



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης

Διάγνωση COVID-19 από ακολουθίες αξονικών τομογραφιών με χρήση βαθιών νευρωνικών δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΒΑΣΙΛΕΙΟΣ Γ. ΜΠΡΙΛΛΑΚΗΣ

Επιβλέπων : Στέφανος Κόλλιας

Καθηγητής Ε.Μ.Π

Αθήνα, Νοέμβριος 2020



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης

Διάγνωση COVID-19 από ακολουθίες αξονικών τομογραφιών με χρήση βαθιών νευρωνικών δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΒΑΣΙΛΕΙΟΣ Γ. ΜΠΡΙΛΛΑΚΗΣ

Επιβλέπων : Στέφανος Κόλλιας

Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 12^η Νοεμβρίου 2020.

.....
Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

.....
Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

.....
Γεώργιος Στάμου
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2020

.....
Βασίλειος Γ. Μπριλλάκης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π

Copyright © Βασίλειος Γ. Μπριλλάκης, 2020.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Με την ανακήρυξη της νόσου του COVID-19 ως πανδημία, κρίθηκε αναγκαία η ανάπτυξη μεθόδων που επιτρέπουν την έγκαιρη και αποτελεσματική ανίχνευση των κρουσμάτων. Η συνηθέστερη μέθοδος διάγνωσης είναι το μοριακό τεστ RT-PCR το οποίο έχει υψηλή ακρίβεια, όμως απαιτεί σημαντικό χρόνο για την εξαγωγή αποτελέσματος. Η ανάλυση των αξονικών τομογραφιών θεωρείται ένα σημαντικό διαγνωστικό εργαλείο που συμπληρώνει την εξέταση RT-PCR, βελτιώνοντας την ακρίβεια διάγνωσης και βοηθώντας τους ασθενείς να λάβουν γρήγορα την κατάλληλη θεραπεία. Ωστόσο, η αξιολόγηση των αξονικών χρειάζεται πολύτιμο χρόνο από τον εξειδικευμένο ιατρό, δείχνοντας την ανάγκη για μια γρήγορη αυτόματη και αξιόπιστη μέθοδο.

Στο πλαίσιο αυτής της διπλωματικής υλοποιήθηκαν συστήματα βαθιάς μάθησης για την έγκυρη και ταχεία ανίχνευση των κρουσμάτων της νόσου από ακολουθίες αξονικών τομογραφιών. Τα συστήματα αυτά εκπαιδεύονται με σειρές αξονικών συνοδευόμενων από το αποτέλεσμα του μοριακού τεστ και είναι ικανά να κατηγοριοποιούν τους ασθενείς COVID-19 με υψηλή ακρίβεια. Η ανάπτυξη αυτών των συστημάτων στηρίχθηκε στην αρχιτεκτονική CNN-RNN σε συνδυασμό με την χρήση τεχνικών και μεθόδων βαθιάς μάθησης.

Λέξεις κλειδιά: Βαθιά Μάθηση, Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ), Αναδρομικά Νευρωνικά Δίκτυα (ΑΝΔ), Κορονοϊός (COVID-19), Αξονική Τομογραφία (CT).

Abstract

When the COVID-19 virus was declared as a pandemic, it was proved a necessity to develop the means for fast and accurate detection of the COVID-19 cases. The most common diagnostic method is the molecular test RT-PCR, which is capable of giving accurate results but it is time consuming. The evaluation of CT scan series is considered as a very important diagnostic tool that complements the RT-PCR test improving the accuracy of the diagnosis, helping the patients to receive the right treatment in a timely manner. However, the evaluation of CT scans requires valuable time from experts proving why a quick accurate and automated method is necessary.

Within this thesis, deep learning systems were developed in order to diagnose fast and accurate the COVID-19 cases from CT scan series. These systems are trained with a series of CT scans accompanied with the results of the molecular test and are capable to classify the COVID-19 patients with high accuracy. The development of those systems was based on the CNN-RNN architecture in conjunction with the use of deep learning techniques and methods.

Keywords: Deep Learning, Convolutional Neural Networks (CNNs), Recursive Neural Networks (RNNs), COVID-19, Computed Tomography (CT).

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου Στέφανο Κόλλα για την εμπιστοσύνη που μου έδειξε δίνοντας μου ένα τόσο επίκαιρο θέμα και για την ευκαιρία που μου έδωσε να συμμετάσχω στην ερευνητική ομάδα του εργαστηρίου.

Επίσης θα ήθελα να ευχαριστήσω τον Μάνο Σεφέρη και τον Δημήτρη Κόλλια για την καθοδήγηση τους κατά την διάρκεια εκπόνησης της διπλωματικής μου.

Τέλος, ένα μεγάλο ευχαριστώ στην οικογένεια μου για την απεριόριστη εμπιστοσύνη και την αδιάκοπη στήριξη τους.

Περιεχόμενα

Περίληψη.....	5
Abstract.....	7
Ευχαριστίες.....	9
Περιεχόμενα	11
Κατάλογος Σχημάτων.....	13
Κατάλογος Πινάκων.....	15
Κεφάλαιο 1.....	17
1.1 Πανδημία SARS-CoV-2.....	17
1.2 Μέθοδοι Διάγνωσης.....	17
1.3 Σκοπός Εργασίας.....	18
1.4 Σχετικές Εργασίες.....	19
1.5 Δομή Διπλωματικής.....	19
Κεφάλαιο 2.....	21
2.1 Εισαγωγή.....	21
2.2 Μηχανική Μάθηση.....	21
2.2.1 Είδη Μηχανικής Μάθησης.....	22
2.2.2 Επαγωγική Μάθηση.....	22
2.2.3 Συνάρτηση Κόστους.....	23
2.2.4 Μετρικές Αξιολόγησης	26
2.3 Βαθιά Μάθηση.....	28
2.3.1 Νευρωνικά Δίκτυα.....	29
2.3.2 Συναρτήσεις Ενεργοποίησης.....	31
2.3.3 Αλγόριθμοι Βελτιστοποίησης.....	33
2.3.4 Οπίσθια Διάδοση	35
2.3.4 Συνελικτικά Νευρωνικά Δίκτυα.....	39
2.3.5 VGG-16.....	41
2.3.6 ResNet-50.....	42
2.3.7 Αναδρομικά Νευρωνικά Δίκτυα.....	43
2.3.8 LSTM.....	44

2.3.9	GRU.....	47
2.3.10	Bidirectional LSTM.....	47
2.2.11	Γενίκευση	47
2.3.12	Transfer Learning.....	48
Κεφάλαιο 3.....		49
3.1	Εισαγωγή.....	49
3.2	Δεδομένα.....	49
3.2.1	Αξονικές Τομογραφίες.....	49
3.2.2	DICOM.....	52
3.3	Προεπεξεργασία Αξονικών.....	54
3.3.1	Εφαρμογή παραθύρου πνεύμονα	54
3.3.2	Προσαρμογή του μήκους της ακολουθίας	55
3.4	Σύνολο Δεδομένων.....	56
		57
Κεφάλαιο 4.....		59
4.1	Εισαγωγή.....	59
4.2	CNN-RNN	59
4.3	Ενσωμάτωση των Δεδομένων.....	60
4.3	Παρουσίαση Αρχιτεκτονικών.....	61
4.3.1	Αρχιτεκτονική με επίπεδο εξόδου	61
4.3.2	Αρχιτεκτονική με πλήρως συνδεδεμένα επίπεδα	62
Κεφάλαιο 5.....		63
5.1	Αρχιτεκτονική με επίπεδο εξόδου.....	63
5.2	Αρχιτεκτονική με πλήρως συνδεδεμένα επίπεδα.....	65
		66
Κεφάλαιο 6.....		67
Βιβλιογραφία.....		69

Κατάλογος Σχημάτων

Σχήμα 2.1 Προβλεπόμενο διάνυσμα P και πραγματικό διάνυσμα Y.....	25
Σχήμα 2.2 Πίνακας Σύγχυσης.....	26
Σχήμα 2.3 Βιολογικός Νευρώνας.....	30
Σχήμα 2.4 Τεχνητός Νευρώνας.....	30
Σχήμα 2.5(α) Καμπύλη με πολλά ακρότατα.....	30
Σχήμα 2.5(β) Έντονες ταλαντώσεις γύρω από ένα τοπικό ελάχιστο.....	30
Σχήμα 2.6 Νευρωνικό Δίκτυο Παραδείγματος 2.5.....	36
Σχήμα 2.7 Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου.....	40
Σχήμα 2.8 Επίπεδο Συνέλιξης.....	40
Σχήμα 2.9 Μέγιστη και Μέση Υποδειγματοληψία.....	41
Σχήμα 2.10 Δομή VGG-16.....	42
Σχήμα 2.11 Residual Block.....	43
Σχήμα 2.12 Ακολουθιακό Δίκτυο.....	44
Σχήμα 2.13 Δομή κυττάρου LSTM.....	46
Σχήμα 3.1 Κλίμακα αριθμών CT.....	50
Σχήμα 3.2(α) Μείωση της αντίθεσης με την αύξηση του πλάτους παραθύρου.....	51
Σχήμα 3.2(β) Αύξηση της φωτεινότητας με μείωση του κέντρου του παραθύρου.....	51
Σχήμα 3.3(α) Παράθυρο Πνεύμονα.....	52
Σχήμα 3.3(β) Παράθυρο Οστού.....	52
Σχήμα 3.3(γ) Παράθυρο Μαλακού Ιστού.....	52
Σχήμα 3.4 Δομή Στοιχείου Δεδομένων.....	53
Σχήμα 3.5 Σύνολο Δεδομένων.....	57
Σχήμα 3.6(α) Αξονική Ασθενούς με COVID-19.....	57
Σχήμα 3.6(β) Αξονική Non_COVID Ασθενούς.....	57
Σχήμα 4.1 Αρχιτεκτονική CNN-RNN.....	60

Σχήμα 5.1 Καμπύλη Ακρίβειας	65
Σχήμα 5.2 Καμπύλη Σφάλματος.....	65
Σχήμα 5.3 Καμπύλη Ακρίβειας.....	66
Σχήμα 5.4 Καμπύλη Σφάλματος.....	66

Κατάλογος Πινάκων

Πίνακας 5.1 Στατιστικά Συνόλου Δεδομένων.....	63
Πίνακας 5.2 Μετρικές χωρίς επαύξηση.....	64
Πίνακας 5.3 Μετρικές με επαύξηση εύρους 10°	64
Πίνακας 5.4 Μετρικές με επαύξηση εύρους 10° και ολίσθηση εύρους 5 px.....	64
Πίνακας 5.5 Μετρικές χωρίς επαύξηση.....	65
Πίνακας 5.6 Μετρικές με επαύξηση εύρους 10°	66
Πίνακας 5.7 Μετρικές με επαύξηση εύρους 10° και ολίσθηση εύρους 5 px.....	66

Κεφάλαιο 1

Εισαγωγή

1.1 Πανδημία SARS-CoV-2

Η νόσος του ιού SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2), ο οποίος έγινε αρχικά γνωστός ως COVID-19, πρωτοεμφανίστηκε στην πόλη Ουχάν, της Κίνας τον Δεκέμβριο του 2019. Από τα πρώτα κιάλας κρούσματα του, παρατηρήθηκαν οι επιπλοκές που μπορεί να προκαλέσει στους νοσούντες με συχνότερες αυτές στο αναπνευστικό σύστημα, όπως πνευμονία και σύνδρομο οξείας αναπνευστικής δυσχέρειας. Στο επόμενο διάστημα ο ιός άρχισε να εξαπλώνεται σε όλη την υφήλιο, με τον παγκόσμιο οργανισμό υγείας να τον ανακηρύσσει ως πανδημία στις 11 Μαρτίου του 2020. Το γεγονός αυτό σε συνδυασμό με την απουσία εμβολίου και φαρμάκων αλλά και την περιορισμένη διαθεσιμότητα διαγνωστικών τεστ, οδήγησαν τα κράτη να λάβουν τα απαραίτητα μέτρα για τον περιορισμό της εξάπλωσης και της σταδιακής μείωσης των κρουσμάτων. Τα μέτρα που λήφθηκαν από το κάθε κράτος και κυρίως ο περιορισμός στην κυκλοφορία με την επιβολή καραντίνας, συντέλεσαν σημαντικά στην αποφόρτιση του συστήματος υγείας, όμως επέφεραν σημαντικές αλλαγές στον εργασιακό τομέα και στις κοινωνικές επαφές των ανθρώπων με ολέθριες συνέπειες για την οικονομία. Ως τις 11 Νοεμβρίου του 2020, έχουν επιβεβαιωθεί 51,6 εκατομμύρια κρούσματα σε περισσότερες από 188 χώρες ενώ έχουν παρατηρηθεί 1,27 εκατομμύρια θάνατοι.

