών δικτύων είναι αυτά που βρίσκονται ανάμεσα στο επίπεδο εισόδου και το επίπεδο εξόδου. Αυτά τα επίπεδα λαμβάνουν ως είσοδο την έξοδο των νευρώνων από άλλα εσωτερικά επίπεδα ή από το επίπεδο εισόδου, και η έξοδός τους διαβιβάζεται ως είσοδος σε άλλα εσωτερικά επίπεδα ή στο επίπεδο εξόδου.

Αυτή η δομή των επιπέδων επιτρέπει την προοδευτική επεξεργασία και την εκμάθηση χαρακτηριστικών σε όλο το νευρωνικό δίκτυο. Κάθε εσωτερικό επίπεδο αποτελείται από έναν σύνολο νευρώνων, και ο κάθε νευρώνας λαμβάνει εισόδους από τους νευρώνες του προηγούμενου επιπέδου ή από το επίπεδο εισόδου.

Στο Σχήμα 11 φαίνεται η συσχέτιση του επιπέδου εισόδου, του επιπέδου εξόδου και των κρυφών επιπέδων.



Σχήμα 11: Η σχέση μεταξύ των επιπέδων των νευρωνικών δικτύων

### 3.2 Deep Learning

Με τον όρο Deep Learning περιγράφεται ως μια ειδίκευση της μηχανικής μάθησης που αποσκοπεί στην ανάπτυξη και την εκμάθηση πιο πολύπλοκων αναπαραστάσεων και χαρακτηριστικών από πολυεπίπεδα νευρωνικά δίκτυα. Η χρήση του Deep Learning είναι εξαιρετικά κατάλληλη για την ανάλυση μεγάλων ποσοτήτων δεδομένων, καθώς οι αλγόριθμοι αυτοί μπορούν να εξάγουν πλούσια και συμπαγή χαρακτηριστικά από τα δεδομένα αυτά. Η ιδέα του Deep Learning είναι να στοιβάζει πολλά επίπεδα νευρωνικών δικτύων, ο καθένας από τα οποία εκτελεί συναρτήσεις αναπαράστασης και εκμάθησης σε διαφορετικό επίπεδο πολυπλοκότητας. Με αυτόν τον τρόπο, τα επίπεδα μπορούν να εκμεταλλευτούν πιο σύνθετες δομές και συσχετίσεις στα δεδομένα και να παράγουν αναπαραστάσεις υψηλού επιπέδου.

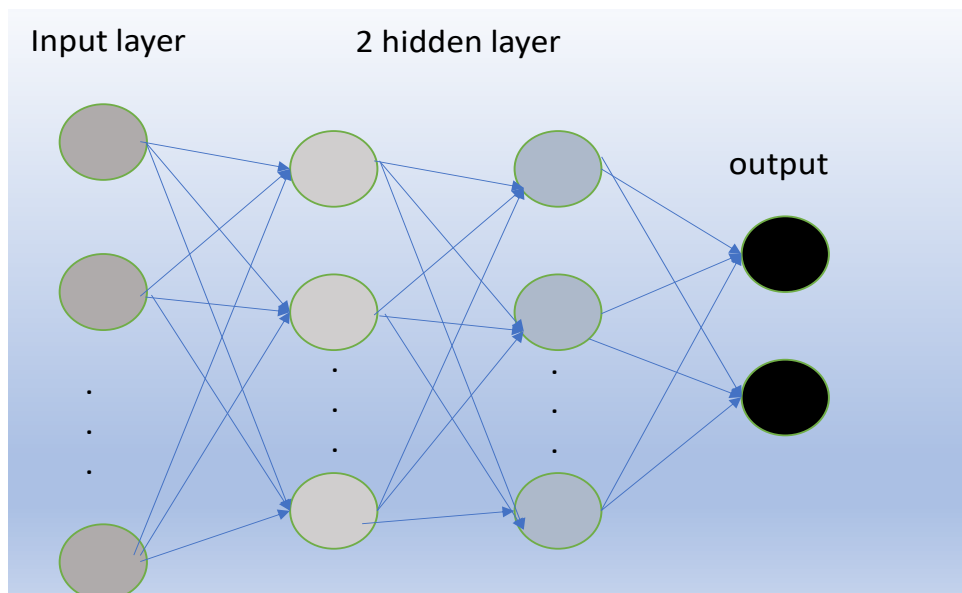
Το Deep Learning έχει αποδειχθεί εξαιρετικά αποτελεσματικό σε πολλούς τομείς, όπως η αναγνώριση εικόνας, η αναγνώριση φωνής, η επεξεργασία φυσικής γλώσσας και άλλες εφαρμογές που απαιτούν την αντιμετώπιση μεγάλων και πολύπλοκων δεδομένων [18].

### 3.3 Feed Forward Τεχνητά Νευρωνικά Δίκτυα

Τα feed-forward τεχνητά νευρωνικά δίκτυα είναι ένα είδος νευρωνικών δικτύων, όπου οι πληροφορίες ρέουν από την είσοδο προς την έξοδο με μία μονή κατεύθυνση. Τα δεδομένα εισέρχονται στο πρώτο επίπεδο, περνούν από διαφορετικά ενδιάμεσα επίπεδα και τελικά εξέρχονται από το τελευταίο επίπεδο εξόδου. Οι ενδιάμεσες επιπέδου αναδράσεις δεν επιτρέπονται σε αυτόν τον τύπο δικτύων, καθιστώντας τα απλά και ευέλικτα στην υλοποίηση και την εκπαίδευση.

Στα feed-forward νευρωνικά δίκτυα, κάθε επίπεδο αποτελείται από ένα σύνολο νευρώνων που συνδέονται με το επόμενο επίπεδο. Κάθε νευρώνας στο ένα επίπεδο είναι συνδεδεμένος με κάποιους νευρώνες στο επόμενο επίπεδο με συναπτικά βάρη. Οι εισόδους των νευρώνων πολλαπλασιάζονται με τα αντίστοιχα συναπτικά βάρη και συνολικά συνεισφέρουν στην υπολογιστική διαδικασία. Στη συνέχεια, η συνολική εισροή των εισόδων περνά από μια συνάρτηση ενεργοποίησης (όπως η σιγμοειδής συνάρτηση ή η συνάρτηση ReLU) που εφαρμόζεται σε κάθε νευρώνα, προκειμένου να παραχθεί η έξοδος του. Η διαδικασία αυτή επαναλαμβάνεται μέχρι η έξοδος να φτάσει στο τελευταίο επίπεδο. [19]

Στο Σχήμα 12 φαίνεται η τυπική μορφή ενός Feed Forward νευρωνικού δικτύου [20], [21].



Σχήμα 12: Τυπική μορφή FFNN

Τα feed-forward νευρωνικά δίκτυα (FFNN) είναι κατάλληλα για περιπτώσεις όπου η λειτουργία της μηχανικής μάθησης είναι ήδη γνωστή. Η βέλτιστη αρχιτεκτονική ενός FFNN εξαρτάται από την πολυπλοκότητα των συσχετίσεων που πρέπει να μοντελοποιηθούν.

Οι μέθοδοι για τον προσδιορισμό του βέλτιστου αριθμού κρυφών επιπέδων μπορούν να χωριστούν σε δύο κατηγορίες. Η μία κατηγορία ξεκινά από ένα πολύ απλό δίκτυο με λίγα κρυφά επίπεδα και σταδιακά αυξάνει την πολυπλοκότητα της αρχιτεκτονικής μέχρι να βρεθεί η βέλτιστη. Η άλλη κατηγορία αναπτύσσει το δίκτυο αντίστροφα, ξεκινώντας από ένα πολύπλοκο δίκτυο και σταδιακά απλοποιεί τη δομή μέχρι να εντοπιστεί η δομή που παρέχει τα βέλτιστα αποτελέσματα.

Και στις δύο περιπτώσεις, η εύρεση της βέλτιστης αρχιτεκτονικής απαιτεί έναν μεγάλο αριθμό υπολογισμών και αξιολογήσεων των αποτελεσμάτων. Ωστόσο, για πρακτικές εφαρμογές όπου η εκτεταμένη αναζήτηση δομής είναι δύσκολη ή ανέφικτη, μπορεί να χρησιμοποιηθεί ένας εμπειρικός κανόνας. Ένας τέτοιος κανόνας λέει ότι ο αριθμός των ρυθμιζόμενων παραμέτρων μπορεί να είναι περίπου το μισό του αριθμού των δειγμάτων που

διατίθενται για εκπαίδευση [22]. Αυτός ο κανόνας παρέχει έναν ευρύ προσανατολισμό για τον καθορισμό του μεγέθους του δικτύου, αλλά πρέπει να λαμβάνονται υπόψη και άλλοι παράγοντες όπως η διαθέσιμη ποσότητα δεδομένων και η πολυπλοκότητα του προβλήματος που προσπαθούμε να λύσουμε.

### 3.4 Backpropagation

Ο αλγόριθμος Backpropagation, αποτελεί μια τεχνική που επιτρέπει την αποδοτική εκπαίδευση των νευρωνικών δικτύων και την εκμάθηση περίπλοκων σχέσεων ανάμεσα στις εισόδους και τις εξόδους. Με την προσέγγιση αυτή, ορίζεται μία συνάρτηση η οποία αξιολογεί την απόδοση του δικτύου ως προς την έξοδο που παράγει. Η συνάρτηση αυτή εκτιμά την απόκλιση της τιμής που παράγει το εκπαιδευμένο (μέχρι την δεδομένη χρονική στιγμή) μοντέλο σε σχέση με την πραγματική. Κατά την ανάπτυξη ενός νευρωνικού δικτύου, ο αλγόριθμος backpropagation χρησιμοποιείται για να υπολογίσει τη διαβάθμιση της συνάρτησης σφάλματος ως προς τα βάρη του δικτύου για ένα δείγμα εισόδου-εξόδου. Ο αλγόριθμος λειτουργεί με τον εξής τρόπο: Αρχικά, το δίκτυο παράγει μια πρόβλεψη για ένα δείγμα εισόδου. Στη συνέχεια, υπολογίζεται η απόκλιση της πρόβλεψης από την πραγματική τιμή εξόδου, χρησιμοποιώντας την επιλεγμένη συνάρτηση σφάλματος.

Έπειτα, ο αλγόριθμος backpropagation χρησιμοποιείται για να υπολογίσει τις μερικές παραγώγους της συνάρτησης σφάλματος ως προς τα βάρη του δικτύου. Αυτό γίνεται με την εφαρμογή του κανόνα της αλυσίδας (chain rule) για τον υπολογισμό των παραγώγων σε κάθε επίπεδο του δικτύου, ξεκινώντας από το τελευταίο επίπεδο και κινούμενοι προς τα πίσω.

Οι υπολογισμένες μερικές παράγωγοι χρησιμοποιούνται στη συνέχεια για την προσαρμογή των βαρών του δικτύου, συνήθως με τη χρήση μιας μεθόδου βελτιστοποίησης, όπως η κλασική μέθοδος της κατάβασης του κλάσματος (gradient descent). Η διαδικασία επαναλαμβάνεται για πολλά δείγματα εισόδου-εξόδου, έτσι ώστε το δίκτυο να μάθει να παράγει ολοένα καλύτερες προβλέψεις. Ο backpropagation βασίζεται σε τέσσερις θεμελιώδεις εξισώσεις που παρέχουν τον τρόπο να υπολογίζεται το σφάλμα και η κλίση της συνάρτησης σφάλματος.

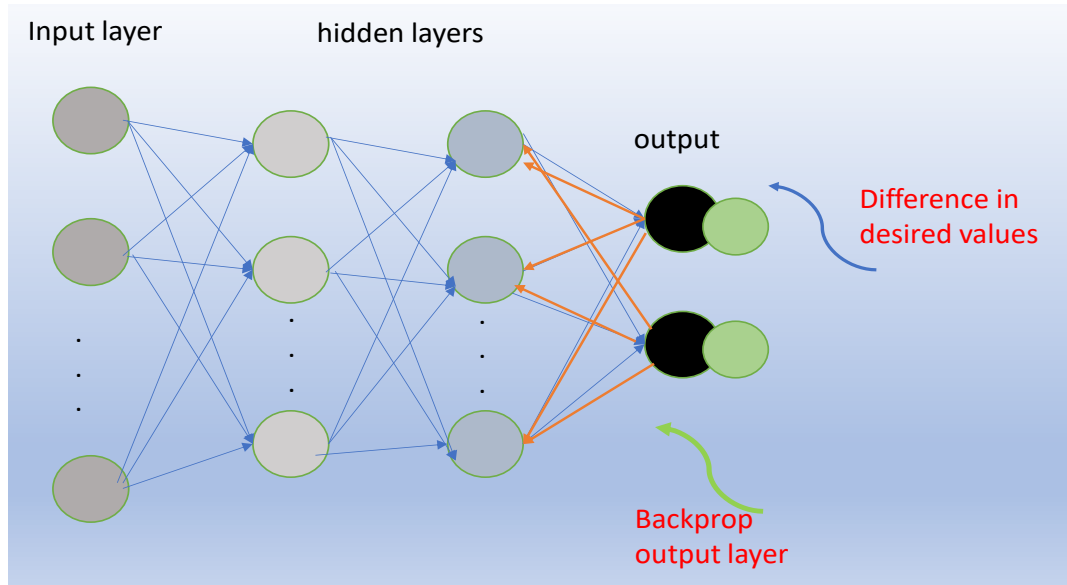
- Εξίσωση για το σφάλμα στο επίπεδο εξόδου
- Εξίσωση που αφορά ένα συγκεκριμένο επίπεδο, για το σφάλμα στο επόμενο επίπεδο
- Εξίσωση για το ρυθμό μεταβολής του σφάλματος σε σχέση με οποιαδήποτε πόλωση στο δίκτυο
- Εξίσωση για το ρυθμό μεταβολής του σφάλματος σε σχέση με οποιαδήποτε βάρος στο δίκτυο



Στους αλγορίθμους που βασίζονται στην αρχιτεκτονική αυτή, χρησιμοποιούνται μέθοδοι  $\text{gradient descent}^2$  ή  $\text{stochastic gradient}^3$ , με σκοπό την ελαχιστοποίηση της συνάρτησης σφάλματος.

Τα νευρωνικά δίκτυα τα οποία αναπτύσσονται με την αρχιτεκτονική του Backpropagation, έχουν την δυνατότητα να παρέχουν λεπτομερείς πληροφορίες για το πώς η αλλαγή των βαρών και της πόλωσης αλλάζει τη συνολική συμπεριφορά του δικτύου.

Στο Σχήμα 13 φαίνεται σχηματικά η αρχιτεκτονική Backpropagation στην λειτουργία των νευρωνικών δικτύων.



Σχήμα 13: Λειτουργία Backpropagation

Η τεχνική του backpropagation βρίσκει αρκετές εφαρμογές στον πραγματικό κόσμο καθώς βελτιώνει σημαντικά την απόδοση των νευρωνικών δικτύων. Η επαναληπτική ρύθμιση των βαρών μέχρι να ελαχιστοποιηθεί το σφάλμα τα καθιστά ικανά να εντοπίζουν μοτίβα συμπεριφοράς των μεγεθών με μεγάλη αξιοπιστία [23]. Η τεχνική του backpropagation χρησιμοποιείται σχεδόν πάντα στην ανάπτυξη των FFNN προκειμένου να αυξάνεται η ακρίβεια των μοντέλων που παράγουν.