1.2 Μέθοδοι Διάγνωσης

Οι συνηθέστερες μέθοδοι διάγνωσης του COVID-19 είναι τα ορολογικά και μοριακά τεστ καθώς και η διάγνωση μέσω ακτινογραφιών και αξονικών τομογραφιών.

Το ορολογικό τεστ, ή αλλιώς τεστ αντισωμάτων, γίνεται με δείγμα αίματος και ενδείκνυται σε περιπτώσεις εξακρίβωσης της έκθεσης ενός ατόμου στον ιό μέσω της ανίχνευσης των αντισωμάτων που έχουν αναπτυχθεί στον οργανισμό. Το τεστ αυτό είναι ιδιαίτερα χρήσιμο αν θέλουμε να προσδιορίσουμε τον βαθμό στον οποίο ένας πληθυσμός έχει “άνοσία της αγέλης” καθώς ανιχνεύει ασθενείς που έχουν αναρρώσει ή πάσχουν αυτή την στιγμή από την νόσο. Αν και υπερτερεί σε ταχύτητα έναντι των υπόλοιπων μεθόδων δεν αποτελεί αξιόπιστη μέθοδος, διότι έχει χαμηλή ακρίβεια ενώ ακόμα και αν ο εξεταζόμενος νοσεί δεν μπορεί να δώσει έγκυρο αποτέλεσμα αν δεν έχουν δημιουργηθεί αντισώματα.

Η πρώτη μέθοδος που κυκλοφόρησε είναι το μοριακό τεστ αλυσιδωτής πολυμεράσης (RT-PCR). Η μέθοδος αυτή είναι η πιο έγκυρη και η συνηθέστερη καθώς εξασφαλίζει μεγάλη ακρίβεια στην διάγνωση. Σ' αυτή την περίπτωση γίνεται λήψη δείγματος από την ρινική ή στοματική κοιλότητα και γίνεται εφαρμογή ειδικών μεθόδων για την απομόνωση του μολυσμένου RNA, με τα αποτελέσματα του να εξάγονται σε 8 ώρες.

Η διάγνωση μπορεί να γίνει ακόμη μέσω της απεικόνισης του πνεύμονα μέσω μιας σειράς αξονικών τομογραφιών από τις οποίες ένας έμπειρος ακτινολόγος είναι ικανός να αντλήσει πληροφορία και να συμπεράνει αν ο ασθενής έχει προσβληθεί από τον ιό. Συγκεκριμένα, στις αξονικές τομογραφίες λόγω της μεγάλης ευαισθησίας τους μπορεί να παρατηρηθούν αλλοιώσεις ακόμα και στο αρχικό στάδιο της νόσου (0-4 ημέρες). Η μέθοδος αυτή, ακολουθείται συμπληρωματικά με το μοριακό τεστ σε ασθενείς με συμπτώματα COVID-19 για την βελτίωση της ακρίβειας διάγνωσης, βοηθώντας τους ασθενείς να λάβουν έγκαιρα την κατάλληλη θεραπεία. Η εξαγωγή των αξονικών τομογραφιών μπορεί να πραγματοποιηθεί σε σύντομο χρονικό διάστημα ενώ η αξιολόγηση τους από έναν έμπειρο ακτινολόγο απαιτεί 15-25 λεπτά ανά ασθενή.

1.3 Σκοπός Εργασίας

Η άφιξη της πανδημίας του SARS-CoV-2 έφερε την κοινωνία αντιμέτωπη με μια νέα πραγματικότητα. Για τον λόγο αυτό κρίθηκε αναγκαία η ανάπτυξη διαγνωστικών μεθόδων για τον περιορισμό των κρουσμάτων συμβάλλοντας στην μείωση της ταχύτητας εξάπλωσης της νόσου. Οι ήδη υπάρχουσες μέθοδοι αδυνατούν να συνδυάσουν ταχύτητα και ακρίβεια, όπου σε κάποιες από αυτές απαιτείται πολύτιμος χρόνος από εξειδικευμένο ιατρικό προσωπικό.

Σκοπός της παρούσας διπλωματικής είναι η ανάπτυξη και υλοποίηση συστημάτων βαθιάς μάθησης για την γρήγορη και έγκυρη διάγνωση κρουσμάτων COVID-19 μέσω αξονικών τομογραφιών θώρακος (Chest CT-scans). Συγκεκριμένα συλλογές δεδομένων με αξονικές τομογραφίες συνοδευόμενες από διάγνωση RT-PCR χρησιμοποιήθηκαν στην εκπαίδευση CNN-RNN αρχιτεκτονικών για την κατηγοριοποίηση τους σε δυο κλάσεις:

- Ασθενής COVID-19
- Μη Ασθενής COVID-19.

1.4 Σχετικές Εργασίες

Στο πλαίσιο της εργασίας μελετήθηκαν μέθοδοι που αφορούν βαθιά νευρωνικά δίκτυα και αρχιτεκτονικές CNN-RNN ([1], [2], [3], [4], [5]), σε θέματα προσαρμογής εκπαιδευμένων δικτύων ([6], [7]), σε διαφάνεια της λειτουργίας των δικτύων ([8], [9]), σε ανίχνευση και χρήση υποσυστημάτων προσοχής ([10], [11]), σε σύνδεση νευρωνικών δικτύων με μεθόδους αναπαράστασης γνώσης ([12], [13], [14]) και σε σύνθεση και επαύξηση δεδομένων ([15], [16]).

1.5 Δομή Διπλωματικής

- Στο **Κεφάλαιο 2** παρουσιάζεται η απαραίτητη θεωρία για την εξοικείωση του αναγνώστη με βασικές έννοιες της μηχανικής μάθησης καθώς επίσης αναλύονται μοντέλα βαθιάς μάθησης.
- Στο **Κεφάλαιο 3** περιγράφεται η διαδικασία που ακολουθήθηκε για την επεξεργασία των δεδομένων.
- Στο **Κεφάλαιο 4** παρουσιάζονται οι αρχιτεκτονικές των τεχνητών νευρωνικών δικτύων που χρησιμοποιήθηκαν.
- Στο **Κεφάλαιο 5** αναλύεται αρχικά η πειραματική διαδικασία που ακολουθήθηκε για κάθε μοντέλο και έπειτα παρουσιάζονται τα αποτελέσματα.
- Στο **Κεφάλαιο 6** παρατίθενται τα βασικά συμπεράσματα της εργασίας καθώς προτείνονται ιδέες για μελλοντική εργασία.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

2.1 Εισαγωγή

Αρχικά, γίνεται η εισαγωγή στην μηχανική μάθηση, δηλαδή την ευρύτερη κατηγορία στην οποία ανήκουν τα τεχνητά νευρωνικά δίκτυα. Παρουσιάζονται οι κατηγορίες στις οποίες διακρίνονται οι αλγόριθμοι μηχανικής μάθησης, δίνοντας έμφαση στην κατηγορία που χρησιμοποιείται για την επίλυση του προβλήματος που πραγματεύεται η συγκεκριμένη διπλωματική. Έπειτα, πραγματοποιείται η μετάβαση στην βαθιά μάθηση όπου εντάσσονται οι αρχιτεκτονικές νευρωνικών δικτύων. Συγκεκριμένα, αναλύεται η δομή και οι μονάδες των νευρωνικών δικτύων, αναφέρονται τεχνικές και μέθοδοι που χρησιμοποιούνται σε αυτά και τέλος παρουσιάζονται εξειδικευμένες κατηγορίες τους.

2.2 Μηχανική Μάθηση

Η μηχανική μάθηση είναι ο κλάδος της επιστήμης των υπολογιστών που ασχολείται με την μελέτη και κατασκευή αλγορίθμων που δίνουν την δυνατότητα στον υπολογιστή να μαθαίνει από την εμπειρία. Το βασικό χαρακτηριστικό της μηχανικής μάθησης είναι ότι αναζητεί πρότυπα και σχέσεις στα δεδομένα με σκοπό να μοντελοποιήσει και να κάνει προβλέψεις πάνω σε αυτά. Σύμφωνα με τον πρώιμο ορισμό που έδωσε ο Arthur Samuel το 1959 ([17]) η μηχανική μάθηση είναι:

Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί.

Ένας πιο επίσημος ορισμός που δόθηκε από τον Tom M. Mitchell ([18]) είναι ο ακόλουθος:

Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία E ως προς μια κλάση εργασιών T και ένα μέτρο επίδοσης P , αν η επίδοση του σε εργασίες μια κλάσης T , όπως αποτιμάται από το μέτρο P , βελτιώνεται με την εμπειρία.

Όλοι οι αλγόριθμοι μηχανικής μάθησης περνάνε από δυο στάδια, το στάδιο της εκπαίδευσης (training) και το στάδιο του ελέγχου (testing). Στο πρώτο στάδιο ο αλγόριθμος δέχεται ένα σύνολο δεδομένων (training set) από τα οποία αποκτάει γνώση

για την επίτευξη συγκεκριμένου σκοπού. Ακόμα μπορεί να υπάρχει υπάρχουσα γνώση από προηγούμενη εκπαίδευση του αλγορίθμου την οποία μπορεί να λάβει υπόψιν του ή να την αγνοήσει κατά την εκπαίδευση. Στην συνέχεια, πραγματοποιείται ο έλεγχος της παραχθείσας γνώσης. Συγκεκριμένα, έχοντας ένα άλλο σύνολο δεδομένων (test set) ξένο ως προς το πρώτο, ελέγχεται η γνώση που αποκτήθηκε κάνοντας ανάκληση του συνόλου μέσω του αλγορίθμου και στην συνέχεια αξιολογείται η απόδοση του στο συγκεκριμένο σύνολο.

2.2.1 Είδη Μηχανικής Μάθησης

Οι αλγόριθμοι μηχανικής μάθησης μπορούν να χωριστούν σε τρεις κατηγορίες ανάλογα με το είδος της εμπειρίας από την οποία μαθαίνουν. Αυτές, είναι ανάλογες με τους τρόπους με τους οποίους μαθαίνει ο άνθρωπος, την μη επιβλεπόμενη μάθηση, την επιβλεπόμενη μάθηση και την ενισχυτική μάθηση.

Μη επιβλεπόμενη μάθηση (Unsupervised Learning), είναι η διαδικασία κατά την οποία δοθέντος ενός κλειστού συνόλου δεδομένων με πολλά χαρακτηριστικά εξάγεται χρήσιμη πληροφορία για την δομή του συνόλου. Χρησιμοποιείται κυρίως σε προβλήματα συσταδιοποίησης (clustering).

Επιβλεπόμενη μάθηση (Supervised Learning), είναι η διαδικασία κατά την οποία δοθέντος ενός κλειστού συνόλου δεδομένων με πολλά χαρακτηριστικά, κάθε στοιχείο του οποίου σχετίζεται με μια ετικέτα (label), γίνεται προσέγγιση μιας συνάρτησης που απεικονίζει εισόδους (στοιχεία) σε εξόδους (labels) βάσει της εμπειρίας που έχει αποκτηθεί από το σύνολο δεδομένων. Χρησιμοποιείται κυρίως σε προβλήματα κατηγοριοποίησης (classification), πρόβλεψης (prediction) και διερμηνείας (interpretation).

Ενισχυμένη μάθηση (Reinforcement Learning), είναι η διαδικασία κατά την οποία κατασκευάζεται μια στρατηγική μέσω της αλληλεπίδρασης με ένα ανοικτό σύνολο δεδομένων.

Είναι φανερό ότι για κάθε πρόβλημα υπάρχει ένας κατάλληλος τρόπος μάθησης και για κάθε τρόπο μάθησης υπάρχει τουλάχιστον ένας κατάλληλος αλγόριθμος.

2.2.2 Επαγωγική Μάθηση

Για το πρόβλημα της συγκεκριμένης εργασίας, χρησιμοποιείται ένας αλγόριθμος επιβλεπόμενης επαγωγικής μάθησης. Η επαγωγική μάθηση στηρίζεται στην υπόθεση επαγωγικής μάθησης (inductive learning hypothesis), σύμφωνα με την οποία “κάθε υπόθεση h που προσεγγίζει καλά την συνάρτηση στόχο για ένα αρκετά μεγάλο σύνολο παραδειγμάτων,

θα προσεγγίζει το ίδιο καλά τη συνάρτηση στόχο και για περιπτώσεις που δεν έχει εξετάσει”. Στην επαγωγική μάθηση το σύστημα μαθαίνει μια συνάρτηση που ονομάζεται συνάρτηση στόχος (target function) και αποτελεί έκφραση του μοντέλου που περιγράφει τα δεδομένα. Η συνάρτηση στόχος απεικονίζει τις τιμές ενός συνόλου μεταβλητών, που ονομάζονται ανεξάρτητες μεταβλητές ή μεταβλητές εισόδου ή χαρακτηριστικά, σε μια τιμή που ονομάζεται εξαρτημένη μεταβλητή ή μεταβλητή εξόδου. Το σύνολο των διαφορετικών δυνατών τιμών εισόδου της συνάρτησης, δηλαδή το πεδίο ορισμού της, ονομάζεται σύνολο περιπτώσεων ή στιγμιοτύπων (instances). Κάθε στιγμιότυπο περιγράφεται από ένα σύνολο χαρακτηριστικών (attributes ή features). Το σύνολο των εισόδων που δέχεται ο αλγόριθμος κατά την εκπαίδευση του ονομάζεται σύνολο εκπαίδευσης (training set) ή παραδείγματα και είναι υποσύνολο του συνόλου στιγμιότυπων. Τα υπόλοιπα στιγμιότυπα αποτελούν το σύνολο ελέγχου (test set) και χρησιμοποιούνται στο στάδιο ελέγχου. Κατά την διαδικασία της εκπαίδευσης προσεγγίζεται η συνάρτηση στόχος μέσω της συνάρτησης πρόβλεψης h (prediction function), η οποία δέχεται ως εισόδους στιγμιότυπα και επιστρέφει μια έξοδο-πρόβλεψη. Για να βελτιωθεί η συμπεριφορά της συνάρτησης πρόβλεψης κατά την μάθηση χρησιμοποιείται η συνάρτηση κόστους (cost ή loss function). Αυτή εντοπίζει την διαφορά της εξόδου της συνάρτησης πρόβλεψης με την μεταβλητή στόχο που είναι η επιθυμητή έξοδος. Σκοπός στο στάδιο εκπαίδευσης είναι η ελαχιστοποίηση της συνάρτησης κόστους. Έπειτα από την ολοκλήρωση της μάθησης του συστήματος, ακολουθεί το στάδιο ελέγχου. Το σύστημα έχει πλέον αποκτήσει νέα γνώση και καλείται να προβλέψει την έξοδο των στιγμιότυπων του συνόλου ελέγχου και στην συνέχεια να αξιολογηθεί η απόδοση του βάσει ορισμένων μετρικών επίδοσης.

2.2.3 Συνάρτηση Κόστους

Όπως προαναφέρθηκε, κατά την διαδικασία μάθησης χρησιμοποιείται η συνάρτηση κόστους για να υπολογιστεί η διαφορά της επιθυμητής εξόδου με την έξοδο της συνάρτησης πρόβλεψης. Η διαφορά αυτή ονομάζεται σφάλμα (loss) και είναι φανερό ότι όσο μικρότερη είναι η τιμή του τόσο καλύτερες είναι οι προβλέψεις του μοντέλου. Η επιλογή της κατάλληλης συνάρτησης κόστους εξαρτάται από την φύση του προβλήματος που καλείται να επιλύσει ένας αλγόριθμος μηχανικής μάθησης. Σε αυτό το σημείο θα ήταν χρήσιμο να αναφερθεί η εντροπία που αποτελεί την βάση για τις συναρτήσεις κόστους που θα παρουσιαστούν στην συνέχεια.

Εντροπία (Entropy), μια τυχαίας μεταβλητής X είναι το μέτρο της μέσης αβεβαιότητας που την χαρακτηρίζει. Για κατανομή πιθανότητας $P(X)$ και τυχαία μεταβλητή X , η εντροπία ορίζεται ως εξής:

$$H(X) = \begin{cases} -\int_x P(x) \log P(x), & \text{για συνεχείς τυχαίες μεταβλητές.} \\ -\sum_x P(x) \log P(x), & \text{για διακριτές τυχαίες μεταβλητές.} \end{cases} \quad (2.1)$$

Η εντροπία παίρνει τιμές από 0 ως 1, όσο πιο κοντά είναι η τιμή της στην μονάδα τόσο μεγαλύτερη είναι η αβεβαιότητα της τυχαίας μεταβλητής. Συνεπώς είναι επιθυμητό η τιμή της να είναι κοντά στο κάτω φράγμα.

Παράδειγμα 2.1. Θεωρείται ένα πείραμα τύχης στο οποίο γίνεται ρίψη ενός αμερόληπτου νομίσματος. Ορίζοντας ως X την τυχαία μεταβλητή που εκφράζει το αποτέλεσμα της ρίψης η μεταβλητή θα παίρνει τις τιμές 0 και 1, όπου η μια αντιστοιχεί στο ενδεχόμενο το αποτέλεσμα να είναι κορώνα ενώ η άλλη στο ενδεχόμενο το αποτέλεσμα να είναι γράμματα. Η μεταβλητή αυτή ακολουθεί την διωνυμική κατανομή Bernoulli. Επειδή το νόμισμα είναι αμερόληπτο οι πιθανότητες των δυο ενδεχομένων είναι ίσες. Η εντροπία της X από την σχέση (2.1) υπολογίζεται ως εξής:

$$H(X) = - \sum_{i=1}^n P(x_i) \log_b p(x_i) \Rightarrow$$

$$H(X) = - \sum_{i=1}^2 P(x_i) \log_2 \frac{1}{2} = 1$$

Μπορεί κάποιος να παρατηρήσει πως το αποτέλεσμα είναι αβέβαιο, το οποίο είναι απολύτως λογικό αν ληφθεί υπόψιν πως τα δύο ενδεχόμενα στο εν λόγω παράδειγμα είναι ισοπίθανα και επομένως είναι δύσκολο να προβλεφθεί το αποτέλεσμα της επόμενης ρίψης. Αν το νόμισμα ήταν μη αμερόληπτο η αβεβαιότητα θα μειωνόταν καθώς το ένα ενδεχόμενο θα ήταν πιθανότερο να συμβεί. Υποθέτοντας ότι το νόμισμα έχει λυγίσει και είναι πλέον μη αμερόληπτο η πιθανότητα το αποτέλεσμα της ρίψης να είναι κορώνα είναι 0.8 ενώ η πιθανότητα να είναι γράμματα είναι 0.2. Από την σχέση (2.1) προκύπτει:

$$H(X) = -0.8 \log_2 0.8 - 0.2 \log_2 0.2 \Rightarrow$$

$$H(X) = 0.72$$

Cross-Entropy Function: Η συνάρτηση σφάλματος εντροπίας ή αλλιώς συνάρτηση λογαριθμικού σφάλματος (log loss function) μετράει τη απόδοση ενός μοντέλου ταξινόμησης του οποίου η έξοδος είναι μια τιμή πιθανότητας μεταξύ 0 και 1 ή και πολλές τιμές μεταξύ 0 και 1 μια για κάθε κλάση του προβλήματος. Το σφάλμα εντροπίας αυξάνει όσο η προβλεπόμενη τιμή πιθανότητας αποκλίνει από την τιμή της πραγματικής κλάσης.

Η κοινή εντροπία και το λογαριθμικό σφάλμα είναι ελαφρώς διαφορετικές έννοιες αλλά στη μηχανική μάθηση όταν υπολογίζονται σφάλματα, οι δύο έννοιες ταυτίζονται. Ο τύπος για περισσότερες από δύο κλάσεις ($M > 2$) είναι ο εξής :

$$L_{ce} = - \sum_{i=1}^M y_i \log(p_i) \quad (2.2)$$

Όπου M είναι ο αριθμός των κλάσεων, y_i η πραγματική πιθανότητα αυτής της κλάσης, προβλεπόμενη πιθανότητα για αυτή την κλάση.

Παράδειγμα 2.2. Θεωρείται ένα πρόβλημα κατηγοριοποίησης με 5 κλάσεις. Για το πρόβλημα αυτό έχει υλοποιηθεί ένα σύστημα που προβλέπει για μια είσοδο σε ποια κλάση ανήκει, επιστρέφοντας μια πιθανότητα για κάθε κλάση.