---

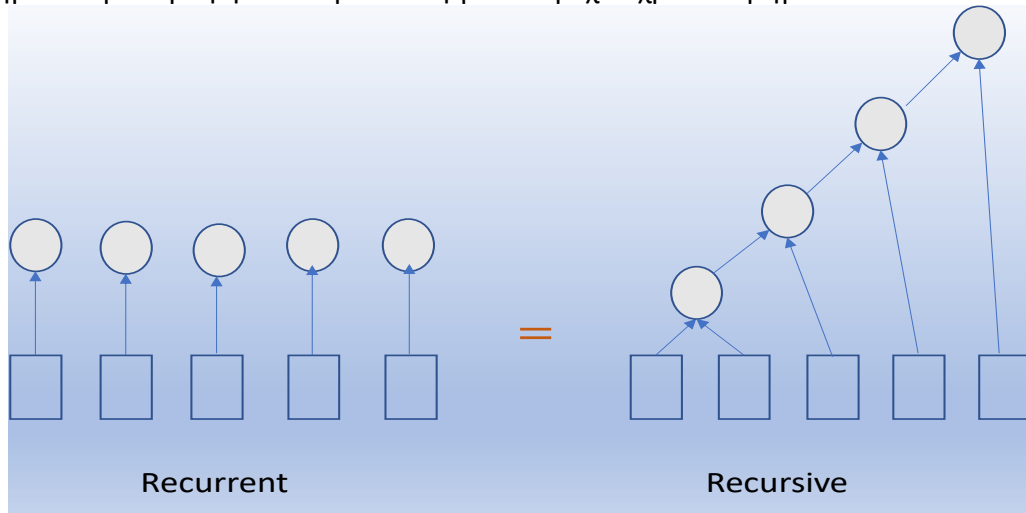
<sup>2</sup> Αναφέρεται και ως αλγόριθμος σύγκλισης με ελάττωση της παραγώγου. Ο αλγόριθμος της απότομης καθόδου ξεκινάει με αρχικές  $\theta_0$  και  $\theta_1$  και αλλάζοντας τιμές σε αυτές τις παραμέτρους προσπαθεί να συγκλίνει σε τιμές που ελαχιστοποιούν την συνάρτηση κόστους  $J(\theta_0, \theta_1)$

<sup>3</sup> Εάν τροποποιηθεί η μέθοδος  $\text{gradient descent}$  και πραγματοποιηθούν οι επαναλήψεις με τον ρυθμό μάθησης να μειώνεται (ξεκινώντας από μια αρχική τιμή  $\theta_0$  κατά  $\theta_0/k$ , ο αλγόριθμος θα συγκλίνει σε τοπικό ελάχιστο του συνολικού σφάλματος.

### 3.5 Recursive Neural Networks

Τα Recursive Neural Networks (RecursiveNN) έχουν την δυνατότητα να παρέχουν δομημένη πρόβλεψη, εφαρμόζοντας τον ίδιο αριθμό συνόλων βαρών σε δομημένες εισόδους. Πρόκειται για μοντέλα που είναι ικανά να αποκτήσουν βαθιά δομημένη γνώση. Ως εκ τούτου, η δομή τους παρομοιάζεται με πολύπλοκες αλυσίδες νευρώνων. Τα RecursiveNN είναι δομημένα με τέτοιο τρόπο που να λειτουργούν αποδοτικά με δομημένες εισόδους.

Ένα RecurrentNN, αποτελεί μία περίπτωση RecursiveNN. Και τα δύο είδη δικτύων διαφοροποιούνται από τα FeedForward NN στο γεγονός ότι η γνώση που αποκτούν δεν βασίζεται αποκλειστικά στο παρελθόν. Το RecursiveNN μαθαίνει από το παρελθόν και επεξεργάζεται νέα δεδομένα με βάση την εμπειρία. Χρησιμοποιώντας αλγόριθμους εκμάθησης προσδιορίζει το είδος των αλλαγών που θα πρέπει να συμβούν στο μέλλον. Τα RecurrentNN λειτουργούν στη γραμμική πρόοδο του χρόνου, συνδυάζοντας το προηγούμενο χρονικό βήμα και μια κρυφή αναπαράσταση για το τρέχον χρονικό βήμα.



Σχήμα 14: Η διαφορά στη δομή μεταξύ RNN και RecursiveNN

Ένα RecursiveNN είναι ένα είδος νευρωνικού δικτύου που δημιουργείται με την εφαρμογή του ίδιου συνόλου βαρών αναδρομικά σε μια δομημένη είσοδο, για την παραγωγή μιας δομημένης πρόβλεψης. Τα RecursiveNN, που μερικές φορές συντομεύονται ως RvNN, ήταν επιτυχή, για παράδειγμα, στη μάθηση ακολουθιών στην επεξεργασία φυσικής γλώσσας, κυρίως σε συνεχείς αναπαραστάσεις φράσεων και προτάσεων με βάση την ενσωμάτωση λέξεων. Τα RvNN εισήχθησαν για πρώτη φορά για να μάθουν κατανεμημένες αναπαραστάσεις δομής, όπως λογικούς όρους. Μοντέλα και γενικά πλαίσια έχουν αναπτυχθεί σε περαιτέρω εργασίες από τη δεκαετία του 1990 [24].

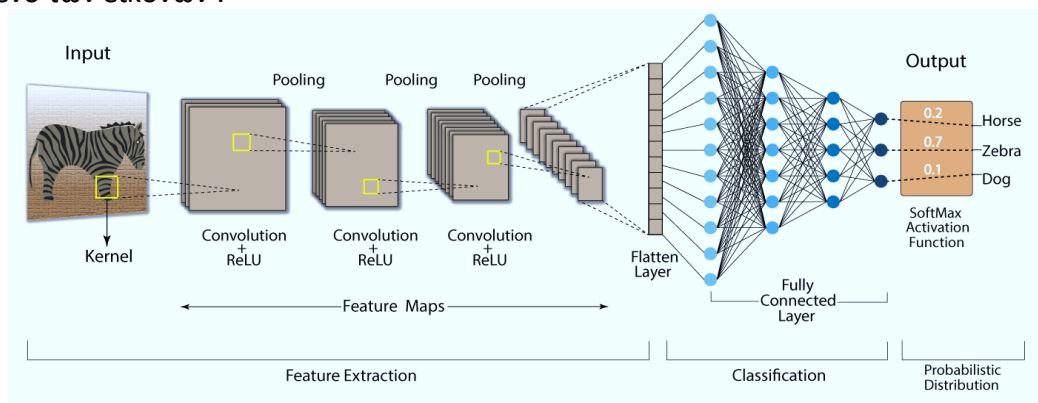
### 3.6 Convolutional Neural Networks

Τα Convolutional Neural Networks (CNNs) αποτελούν έναν ειδικό τύπο νευρωνικών δικτύων που χρησιμοποιούνται ευρέως για την ανάλυση εικόνων και την επεξεργασία σήματος. Έχουν σχεδιαστεί με στόχο να αντιμετωπίσουν τις ειδικές απαιτήσεις και δομή των εικόνων.

Τα CNN αποτελούνται από επίπεδα νευρώνων, συμπεριλαμβανομένων των επιπέδων συνέλιξης (convolutional layers), των επιπέδων συγκέντρωσης (pooling layers) και των πλήρως συνδεδεμένων επιπέδων (fully connected layers). Οι συνέλιξεις και οι συγκεντρώσεις επιτρέπουν στο δίκτυο να αντλήσει χαρακτηριστικά και πληροφορίες από την εικόνα σε διάφορα επίπεδα αφαιρώντας τη χωρική δομή (διάταξη δεδομένων στο χώρο) και μειώνοντας την πολυπλοκότητα των δεδομένων.

Η διαδικασία εκπαίδευσης των CNN περιλαμβάνει την προσαρμογή των βαρών του δικτύου με βάση τα δεδομένα εκπαίδευσης και την ελαχιστοποίηση του σφάλματος (loss) της προς τα εμπρός διάδοσης (forward propagation). Τα τελικά επίπεδα του δικτύου συνήθως περιέχουν συναρτήσεις σφάλματος που αξιολογούν την απόδοση του δικτύου σε σχέση με την επιθυμητή έξοδο (όπως η κατηγοριοποίηση εικόνων σε διάφορες κλάσεις).

Έτσι, τα CNN είναι ικανά να αναλύουν εικόνες και να εξάγουν χαρακτηριστικά από διάφορα επίπεδα, επιτρέποντας τους να μοντελοποιούν και να κατανοούν τη δομή και το περιεχόμενο των εικόνων .



Σχήμα 15: Παράδειγμα CNN[25]

Στο Σχήμα 15 φαίνεται ότι η εικόνα περνάει από διαδοχικά στάδια κωδικοποίησης και δειγματοληψίας με αποτέλεσμα να δημιουργείται μία γραμμική κωδικοποίηση της Σχήμας σε ένα διάνυσμα μίας διάστασης. Η όλη αυτή επεξεργασία πραγματοποιείται σε διαδοχικά επίπεδα του νευρωνικού δικτύου. Στην συνέχεια το διάνυσμα περνάει από μία σειρά επιπέδων μέχρι να προκύψει ένα διάνυσμα στην έξοδο με τόσες συντεταγμένες όσες και οι κλάσεις στις οποίες μπορεί να ενταχθεί η Σχήμα. Οι τιμές στο διάνυσμα αντιστοιχούν με την πιθανότητα η Σχήμα να ανήκει στην συγκεκριμένη κλάση[26] .

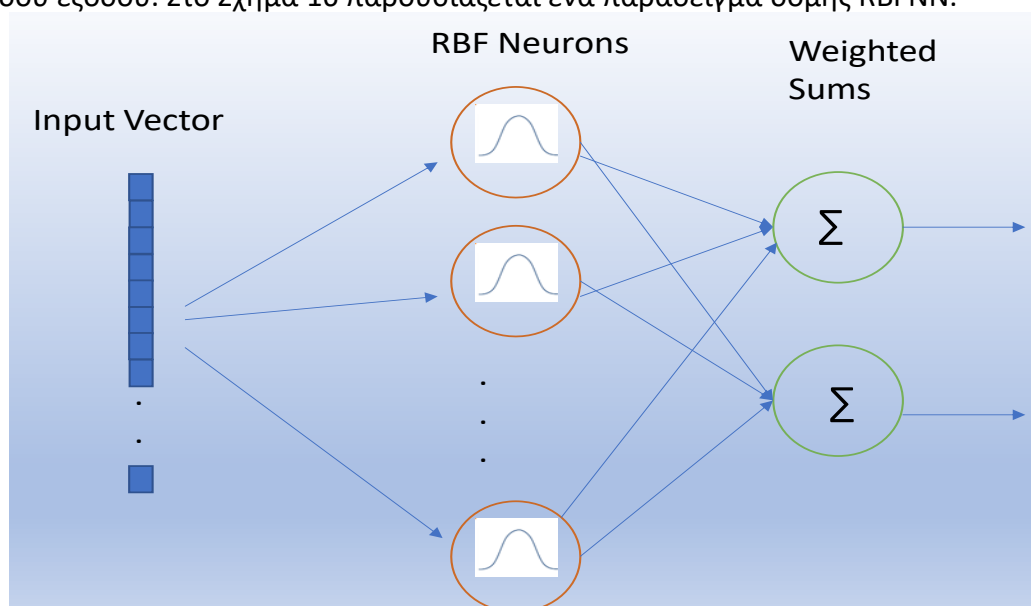
### 3.7 Radial Basis Function Networks

Η βάση για την προσέγγιση των Radial Basis Function Networks (RBNF) αποτέλεσε η μελέτη της τοπικά συντονισμένης απόκρισης των βιολογικών νευρώνων [27]. Η ιδέα μοντελοποίησης πίσω από τα RBF είναι ένας συνδυασμός μάθησης χωρίς επίβλεψη και με επίβλεψη [28]. Πρόκειται για μία σχετικά νέα προσέγγιση. Η διαδικασία εκπαίδευσης των αρχιτεκτονικών RBF χρησιμοποιεί μια βελτιστοποίηση μέτρησης απόστασης ομοιότητας για την εύρεση παρόμοιων και ανόμοιων δειγμάτων [29]. Τα RBF εισήχθησαν αρχικά ως μια εποπτευόμενη προσέγγιση για εργασίες ταξινόμησης και παλινδρόμησης . Οι Broomhead και

Lowe<sup>4</sup> πρότειναν να αντληθούν τα κέντρα είτε από μια ομοιόμορφη κατανομή είτε τυχαία από τα δείγματα εκπαίδευσης και στη συνέχεια να βελτιστοποιήσουν τα βάρη εξόδου χρησιμοποιώντας μια ψευδο-αντίστροφη αναλυτική λύση. Η εκκίνηση των κέντρων συμπλέγματος τυχαία και η εκπαίδευση μόνο των βαρών εξόδου είναι μια διαδικασία εκπαίδευσης μιας φάσης για τα RBF. Η εκπαίδευση δύο φάσεων για τα RBF χρησιμοποιεί διάφορες μεθόδους για την προετοιμασία των κέντρων καθώς και τη βελτιστοποίηση των βαρών εξόδου [30] [31]. Η έρευνα χρησιμοποίησε εποπτευόμενες και μη εποπτευόμενες μεθόδους για την προετοιμασία των κέντρων [32]. Όλες αυτές οι μέθοδοι για την προετοιμασία του κέντρου συμπλέγματος προϋποθέτουν έναν σταθερό χώρο χαρακτηριστικών για το επίπεδο εισόδου [33]. Τα RBF έχουν εφαρμοστεί σε ένα ευρύ φάσμα τομέων και για ταξινόμηση προτύπων και παλινδρόμηση τα τελευταία χρόνια, με εφαρμογές που περιλαμβάνουν τον αυτόματο έλεγχο [27], [33], [34], [35] την online εκπαίδευση [36], [11] τα ενεργειακά συστήματα [12], την μηχανική διεργασιών [37], την πρόβλεψη σεισμικής δραστηριότητας [22] κτλ.

Τα Radial Basis Function Networks (RBFN) αποτελούν μια μέθοδο για την εκμάθηση μιας χαρτογράφησης οντοτήτων από έναν χώρο χαρακτηριστικών σε ένα νέο χώρο διαστάσεων. Η αρχιτεκτονική των RBFN περιλαμβάνει ένα επίπεδο εισόδου, ένα ενιαίο κρυφό στρώμα με κέντρα εκπαίδευσης και ένα στρώμα εξόδου.

Κατά τη διάρκεια της αξιολόγησης, το RBFN υπολογίζει την απόσταση μεταξύ των εισόδων του δικτύου και των κέντρων (το κέντρο είναι μια μέγιστη τιμή που αντιστοιχεί σε έναν νευρώνα του κρυφού στρώματος) και εφαρμόζει μια συνάρτηση ενεργοποίησης (όπως η Gaussian) σε αυτήν την απόσταση. Η ενεργοποίηση αυτή αναπαριστά το μέτρο της ομοιότητας μεταξύ της εισόδου και του κέντρου. Στη συνέχεια, οι έξοδοι του δικτύου υπολογίζονται πολλαπλασιάζοντας τις ενεργοποιήσεις του κρυφού στρώματος με τα βάρη του επιπέδου εξόδου. Στο Σχήμα 16 παρουσιάζεται ένα παράδειγμα δομής RBFNN.



Σχήμα 16: Παράδειγμα RBFNN

<sup>4</sup> Ήταν οι πρώτοι που παρουσίασαν την αρχιτεκτονική RBFN

Η πλήρης διαδικασία εκπαίδευσης των RBF είναι μία διαδικασία τριών φάσεων:

- Εκμάθηση χωρίς επίβλεψη: Σε αυτό το βήμα υπολογίζονται τα κέντρα των συστάδων που είναι αντιπροσωπευτικά των δεδομένων. Ο αλγόριθμος ομαδοποίησης k-means χρησιμοποιείται ευρέως για το σκοπό αυτό. Το k-means βρίσκει επαναληπτικά ένα σύνολο κέντρων και ελαχιστοποιεί τη συνολική απόσταση μεταξύ των κέντρων και των μελών σε ολόκληρο το σύνολο δεδομένων.
- Υπολογισμός βαρών του δικτύου RBF.
- Βελτιστοποίηση από άκρο σε άκρο: Μετά την προετοιμασία των βαρών και των κέντρων των συστάδων του RBF, πραγματοποιείται η βελτιστοποίηση του δικτύου από άκρο σε άκρο μέσω backpropagation και gradient descent [13].

## Recurrent Neural Networks

---

### 4. Recurrent Neural Networks

#### 4.1 Χαρακτηριστικά

Τα Recurrent Neural Networks (RNN) είναι ένας τύπος νευρωνικών δικτύων που έχουν τη δυνατότητα να αντιμετωπίζουν ακολουθιακά δεδομένα και να διατηρούν τη γνώση απ' το παρελθόν. Οι βασικές χαρακτηριστικές ιδιότητες των RNN περιλαμβάνουν:

**Χρονική Εξάρτηση:** Τα RNN αναγνωρίζουν και επεξεργάζονται ακολουθιακά δεδομένα, όπως χρονοσειρές ή διαδοχικές παρατηρήσεις. Ο παράγοντας του χρόνου είναι κρίσιμος για τη διάκριση μεταξύ των στοιχείων της ακολουθίας.

**Κοινή Κατανομή Παραμέτρων:** Τα RNN χρησιμοποιούν την ίδια συνάρτηση ενεργοποίησης και το ίδιο σύνολο βαρών για κάθε χρονικό βήμα της ακολουθίας. Αυτό επιτρέπει την κοινή μάθηση και την κατανομή πληροφορίας σε όλα τα βήματα.

**Αντίστροφη Σύνδεση:** Τα RNN έχουν σύνδεση προς τα πίσω που επιτρέπει τη διάδοση πληροφορίας από προηγούμενα χρονικά βήματα στο τρέχον βήμα. Αυτό επιτρέπει την κατανόηση και την αναπαράσταση των μακροπρόθεσμων συσχετίσεων στην ακολουθία.

**Μνήμη Κατάστασης:** Τα RNN διατηρούν μια εσωτερική κατάσταση ή μνήμη που αποθηκεύει πληροφορία από προηγούμενα βήματα. Αυτή η μνήμη επιτρέπει την αποθήκευση συνολικού περιεχομένου της ακολουθίας και την αξιοποίησή του για προβλέψεις ή επεξεργασία στοιχείων.

Με τη συνδυασμένη χρήση αυτών των χαρακτηριστικών, τα RNN είναι ιδιαίτερα αποτελεσματικά για προβλήματα που σχετίζονται με την ακολουθιακή επεξεργασία δεδομένων, και μπορούν να εκπαιδευτούν να αντιληφθούν και να μάθουν τα μοτίβα και τις εξαρτήσεις που υπάρχουν στις ακολουθίες.[38]

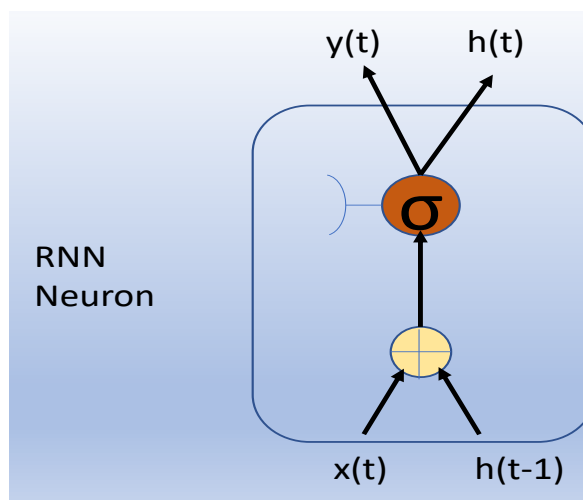
Βασικό χαρακτηριστικό των RNN είναι η κυκλική συνδεσιμότητα των νευρώνων τους, που επιτρέπει την ενημέρωση της τρέχουσας κατάστασης με βάση παρελθούσες καταστάσεις και την τρέχουσα είσοδο. Στα τεχνητά νευρωνικά δίκτυα αυτής της μορφής περιέχεται τουλάχιστον μία σύνδεση η οποία ανατροφοδοτεί την έξοδο ενός νευρώνα είτε σε νευρώνα του ίδιου επιπέδου, είτε σε νευρώνα προηγούμενου επιπέδου. Με τον τρόπο αυτό καταφέρνουν, μέσα από μία επαναληπτική διαδικασία, μετά από κύκλους ανατροφοδότησης να φθάνουν – πιθανότατα - σε μία σταθερή κατάσταση. Ζητούμενο στην διαδικασία μάθησης είναι να υπολογιστεί εκείνος ο συνδυασμός των συναπτικών βαρών, που οδηγεί σε σταθεροποίηση των εξόδων του δικτύου. Τα RNN είναι κατάλληλα για την

επεξεργασία δεδομένων χρονοσειρών ή άλλων τύπων δεδομένων τα οποία εξελίσσονται διαδοχικά. Στην ουσία, αποτελούν μία επέκταση των Feedforward νευρωνικών δικτύων, προκειμένου να είναι δυνατή η επεξεργασία ακολουθιών μεταβλητού μήκους (ή ακόμα και απεριόριστου μήκους). Μερικές από τις πιο δημοφιλείς RNN αρχιτεκτονικές που χρησιμοποιούνται, είναι τα Long Short-term memory (LSTM) και τα Gated Recurrent Units (GRU).[39]

Τα FFNN γενικά δέχονται εισόδους σταθερού μεγέθους και παρέχουν εξόδους, επίσης, σταθερού μεγέθους. Αντίθετα, τα RNN δέχονται σαν είσοδο ακολουθίες δεδομένων μεταβλητού μεγέθους και αποδίδουν ακολουθίες μεταβλητού μήκους. Αυτό επιτυγχάνεται μέσα από την κοινή χρήση παραμέτρων και μετασχηματισμών με την πάροδο του χρόνου. Μία άλλη διαφορά των ζητημάτων που μπορούν να χειριστούν τα δύο είδη νευρωνικών δικτύων είναι η φύση της εισόδου που είναι διαθέσιμη. Τα FFNN αντιμετωπίζουν μία είσοδο για κάθε έξοδο που παράγουν, και αυτό μπορεί να περιορίζει την ικανότητά τους να χειριστούν πολλαπλές διαδοχικές εισόδους που συσχετίζονται μεταξύ τους. Αυτό είναι ιδιαίτερα προφανές στην ανάλυση χρονοσειρών, όπου ο χρόνος είναι μια σημαντική παράμετρος και η σειρά των εισόδων έχει σημασία. Χαρακτηριστικό παράδειγμα είναι οι μελέτες που σχετίζονται με την εκτίμηση και την πρόβλεψη της φυσικής κατάστασης των ανθρώπων. Η φυσική κατάσταση σε μία δεδομένη χρονική συγκυρία δεν εξαρτάται αποκλειστικά από μία δεδομένη παρελθούσα κατάσταση αλλά από μία σειρά καταστάσεων που μπορεί να φθάνει σε κάποιο δεδομένο χρονικό παράθυρο. Η επέκταση των FFNN σε RNN αποδίδει την αντιμετώπιση ακριβώς αυτών των αδυναμιών.

## 4.2 Δομή

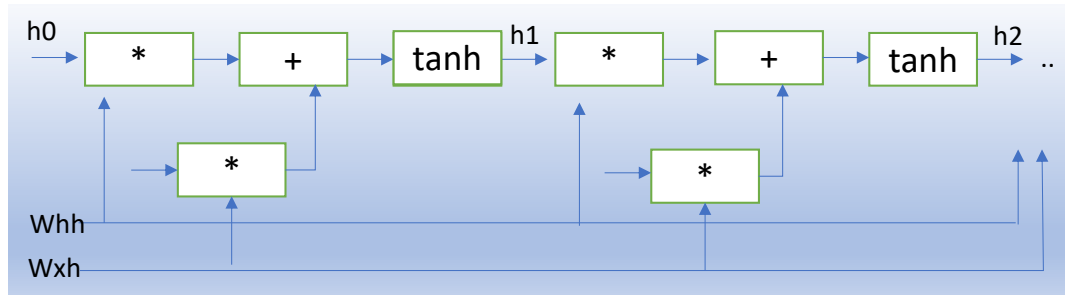
Η απλούστερη προσέγγιση για την δημιουργία RNN, αφορά την τροποποίηση των FFNN, ώστε να είναι εφικτή η μεταφορά λανθάνουσας πληροφορίας από ένα τμήμα δεδομένων (χρονικό βήμα) σε άλλο τμήμα. Μετά την εκπαίδευση του μοντέλου στο αντίστοιχο επίπεδο, αυτές οι λανθάνουσες καταστάσεις μπορούν να ερμηνευτούν ως αναπαράσταση αποκτηθείσας γνώσης. Στο Σχήμα 17 φαίνεται η γενική μορφή των νευρώνων σε ένα RNN δίκτυο.



Σχήμα 17: Τυπικός Νευρώνας σε RNN

Στο παραπάνω σχήμα φαίνεται ότι η συνάρτηση ενεργοποίησης δέχεται σαν είσοδο τον συνδυασμό της αναδρομικής πληροφορίας  $h(t-1)$ , της εισόδου  $x(t)$  και την σύναψη. Στην έξοδο παρέχει την έξοδο του νευρώνα και την αντίστοιχη αναδρομική πληροφορία [40].

Με την ενσωμάτωση της αναδρομικής πληροφορίας, είναι δυνατό να αλλάξει ο γραμμικός μετασχηματισμός που χρησιμοποιείται στον υπολογισμό της κατάστασης που προκύπτει από τη επεξεργασία στα κρυφά επίπεδα του νευρωνικού δικτύου, ώστε να μην εξαρτάται μόνο από την είσοδο  $x$  αλλά και από πληροφορίες που προέρχονται από προηγούμενα τμήμα των δεδομένων, που μεταφέρονται μέσω της κρυφής κατάστασης. Με τον τρόπο αυτό γίνεται δυνατή η χρησιμοποίηση δεδομένων που αποτελούνται από σύνολα τμημάτων διαφορετικού μεγέθους. Στο Σχήμα 18 φαίνεται ένα παράδειγμα ενός τέτοιου νευρωνικού δικτύου.



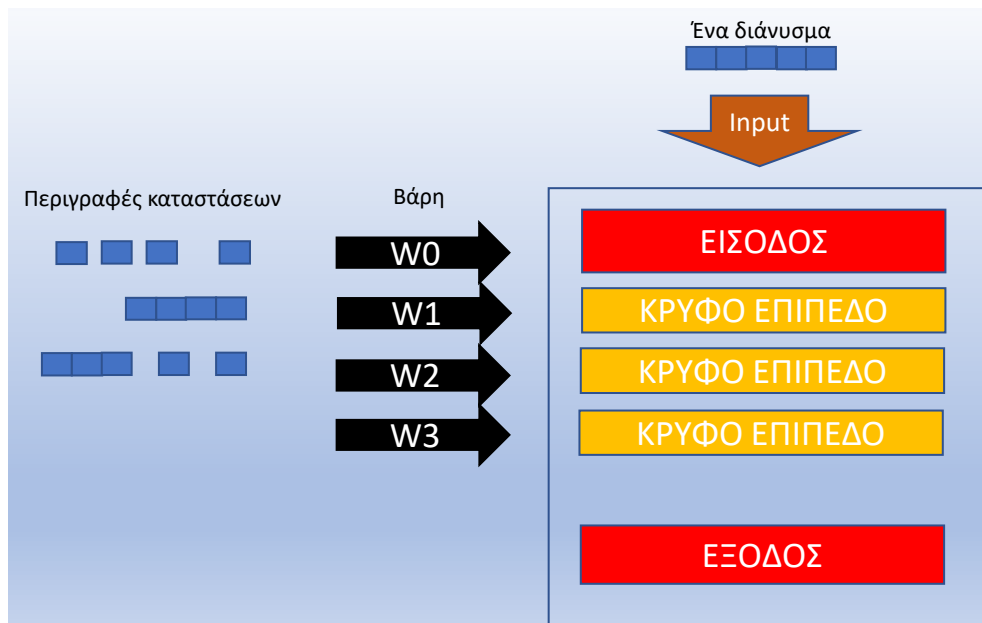
Σχήμα 18: Αναπαράσταση απλού RNN δικτύου

Σε κάθε χρονικό βήμα, η είσοδος (συνήθως συμβολίζεται ως  $h_0$ ) και η κατάσταση από το προηγούμενο χρονικό βήμα συνδυάζονται με έναν συνδυασμό σταθμισμένων παραμέτρων. Αυτός ο συνδυασμός υποβάλλεται σε μια συνάρτηση ενεργοποίησης, η οποία αντιστοιχεί στους νευρώνες του επόμενου επιπέδου.

Η διαδικασία αυτή επαναλαμβάνεται σε κάθε χρονικό βήμα και σε κάθε κρυφό επίπεδο του νευρωνικού δικτύου, μέχρι να φτάσουμε στο επίπεδο εξόδου. Κατά αυτόν τον τρόπο, η επεξεργασία σε κάθε επίπεδο εξαρτάται από την επεξεργασία στο προηγούμενο επίπεδο και την κατάσταση σε ένα συγκεκριμένο τμήμα των δεδομένων.

Αυτή η αρχιτεκτονική επιτρέπει στο RNN να διαχειρίζεται καταστάσεις που περιλαμβάνονται στα δεδομένα και περιγράφονται από διαφορετικά πλήθη παραμέτρων. Επίσης, μπορεί να αντιμετωπίσει την έλλειψη ορισμένων τμημάτων των δεδομένων σε συγκεκριμένες περιπτώσεις. Η επεξεργασία δεδομένων που έχουν τη μορφή χρονοσειρών, για την δημιουργία μοντέλου πρόβλεψης για το ύψος του αριθμού των θανάτων από COVID-19, με την προσέγγιση αυτή θα μπορούσε να αναθέτει την επεξεργασία της κάθε κατάστασης της χρονοσειράς (τμήμα των δεδομένων) σε διαφορετικό επίπεδο του δικτύου. Στο Σχήμα 19 αποδίδεται ο τρόπος με τον οποίο μπορεί να πραγματοποιηθεί η είσοδος των καταστάσεων στο νευρωνικό δίκτυο, σε διαφορετικό επίπεδο και υπό διαφορετική βαρύτητα.





Σχήμα 19: Περίπτωση απλού RNN

#### 4. 3 Εκπαίδευση RNN

Νωρίτερα στο κεφάλαιο 6 αναλύθηκε η μέθοδος Backpropagation. Όπως αναφέρθηκε ο αλγόριθμος Backpropagation είναι κατάλληλος για την εκπαίδευση Feed Forward Τεχνητών Νευρωνικών Δικτύων. Ο απώτερος στόχος του αλγορίθμου Backpropagation είναι να ελαχιστοποιήσει το σφάλμα των εξόδων του δικτύου.

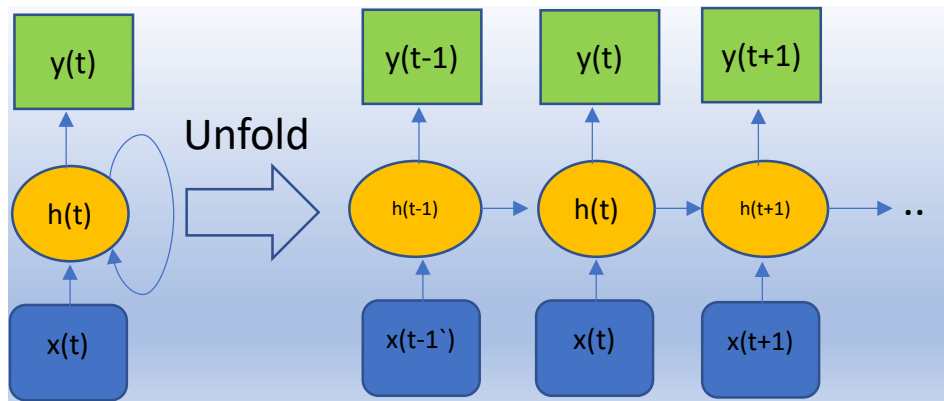
Το Backpropagation Through Time (BPTT) είναι μια προσαρμογή του αλγορίθμου εκπαίδευσης Backpropagation που εφαρμόζεται σε ακολουθιακά δεδομένα, όπως οι χρονοσειρές. Στα αναδρομικά νευρωνικά δίκτυα, κάθε βήμα εισόδου συνδέεται με την αντίστοιχη έξοδο, και το BPTT λειτουργεί για όλα τα χρονικά βήματα εισόδου.