P	Y
0.024	0
0.925	1
0.014	0
0.011	0
0.026	0

Σχήμα 2.1 Προβλεπόμενο διάνυσμα P και πραγματικό διάνυσμα Y

Παρατηρείται ότι το μοντέλο έχει προβλέψει σωστά την κλάση της εισόδου. Αυτό φαίνεται και στην μικρή τιμή που επιστρέφει η συνάρτηση κόστους. Αντικαθιστώντας στην σχέση (2.2) έχουμε:

$$\begin{aligned} L_{ce} &= -0 \log(0.024) - 1 \log(0.925) + \\ &\quad -0 \log(0.014) - 0 \log(0.011) + \\ &\quad -0 \log(0.026) = 0.0339 \end{aligned}$$

Binary Cross-Entropy Function: Η συνάρτηση αυτή χρησιμοποιείται για προβλήματα δυαδικής κατηγοριοποίησης. Υπολογίζεται ως το μέσο σφάλμα εντροπίας μεταξύ όλων των στιγμιοτύπων.

$$L_{Bin CE} = -\frac{1}{N} \left[\sum_{i=1}^N y_i \log p_i + (1 - y_i) \log(1 - p_i) \right] \quad (2.3)$$

2.2.4 Μετρικές Αξιολόγησης

Για την αξιολόγηση του συστήματος θα γίνει χρήση ορισμένων δεικτών που ονομάζονται μετρικές αξιολόγησης (evaluation metrics). Η επιλογή των κατάλληλων μετρικών εξαρτάται από το πρόβλημα που επιλύεται, στην περίπτωση της κατηγοριοποίησης οι συνηθέστερες μετρικές είναι η ακρίβεια, το precision, το recall και το F1-score. Είναι σημαντικό εδώ πριν οριστούν οι μετρικές, να αναφερθούν για ένα πρόβλημα κατηγοριοποίησης με δυο κλάσεις οι εξής έννοιες:

Αληθές Θετικό (true positive): Ο αριθμός των στιγμιότυπων της πρώτης κλάσης για τα οποία η πρόβλεψη ήταν σωστή.

Αληθές Αρνητικό (true negative): Ο αριθμός των στιγμιότυπων της δεύτερης κλάσης για τα οποία η πρόβλεψη ήταν σωστή.

Ψευδές Θετικό ή Σφάλμα τύπου I (false positive): Ο αριθμός των στιγμιότυπων της δεύτερης κλάσης για τα οποία η πρόβλεψη ήταν λανθασμένη, δηλαδή προβλέφθηκαν ως στιγμιότυπα της πρώτης κλάσης.

Ψευδές Αρνητικό Σφάλμα τύπου II (false negative): Ο αριθμός των στιγμιότυπων της πρώτης κλάσης για τα οποία η πρόβλεψη ήταν λανθασμένη, δηλαδή προβλέφθηκαν ως στιγμιότυπα της δεύτερης κλάσης.

Οι παραπάνω τιμές διαμορφώνουν τον **πίνακα σύγχυσης** (confusion matrix).

		Actual Class	
		P	N
Predicted Class	P	TP	FP
	N	FN	TN

Σχήμα 2.2 Πίνακας Σύγχυσης

Η χρήση αυτή του πίνακα επιτρέπει μια πιο λεπτομερή ανάλυση στην απόδοση της κατηγοριοποίησης. Η ανάλυση αυτή είναι απαραίτητη σε περιπτώσεις που συναντάται μη ισορροπημένο (imbalanced data) σύνολο δεδομένων, δηλαδή τα πλήθη των στιγμιότυπων των κλάσεων διαφέρουν σε μεγάλο βαθμό. Επιπλέον χρήσιμος θεωρείται ο πίνακας όταν

λαμβάνονται υπόψιν οι λανθασμένες προβλέψεις των κλάσεων. Αξιόπιστες μετริกές για τέτοια σύνολα θεωρούνται το precision, το recall και το F1-score.

Precision: Είναι ο λόγος των στιγμιοτύπων μιας κλάσης που σωστά προβλέφθηκαν προς το πλήθος των στιγμιοτύπων που προβλέφθηκαν ως στιγμιότυπα αυτής της κλάσης. Με αυτή την μετρική παρατηρείται πόσο ακριβές είναι το μοντέλο στην πρόβλεψη θετικών και χρησιμοποιείται όταν η απόδοση εξαρτάται από το πλήθος των λανθασμένα προβλεφθέντων θετικών.

$$\begin{aligned} \text{precision} &= \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \\ \text{precision} &= \frac{\text{True Positive}}{\text{Total Predicted Positive}} \end{aligned} \quad (2.4)$$

Recall: Είναι ο λόγος των στιγμιοτύπων μιας κλάσης που σωστά προβλέφθηκαν προς το πλήθος των πραγματικών στιγμιοτύπων της κλάσης. Από την μετρική αυτή μπορεί να διαπιστωθεί πόσο ακριβές είναι το μοντέλο στην πρόβλεψη θετικών και χρησιμοποιείται όταν η απόδοση εξαρτάται από το πλήθος των λανθασμένα προβλεφθέντων αρνητικών.

$$\begin{aligned} \text{recall} &= \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \\ \text{recall} &= \frac{\text{True Positive}}{\text{Total Actual Positive}} \end{aligned} \quad (2.5)$$

F1-score: Είναι ο αρμονικός μέσος του precision και του recall.

$$F_1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (2.6)$$

Ακολουθούν μερικά παραδείγματα στα οποία η μετρική της ακρίβειας δεν είναι αντιπροσωπευτική για την αξιολόγηση του συστήματος:

Παράδειγμα 2.3. Έστω ένα σύστημα διάγνωσης που διακρίνει τους ανθρώπους σε δυο κλάσεις υγιής και μη υγιής. Το σύστημα μεταξύ 1000 ανθρώπων προβλέπει σωστά ότι οι 989 είναι υγιείς (TN) και ότι 1 είναι ασθενής (TP) ενώ προβλέπει λανθασμένα ως υγιείς 10 ασθενείς (FN). Αξιολογώντας το σύστημα ως προς την ακρίβεια (accuracy) παρατηρείται ότι το σύστημα είναι πολύ καλό καθώς πετυχαίνει ακρίβεια 99%. Στην συγκεκριμένη περίπτωση όμως το σύνολο δεδομένων είναι μη ισορροπημένο οπότε η ακρίβεια δεν είναι αντιπροσωπευτική για την αξιολόγηση του συστήματος. Επίσης δεν μπορεί να αγνοηθεί η αδυναμία του στην διάκριση των νοσούντων καθώς ένα τέτοιο λάθος θα κόστιζε. Εδώ η μετρική recall θα ήταν αξιόπιστη για την αξιολόγηση του συστήματος, καθώς λαμβάνεται υπόψιν το πλήθος των εσφαλμένων αρνητικών (FN).

Παράδειγμα 2.4. Έστω ότι το ηλεκτρονικό ταχυδρομείο διαθέτει ένα σύστημα για να διακρίνει την αλληλογραφία σε δυο κλάσεις σημαντικά και ανεπιθύμητα. Το σύστημα αυτό μεταξύ 100 μηνυμάτων προβλέπει σωστά ότι τα 90 είναι σημαντικά (TN) και ότι τα 2 είναι ανεπιθύμητα (TP) ενώ προβλέπει λανθασμένα 8 σημαντικά ως ανεπιθύμητα (FP). Ομοίως με το προηγούμενο παράδειγμα παρά την υψηλή ακρίβεια, εδώ πρέπει να ληφθεί υπόψιν για την αξιολόγηση του συστήματος ο αριθμός των σημαντικών μηνυμάτων που κατηγοριοποιήθηκαν ως ανεπιθύμητα. Χρήσιμη σε αυτή την περίπτωση θα ήταν η μετρική precision.

2.3 Βαθιά Μάθηση

Οι κλασσικές τεχνικές μηχανικής μάθησης εμφανίζουν δυσκολία στην επεξεργασία ακατέργαστων δεδομένων. Για δεκαετίες η ανάπτυξη συστημάτων αναγνώρισης προτύπων και μηχανικής μάθησης απαιτούσε αυστηρή μελέτη αλλά και εξειδίκευση στο πεδίο έρευνας για την σχεδίαση μεθόδων εξαγωγής χαρακτηριστικών που μετατρέπουν ακατέργαστα δεδομένα σε κατάλληλες αναπαραστάσεις ώστε το σύστημα να είναι ικανό να εντοπίζει ή να κατηγοριοποιεί πρότυπα για τις εισόδους που του δίνονται. Η δυσκολία αυτή των παραδοσιακών συστημάτων μηχανικής μάθησης αντιμετωπίστηκε με την χρήση των τεχνικών βαθιάς μάθησης.

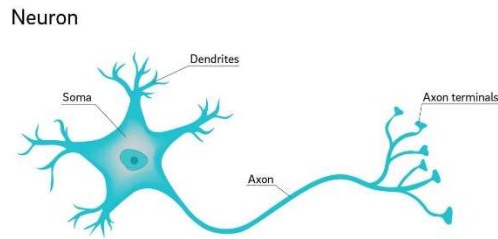
Η βαθιά μάθηση βασίζεται σε μεθόδους αναπαραστασιακής μάθησης (representational learning), οι οποίες επιτρέπουν στο σύστημα να βρίσκει αυτόματα την αναπαράσταση που χρειάζεται ώστε να είναι εφικτή η κατηγοριοποίηση ή ο εντοπισμός των προτύπων. Συγκεκριμένα οι μέθοδοι βαθιάς μάθησης χρησιμοποιούν πολλά επίπεδα αναπαράστασης που κατασκευάζονται από μη γραμμικές μονάδες, μετασχηματίζοντας την

αναπαράσταση ενός επιπέδου σε αναπαράσταση ενός υψηλότερου και πιο αφαιρετικού επιπέδου ξεινώνοντας από την αιατέργαστη είσοδο. Με αυτό τον τρόπο τα συστήματα μπορούν να μάθουν πολύ σύνθετες συναρτήσεις. Στην κατηγοριοποίηση τα υψηλά επίπεδα αναπαράστασης δίνουν έμφαση σε χαρακτηριστικά της εισόδου που είναι σημαντικά για την διάκριση της. Για παράδειγμα, σε μια εικόνα που αναπαρίσταται με ένα πίνακα τιμών pixel τα χαρακτηριστικά που μαθαίνονται στο πρώτο επίπεδο συνήθως σχετίζονται την παρουσία ή την απουσία ακμών σε συγκεκριμένα σημεία της εικόνας. Στο δεύτερο επίπεδο συνηθίζεται να εντοπίζονται μοτίβα που σχηματίζουν οι ακμές. Το τρίτο επίπεδο συνδυάζονται τα μοτίβα σε κομμάτια γνωστών μας αντικειμένων και στην συνέχεια ανώτερα επίπεδα συνδυάζουν τα αντικείμενα αυτά εντοπίζοντας τα αντικείμενα τα οποία αποτελούν. Το βασικό γνώρισμα της βαθιάς μάθησης είναι ότι αυτά τα επίπεδα δεν έχουν σχεδιαστεί από μηχανικούς αλλά αναπτύσσονται από τα δεδομένα μέσω μιας διαδικασίας μάθησης γενικού σκοπού ([19]).

Αν και σήμερα αποτελεί ίσως τον πιο δραστήριο τομέα της μηχανικής μάθησης, από το 1943 οπότε παρουσιάστηκε το πρώτο μοντέλο νευρωνικού δικτύου από τους McCulloch και Pitts μέχρι το 2006 ([20], [21]) δεν υπήρχε ενδιαφέρον για τις αρχιτεκτονικές βαθιάς μάθησης. Ωστόσο από τότε οι τεχνικές βαθιάς μάθησης έχουν κάνει μεγάλη πρόοδο, βοήθησαν ακόμα και στην επίλυση προβλημάτων τα οποία ήταν άλυτα από την κοινότητα της τεχνητής νοημοσύνης για πολλά χρόνια. Σήμερα βρίσκουν εφαρμογή σε πολλούς επιστημονικούς τομείς μέσω των νευρωνικών δικτύων.