Κατά τη διάρκεια του BPTT, για κάθε χρονικό βήμα, πραγματοποιείται ένα αντίγραφο του δικτύου, με τα ίδια βάρη, ώστε να μπορεί να αξιολογηθεί η επίδοσή του. Τα σφάλματα υπολογίζονται για κάθε χρονικό βήμα, και συσσωρεύονται από το τελευταίο χρονικό βήμα προς τα πίσω. Αυτό σημαίνει ότι τα σφάλματα από την τελική έξοδο επιστρέφουν στο προηγούμενο βήμα εισόδου, και έτσι συνεχίζουν να επιστρέφουν προς τα πίσω για κάθε χρονικό βήμα.

Μετά τον υπολογισμό των σφαλμάτων, πραγματοποιείται η ενημέρωση των βαρών του δικτύου, χρησιμοποιώντας τη μέθοδο gradient descent, όπως στο απλό Backpropagation. Αυτή η ενημέρωση των βαρών συμβάλλει στη βελτίωση της απόδοσης του δικτύου στην πρόβλεψη ακολουθιακών δεδομένων.

Η διαδικασία του BPTT επαναλαμβάνεται, με στόχο τη σύγκλιση του δικτύου και τη βελτίωση της ακρίβειάς των προβλέψεων στα ακολουθιακά δεδομένα.

Τα δεδομένα εκπαίδευσης σε ένα αναδρομικό νευρωνικό δίκτυο αποτελούνται από ζεύγη διανυσμάτων εισόδων -εξόδων. (Σχήμα 20)



Σχήμα 20: Γενική λειτουργία των αλγορίθμων εκπαίδευσης RNN

Πρώτο βήμα για την εκπαίδευση του αναδρομικού δικτύου είναι ο υπολογισμός του σφάλματος(L)

$$L = \sum_{t=1}^T L^{(t)} \quad (4-1)$$

Στην συνέχεια η εύρεση της μερικής παραγώγου του L(σφάλμα) με βάση τα τρία βάρη

$$W_h, W_{hh}, W_{hx}$$

για κάθε χρονικό βήμα.

$$\frac{\partial L^{(t)}}{\partial W_{hh}} = \frac{\partial L^{(t)}}{\partial y^{(t)}} \cdot \frac{\partial y^{(t)}}{\partial h^{(t)}} \cdot \left( \sum_{k=1}^t \frac{\partial h^{(t)}}{\partial h^{(k)}} \cdot \frac{\partial h^{(k)}}{\partial W_{hh}} \right) \quad (4-2)$$

Μετά τον υπολογισμό του σφάλματος τα βάρη διορθώνονται με βάση τον κανόνα Widrow-Hoff:

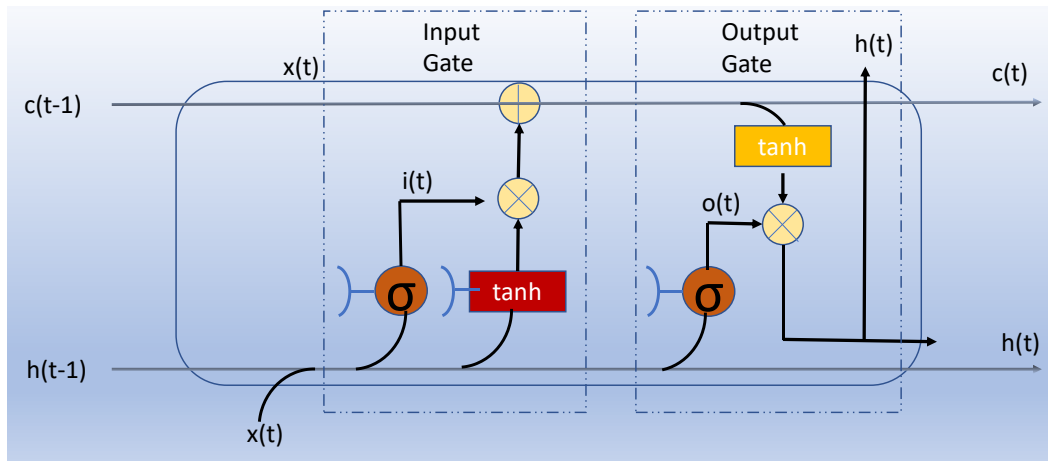
$$W^{i+1} = W^i + \eta(y_i - W^i x_i) x_i^T \quad (4-3)$$

Όπου  $\eta$  ο ρυθμός μάθησης.

Όπως και στο απλό Back Propagation έτσι κι εδώ οι διορθώσεις των βαρών επαναλαμβάνονται μέχρι να ελαχιστοποιηθεί το σφάλμα [40].

#### 4. 4 Long-Short Term Τεχνητά Νευρωνικά Δίκτυα (LSTMNN)

Τα RNN όπως περιεγράφηκαν στην παραπάνω παράγραφο, παρουσιάζουν μία σημαντική αδυναμία, καθώς δεν έχουν την δυνατότητα να συνδέσουν σχετικές πληροφορίες, όταν μεταξύ αυτών παρεμβάλλεται μεγάλο κενό. Το πρόβλημα αυτό καλύπτεται με τα LSTMNN. Προκειμένου να αντιμετωπιστεί το πρόβλημα των μακροπρόθεσμων εξαρτήσεων, επιδιώχθηκε να βελτιωθεί η ικανότητα απομνημόνευσης του τυπικού αναδρομικού νευρώνα εισάγοντας μια πύλη σε αυτό. Στο Σχήμα 21 φαίνεται η βασική αρχιτεκτονική των LSTMNN.



Σχήμα 21: Βασική αρχιτεκτονική των LSTMNN

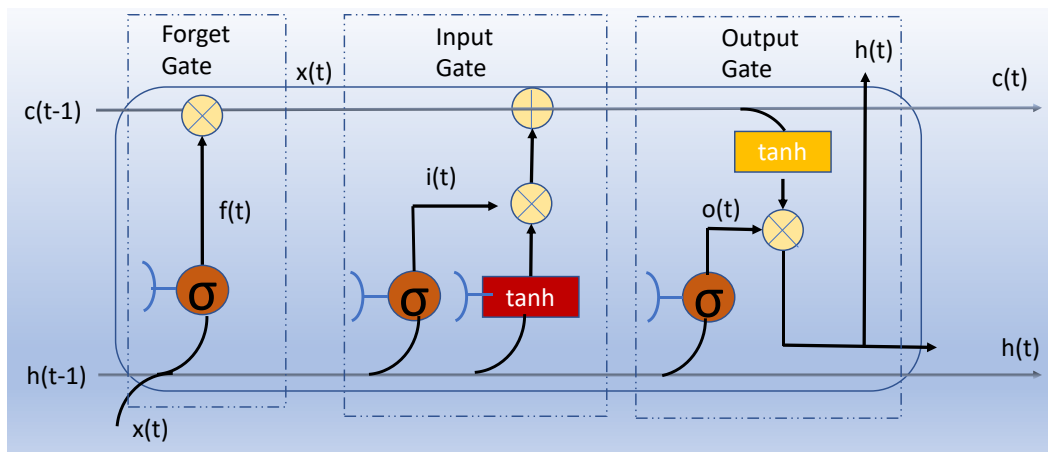
Στην αρχή, η είσοδος  $x(t)$  περνάει μέσω μιας πύλης εισόδου (input gate), η οποία αποφασίζει ποιες νέες πληροφορίες θα αποθηκευτούν στην κατάσταση του κελιού (cell state). Η πύλη εισόδου χρησιμοποιεί διάφορες συναρτήσεις ενεργοποίησης και πόλωσης για να ρυθμίσει την επίδραση της εισόδου.

Το αποτέλεσμα της πύλης εισόδου συνδυάζεται με την αναδρομική πληροφορία  $h(t)$  από το προηγούμενο χρονικό βήμα, καθώς και με την προηγούμενη κατάσταση του κελιού, και περνάει μέσω νευρώνων με διάφορες συναρτήσεις ενεργοποίησης. Οι έξοδοι αυτών των νευρώνων συνδυάζονται και περνούν μέσω μιας συνάρτησης για να παραχθεί η νέα κατάσταση του δικτύου και η αναδρομική πληροφορία.

Στη συνέχεια, η αναδρομική πληροφορία και η νέα κατάσταση του δικτύου περνάνε μέσω μιας πύλης εξόδου (output gate). Η πύλη εξόδου αποφασίζει ποιες πληροφορίες μπορούν να εξαχθούν από την κατάσταση του κελιού, βάσει της τρέχουσας κατάστασης του νευρώνα. Το αποτέλεσμα αυτό συνδυάζεται με το αποτέλεσμα της πύλης εισόδου και παράγεται η έξοδος του δικτύου για το συγκεκριμένο χρονικό βήμα.

Συνολικά, η αρχιτεκτονική των LSTM δικτύων επιτρέπει τη διατήρηση και την ανανέωση πληροφορίας από προηγούμενα χρονικά βήματα, επιτρέποντας την αντιμετώπιση μακροπρόθεσμων εξαρτήσεων σε ακολουθιακά δεδομένα. Αυτό τους καθιστά ιδιαίτερα αποτελεσματικά σε προβλήματα όπως η πρόβλεψη χρονοσειρών. Συνήθως τα LSTMNN εφαρμόζονται με παραλλαγές στα έργα που χρησιμοποιούνται. Οι κυριότερες παραλλαγές LSTM με πύλη FORGET και LSTM με πύλη PEEP-HOLE.

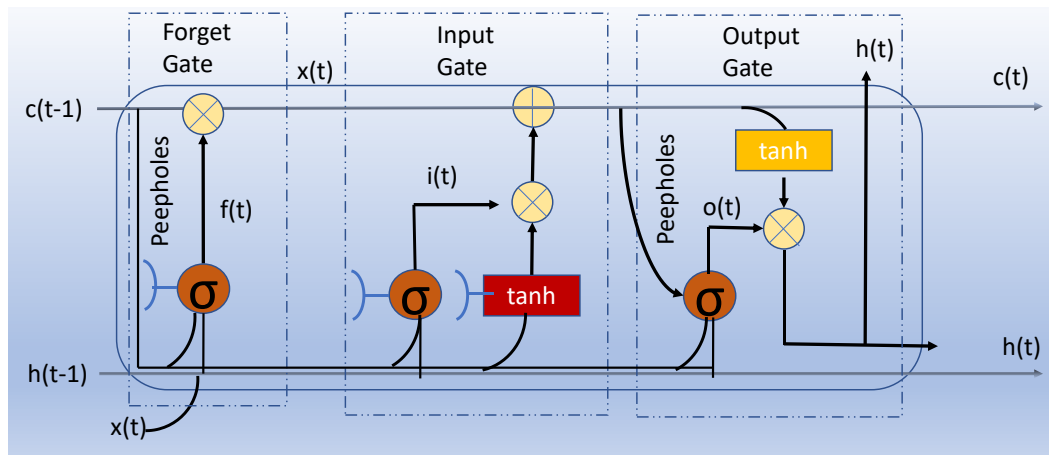
Στο Σχήμα 22 φαίνεται η τυπική δομή ενός νευρωνικού δικτύου LSTM με πύλη FORGET.



Σχήμα 22: LSTMNN με FORGET πύλη.

Η πύλη FORGET προηγείται της πύλης εισόδου. Σε αυτή υπάρχει ένας νευρώνας που δέχεται την αναδρομική πληροφορία και την είσοδο (και την πόλωση). Στην πύλη εισόδου περνάνε τα ίδια και το αποτέλεσμα του συνδυασμού της εξόδου του νευρώνα της πύλης FORGET και της κατάστασης του δικτύου της προηγούμενης κατάστασης. Η αρχιτεκτονική του δικτύου μετά την πύλη FORGET είναι όπως του LSTMN χωρίς πύλη FORGET. Η πύλη FORGET μπορεί να καθορίζει ποιες πληροφορίες θα απορριφθούν από την κατάσταση του νευρώνα. Όταν η τιμή της πύλης είναι 1, διατηρεί αυτές τις πληροφορίες ενώ αν είναι 0, απορρίπτονται. Σε μελέτες αποδείχθηκε ότι όσο αύξανε η πόλωση της πύλης FORGET, η απόδοση του δικτύου LSTM συνήθως γινόταν καλύτερη.

Καθώς οι πύλες των δικτύων LSTM που ακολουθούν το παραπάνω πρότυπο, δεν έχουν απευθείας συνδέσεις μεταξύ της προηγούμενης κατάστασης με τους νευρώνες, παρατηρείται έλλειψη βασικών πληροφοριών που επηρεάζουν αρνητικά την απόδοση του δικτύου. Η λύση στο ζήτημα αυτό είναι η ενσωμάτωση στο νευρωνικό δίκτυο των συνδέσεων PEEPHOLE. Στο Σχήμα 23 φαίνεται η αρχιτεκτονική των LSTMN που περιλαμβάνουν συνδέσεις PEEPHOLE.



Σχήμα 23: Γενική μορφή LSTMN με συνδέσεις PEEPHOLE

Η προηγούμενη κατάσταση του δικτύου μπορεί να περάσει σαν είσοδος στους νευρώνες των πυλών του δικτύου (FORGET, εισόδου και εξόδου). Αυτή είναι και η βασική διαφορά σε σχέση με τα LSTMN με FORGET πύλη. Οι συνδέσεις PEEPHOLE επιτρέπουν στο LSTMN να ελέγχει τις τρέχουσες εσωτερικές του καταστάσεις να καθίσταται ικανό να εκπαιδευτεί από σταθερούς και ακριβείς αλγόριθμους χρονισμού χωρίς εξωτερική παρέμβαση. Σε δοκιμές διαφόρων συνδυασμών των παραπάνω αναφερομένων παραλλαγών του LSTM αποδείχθηκε ότι οι πύλες εισόδου και εξόδου είναι τα πιο κρίσιμα στοιχεία και η αφαίρεση οποιουδήποτε από αυτά προφανώς μειώνει την απόδοση δικτύου. Επιπλέον, η τροποποιημένη συνδεδεμένη είσοδος και η πύλη FORGET θα μπορούσαν να μειώσουν τον αριθμό των παραμέτρων και να μειώσουν το υπολογιστικό κόστος χωρίς σημαντική μείωση στην απόδοση του δικτύου. Λόγω αυτής της ισχυρής ικανότητας, το LSTM έχει γίνει το επίκεντρο της βαθιάς μάθησης και έχει εφαρμοστεί σε πολλαπλές εργασίες [38] [41].

# ΚΕΦΑΛΑΙΟ 5

## Υπολογιστικά Εργαλεία για Διαδικασίες Μηχανικής Μάθησης

---

### 5. Υπολογιστικά Εργαλεία για Διαδικασίες Μηχανικής Μάθησης

Για την υλοποίηση διαδικασιών μάθησης είναι δυνατό να χρησιμοποιηθούν οι γλώσσες προγραμματισμού υψηλού επιπέδου. Ωστόσο η ανάπτυξη εφαρμογών με αυτόν το προσανατολισμό είναι επίπονη διαδικασία που απαιτεί λεπτομερή σχεδίαση και την εκτέλεση αρκετών δοκιμών. Για το λόγο αυτό είναι προτιμότερο να χρησιμοποιηθούν πιο εξειδικευμένες λύσεις οι οποίες παρέχουν έτοιμη λειτουργικότητα για ένα μεγάλο πλήθος διεργασιών που είναι παρούσες στις περισσότερες διαδικασίες μηχανικής μάθησης. Τέτοιου είδους λύσεις είναι η γλώσσα *python* ή η γλώσσα *R*. Πρόκειται για κάτι περισσότερο από γλώσσες προγραμματισμού. Μπορεί να θεωρηθεί ότι βρίσκονται ένα επίπεδο πάνω από τις γλώσσες προγραμματισμού υψηλού επιπέδου (*C*, *Java* κλπ). Οι γλώσσες αυτές παρέχουν έναν αριθμό από εξειδικευμένες βιβλιοθήκες προσανατολισμένες σε εξειδικευμένους σκοπούς. Στην παρούσα εφαρμογή χρησιμοποιήθηκε η γλώσσα *python* η οποία χρησιμοποιείται κατά κόρον για την ανάπτυξη μοντέλων μηχανικής μάθησης, με πολύ καλά αποτελέσματα.

#### 5.1 Η γλώσσα προγραμματισμού *Python*

Η *Python* είναι μια αντικειμενοστραφής (μπορεί ωστόσο να χρησιμοποιηθεί και στην ανάπτυξη προγραμμάτων με δομημένο προγραμματισμό) γλώσσα προγραμματισμού υψηλού επιπέδου. Το συντακτικό της δίνει την δυνατότητα να παρέχει στις εφαρμογές που χρησιμοποιείται δυναμική σημασιολογία. Τα βασικά της χαρακτηριστικά είναι οι δομές δεδομένων υψηλού επιπέδου που ενσωματώνει στον πυρήνα της και η δυνατότητα για δυναμική δημιουργία τύπων δεδομένων. Επιπλέον δύναται να δεσμεύει μνήμη δυναμικά

κάτι που την κάνει κατάλληλη για ανάπτυξη εφαρμογών μέσα σε σύντομο (σχετικά) χρονικό διάστημα. Αυτό το καταφέρνει καθώς δίνει στον προγραμματιστή να αναπτύσσει δέσμες λειτουργιών ή σεναρίων που συνδυάζουν έτοιμες λειτουργίες.

Το συντακτικό της Python δεν διαφέρει πολύ από το συντακτικό της C, της C++ ή της Java. Αυτό κάνει τους προγραμματιστές να εξοικειώνονται εύκολα με αυτό και να μπορούν γρήγορα να εντοπίζουν τις δυνατότητες της στην ανάπτυξη εφαρμογών. Καθώς η Python μπορεί εύκολα να γίνει «κτήμα» μεγάλου συνόλου προγραμματιστών, η χρήση της στην δημιουργία λογισμικού, μειώνει το κόστος του.

Το κυριότερο πλεονέκτημα της είναι η διαθεσιμότητα μεγάλου αριθμού βιβλιοθηκών έτοιμων λειτουργικών μονάδων και δομών δεδομένων για την υποστήριξη της ανάπτυξης εφαρμογών. Με τον τρόπο αυτό δίνεται η δυνατότητα να δημιουργούνται εφαρμογές που ακολουθούν την αρθρωτή αρχιτεκτονική ανάπτυξης και εκμεταλλεύονται την επαναχρησιμοποίηση ορθού και τεκμηριωμένου κώδικα. Ο μεταφραστής της Python και η εκτεταμένη βιβλιοθήκη του πυρήνα διατίθενται σε (ανοικτό) κώδικα αλλά και σε εκτελέσιμη μορφή, μέσω του διαδικτύου δωρεάν, κάτι που διευκολύνει τη διάδοση της υιοθέτησης της από την κοινότητα των προγραμματιστών.

Τα προγράμματά που αναπτύσσονται μπορούν να εκτελούνται σε κάθε λειτουργικό σύστημα χωρίς να χρειάζεται ο προγραμματιστής να μεριμνήσει να προσαρμόσει τον κώδικα με κάποιο τρόπο [44].

Η Python περιλαμβάνει κάποια χαρακτηριστικά τα οποία συμβάλλουν στην διαρκή και αυξανόμενη υιοθέτηση της από τους προγραμματιστές. Επιγραμματικά τα χαρακτηριστικά αυτά είναι [45]:

- Είναι εύκολη στην εκμάθηση του συντακτικού της και της χρήσης της: Η εκμάθηση της Python είναι ευκολότερη σε σύγκριση με άλλες γλώσσες προγραμματισμού. Ο βασικότερος λόγος για αυτό είναι η ομοιότητα της με τη C. Όντας σε επίπεδο ανώτερο από τη C, η σύνταξή της είναι απλή και ομοιάζει την αγγλική γλώσσα. Η δομή του κώδικα στα προγράμματα της γίνεται με πιο συμβολισμό που είναι προφανής – ακόμα και για ανθρώπους που ασχολούνται με τους αλγορίθμους σε θεωρητικό επίπεδο. Επιτρέπει στον προγραμματιστή να χρησιμοποιεί ένα περιορισμένο σύνολο συμβολισμών και επιπλέον δεν τον αναγκάζει να προβαίνει συχνά σε ανάπτυξη δικών του δομών δεδομένων.
- Είναι γλώσσα με υψηλή εγγενή εκφραστικότητα: Με την Python μπορούν να αναπτυχθούν προγράμματα που εκτελούν πολύπλοκες διαδικασίες με λίγες γραμμές κώδικα. Αυτό οφείλεται στο γεγονός ότι στον πυρήνα της περιλαμβάνει μεγάλο αριθμό έτοιμων λειτουργιών που χρησιμοποιούνται συχνά σε έργα διαφορετικών προσανατολισμών. Κατά συνέπεια, απαλλασσόμενος ο προγραμματιστής από μέριμνα για την ανάπτυξη βασικών λειτουργικών μονάδων, έχει περισσότερο χρόνο να ασχοληθεί με την αλγοριθμική βελτιστοποίηση των προγραμμάτων του αλλά και να επιταχύνει την εργασία του.
- Είναι ερμηνευτική γλώσσα: Οι εντολές της Python εκτελούνται γραμμή – γραμμή. Ο κώδικας εκτελείται είτε μέχρι το τέλος του είτε μέχρι να εντοπιστεί σφάλμα. Με τον τρόπο αυτό τα σφάλματα εντοπίζονται ευκολότερα και με μεγαλύτερη ακρίβεια.
- Λειτουργεί σε όλα τα δημοφιλή λειτουργικά συστήματα: Τα προγράμματα της Python τρέχουν σε όλα τα γνωστά λειτουργικά συστήματα χωρίς διαφοροποιήσεις στον κώδικα. Με τον τρόπο αυτό οι προγραμματιστές δεν είναι αναγκασμένοι να γνωρίζουν το που θα τρέξει το πρόγραμμα που αναπτύσσουν ή να δημιουργούν διαφορετικές εκδόσεις για κάθε λειτουργικό σύστημα.

- Ο πυρήνας του και οι βιβλιοθήκες του είναι ανοικτού κώδικα και διατίθεται δωρεάν: Τόσο ο πυρήνας της python όσο και οι βιβλιοθήκες της είναι ανοικτού κώδικα και δωρεάν διαθέσιμες μέσω του διαδικτύου. Αυτό δίνει την δυνατότητα στους προγραμματιστές να προσαρμόζουν τις λειτουργίες της στις απαιτήσεις των εφαρμογών που αναπτύσσουν. Ο ανοικτός κώδικας γενικότερα δίνει τη δυνατότητα στους προγραμματιστές να ενσωματώνουν νέες δυνατότητες μειώνοντας τόσο τον χρόνο όσο και το κόστος ανάπτυξης των εφαρμογών.
- Είναι (και) αντικειμενοστραφής γλώσσα: Η python επιτρέπει στους προγραμματιστές να δημιουργούν εφαρμογές και με δομημένο και με αντικειμενοστραφή προγραμματισμό. Μπορούν δηλαδή να εκμεταλλεύονται – ακόμα και στην ίδια εφαρμογή συνδυασμένα – τα πλεονεκτήματα και των δύο βασικών προσεγγίσεων του προγραμματισμού.
- Υποστήριξη γραφικών διεπαφών και διαδικτυακών εφαρμογών: Εσφαλμένα επικρατεί η άποψη ότι η python δεν μπορεί να χρησιμοποιηθεί για την ανάπτυξη εφαρμογών με γραφικές διεπαφές ή για χρήση στο διαδίκτυο. Αυτό μπορεί να γίνει με την ενσωμάτωση κατάλληλων βιβλιοθηκών. Ωστόσο η χρήση της για αυτούς τους προσανατολισμούς είναι σχετικά περιορισμένη.
- Ενσωμάτωση σε άλλες εφαρμογές: Προγράμματα της python μπορούν να κληθούν από εφαρμογές που αναπτύσσονται με άλλες γλώσσες προγραμματισμού. Αντιστρόφως μπορεί και κώδικας της python να καλεί εκτελέσιμα προγράμματα γραμμένα σε άλλη γλώσσα προγραμματισμού.
- Δυναμική κατανομή μνήμης: Η μνήμη εκχωρείται στις μεταβλητές κατά τον χρόνο εκτέλεσης. Για τον λόγο αυτό δεν είναι απαραίτητο στα προγράμματα python να καθορίζεται ο τύπος δεδομένων της μεταβλητής εξ αρχής.
- Είναι γλώσσα υψηλού επιπέδου: Η Python έχει σχεδιαστεί να είναι μια γλώσσα προγραμματισμού υψηλού επιπέδου. Ο προγραμματιστής δεν χρειάζεται να μεριμνά με θέματα όπως τη δομή κωδικοποίησης, την αρχιτεκτονική και τη διαχείριση μνήμης.

Η αρχιτεκτονική της Python περιλαμβάνει μία εικονική μηχανή Python (Python virtual machine) η οποία τοποθετείται στο ενδιάμεσο μεταξύ του ηλεκτρονικού υπολογιστή και της γλώσσας μηχανής. Αυτή η αρχιτεκτονική επιτρέπει στον κώδικα Python να εκτελείται ανεξάρτητα από το λειτουργικό σύστημα ή την αρχιτεκτονική του υπολογιστή.

Ο τρόπος που λειτουργεί η αρχιτεκτονική αυτή περιγράφεται ως εξής :

- Ο μεταφραστής διαβάζει το κώδικα python και ελέγχει αν υπάρχουν συντακτικά σφάλματα. Σε θετική περίπτωση η εκτέλεση του κώδικα διακόπτεται και εμφανίζεται κατάλληλο μήνυμα.
- Εάν δεν υπάρχει συντακτικό σφάλμα, τότε ο μεταφραστής το μεταφράζει σε δυαδικό κώδικα.
- Ο δυαδικός κώδικας αποστέλλεται στην εικονική μηχανή python και εκτελείται. Αν παρουσιαστεί σφάλμα κατά τη διάρκεια της εκτέλεσης, διακόπτεται και εμφανίζεται κατάλληλο μήνυμα σφάλματος.[42]

## 5.2 Η πλατφόρμα Tensorflow

Η βιβλιοθήκη TensorFlow είναι μια από τις πολυδάπανες βιβλιοθήκες μηχανικής μάθησης και αναπτύσσεται από την Google. Αρχικά, δημιουργήθηκε για ανάπτυξη μοντέλων μηχανικής μάθησης που αναφέρονται στην τεχνική του "deep learning" και συγκεκριμένα στα νευρωνικά δίκτυα.

Η TensorFlow παρέχει ένα ευέλικτο πλαίσιο για τη δημιουργία, την εκπαίδευση και την εκτέλεση μοντέλων μηχανικής μάθησης. Περιλαμβάνει μια ευρεία γκάμα από λειτουργίες για την ανάπτυξη των μοντέλων, συμπεριλαμβανομένων των επιπέδων νευρωνικών δικτύων, είναι ανοικτού κώδικα και διατίθεται δωρεάν. Τα βασικά της χαρακτηριστικά είναι:

- Μπορεί να εκτελέσει αποτελεσματικά λειτουργίες χαμηλού επιπέδου
- Έχει την ικανότητα να υπολογίζει την τάση αυθαίρετων εκφράσεων.
- Περιλαμβάνει λειτουργίες για την κλιμάκωση των υπολογισμών σε πολλές συσκευές.
- Παρέχει λειτουργίες για τη γραφική αναπαράσταση των αποτελεσμάτων της μηχανικής μάθησης.

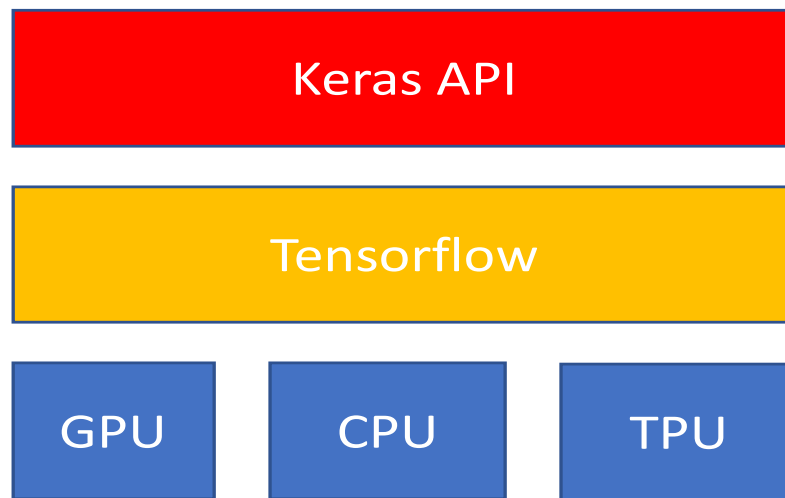
Ο βασικός λόγος για την ταχεία υιοθέτηση της TensorFlow από τους ανθρώπους που ασχολούνται με την μηχανική μάθηση, είναι ότι η δημιουργία και η ανάπτυξη εφαρμογών γίνεται εύκολα και γρήγορα με πολύ καλά αποτελέσματα. Παρέχει Application Programming Interfaces (APIs) για τις γνωστότερες γλώσσες προγραμματισμού και περιβάλλοντα ανάπτυξης (πχ python, C, C ++, Rust, Haskell, Go, Java, Android, iOS, Mac OS, Windows, Linux και Raspberry Pi). Η TensorFlow λαμβάνει υπ' όψη πιθανούς περιορισμούς σε διαθέσιμους πόρους και μπορεί να υποστηρίξει εφαρμογές σε όλους τους τύπους υπολογιστών, ανεξάρτητα από τις δυνατότητες επεξεργασίας που έχουν. Καθώς χρησιμοποιείται από εκτεταμένη κοινότητα προγραμματιστών, διατίθενται στο διαδίκτυο μεγάλος αριθμός τεκμηρίων, ικανός να λύσει κάθε πρόβλημα και απορία των προγραμματιστών [43].

## 5.3 Η βιβλιοθήκη Keras

Η βιβλιοθήκη Keras παρέχει ένα API για την εκτέλεση διαδικασιών Deep Learning. Είναι σε έναν επίπεδο πάνω την Tensorflow παρέχοντας βασικά λειτουργικά στοιχεία δομές δεδομένων για την δημιουργία εφαρμογών μηχανικής μάθησης. Με την Keras οι μηχανικοί και οι ερευνητές να δύνανται να εκμεταλλεύονται τις δυνατότητες της TensorFlow στην κλίμακα που απαιτεί το κάθε έργο. Παρέχει υψηλό επίπεδο αφαίρεσης σε θέματα μηχανικής μάθησης, δίνοντας στον προγραμματιστή την δυνατότητα να αναπτύσσει εξειδικευμένες εφαρμογές χωρίς να είναι γνώστης λεπτομερειών των διαφορετικών αλγορίθμων μηχανικής μάθησης. Αυτό είναι ένα πλεονέκτημα που συμβάλλει στην υιοθέτηση των πρακτικών της μηχανικής μάθησης για την λήψη αποφάσεων σε οργανισμούς διαφορετικής κλίμακας. Υποστηρίζει όλα τα δημοφιλή μοντέλα νευρωνικών δικτύων και συνεχώς η ομάδα ανάπτυξης ενσωματώνει και νέες δυνατότητες. Ο εντοπισμός σφαλμάτων γίνεται εύκολα καθώς ο



κώδικας της είναι γραμμένος σε python. Το που τοποθετείται η βιβλιοθήκη σε σχέση με την Tensorflow φαίνεται στο Σχήμα 24.