2.3.1 Νευρωνικά Δίκτυα

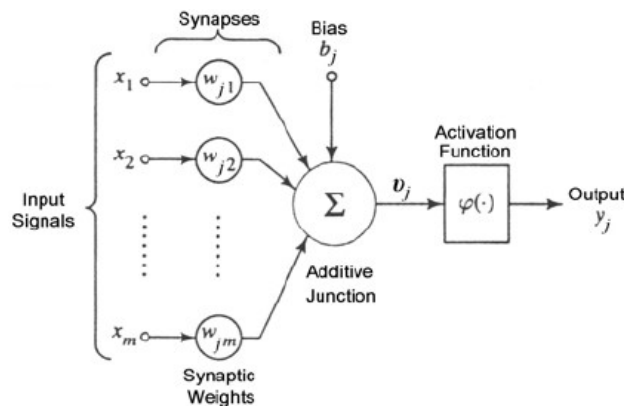
Η βασικότερη αρχιτεκτονική βαθιάς μάθησης είναι αυτή των νευρωνικών δικτύων. Τα νευρωνικά δίκτυα είναι υπολογιστικά μοντέλα εμπνευσμένα από την δομή και την λειτουργία των νευρώνων του εγκεφάλου. Ένας βιολογικός νευρώνας αποτελείται από το σώμα που αποτελεί τον πυρήνα του, τους δενδρίτες μέσω των οποίων λαμβάνει σήματα από γειτονικούς νευρώνες (σήματα εισόδου) και τον άξονα που είναι η έξοδος του νευρώνα με τον οποίο συνδέεται με άλλους νευρώνες. Σε κάθε δενδρίτη υπάρχει ένα απειροελάχιστο κενό που ονομάζεται σύναψη. Οι συνάψεις μέσω χημικών διαδικασιών επιβραδύνουν ή επιταχύνουν τη ροή ηλεκτρονίων προς το σώμα του νευρώνα. Η ικανότητα μάθησης και μνήμης του εγκεφάλου στηρίζεται στην ικανότητα των συνάψεων να μεταβάλλουν την αγωγιμότητα τους. Τα ηλεκτρικά σήματα του που εισέρχονται στο σώμα των νευρώνων συνδυάζονται και αν το αποτέλεσμα ξεπερνά κάποια τιμή κατωφλίου το σήμα διαδίδεται προς άλλους νευρώνες μέσω του άξονα.



Σχήμα 2.3 Βιολογικός Νευρώνας

Ένας ανθρώπινος εγκέφαλος αποτελείται από περίπου 100 δισεκατομμύρια νευρώνες με κάθε νευρώνα να συνδέεται με περίπου 1000 νευρώνες. Αυτό πραγματοποιείται μέσω του άξονα του νευρώνα στον οποίο καταλήγουν ισάριθμοι δενδρίτες άλλων νευρώνων. Δεδομένου ότι κάθε δενδρίτης έχει μια σύναψη, στον εγκέφαλο υπάρχουν 100 τρισεκατομμύρια συνάψεις οι οποίες επηρεάζουν την λειτουργία του. Ο χρόνος απόκρισης των νευρώνων είναι της τάξης των χιλιοστών του δευτερολέπτου (ms), έτσι ο εγκέφαλος είναι σε θέση να λαμβάνει πολύπλοκες αποφάσεις. Αυτό οφείλεται στην υπολογιστική ισχύ του και στο γεγονός ότι η πληροφορία είναι διαμοιρασμένη σε όλο του τις μονάδες. Από τα παραπάνω είναι εμφανής η επιθυμία μοντελοποίησης της δομής και της λειτουργίας του ανθρώπινου εγκεφάλου, προφανώς σε μικρότερη κλίμακα καθώς είναι αδύνατο να απεικονιστεί σε πλήρη κλίμακα. Τα μοντέλα αυτά ονομάζονται τεχνητά νευρωνικά δίκτυα.

Τα τεχνητά νευρωνικά δίκτυα είναι υπολογιστικά μοντέλα τα οποία σε αντιστοιχία με τους βιολογικούς νευρώνες του εγκεφάλου αποτελούνται από τεχνητούς νευρώνες (perceptron's). Όπως φαίνεται και στο σχήμα, ένας τεχνητός νευρώνας δέχεται σήματα εισόδου $x_1, x_2, x_3, \dots, x_n$ τα οποία σε αντίθεση με τους ηλεκτρικούς παλμούς είναι συνεχείς μεταβλητές. Σε αντιστοιχία με τις σύναψη στην είσοδο του νευρώνα έχουμε το βάρος w_i το οποίο μεταβάλλει την τιμή της εισόδου.



Σχήμα 2.4 Τεχνητός Νευρώνας

Στο σώμα του τεχνητού νευρώνα γίνεται η άθροιση των εισόδων επηρεασμένων από τα βάρη ενώ στην συνέχεια το άθροισμα περνάει από μια συνάρτηση ενεργοποίησης η οποία διαμορφώνει την τελική τιμή της εξόδου y ανάλογα με την τιμή κατωφλίου που έχει η συνάρτηση ενεργοποίησης. Ένας τεχνητός νευρώνας μπορεί να έχει πολλές εξόδους όλες όμως θα πρέπει να έχουν την ίδια τιμή. Ορισμένες φορές εκτός από τις εισόδους τους και τα αντίστοιχα βάρη, ο νευρώνας μπορεί να έχει και κάποιο βάρος w_0 το οποίο ονομάζεται πόλωση (bias) ή παράγοντας προδιάθεσης του νευρώνα. Η διαφορά της πόλωσης με τα υπόλοιπα βάρη είναι ότι επιδρά σε όλες τις τιμές τις εισόδου.

Οι τεχνητοί νευρώνες είναι οργανωμένοι σε δομές παρόμοιες αυτών του ανθρώπινου εγκεφάλου. Συγκεκριμένα οργανώνονται σε μια σειρά από στρώματα ή επίπεδα (layers). Το πρώτο επίπεδο του δικτύου ονομάζεται επίπεδο εισόδου (input layer) και χρησιμοποιείται για την εισαγωγή δεδομένων. Τα στοιχεία αυτό του επιπέδου δεν είναι ουσιαστικά νευρώνες καθώς δεν υπολογίζουν κάτι απλώς προωθούν την είσοδο στο επόμενο επίπεδο, συνεπώς δεν έχουν βάρη και συνάρτηση ενεργοποίησης. Στην συνέχεια ακολουθούν προαιρετικά ένα ή και περισσότερα ενδιάμεσα ή κρυφά επίπεδα (hidden layers) όπου στην περίπτωση που υπάρχουν περισσότερα από ένα κρυφά επίπεδα το δίκτυο ονομάζεται βαθύ νευρωνικό δίκτυο (deep neural network). Στο τέλος του δικτύου υπάρχει το επίπεδο εξόδου (output layer).

Ιδιαίτερη σημασία σε ένα νευρωνικό δίκτυο έχει η σύνδεση μεταξύ των επιπέδων, ανάλογα με την σύνδεση τα επίπεδα διακρίνονται σε πλήρως και μερικώς συνδεδεμένα. Πλήρως συνδεδεμένα (fully connected) ονομάζονται 2 επίπεδα στα οποία όλοι οι νευρώνες του πρώτου επιπέδου συνδέονται με όλους τους νευρώνες του δεύτερου. Σε κάθε άλλη περίπτωση είναι μερικώς συνδεδεμένα (partially connected). Όταν η εξοδοί των νευρώνων ενός επιπέδου δεν συνδέονται με καμία είσοδο του ίδιου ή προηγούμενου επιπέδου το δίκτυο χαρακτηρίζεται ως πρόσθιας διάδοσης (feedforward network). Ενώ στην αντίθετη περίπτωση όπου υπάρχει σύνδεση με την είσοδο του ίδιου ή προηγούμενου επιπέδου το δίκτυο χαρακτηρίζεται ως αναδρομικό (recurrent network).

2.3.2 Συναρτήσεις Ενεργοποίησης

Η συνάρτηση ενεργοποίησης είναι μια μαθηματική συνάρτηση η οποία καθορίζει την έξοδο του τεχνητού νευρώνα. Απλή μορφή τέτοιας συνάρτησης είναι η δυαδική συνάρτηση που επιστρέφει 0 ή 1 ανάλογα με την είσοδο, ενεργοποιώντας ή απενεργοποιώντας κατά αυτό τον τρόπο τον νευρώνα. Επιπλέον η συνάρτηση ενεργοποίησης χρησιμοποιείται σε περιπτώσεις όπου είναι επιθυμητή η κανονικοποίηση της εξόδου του, καθώς και σε περιπτώσεις που κρίνεται αναγκαία η εισαγωγή μη γραμμικότητας στο σύστημα. Οι συναρτήσεις ενεργοποίησης χωρίζονται σε γραμμικές και μη γραμμικές. Οι βασικότερες από αυτές είναι οι εξής :

Σιγμοειδής συνάρτηση (sigmoid function). Είναι μια παραγωγίσιμη συνάρτηση με πεδίο ορισμού το σύνολο των πραγματικών αριθμών και σύνολο τιμών το κλειστό σύνολο 0 ως 1. Χρησιμοποιείται συνήθως στο επίπεδο εξόδου νευρωνικών κατηγοριοποίησης δυο κλάσεων. Η συνάρτηση ορίζεται από τον παρακάτω τύπο:

$$S(x) = \frac{e^x}{e^x + 1} \quad (2.7)$$

Συνάρτηση Softmax. Η συνάρτηση αυτή συναντάται συχνά ως συνάρτηση ενεργοποίησης στο επίπεδο εξόδου νευρωνικών δικτύων κατηγοριοποίησης πολλών κλάσεων. Συγκεκριμένα δέχεται ως είσοδο ένα διάνυσμα από πραγματικές τιμές και επιστρέφει ως έξοδο ένα διάνυσμα με τις κανονικοποιημένες εξόδους για κάθε κλάση όπου κάθε μια διαιρείται με το άθροισμα όλων των κανονικοποιημένων εξόδων (βλέπε σχέση 2.8). Η συνάρτηση ορίζεται από τον παρακάτω τύπο:

$$\text{Softmax}(y_i) = \frac{e^{y_i}}{\sum_{i=1}^N e^{y_i}} \quad (2.8)$$

Συνάρτηση ReLU (Rectified Linear Unit). Αποτελεί την συνηθέστερη συνάρτηση ενεργοποίησης στα νευρωνικά δίκτυα. Επιστρέφει την τιμή της εισόδου ή μηδέν στην περίπτωση που η είσοδος είναι μικρότερη του μηδενός. Η συνάρτηση ορίζεται από τον παρακάτω τύπο:

$$\text{ReLU}(x) = \max(x, 0) \quad (2.9)$$

Συνάρτηση υπερβολικής εφαπτομένης (Hyperbolic Tangent Function). Η συνάρτηση αυτή μας επιστρέφει τον λόγο της συνάρτησης υπερβολικού ημιτόνου προς την συνάρτηση υπερβολικού συνημιτόνου. Η συνάρτηση ορίζεται από τον παρακάτω τύπο:

$$\tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1} \quad (2.10)$$

2.3.3 Αλγόριθμοι Βελτιστοποίησης

Οι αλγόριθμοι μηχανικής μάθησης συνδέονται άμεσα με την βελτιστοποίηση καθώς έχουν στόχο την μεγιστοποίηση ή την ελαχιστοποίηση της συνάρτησης κόστους $L(\theta)$, όπου θ είναι το διάνυσμα των παραμέτρων του νευρωνικού δικτύου. Η βελτιστοποίηση συνάρτησης κόστους επιτυγχάνεται κατά την διάρκεια της εκπαίδευσης του συστήματος με την ανανέωση των παραμέτρων του. Οι μέθοδοι που καθορίζουν την ανανέωση ονομάζονται αλγόριθμοι βελτιστοποίησης. Είναι φανερό ότι για την αποτελεσματική εκπαίδευση του συστήματος έχει ιδιαίτερη σημασία η επιλογή του κατάλληλου αλγορίθμου. Παρακάτω παρουσιάζονται ορισμένοι από τους συνηθέστερους αλγορίθμους βελτιστοποίησης:

Αλγόριθμος Επικλινούς Καθόδου (Gradient Descent Algorithm). Είναι ένας από τους πιο δημοφιλείς αλγόριθμους βελτιστοποίησης που χρησιμοποιούνται στα νευρωνικά δίκτυα ειδικά σε προβλήματα κατηγοριοποίησης και παλινδρόμησης. Ο αλγόριθμος αυτός ελαχιστοποιεί την συνάρτηση κόστους ακολουθώντας το ελάχιστο της κλίσης (gradient) με την κλίση να υπολογίζεται από τον αλγόριθμο οπίσθιας διάδοσης. Επιπλέον μέσω της οπίσθιας διάδοσης το σφάλμα διαδίδεται από τα τελευταία επίπεδα του δικτύου προς τα αρχικά, με τα βάρη κάθε επιπέδου να ανανεώνονται με τρόπο τέτοιο ώστε να ελαχιστοποιηθεί η συνάρτηση κόστους. Αν και είναι εύκολος στην υλοποίηση και στους υπολογισμούς, ο αλγόριθμος αυτός δεν συνιστάται για μεγάλα σύνολα δεδομένων εκπαίδευσης καθώς τα βάρη ανανεώνονται μετά τον υπολογισμό της κλίσης πάνω σε ολόκληρο το σύνολο δεδομένων. Αυτό έχει ως αποτέλεσμα ο αλγόριθμος να συγκλίνει πολύ αργά και να απαιτεί μεγάλο χώρο στην μνήμη. Τα ανανεωμένα βάρη του νευρωνικού υπολογίζονται ως εξής:

$$\theta_{new} = \theta_{old} - \varepsilon \frac{1}{n} \nabla_{\theta} \sum_{i=1}^n L(f(x_i, \theta_{old}), y_i) \quad (2.11)$$

Όπου $\varepsilon \in R$, είναι μια παράμετρος που ονομάζεται ρυθμός μάθησης (learning rate) και καθορίζει το βήμα σύγκλισης σε ένα ελάχιστο, f ή h είναι η συνάρτηση που προσεγγίζει την συνάρτηση στόχο (βλέπε υποενότητα 2.2.2), n είναι το πλήθος των στιγμιотύπων x_i του συνόλου εκπαίδευσης και y_i η μεταβλητή στόχος.

Αλγόριθμος Στοχαστικής Επικλινούς Καθόδου (Stochastic Gradient Descent Algorithm). Ο αλγόριθμος αυτός είναι παραλλαγή του αλγορίθμου επικλινούς καθόδου με την διαφορά ότι τα βάρη ανανεώνονται έπειτα από τον υπολογισμό της κλίσης ενός στιγμιотύπου ή μιας παρτίδας

στιγμιοτύπων του συνόλου δεδομένων. Έτσι συγκλίνει γρηγορότερα σε ένα ελάχιστο και δεν έχει μεγάλες απαιτήσεις σε μνήμη καθώς δεν χρειάζεται να αποθηκεύει τα σφάλματα όλων των στιγμιοτύπων μέχρι να γίνει αλλαγή στα βάρη του μοντέλου. Συνεπώς ο αλγόριθμος αυτός είναι κατάλληλος για μεγάλα σύνολα δεδομένων. Επιλέγοντας μια παρτίδα m στιγμιοτύπων από το σύνολο εκπαίδευσης και τις αντίστοιχες μεταβλητές στόχους y_i η ανανέωση των βαρών υπολογίζεται από τον παρακάτω τύπο:

$$\theta_{new} = \theta_{old} - \varepsilon \frac{1}{m} \nabla_{\theta} \sum_{i=1}^m L(f(x_i, \theta_{old}), y_i) \quad (2.12)$$

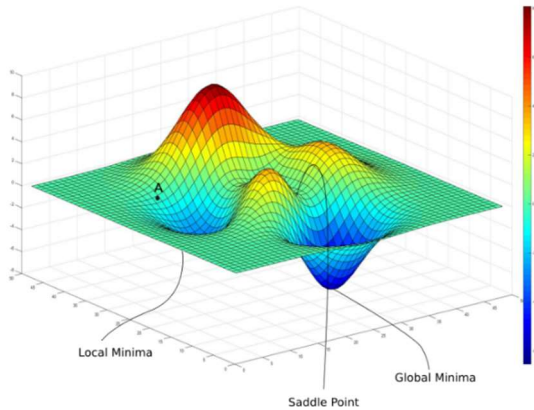
Αλγόριθμοι με ορμή (Momentum) ([22]). Όταν οι καμπύλες των συναρτήσεων κόστους περιέχουν πολλά ακρότατα (βλέπε σχήμα 2.5(α)) τότε η εκπαίδευση του μοντέλου αργεί ακόμα και με τον αλγόριθμο στοχαστικής επικλινούς καθόδου. Αυτό συμβαίνει λόγω των έντονων ταλαντώσεων γύρω από τα τοπικά ελάχιστα που συναντάει ο αλγόριθμος (βλέπε σχήμα 2.5(β)). Το πρόβλημα αυτό αντιμετωπίζεται με την εισαγωγή της μεθόδου της ορμής $\mathbf{u}$ ([23]). Στους αλγορίθμους με ορμή λαμβάνονται υπόψιν στον υπολογισμό των βαρών οι κλίσεις που υπολογίστηκαν σε προηγούμενα βήματα. Κατά αυτό το τρόπο επιταχύνεται η διαδικασία σύγκλισης οδηγώντας τις παραμέτρους προς την σχετική κατεύθυνση και εξομαλύνοντας τις ταλαντώσεις στις άσχετες κατευθύνσεις. Επιλέγοντας μια παρτίδα m στιγμιοτύπων από το σύνολο εκπαίδευσης και τις αντίστοιχες μεταβλητές στόχους y_i , η ανανέωση των βαρών ενός μοντέλου προσθέτοντας ορμή στον αλγόριθμο της στοχαστικής επικλινούς καθόδου υπολογίζεται ως εξής:

$$\mathbf{g} = \frac{1}{m} \nabla_{\theta} \sum_{i=1}^m L(f(x_i, \theta_{old}), y_i) \quad (2.13)$$

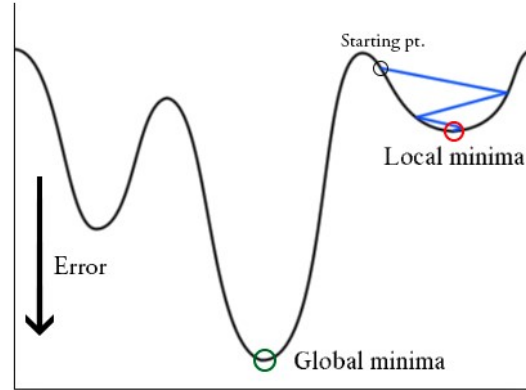
$$\mathbf{u}_{new} = \alpha \mathbf{u}_{old} - \varepsilon \mathbf{g} \quad (2.14)$$

$$\theta_{new} = \theta_{old} + \mathbf{u}_{new} \quad (2.15)$$

Όπου $\alpha \in [0,1]$ μια παράμετρος η τιμή της οποία καθορίζει κατά πόσο συνυπολογίζεται η τιμή της ορμής στα βάρη.



Σχήμα 2.5(α) Καμπύλη με πολλά ακρότατα.



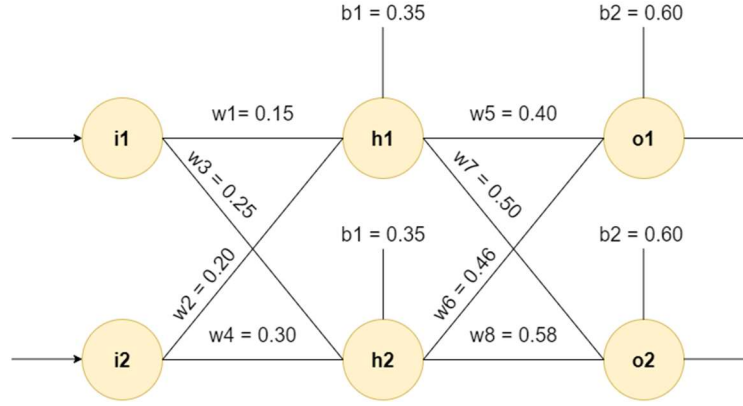
Σχήμα 2.5(β) Έντονες ταλαντώσεις γύρω από ένα τοπικό ελάχιστο.

Αλγόριθμος Adam. Ο αλγόριθμος Adam (Adaptive Moment Estimation) ([24]) είναι ένας αλγόριθμος που περιλαμβάνει την μέθοδο της ορμής. Ο αλγόριθμος χρησιμοποιεί τις τιμές των ροπών πρώτης (μέσος όρος) και δεύτερης (διακύμανσης) τάξης για την ανανέωση των παραμέτρων. Σε πολλά σύγχρονα συστήματα μηχανικής μάθησης γίνεται χρήση του αλγόριθμου Adam λόγω της ικανότητάς του να συγκλίνει πολύ γρήγορα και να αντιμετωπίζει τα προβλήματα που εμφανίζουν οι προηγούμενες τεχνικές.

2.3.4 Οπίσθια Διάδοση

Η οπίσθια διάδοση είναι ο αλγόριθμος με τον οποίο εκπαιδεύονται τα νευρωνικά δίκτυα, αρχικά υπολογίζοντας την κλίση της συνάρτησης κόστους και στην συνέχεια εφαρμόζοντας έναν αλγόριθμο βελτιστοποίησης. Για την κατανόηση της μεθόδου θα γίνει περιγραφή της εκπαίδευσης του νευρωνικού δικτύου. Δίνοντας μια είσοδο $\mathbf{x}$ σε ένα νευρωνικό δίκτυο η απαραίτητη πληροφορία περνάει από όλους τους νευρώνες του διασχίζοντας όλα τα επίπεδα του και στο τέλος επιστρέφεται μια έξοδος $\mathbf{y}$. Η διαδικασία αυτή ονομάζεται πρόσθια διάδοση ή πρόσθια τροφοδότηση (forward propagation). Κατά την εκπαίδευση η πρόσθια διάδοση συνεχίζεται μέχρι να υπολογιστεί το σφάλμα της συνάρτησης κόστους $J(\boldsymbol{\theta})$. Ο αλγόριθμος οπίσθιας διάδοσης επιτρέπει την πληροφορία του σφάλματος να μεταδοθεί προς τα πίσω ξεκινώντας από το τελευταίο επίπεδο με σκοπό τον υπολογισμό της κλίσης της συνάρτησης κόστους $\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$. Είναι σημαντικό εδώ να τονιστεί ότι ο αλγόριθμος οπίσθιας διάδοσης υπολογίζει μόνο την κλίση της συνάρτησης κόστους, η οποία χρησιμοποιείται στην συνέχεια για την ανανέωση των παραμέτρων του μοντέλου από έναν αλγόριθμο βελτιστοποίησης. Για τον υπολογισμό της κλίσης γίνεται χρήση του κανόνα της αλυσίδας ώστε να υπολογιστούν οι εξαρτήσεις όλων των παραγόντων που επιδρούν στην συνάρτηση κόστους.