Σχήμα 24: Σχέση των βιβλιοθηκών keras - tensorflow

Η βιβλιοθήκη τεκμηριώνεται και υποστηρίζεται στο διαδίκτυο σε βαθμό τέτοιο που να επιτρέπει στους προγραμματιστές να αναζητούν και να εντοπίζουν λύσεις σε κάθε πρόβλημα που μπορεί να αντιμετωπίσουν. Υπάρχει επίσης μεγάλος αριθμός και ποικιλία διαδικτυακών τεκμηρίων για τον τρόπο ανάπτυξης διαφορετικών μοντέλων μηχανικής μάθησης [44].

### 6. Case study

#### 6.1 Περιγραφή

Στην παρούσα μελέτη επιχειρείται να προσεγγιστούν οι συσχετίσεις που επηρεάζουν την διασπορά του ιού COVID-19 και η δημιουργία μοντέλου πρόβλεψης της διασποράς σε μελλοντικό χρόνο, μέσα από την επεξεργασία δεδομένων από νευρωνικά δίκτυα αρχιτεκτονικής RNN. Το μέτρο της διασποράς επιλέχθηκε να είναι ο αριθμός των θανάτων που καταγράφονται εντός 24ώρου στην Ελλάδα.

#### 6.2 Χαρακτηριστικά της διασποράς του Covid-19

Η Covid-19 είναι μια μολυσματική ασθένεια, η νόσος αντίστοιχου κορονοϊού, που προκαλείται από τον ιό Corona SARS-CoV-2. Εντοπίστηκε για πρώτη φορά στις 17 Νοεμβρίου 2019 στην πόλη Wuhan στην κεντρική Κίνα και από τότε ο ιός έχει εξαπλωθεί σε όλο τον κόσμο. Συνολικά, περίπου 200 χώρες επλήγησαν από αυτή την παγκόσμια κρίση υγείας, με περισσότερα από 20.000 επιβεβαιωμένα κρούσματα στις ΗΠΑ, Ισπανία, Ιταλία, Γερμανία, Κίνα, Γαλλία, Ιράν, Ηνωμένο Βασίλειο, Τουρκία και Ελβετία. Σε αυτές τις χώρες, το ποσοστό θανάτων (αριθμός θανάτων/αριθμός επιβεβαιωμένων κρουσμάτων) κυμαίνεται από 12,5% (Ηνωμένο Βασίλειο) ε έως 1,5% (Γερμανία). Από αυτά τα ενδεικτικά στατιστικά στοιχεία, διαφαίνεται η ικανότητα του ιού να διασπείρεται με γρήγορους ρυθμούς. Η κατάσταση που διαμορφώθηκε σε όλο τον κόσμο, ανάγκασε τις κυβερνήσεις να λάβουν αποφασιστικά μέτρα για την αναστολή της διασποράς τους. Σε αυτό το πλαίσιο και σε παγκόσμια κλίμακα αποφασίστηκαν:

- Περιορισμοί στην αναγκαιότητα μετάβασης στους χώρους εργασίας και εκπαίδευσης.
- Απαγορεύσεις κυκλοφορίας σε ολόκληρες περιοχές για μεγάλο μέρος ή και για ολόκληρη τη μέρα.

- Περιορισμοί και απαγορεύσεις στις μετακινήσεις, όχι μόνο από χώρα σε χώρα αλλά ακόμα και εντός της ίδιας χώρας.
- Καθορίστηκαν πρωτόκολλα συμπεριφοράς τόσο για ανθρώπους που παρουσίαζαν συμπτώματα της νόσου ή είχε επιβεβαιωθεί η προσβολή τους από τον ιό, όσο και για απολύτως υγιείς ανθρώπους [10].

Τα παραπάνω μέτρα είναι τα κυριότερα που επιβλήθηκαν από τον ξέσπασμα της πανδημίας και είναι ενδεικτικά της αναταραχής που επέφερε ο Covid-19 στις κοινωνίες σε όλον τον πλανήτη. Είναι σημαντικό να υπάρχουν μέθοδοι έγκαιρης προειδοποίησης μέσω των οποίων να μπορεί να προβλεφθεί πόσο θα επηρεάσει η ασθένεια την κοινωνία, ώστε να εκτιμώνται τα μέτρα που θα πρέπει κάθε φορά να εφαρμόζονται. Είναι προφανές ότι μεταξύ άλλων, ένα εργαλείο που θα μπορούσε να εκτιμά με σχετική αξιοπιστία τον αριθμό των κρουσμάτων ασθενών με Covid-19 σε μελλοντικούς χρόνους – βάση αντικειμενικών παραμέτρων – θα μπορούσε να βοηθήσει τις κυβερνήσεις στις διαδικασίες λήψης αποφάσεων για τα μέτρα που είναι κατάλληλα να εφαρμόζονται ανά περίπτωση.

Είναι αδιαμφισβήτητο γεγονός ότι η γνώση περί του ιού είναι ακόμα περιορισμένη – αν και έχει ήδη γίνει αρκετή πρόοδος από τις αρχές του 2020 μέχρι σήμερα. Υπάρχουν σημεία στα οποία έχουν εδραιωθεί σημαντικά συμπεράσματα για τα χαρακτηριστικά της νόσου και τον τρόπο αντιμετώπισης της ενώ υπάρχουν ακόμα και σήμερα έντονες διαφωνίες (πχ χρησιμότητα και αποτελεσματικότητα των εμβολιασμών, σκοπιμότητα και αποτελεσματικότητα των περιοριστικών μέτρων). Η απουσία κοινά αποδεκτών στοιχείων που να συνθέτουν το προφίλ τόσο της ίδιας της νόσου όσο και της αντιμετώπισης της, επηρεάζει αρνητικά και την έρευνα για την ανάπτυξη εργαλείων πρόβλεψης. Η δυσκολία δεν έγκειται στον τεχνικό τομέα αλλά κυρίως στην ουσία της αντιμετώπισης καθώς είναι δύσκολο να εντοπιστούν με υψηλό βαθμό αντικειμενικότητας οι παράμετροι που θα πρέπει να ελέγχονται και να αξιολογούνται για την ανάπτυξη εργαλείων πρόβλεψης.

### 6.3 Σύνολα Δεδομένων

Τα δεδομένα που χρησιμοποιήθηκαν για την συγκεκριμένη εργασία είναι οι ημερήσιοι θάνατοι που προκλήθηκαν από Covid-19 στην Ελλάδα από την εμφάνιση των πρώτων κρουσμάτων στην χώρα μας (20 Μαρτίου του 2020) έως και 22 Απριλίου του 2022.

Τα δεδομένα αντλήθηκαν από την σελίδα [ourworldindata.org](http://ourworldindata.org) [45]. Η συγκεκριμένη σελίδα δημοσιεύει έρευνες και δεδομένα ώστε να σημειωθεί πρόοδος ενάντια στα μεγαλύτερα προβλήματα του κόσμου όπως είναι η φτώχεια, οι ασθένειες, η πείνα, η κλιματική αλλαγή, ο πόλεμος, και η ανισότητα.

Τα δεδομένα αποθηκεύτηκαν απ'την σελίδα σε μορφή csv και στην συνέχεια έγινε επεξεργασία μέσω ενός script της python για να απομονωθούν οι θάνατοι από τα υπόλοιπα δεδομένα και να μπορέσει να γίνει μοντελοποίηση.

### 6.4 Υλοποίηση

Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο η υλοποίηση του πειραματικού μέρους έγινε με χρήση της γλώσσας python. Η python έχει αναπτύξει βιβλιοθήκες για την εκπαίδευση νευρωνικών δικτύων όπως το keras, ωστόσο για την βαθύτερη κατανόηση της αρχιτεκτονικής των αναδρομικών νευρωνικών δικτύων έγινε κωδικοποίηση του αλγορίθμου

από το μηδέν. Παρ' όλ' αυτά τρέχοντας τα δεδομένα και μέσω της βιβλιοθήκης keras δόθηκε η δυνατότητα σύγκρισης των αποτελεσμάτων μας και εξαγωγή ασφαλών συμπερασμάτων ως προς την αποτελεσματικότητα του κώδικα.

Όπως αναφέρθηκε πιο πριν τα δεδομένα είναι οι θάνατοι από Covid-19. Για την εκπαίδευση του αλγορίθμου RNN τέθηκαν είσοδοι και έξοδοι ώστε να γίνει η εκπαίδευση με επίβλεψη. Ως είσοδος (x) χρησιμοποιήθηκαν τα δεδομένα πέντε ημερών και ως έξοδος (y) τα δεδομένα για την έκτη μέρα.

Για να μπορέσει να γίνει αποτελεσματικά η εκπαίδευση του δικτύου το σύνολο δεδομένων χωρίστηκε σε τρία υποσύνολα.

- Σύνολο εκπαίδευσης
- Σύνολο αξιολόγησης
- Σύνολο ελέγχου

Από το σύνολο των δεδομένων (763 ημέρες) το 50% (383 ημέρες-Από 20 Μαρτίου του 2020 έως 7 Απριλίου του 2020) χρησιμοποιήθηκε για την εκπαίδευση του δικτύου, το 25% (190 ημέρες-Από 7 Απριλίου του 2021 έως 14 Οκτωβρίου του 2021) για την αξιολόγηση και το υπόλοιπο 25% (191 ημέρες-Από 14 Οκτωβρίου του 2021 έως 22 Απριλίου του 2022) για τον έλεγχο.

Στη συνέχεια, η συνάρτηση εκπαιδεύει το RNN χρησιμοποιώντας stochastic gradient descent (SGD) για τον καθορισμένο αριθμό εποχών, ενώ παρακολουθεί το σφάλμα στα σύνολα εκπαίδευσης και αξιολόγησης. Αξιολογεί επίσης την απόδοση του εκπαιδευμένου μοντέλου στο δοκιμαστικό σύνολο.

Η διαδικασία εκπαίδευσης περιλαμβάνει, τον υπολογισμό των κλίσεων του σφάλματος σε σχέση με τις παραμέτρους του μοντέλου χρησιμοποιώντας το BPTT και την ενημέρωση των παραμέτρων του μοντέλου χρησιμοποιώντας το βήμα stochastic gradient descent. Ο ρυθμός μάθησης συνήθως μειώνεται με την πάροδο του χρόνου για να βοηθήσει το μοντέλο να συγκλίνει. Το σύνολο αξιολόγησης χρησιμοποιείται για την παρακολούθηση του σφάλματος κατά τη διάρκεια της εκπαίδευσης και για την αποφυγή υπερβολικής προσαρμογής στο σύνολο εκπαίδευσης. Το σύνολο ελέγχου χρησιμοποιείται για την αξιολόγηση της απόδοσης του τελικού εκπαιδευμένου μοντέλου.

Για να μπορέσει να γίνει η εκπαίδευση σωστά έγινε κανονικοποίηση των δεδομένων. Με την βοήθεια της βιβλιοθήκης sklearn τα δεδομένα κανονικοποιήθηκαν από -1 έως 1.

Η εκπαίδευση του μοντέλου επαναλήφθηκε για 200 epochs με τη χρήση του SGD με σταθερό ρυθμό εκμάθησης 0,05 ωστόσο μέσω μιας συνάρτησης η εκπαίδευση τερματίζεται αν το σφάλμα δεν παρουσιάζει μείωση 10 epochs μετά από την ελάχιστη τιμή του.

Τα βάρη (U, W, V) αρχικοποιήθηκαν με τυχαίες τιμές. Όπου U πίνακας βαρών που πολλαπλασιάζονται με την είσοδο, W πίνακας βαρών όπου πολλαπλασιάζονται με τις recurrent στοιβάδες και V πίνακας βαρών όπου πολλαπλασιάζονται με την έξοδο.

Για την καταγραφή των σφαλμάτων του δικτύου έγινε χρήση της συνάρτησης μέσου απόλυτου σφάλματος (mean absolute error):

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (6-1)$$

Για την αποτελεσματικότητα του μοντέλου σε σχέση με τα δεδομένα έγινε χρήση του Συντελεστή προσδιορισμού  $R^2$

$$R^2 = 1 - \frac{\sum (y_j - \hat{y}_j)^2}{\sum (y_j - \bar{y})^2} \quad (6-2)$$

Οι λειτουργικές παράμετροι για τον Αλγόριθμο RNN που εφαρμόζονται σε όλα πειράματα παρουσιάζονται στον Πίνακα(I).

Πίνακας(I): Λειτουργικές παράμετροι RNN

Παράμετροι αλγορίθμου	Τιμές παραμέτρων
Hidden Units	30,50,100, 150,250
Learning rate	0.05
BPPT	5
Activation function	Tanh
Epochs	200

Ο ρυθμός εκμάθησης είναι μια διαμορφώσιμη παράμετρος που χρησιμοποιείται στην εκπαίδευση νευρωνικών δικτύων και έχει μικρή θετική τιμή, συχνά στην περιοχή μεταξύ 0,0 και 1,0.

Ο ρυθμός εκμάθησης ελέγχει πόσο γρήγορα το μοντέλο προσαρμόζεται στο πρόβλημα. Χρησιμοποιώντας μικρό ρυθμό εκμάθησης το μοντέλο απαιτεί πιο πολλά epochs για να εκπαιδευτεί ενώ τα μεγαλύτερα ποσοστά μάθησης οδηγούν σε γρήγορες αλλαγές και απαιτούν λιγότερες εποχές εκπαίδευσης. Η επιλογή του αριθμού της παραμέτρου επιλέχθηκε βάσει των αποτελεσμάτων της πειραματικής διαδικασίας [46].

Ο καθορισμός του αριθμού των νευρώνων στα κρυφά επίπεδα είναι ένα πολύ σημαντικό μέρος της απόφασης της συνολικής αρχιτεκτονικής του νευρωνικού δικτύου. Τόσο ο αριθμός των κρυφών επιπέδων όσο και ο αριθμός των νευρώνων σε καθένα από αυτά τα κρυφά στρώματα πρέπει να ληφθούν προσεκτικά υπόψη.

Η χρήση πολύ λίγων νευρώνων στα κρυφά στρώματα θα έχει ως αποτέλεσμα κάτι που ονομάζεται υποπροσαρμογή (underfitting). Η υποπροσαρμογή συμβαίνει όταν υπάρχουν πολύ λίγοι νευρώνες στα κρυφά στρώματα για να ανιχνεύσουν επαρκώς τα σήματα σε ένα περίπλοκο σύνολο δεδομένων.

Η χρήση πάρα πολλών νευρώνων στα κρυφά στρώματα από την άλλη μπορεί να οδηγήσει σε πολλά προβλήματα. Πρώτον, πάρα πολλοί νευρώνες στα κρυφά στρώματα μπορεί να οδηγήσουν σε overfitting [47]. Αυτό συμβαίνει όταν το νευρωνικό δίκτυο έχει τόση ικανότητα επεξεργασίας πληροφοριών που η περιορισμένη ποσότητα πληροφοριών που περιέχεται στο σετ εκπαίδευσης δεν είναι αρκετή για να εκπαιδεύσει όλους τους νευρώνες στα κρυφά επίπεδα. Ένα δεύτερο πρόβλημα μπορεί να παρουσιαστεί ακόμα και όταν τα δεδομένα εκπαίδευσης είναι επαρκή. Ένας υπερβολικά μεγάλος αριθμός νευρώνων στα κρυφά στρώματα μπορεί να αυξήσει τον χρόνο που απαιτείται για την εκπαίδευση του δικτύου. Ο χρόνος εκπαίδευσης μπορεί να αυξηθεί σε σημείο που να είναι αδύνατη η επαρκής εκπαίδευση του νευρωνικού δικτύου. Η “χρυσή τομή” μεταξύ πολλών και λίγων νευρώνων στα κρυφά στρώματα ανιχνεύεται μέσα από την πειραματική διαδικασία.

Όπως αναφέρθηκε και νωρίτερα τα RNN χρησιμοποιούν έναν αλγόριθμο backpropagation για εκπαίδευση, αλλά εφαρμόζεται για κάθε timestamp. Είναι κοινός γνωστό ως Back-propagation Through Time (BTT).

Υπάρχουν ορισμένα ζητήματα με την Back-propagation όπως:

Vanishing Gradient:

Όταν η ενημέρωση των βαρών είναι πολύ μικρή, ο χρόνος εκπαίδευσης διαρκεί πολύ περισσότερο και στη χειρότερη περίπτωση, αυτό μπορεί να σταματήσει εντελώς την εκπαίδευση του νευρωνικού δικτύου.

### Exploding Gradient:

Το ακριβώς αντίθετο από το ζήτημα του vanishing gradient. Η ενημέρωση των βαρών δημιουργεί μεγάλες αλλαγές οι οποίες αυξάνονται εκθετικά κατά την εκπαίδευση και δεν συγκλίνουν ποτέ. Για να ξεπεράσουμε το vanishing gradient problem, χρειαζόμαστε μια συνάρτηση της οποίας η δεύτερη παράγωγος να μπορεί να διατηρηθεί για μεγάλο εύρος πριν πάει στο μηδέν. Η συνάρτηση ενεργοποίησης tanh είναι μια καλή συνάρτηση με την παραπάνω ιδιότητα γι' αυτό και επιλέχθηκε έναντι άλλων συναρτήσεων ενεργοποίησης.[48][43]

Επίσης, ένα από τα κύρια προβλήματα του BPTT είναι το υψηλό κόστος της ενημέρωσης μίας παραμέτρου, που καθιστά αδύνατη τη χρήση μεγάλου αριθμού επαναλήψεων. Γι'αυτό τον λόγο υπάρχει το TBPTT (Truncated Backpropagation Through Time) όπου μέσω της μείωσης των βημάτων κατά την εκπαίδευση μειώνει το κόστος και τον χρόνο που απαιτείται.

Η επιλογή της καταλληλότερης τιμής έγινε και πάλι βάση των βέλτιστων αποτελεσμάτων κατά την εκπαίδευση του μοντέλου.

Για να μπορέσουμε να βγάλουμε ασφαλέστερα συμπεράσματα για το κάθε μοντέλο, οι αλγόριθμοι των RNN δικτύων δοκιμάστηκαν 20 φορές.

Όσον αφορά τα δίκτυα RBF, υπάρχουν δύο σεν παραμέτρων: τα συναπτικά βάρη που ενώνουν το κρυφό επίπεδο με το επίπεδο εξόδου, και τα κέντρα των κόμβων RBF. Στην συγκεκριμένη εργασία έχει χρησιμοποιηθεί ο αλγόριθμος fuzzy means ο οποίος βρίσκει τον αριθμό και τη θέση των κόμβων RBF σε μία αυτοματοποιημένη, μη - επαναληπτική διαδικασία χωρίς καμία τυχαιότητα. Τα βάρη, αφού βρεθούν τα κέντρα, υπολογίζονται από το πρόβλημα ελαχίστων τετραγώνων. Μάλιστα αυτά τα βάρη είναι τα βέλτιστα για τα δεδομένα κέντρα, αφού προέρχονται από τη λύση του προβλήματος ελαχίστων τετραγώνων. Όσες φορές και να τρέξει το δίκτυο, θα βγάλει τα ίδια αποτελέσματα, μιας και δεν υπάρχει καμία τυχαιότητα γι' αυτό τον λόγο δεν έχει επαναληφθεί η διαδικασία όπως στα RNN.[46]

## 6.5 Αποτελέσματα

Σ' αυτό το κεφάλαιο παρουσιάζονται στους πίνακες (II, III) ο μέσος όρος των αποτελεσμάτων με διαφορετικές τιμές των hidden units για τα RNN μοντέλα καθώς και οι μέσοι όροι των χρόνων που χρειάστηκε σε κάθε "τρέξιμο".

Στη συνέχεια παρουσιάζονται τα αποτελέσματα και των τριών αρχιτεκτονικών ώστε να συγκριθούν και να γίνει επιλογή του αποδοτικότερου μοντέλου.

Πίνακας II: Μέσοι όροι αποτελεσμάτων του αλγορίθμου RNN

(MAE,  $R^2$ ) καθώς και ο υπολογιστικός χρόνος.

Hidden units	MAE			$R^2$			Time (s)
	Training	validation	testing	Training	Validation	Testing	
30	6.389	5.366	6.390	0.922	0.939	0.918	12.69
50	6.130	5.278	6.453	0.926	0.941	0.919	13.21
100	5.826	5.175	6.397	0.931	0.942	0.920	27.58

150	5.486	5.111	6.295	0.936	0.943	0.922	46.26
250	5.485	5.124	6.319	0.937	0.943	0.921	71.73

Πίνακας III: Μέσοι όροι αποτελεσμάτων (MAE,  $R^2$ ) του αλγορίθμου RNN με την χρήση της βιβλιοθήκης keras καθώς και ο υπολογιστικός χρόνος.

Hidden units	MAE			$R^2$			Time (s)
	Training	validation	testing	Training	Validation	Testing	
30	5.487	6.660	7.596	0.925	0.903	0.888	45.2
50	6.016	6.767	7.066	0.916	0.902	0.902	48.6
100	5.820	6.145	6.711	0.918	0.919	0.907	60.6
150	5.674	6.010	6.549	0.921	0.921	0.921	63.2
250	5.597	5.863	6.460	0.932	0.925	0.912	78.0

Όπως φαίνεται στους παραπάνω πίνακες αυξάνοντας τον αριθμό των κρυφών νευρώνων αυξάνεται και ο χρόνος που απαιτείται για την κάθε εκπαίδευση το οποίο είναι αναμενόμενο αφού αυξάνεται η υπολογιστική διαδικασία.

Παρατηρούμε ότι με την χρήση μικρού αριθμού νευρώνων το μοντέλο δεν εκπαιδεύεται καλά διότι οι νευρώνες στα κρυφά στρώματα δεν ήταν αρκετοί για να ανιχνεύσουν επαρκώς τα σήματα σε ένα περίπλοκο σύνολο δεδομένων.

Ο αλγόριθμος και στις δύο περιπτώσεις αποδίδει καλύτερα όταν οι παράμετροι hidden units έχουν τιμή 150.

Αυξάνοντας την παράμετρο κι άλλο το μοντέλο άρχισε να μην αποδίδει καλά με τα δεδομένα ελέγχου διότι υπερεκπαιδεύτηκε και έτσι απομνημόνευσε πληροφορίες οι οποίες δεν ήταν επιθυμητές.

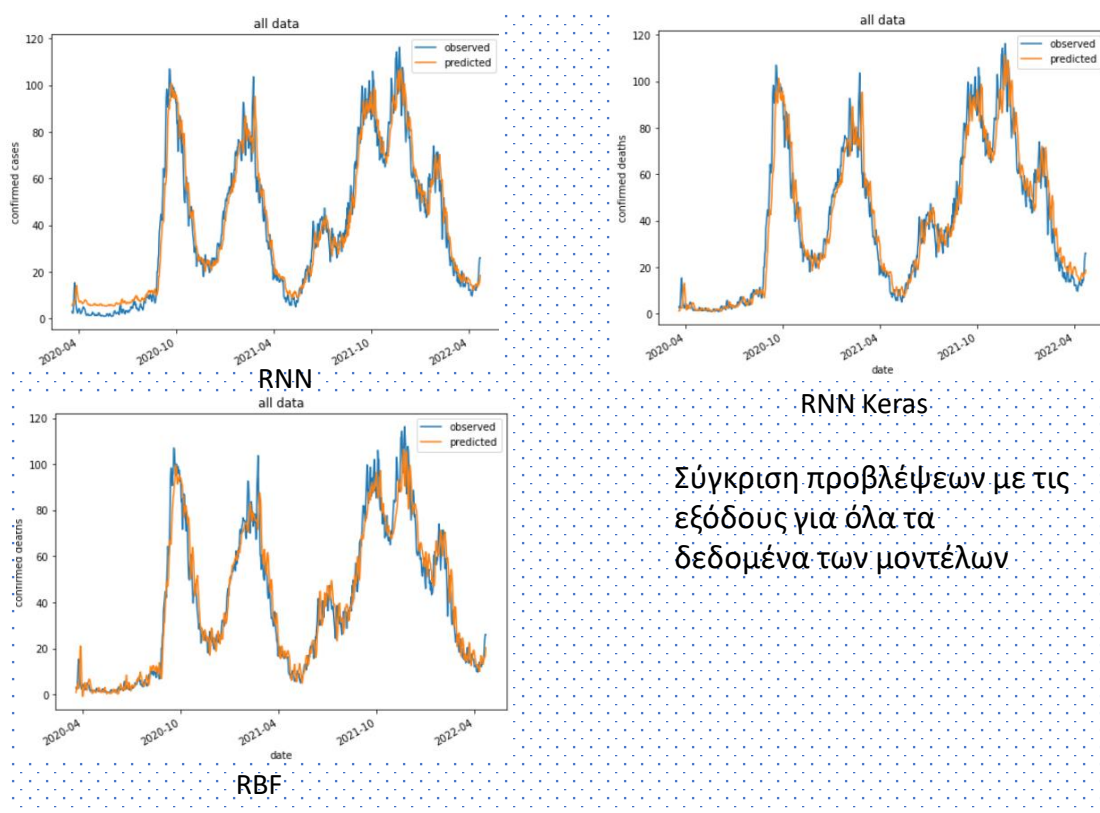
Το MAE μετράει το σφάλμα της επιθυμητής τιμής σε σχέση με την πραγματική. Κατά συνέπεια μια μικρή τιμή MAE υποδεικνύει ότι η έξοδος του μοντέλου έχει μικρή απόκλιση με τα πραγματικά αποτελέσματα και η πρόβλεψη είναι κοντά στην επιθυμητή έξοδο. Στο τέλος επιλέγεται το μοντέλο με το μικρότερο MAE στο σύνολο αξιολόγησης.

Στον Πίνακα IV κάνουμε σύγκριση των αποτελεσμάτων των τριών μοντέλων συγκρίνοντας τα βέλτιστα τρεξίματα του κάθε μοντέλου καθώς και τους μέσους όρους.

Πίνακας IV: Αποτελέσματα αλγορίθμων και των τριών μοντέλων (MAE,  $R^2$ ) καθώς και χρόνος υπολογισμών

Model	MAE			$R^2$			Time (s)
	Training	Validation	Testing	Training	Validation	Testing	

RNN <sup>5</sup>	5.300 (5.486 ± 0.204)	5.042 (5.111 ± 0.036)	6.272 (6.295 (± 0.135)	0.939 (0.936 ± 0,001)	0.945 (0.943 ± 0.001)	0.925 ( 0.922 ± 0.002)	62,4 (46.26 )
RNN (Keras)	5.596 (5.674 ± 0.091)	5.952 (6.010 ± 0.193)	6.422 (6.549 ± 0.069)	0.924 (0.921 ± 0.001)	0.922 (0.921 ± 0.001)	0.922 (0.921 ± 0.004)	75.1 (63.2)
RBF	3.926	6.020	7.138	0.956	0.900	0.888	27.58



Σχήμα 25: Σύγκριση αποτελεσμάτων των τριών μοντέλων σε σχέση τις πραγματικές εξόδους

<sup>5</sup> Οι αλγόριθμοι RNN “έτρεξαν” 20 φορές. Κατά σειρά παρουσιάζονται τα καλύτερα αποτελέσματα (βάση του μικρότερου MAE στο validation set) και στην συνέχεια στην παρένθεση οι μέσοι όροι των 20 τρεξιμάτων καθώς και η τυπική απόκλιση.



### Σύγκριση αποτελεσμάτων με τις τρεις μεθόδους:

Στην παραπάνω σχηματική απεικόνιση των αποτελεσμάτων παρατηρήθηκαν διαφοροποιήσεις μεταξύ των μεθόδων.

Πιο συγκεκριμένα:

Βλέπουμε ότι το μοντέλο RBF κάνει πολύ καλό fit με τα δεδομένα εκπαίδευσης τόσο σχηματικά όσο και βάση αποτελεσμάτων, παρ' όλ' αυτά με τα δεδομένα ελέγχου η συμπεριφορά του μοντέλου δεν δείχνει το ίδιο αποδοτική. Αυτό το φαινόμενο λέγεται overfitting και σημαίνει ότι το μοντέλο "μαθαίνει" το σύνολο εκπαίδευσης τόσο καλά που είναι "απομνημονευμένα" χαρακτηριστικά που δεν λειτουργούν τόσο καλά στο σύνολο ελέγχου. Δηλαδή, το μοντέλο δεν παρουσιάζει καλή γενίκευση.

Στην ιδανική περίπτωση, όταν επιλέγουμε ένα μοντέλο, θέλουμε το μοντέλο μας 1) να μαθαίνει κατά τη διάρκεια της εκπαίδευσης και 2) να "γενικεύει" κατά το διάστημα της αξιολόγησης (σύνολο αξιολόγησης) ώστε να αποφευχθεί το φαινόμενο overfitting.

Επίσης, παρατηρούμε μια καθυστέρηση μεταξύ των εξόδων και των προβλέψεων στη γραφική παράσταση του RBF μοντέλου περισσότερο και λιγότερο στις άλλες δύο μεθόδους. Αυτό σημαίνει ότι το μοντέλο δεν είναι σε θέση να κατανοήσει επαρκώς τα μοτίβα στα δεδομένα και δεν είναι σε θέση να κάνει ακριβείς προβλέψεις για μελλοντικά χρονικά σημεία. Αυτό μπορεί να είναι μία ένδειξη της μειωμένης ικανότητας του μοντέλου RBF στην εκπαίδευση τέτοιας μορφής δεδομένων.

Κάτι άλλο που αποτυπώνεται στην γραφική παράσταση είναι ότι οι προβλέψεις υποτιμούν σταθερά τις πραγματικές τιμές, αυτό θα μπορούσε να σημαίνει ότι το μοντέλο δεν καταγράφει όλα τα σημαντικά χαρακτηριστικά ή μοτίβα στα δεδομένα ή ότι η πολυπλοκότητα του μοντέλου είναι πολύ χαμηλή για να μοντελοποιήσει με ακρίβεια τα δεδομένα. Αυτό συμβαίνει διότι το RBF δεν είναι σε θέση να μάθει πολύπλοκες μη γραμμικές σχέσεις μεταξύ των χαρακτηριστικών εισόδου και της μεταβλητής εξόδου.

Είναι σαφές ότι η συμπεριφορά του εκπαιδευμένου RNN μοντέλου έχει καλύτερα αποτελέσματα με άγνωστα δεδομένα συγκριτικά με το RBF.

Το προτεινόμενο μοντέλο RNN πέτυχε τελικό MAE 5.3 στη φάση εκπαίδευσης 5.04 στη φάση αξιολόγησης και 6.27 στη φάση ελέγχου.

Το μοντέλο RNN με χρήση της βιβλιοθήκης Keras πέτυχε τελικό MAE 5.59 στη φάση εκπαίδευσης 5.95 στη φάση αξιολόγησης και 6.42 στη φάση ελέγχου.

Αντίθετα, το μοντέλο RBF πέτυχε καλύτερο MAE 3.926 στη φάση εκπαίδευσης αλλά χειρότερο στη φάση αξιολόγησης 6.02 και ελέγχου 7.13.

Οι χρόνοι που χρειάστηκε το κάθε μοντέλο δείχνουν ότι το RBF δίκτυο είναι πιο γρήγορο κάτι που είναι αναμενόμενο αν αναλογιστούμε την πολυπλοκότητα της αρχιτεκτονικής του RNN δικτύου.

Τέλος, τα συγκριτικά αποτελέσματα του MAE για την τρία μοντέλα δείχνουν ότι το μοντέλο RNN πέτυχε το καλύτερο αποτέλεσμα 5.11.