Παράδειγμα 2.5. Έστω ένα νευρωνικό δίκτυο με ένα επίπεδο εισόδου, ένα κρυφό επίπεδο και ένα επίπεδο εξόδου. Η τοπολογία του δικτύου καθώς και τα βάρη των εισόδων κάθε νευρώνα δίνονται από το παρακάτω σχήμα:



Σχήμα 2.6 Νευρωνικό Δίκτυο Παραδείγματος 2.5

Δίνετε στο δίκτυο το στιγμιότυπο $(i1, i2) = (0.05, 0.10)$ το οποίο έχει επιθυμητή έξοδο $(o1, o2) = (0.01, 0.99)(*)$.

Αρχικά πραγματοποιείται η διαδικασία της πρόσθιας διάδοσης δίνοντας την είσοδο στο δίκτυο. Κάθε νευρώνας εκτός αυτών του επιπέδου εισόδου, αθροίζει τις εισόδους του και στην συνέχεια χρησιμοποιεί την συνάρτηση ενεργοποίησης για να καθορίσει την έξοδο του, θεωρείται ότι όλοι οι νευρώνες έχουν συνάρτηση ενεργοποίησης την σιγμοειδή. Αναλυτικά:

$$h1_{sum} = w_1 i_1 + w_2 i_2 + b_1 \quad (1)$$

$$h1_{sum} = 0.3775 \quad (2)$$

$$h1_{out} = \frac{e^{h1_{sum}}}{e^{h1_{sum}} + 1} = \frac{1}{e^{-h1_{sum}} + 1} = 0.593269992 \quad (3)$$

$$h2_{sum} = w_3 i_1 + w_4 i_2 + b_1 \quad (4)$$

$$h2_{sum} = 0.3925 \quad (5)$$

$$h2_{out} = 0.596884378 \quad (6)$$

$$o1_{sum} = w_5 h1_{out} + w_6 h2_{out} + b2 = 1.105905967 \quad (7)$$

$$o1_{out} = \frac{e^{o1_{sum}}}{e^{o1_{sum}} + 1} = \frac{1}{e^{-o1_{sum}} + 1} = 0.75136507 \quad (8)$$

$$o2_{sum} = w_7 h1_{out} + w_8 h2_{out} + b2 = 1.224921404 \quad (9)$$

$$o2_{out} = 0.7729284653 \quad (10)$$

Για τον υπολογισμό του σφάλματος, έστω ότι η συνάρτηση κόστους είναι η συνάρτηση μέσου τετραγωνικού σφάλματος (Mean Squared Error):

$$L_{total} = \sum_{i=1}^2 \frac{1}{2} (target - out)^2 \quad (2.16)$$

Τότε από την (2.16):

$$\begin{aligned} L_{total} &= L_{o1} + L_{o2} \Rightarrow \\ L_{total} &= \frac{1}{2} (target - o1_{out})^2 + \\ &\quad + \frac{1}{2} (target - o2_{out})^2 \xrightarrow{(*),(8),(10)} \\ L_{total} &= 0.274811083 + 0.023560026 = 0.298371109 \end{aligned} \quad (11)$$

Τώρα με την μέθοδο της οπίσθιας διάδοσης θα υπολογιστεί η κλίση της συνάρτησης κόστους ώστε στην συνέχεια με τον αλγόριθμο στοχαστικής επικλινούς καθόδου να ανανεωθούν τα βάρη του δικτύου.

Για την παράμετρο w_5 με χρήση του κανόνα της αλυσίδας:

$$\frac{\partial L_{total}}{\partial w_5} = \frac{\partial L_{total}}{\partial o1_{out}} \frac{\partial o1_{out}}{\partial o1_{sum}} \frac{\partial o1_{sum}}{\partial w_5} \quad (12)$$

$$\begin{aligned} \frac{\partial L_{total}}{\partial o1_{out}} &= \frac{\partial (\frac{1}{2} (target - o1_{out})^2 + \frac{1}{2} (target - o2_{out})^2)}{\partial o1_{out}} \\ &= -(target - o1_{out}) = 0.74136507 \end{aligned} \quad (13)$$

$$\begin{aligned} \frac{\partial o1_{out}}{\partial o1_{sum}} &= \frac{\partial (\frac{1}{e^{-o1_{sum}} + 1})}{\partial o1_{sum}} \Rightarrow \\ \frac{\partial o1_{out}}{\partial o1_{sum}} &= o1_{out}(1 - o1_{out}) = 0.18681502 \end{aligned} \quad (14)$$

$$\begin{aligned} \frac{\partial o1_{sum}}{\partial w_5} &= \frac{\partial (w_5 h1_{out} + w_6 h2_{out} + b2)}{\partial w_5} \Rightarrow \\ \frac{\partial o1_{sum}}{\partial w_5} &= h1_{out} = 0.593269992 \end{aligned} \quad (15)$$

Συνεπώς από (12),(13),(14),(15):

$$\frac{\partial L_{total}}{\partial w_5} = \frac{\partial L_{total}}{\partial o1_{out}} \frac{\partial o1_{out}}{\partial o1_{sum}} \frac{\partial o1_{sum}}{\partial w_5} = 0.593269992 \quad (16)$$

Η ανανέωση της παραμέτρου για ρυθμό μάθησης $\varepsilon = 0.5$ με τον αλγόριθμο βελτιστοποίησης της επικλινούς καθόδου υπολογίζεται ως:

$$w_{5_new} = w_{5_old} - 0.5 \frac{\partial L_{total}}{\partial w_{5_old}} = 0.3589648 \quad (17)$$

Ομοίως υπολογίζονται τα βάρη του επιπέδου εξόδου:

$$w_{6_new} = 0,408666186 \quad (18)$$

$$w_{7_new} = 0,511301270 \quad (19)$$

$$w_{8_new} = 0,561370121 \quad (20)$$

Για το κρυφό επίπεδο υπολογίζεται το w_1 :

$$\begin{aligned} \frac{\partial L_{total}}{\partial w_1} &= \frac{\partial L_{total}}{\partial h_{1_out}} \frac{\partial h_{1_out}}{\partial h_{1_sum}} \frac{\partial h_{1_sum}}{\partial w_1} \xrightarrow{(25)(26)(27)} \\ \frac{\partial L_{total}}{\partial w_1} &= 0.000438568 \end{aligned} \quad (20)$$

$$\frac{\partial L_{total}}{\partial h_{1_out}} = \frac{\partial L_1}{\partial h_{1_out}} + \frac{\partial L_2}{\partial h_{1_out}} \quad (21)$$

$$\begin{aligned} \frac{\partial L_1}{\partial h_{1_out}} &= \frac{\partial L_1}{\partial o_{1_sum}} \frac{\partial o_{1_sum}}{\partial h_{1_out}} = 0.138498562 * 0.4 \Rightarrow \\ \frac{\partial L_1}{\partial h_{1_out}} &= 0.055399425 \end{aligned} \quad (22)$$

$$\begin{aligned} \frac{\partial L_1}{\partial o_{1_sum}} &= \frac{\partial L_1}{\partial o_{1_out}} \frac{\partial o_{1_out}}{\partial o_{1_sum}} \Rightarrow \\ \frac{\partial L_1}{\partial o_{1_sum}} &= 0.74136507 * 0.18681502 \Rightarrow \\ \frac{\partial L_1}{\partial o_{1_sum}} &= 0.1384981304 \end{aligned} \quad (23)$$

$$\begin{aligned} \frac{\partial o_{1_sum}}{\partial h_{1_out}} &= w_5 = 0.4 \\ \frac{\partial L_2}{\partial h_{1_out}} &= \frac{\partial L_2}{\partial o_{2_out}} \frac{\partial o_{2_out}}{\partial o_{2_sum}} \frac{\partial o_{2_sum}}{\partial h_{1_out}} \Rightarrow \\ \frac{\partial L_2}{\partial h_{1_out}} &= -0.019049119 \end{aligned} \quad (24)$$

$$\begin{aligned} \frac{\partial L_{total}}{\partial h_{1_out}} &= 0.055399425 + (-0.019049119) \Rightarrow \\ \frac{\partial L_{total}}{\partial h_{1_out}} &= 0.036350306 \end{aligned} \quad (25)$$

$$\frac{\partial h_{1_out}}{\partial h_{1_sum}} = h_{1_out}(1 - h_{1_out}) \Rightarrow$$

$$\begin{aligned}\frac{\partial h_{1_{out}}}{\partial h_{1_{sum}}} &= 0.593269992(1 - 0.593269992) \Rightarrow \\ \frac{\partial h_{1_{out}}}{\partial h_{1_{sum}}} &= 0.2413007086\end{aligned}\quad (26)$$

$$\frac{\partial h_{1_{sum}}}{\partial w_1} = i_1 = 0.05 \quad (27)$$

Η νέα τιμή της παραμέτρου είναι:

$$w_{1_{new}} = w_{1_{old}} - 0.5 \frac{\partial L_{total}}{\partial w_{1_{old}}} = 0.149780716 \quad (28)$$

Ομοίως υπολογίζονται τα βάρη του κρυφού επιπέδου:

$$w_{2_{new}} = 0.19956143 \quad (29)$$

$$w_{3_{new}} = 0.24975114 \quad (30)$$

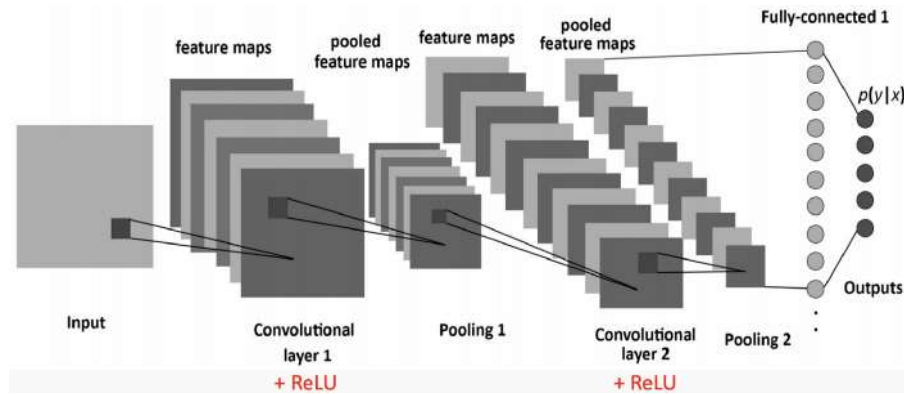
$$w_{4_{new}} = 0.29959229 \quad (31)$$

Ανανεώθηκαν όλα τα βάρη του δικτύου. Το σφάλμα τώρα είναι 0.291027924. Επαναλαμβάνοντας την διαδικασία αυτή 10000 φορές για το παραπάνω παράδειγμα το σφάλμα φτάνει την τιμή 0.0000351085. Σε αυτό το σημείο αν δοθεί η είσοδος (0.05, 0.1) η έξοδος του νευρωνικού θα είναι (0.015912196, 0.984065734) που είναι πολύ κοντά στην επιθυμητή έξοδο (0.15, 0.99).