Τα συγκεκριμένα αποτελέσματα έρχονται να επιβεβαιώσουν την θεωρία ότι στα RNN, τα σήματα που διέρχονται από ανατροφοδοτούμενες συνδέσεις αποτελούν μια αποτελεσματική μνήμη για το δίκτυο, το οποίο μπορεί στη συνέχεια να χρησιμοποιήσει τις πληροφορίες στη μνήμη για να προβλέψει καλύτερα τις μελλοντικές τιμές χρονοσειρών. Αντίθετα τα RBF δεν ανταποκρίθηκαν το ίδιο καλά για τις προβλέψεις των συγκεκριμένων χρονοσειρών. Τα RBF αντιμετωπίζουν κάθε είσοδο ανεξάρτητα και εφαρμόζουν συναρτήσεις ακτινικού τύπου για να χαρτογραφήσουν την είσοδο σε κρυφές μονάδες Αυτό μπορεί να περιορίσει την ικανότητά τους να καταγράφουν τα διαδοχικά μοτίβα και τη δυναμική που υπάρχουν στα δεδομένα χρονοσειρών.

Συγκρίνοντας τις τυπικές αποκλίσεις μεταξύ των RNN μοντέλων, οι τιμές τους βρίσκονται σε αρκετά καλά επίπεδα οπότε οι αλγόριθμοι είναι αξιόπιστοι ως προς τη

σταθερότητα των αποτελεσμάτων τους. Για τα RBF δεν παρήχθησαν διαφορετικά αποτελέσματα γι' αυτό και δεν υπάρχουν δεδομένα τυπικής απόκλισης (βλ. ενότητα 6.4)

# ΚΕΦΑΛΑΙΟ 7

## Συμπεράσματα

---

Στην διπλωματική εργασία μελετήθηκαν αλγόριθμοι νευρωνικών δικτύων που ανήκουν στο ευρύτερο πλαίσιο της Μηχανικής Μάθησης. Στην συνέχεια έγινε εκτενέστερη ανάλυση των αναδρομικών νευρωνικών δικτύων. Η συγκεκριμένη αρχιτεκτονική χρησιμοποιήθηκε για την ανάπτυξη του αλγορίθμου προβλέψεων θανάτων από Covid-19.

Η πολυπλοκότητα και ο αυξημένος θόρυβος των συγκεκριμένων δεδομένων μας οδήγησε σε πολλές δοκιμές και παραμετροποιήσεις του κώδικα ώστε να επιτευχθούν τα βέλτιστα αποτελέσματα. Η αποδοτικότερη μοντελοποίηση των αναδρομικών νευρωνικών δικτύων για δεδομένα χρονοσειρών σε σχέση με άλλες αρχιτεκτονικές επαληθεύτηκε κατά την σύγκριση με τα αποτελέσματα του αλγορίθμου RBF.

Ενώ κατά την εκπαίδευση τα RBF δείχνουν να ενσωματώνονται αρκετά καλά, όταν καλούνται να κάνουν προβλέψεις σε άγνωστα δεδομένα τα αποτελέσματα είναι πολύ χειρότερα από των RNN.

Ως γενικό συμπέρασμα μπορούμε να καταλήξουμε ότι για τα συγκεκριμένα δεδομένα είναι αρκετά δύσκολο να πάρουμε ασφαλείς προβλέψεις λόγω των πολλών παραμέτρων που επηρεάζουν το φαινόμενο και δεν μπορούμε να γνωρίζουμε, ωστόσο αν έπρεπε να χρησιμοποιήσουμε ένα εργαλείο πρόβλεψης τα αναδρομικά νευρωνικά δίκτυα δείχνουν να λειτουργούν αποδοτικότερα σε σχέση με τα RBF.

## ΒΙΒΛΙΟΓΡΑΦΙΑ - ΠΗΓΕΣ

---

- [1] B. Mehlig, *Machine Learning with Neural Networks*. 2021. doi: 10.1017/9781108860604.
- [2] P. Lara-Benítez, M. Carranza-García, J. M. Luna-Romera, and J. C. Riquelme, “Short-term solar irradiance forecasting in streaming with deep learning,” *Neurocomputing*, vol. 546, p. 126312, 2023, doi: <https://doi.org/10.1016/j.neucom.2023.126312>.
- [3] P. Lara-Benítez, M. Carranza-García, and J. Riquelme, “An Experimental Review on Deep Learning Architectures for Time Series Forecasting,” *Int J Neural Syst*, vol. 31, Nov. 2020, doi: 10.1142/S0129065721300011.
- [4] C. L. Giles, S. Lawrence, and A. C. Tsoi, “Noisy Time Series Prediction using Recurrent Neural Networks and Grammatical Inference,” *Mach Learn*, vol. 44, no. 1, pp. 161–183, 2001, doi: 10.1023/A:1010884214864.
- [5] H. Zhang *et al.*, “Dynamic Estimation of Epidemiological Parameters of COVID-19 Outbreak and Effects of Interventions on Its Spread,” *medRxiv*, p. 2020.04.01.20050310, Jan. 2020, doi: 10.1101/2020.04.01.20050310.
- [6] C. E. Madubueze, S. Dachollom, and I. O. Onwubuya, “Controlling the Spread of COVID-19: Optimal Control Analysis,” *Comput Math Methods Med*, vol. 2020, p. 6862516, 2020, doi: 10.1155/2020/6862516.
- [7] H. Batool and L. Tian, “Correlation Determination between COVID-19 and Weather Parameters Using Time Series Forecasting: A Case Study in Pakistan,” *Math Probl Eng*, vol. 2021, p. 9953283, 2021, doi: 10.1155/2021/9953283.
- [8] S. Y. Jung, H. Jo, H. Son, and H. Hwang, “Real-World Implications of Rapidly Responsive COVID-19 Spread Model with Time Dependent Parameters Via Deep Learning: Algorithm Development and Validation,” *J Med Internet Res*, vol. 22, Sep. 2020, doi: 10.2196/19907.
- [9] D. Painuli, D. Mishra, S. Bhardwaj, and M. Aggarwal, “20 - Forecast and prediction of COVID-19 using machine learning,” in *Data Science for COVID-*

- 19, U. Kose, D. Gupta, V. H. C. de Albuquerque, and A. Khanna, Eds., Academic Press, 2021, pp. 381–397. doi: <https://doi.org/10.1016/B978-0-12-824536-1.00027-7>.
- [10] M. A. Lmater, M. Eddabbah, T. Elmoussaoui, and S. Boussaa, “Modelization of Covid-19 pandemic spreading: A machine learning forecasting with relaxation scenarios of countermeasures,” *J Infect Public Health*, vol. 14, no. 4, pp. 468–473, 2021, doi: <https://doi.org/10.1016/j.jiph.2021.01.004>.
- [11] A. ALEXANDRIDIS, “EVOLVING RBF NEURAL NETWORKS FOR ADAPTIVE SOFT-SENSOR DESIGN,” *Int J Neural Syst*, vol. 23, no. 06, p. 1350029, Dec. 2013, doi: 10.1142/S0129065713500299.
- [12] D. Karamichailidou, V. Kaloutsas, and A. Alexandridis, “Wind turbine power curve modeling using radial basis function neural networks and tabu search,” *Renew Energy*, vol. 163, 2021, doi: 10.1016/j.renene.2020.10.020.
- [13] M. Amirian and F. Schwenker, “Radial Basis Function Networks for Convolutional Neural Networks to Learn Similarity Distance Metric and Improve Interpretability,” *IEEE Access*, vol. 8, pp. 123087–123097, 2020, doi: 10.1109/ACCESS.2020.3007337.
- [14] M. Stogiannos, A. Alexandridis, and H. Sarimveis, “An enhanced decentralized artificial immune-based strategy formulation algorithm for swarms of autonomous vehicles,” *Applied Soft Computing Journal*, vol. 89, 2020, doi: 10.1016/j.asoc.2020.106135.
- [15] A. Alexandridis, E. Paizis, E. Chondrodima, and M. Stogiannos, “A particle swarm optimization approach in printed circuit board thermal design,” *Integr Comput Aided Eng*, vol. 24, pp. 143–155, 2017, doi: 10.3233/ICA-160536.
- [16] I. Famelis, A. Alexandridis, and C. Tsitouras, “A highly accurate differential evolution–particle swarm optimization algorithm for the construction of initial value problem solvers,” *Engineering Optimization*, pp. 1–16, Nov. 2017, doi: 10.1080/0305215X.2017.1400545.
- [17] A. Kapnopoulos and A. Alexandridis, “A cooperative particle swarm optimization approach for tuning an MPC-based quadrotor trajectory tracking scheme,” *Aerosp Sci Technol*, vol. 127, p. 107725, 2022, doi: <https://doi.org/10.1016/j.ast.2022.107725>.

- [18] L. Alzubaidi *et al.*, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *J Big Data*, vol. 8, no. 1, p. 53, 2021, doi: 10.1186/s40537-021-00444-8.
- [19] D. Svozil, V. Kvasnicka, and J. Pospichal, “Introduction to multi-layer feed-forward neural networks,” *Chemometrics and Intelligent Laboratory Systems*, vol. 39, no. 1, pp. 43–62, 1997, doi: [https://doi.org/10.1016/S0169-7439\(97\)00061-0](https://doi.org/10.1016/S0169-7439(97)00061-0).
- [20] D. A. Cirovic, “Feed-forward artificial neural networks: applications to spectroscopy,” *TrAC Trends in Analytical Chemistry*, vol. 16, no. 3, pp. 148–155, 1997, doi: [https://doi.org/10.1016/S0165-9936\(97\)00007-1](https://doi.org/10.1016/S0165-9936(97)00007-1).
- [21] K. Socha and C. Blum, “An ant colony optimization algorithm for continuous optimization: application to feed-forward neural network training,” *Neural Comput Appl*, vol. 16, no. 3, pp. 235–247, 2007, doi: 10.1007/s00521-007-0084-z.
- [22] A. Alexandridis, E. Chondrodima, E. Efthimiou, G. Papadakis, F. Vallianatos, and D. Triantis, “Large Earthquake Occurrence Estimation Based on Radial Basis Function Neural Networks,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, pp. 5443–5453, Sep. 2014, doi: 10.1109/TGRS.2013.2288979.
- [23] J. Kumar Jaiswal and R. Das, “Application of artificial neural networks with backpropagation technique in the financial data,” *IOP Conf Ser Mater Sci Eng*, vol. 263, no. 4, p. 042139, 2017, doi: 10.1088/1757-899X/263/4/042139.
- [24] O. Irsoy and C. Cardie, “Deep recursive neural networks for compositionality in language,” *Adv Neural Inf Process Syst*, vol. 3, pp. 2096–2104, Jan. 2014.
- [25] Swapna K E, “<https://developersbreach.com/convolution-neural-network-deep-learning/>.”
- [26] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” *ArXiv e-prints*, Nov. 2015.
- [27] A. Alexandridis, H. Sarimveis, and K. Ninos, “A Radial Basis Function network training algorithm using a non-symmetric partition of the input space – Application to a Model Predictive Control configuration,” *Advances in Engineering Software*, vol. 42, no. 10, pp. 830–837, 2011, doi: <https://doi.org/10.1016/j.advengsoft.2011.05.026>.
- [28] H. Sarimveis, A. Alexandridis, G. Tsekouras, and G. Bafas, “A Fast and Efficient Algorithm for Training Radial Basis Function Neural Networks Based on a Fuzzy

- Partition of the Input Space,” *Ind Eng Chem Res*, vol. 41, Jan. 2002, doi: 10.1021/ie010263h.
- [29] A. Alexandridis, H. Sarimveis, and G. Bafas, “A new algorithm for online structure and parameter adaptation of RBF networks,” *Neural Networks*, vol. 16, no. 7, pp. 1003–1017, 2003, doi: [https://doi.org/10.1016/S0893-6080\(03\)00052-2](https://doi.org/10.1016/S0893-6080(03)00052-2).
- [30] A. Alexandridis, E. Chondrodima, and H. Sarimveis, “Cooperative learning for radial basis function networks using particle swarm optimization,” *Appl Soft Comput*, vol. 49, pp. 485–497, 2016, doi: <https://doi.org/10.1016/j.asoc.2016.08.032>.
- [31] A. Alexandridis and E. Chondrodima, “A medical diagnostic tool based on radial basis function classifiers and evolutionary simulated annealing,” *J Biomed Inform*, vol. 49, pp. 61–72, 2014, doi: <https://doi.org/10.1016/j.jbi.2014.03.008>.
- [32] M. Stogiannos, A. Alexandridis, and H. Sarimveis, “Model predictive control for systems with fast dynamics using inverse neural models,” *ISA Trans*, vol. 72, 2018, doi: 10.1016/j.isatra.2017.09.016.
- [33] A. Alexandridis, M. Stogiannos, A. Kyriou, and H. Sarimveis, “An offset-free neural controller based on a non-extrapolating scheme for approximating the inverse process dynamics,” *J Process Control*, vol. 23, no. 7, 2013, doi: 10.1016/j.jprocont.2013.04.008.
- [34] M. Papadimitrakis, M. Stogiannos, H. Sarimveis, and A. Alexandridis, “Multi-ship control and collision avoidance using mpc and rbf-based trajectory predictions,” *Sensors*, vol. 21, no. 21, 2021, doi: 10.3390/s21216959.
- [35] M. Papadimitrakis and A. Alexandridis, “Active vehicle suspension control using road preview model predictive control and radial basis function networks,” *Appl Soft Comput*, vol. 120, 2022, doi: 10.1016/j.asoc.2022.108646.
- [36] D. Karamichailidou, S. Koletsios, and A. Alexandridis, “An RBF online learning scheme for non-stationary environments based on fuzzy means and Givens rotations,” *Neurocomputing*, vol. 501, 2022, doi: 10.1016/j.neucom.2022.06.016.
- [37] D. Karamichailidou, A. Alexandridis, G. Anagnostopoulos, G. Syriopoulos, and O. Sekkas, “Modeling biogas production from anaerobic wastewater treatment plants using radial basis function networks and differential evolution,” *Comput Chem Eng*, vol. 157, p. 107629, 2022, doi: <https://doi.org/10.1016/j.compchemeng.2021.107629>.

- [38] Z. Lipton, “A Critical Review of Recurrent Neural Networks for Sequence Learning,” May 2015.
- [39] R. DiPietro and G. D. Hager, “Chapter 21 - Deep learning: RNNs and LSTM,” in *Handbook of Medical Image Computing and Computer Assisted Intervention*, S. K. Zhou, D. Rueckert, and G. Fichtinger, Eds., Academic Press, 2020, pp. 503–519. doi: <https://doi.org/10.1016/B978-0-12-816176-0.00026-0>.
- [40] R. Pascanu, T. Mikolov, and Y. Bengio, “On the difficulty of training recurrent neural networks,” in *30th International Conference on Machine Learning, ICML 2013*, 2013.
- [41] R. DiPietro and G. D. Hager, “Chapter 21 - Deep learning: RNNs and LSTM,” in *Handbook of Medical Image Computing and Computer Assisted Intervention*, S. K. Zhou, D. Rueckert, and G. Fichtinger, Eds., Academic Press, 2020, pp. 503–519. doi: <https://doi.org/10.1016/B978-0-12-816176-0.00026-0>.
- [42] A. N. Kamthane and A. A. Kamthane, *Programming and Problem Solving with Python*. McGraw Hill Education. [Online]. Available: <https://books.google.gr/books?id=48lcDwAAQBAJ>
- [43] M. J. J. Douglass, “Book Review: Hands-on Machine Learning with Scikit-Learn, Keras, and Tensorflow, 2nd edition by Aurélien Géron,” *Phys Eng Sci Med*, vol. 43, no. 3, 2020, doi: 10.1007/s13246-020-00913-z.
- [44] Martin Heller, “<https://www.infoworld.com/article/3336192/what-is-keras-the-deep-neural-network-api-explained.html>.”
- [45] E. Mathieu *et al.*, “Coronavirus Pandemic (COVID-19),” *Our World in Data*.
- [46] J. Brownlee, *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions*. Machine Learning Mastery, 2018. [Online]. Available: <https://books.google.gr/books?id=T1-nDwAAQBAJ>
- [47] S. Salman and X. Liu, “Overfitting Mechanism and Avoidance in Deep Neural Networks,” *ArXiv*, vol. abs/1901.06566, 2019.
- [48] Z. Hu, J. Zhang, and Y. Ge, “Handling Vanishing Gradient Problem Using Artificial Derivative,” *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3054915.