2.3.4 Συνελικτικά Νευρωνικά Δίκτυα

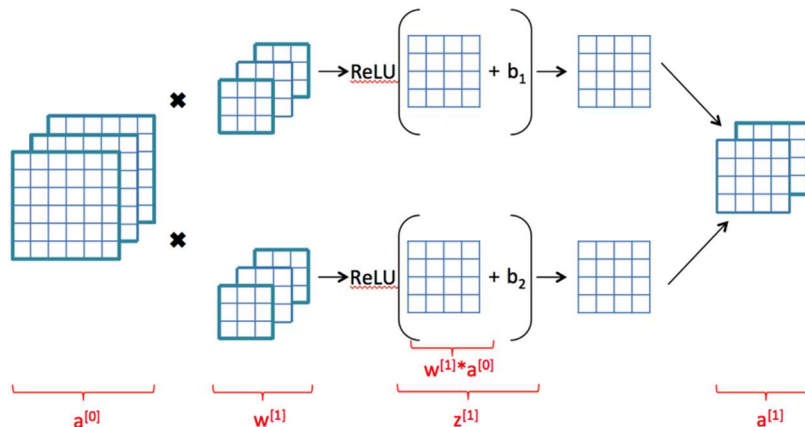
Τα συνελικτικά νευρωνικά δίκτυα ([20], [21]) είναι μια ειδική κατηγορία πρόσθιου νευρωνικού δικτύου, σχεδιασμένο για την επεξεργασία δεδομένων τα οποία είναι σε μορφή πίνακα, όπως είναι τα σήματα, οι εικόνες και τα βίντεο. Η λειτουργία των συνελικτικών δικτύων βασίζεται στην ύπαρξη συνδέσεων μεταξύ γειτονικών νευρώνων, στον διαμοιρασμό βαρών, στην χωρική υποδειγματοληψία (pooling) και στη χρήση πολλών επιπέδων.

Αναλυτικότερα η αρχιτεκτονική (βλέπε σχήμα 2.7) ενός συνελικτικού δικτύου αποτελείται από στάδια κάθε ένα από τα οποία περιέχει επίπεδα. Τα πρώτα στάδια περιλαμβάνουν δυο είδη επιπέδων: τα συνελικτικά επίπεδα και τα επίπεδα pooling.



Σχήμα 2.7 Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου

Οι νευρώνες ενός συνελικτικού επιπέδου χωρίζονται σε ομάδες οι οποίες είναι υπεύθυνες για την εξαγωγή ενός χάρτη χαρακτηριστικών (feature map). Συγκεκριμένα κάθε νευρώνας υπολογίζει την συνέλιξη ενός τμήματος από τους χάρτες χαρακτηριστικών που εξήγαγε το προηγούμενο επίπεδο με μια μήτρα φίλτρων (kernel). Το αποτέλεσμα της συνέλιξης περνάει από μια μη γραμμική συνάρτηση ενεργοποίησης όπως η ReLU (σχήμα 2.8) πριν γίνει η εισαγωγή στον χάρτη χαρακτηριστικών. Όλοι οι νευρώνες μιας ομάδας μοιράζονται την ίδια μήτρα φίλτρων, συνεπώς διαφορετικοί χάρτες χαρακτηριστικών χρησιμοποιούν διαφορετικές μήτρες.

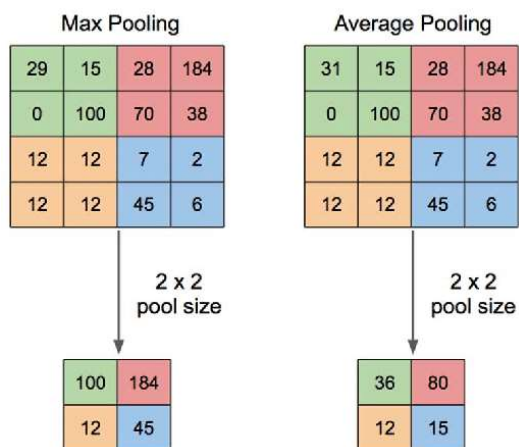


Σχήμα 2.8 Επίπεδο Συνέλιξης

Η αρχιτεκτονική αυτή στηρίζεται στο γεγονός ότι σε πίνακες δεδομένων όπως οι εικόνες, οι γειτονικές τιμές εικονοστοιχείων (pixels) συσχετίζονται μεταξύ τους σχηματίζοντας μοτίβα που μπορούν να εντοπιστούν εύκολα. Αυτό σε συνδυασμό με ότι τα μοτίβα αυτά δεν συναντώνται αποκλειστικά σε μια θέση ή μπορεί και να επαναλαμβάνονται σε μια εικόνα

καθιστά την χρήση κοινών φίλτρων αποδοτική. Ενώ η χρονική πολυπλοκότητα της πρόσθιας μετάδοσης παραμένει ίδια μειώνονται σημαντικά οι σημαντικά οι παράμετροι του μοντέλου.

Το επίπεδο που ακολουθεί το επίπεδο της συνέλιξης είναι το επίπεδο του pooling. Εκεί γίνεται διαλογή των χαρακτηριστικών που παρήγαγε το προηγούμενο επίπεδο διατηρώντας τα σημαντικότερα χαρακτηριστικά και συμπύσσοντας εκείνα που είναι σημασιολογικά παρόμοια. Οι νευρώνες για να κάνουν την διαλογή (βλέπε σχήμα 2.9) υπολογίζουν το μέγιστο (max) ή τον μέσο (average) του χάρτη χαρακτηριστικών. Ακόμα υπάρχει η δυνατότητα από τους γειτονικούς τους νευρώνες να επιλέξουν τμήμα του χάρτη προσπερνώντας έναν αριθμό γραμμών ή στηλών του χάρτη χαρακτηριστικών. Κατά αυτό το τρόπο επιτυγχάνεται μείωση στην διάσταση της αναπαράστασης.



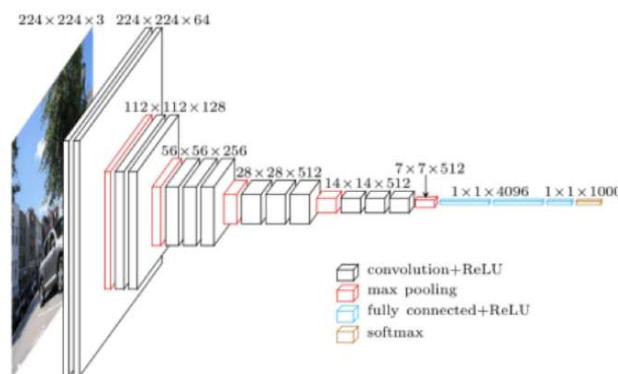
Σχήμα 2.9 Μέγιστη και Μέση Υποδειγματοληψία

Από όσα αναφέρθηκαν είναι φανερό ότι γίνεται αποτελεσματικά η εξαγωγή των απαραίτητων χαρακτηριστικών χωρίς να επιβαρύνεται το σύστημα με παραμέτρους. Η προσθήκη σταδίων αποτελούμενων από επίπεδα συνέλιξης και pooling επιτρέπει την ιεραρχική αναπαράσταση των χαρακτηριστικών, ξεκινώντας από χαρακτηριστικά χαμηλού επιπέδου και καταλήγοντας σε υψηλά χαρακτηριστικά. Τα στάδια αυτά μπορεί να ακολουθούνται από ένα στάδιο με πλήρως συνδεδεμένα επίπεδα στην περίπτωση που το δίκτυο προορίζεται για κατηγοριοποίηση. Τα συνελικτικά δίκτυα βρίσκουν εφαρμογές σε πολλά επιστημονικά πεδία που αφορούν ήχο, εικόνα και βίντεο. Ένα από αυτά είναι η ιατρική απεικόνιση, η οποία αποτελεί αντικείμενο μελέτης αυτής της εργασίας. Θα αναλυθούν παρακάτω δυο τύποι συνελικτικών νευρωνικών που θα χρησιμοποιηθούν στην συνέχεια.

2.3.5 VGG-16

Το VGG-16 είναι ένα συνελικτικό νευρωνικό δίκτυο το οποίο προτάθηκε από τους K.Simonyan και A.Zisserman ([25]) και θεωρείται μια από τις κορυφαίες αρχιτεκτονικές

για προβλήματα όρασης υπολογιστών. Η αρχιτεκτονική αυτή έκανε πρόοδο σε σχέση με το AlexNet ([26]) αντικαθιστώντας τα επίπεδα συνέλιξης με μεγάλες μήτρες φίλτρων με πολλά συνεχόμενα επίπεδα συνέλιξης που χρησιμοποιούν μήτρες φίλτρων 3x3. Με την χρήση αυτής της τεχνικής είναι δυνατή η προσομοίωση συνελικτικών επιπέδων με μεγάλες μήτρες διατηρώντας τα υπολογιστικά οφέλη των μικρότερων φίλτρων.

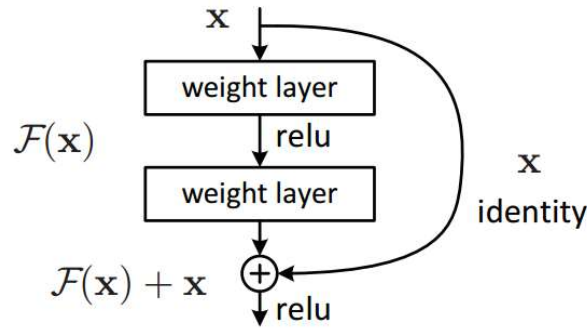


Σχήμα 2.10 Δομή VGG-16

Η δομή του VGG-16 ή VGG 16 επιπέδων (Σχήμα 2.10) περιλαμβάνει 13 συνελικτικά και 3 πλήρως συνδεδεμένα επίπεδα. Τα συνελικτικά συνελίσουν συνεχόμενα τα δεδομένα μετασχηματίζοντας τις δομές τους από την αρχική των διαστάσεων 224x224x3 σε μία αντίστοιχη διαστάσεων 14x14x512, ενώ τα 3 τελευταία πλήρως συνδεδεμένα επίπεδα υλοποιούν την ταξινόμηση και παράγουν το επιθυμητό 1x1000 διάνυσμα πιθανοτήτων.

2.3.6 ResNet-50

Το Residual Net ([27]) ή Resnet δημιουργήθηκε από την Microsoft και ήταν ο νικητής του ILSVRC το 2015. Το Resnet είναι βασισμένο στην δομική μονάδα που ονομάζεται residual block (Σχήμα 2.11). Σε αυτή παρατηρείται διαφορά από την συμβατική ροή επεξεργασίας των δικτύων στην οποία το κάθε επίπεδο λαμβάνει μία x είσοδο και παράγει μία έξοδο $F(x)$, με την $F(\cdot)$ να εξαρτάται από το είδος του επιπέδου. Στην περίπτωση του Resnet κάθε residual block παράγει μια έξοδο $H(x) = F(x) + x$.



Σχήμα 2.11 Residual Block

Έτσι, με αυτόν τον τρόπο αντί μετά από κάθε επίπεδο να παράγεται μία αντιπροσώπευση των δεδομένων η οποία δεν συνδέεται καθόλου με την προηγούμενη, τώρα κάθε block παράγει μία μεταβολή της εισόδου, η οποία προστίθεται στην πρώτη για να παραχθεί τροποποιημένη έκδοση της εισόδου. Ουσιαστικά, πρόκειται για μία εξέλιξη των δεδομένων καθώς περνούν μέσα από τα blocks για να καταλήξουν στο τελικό πλήρως συνδεδεμένο επίπεδο. Σύμφωνα με τους δημιουργούς του δικτύου είναι ευκολότερη η βελτιστοποίηση του συνδυασμού των επιπέδων με αυτό τον τρόπο παρά με τον συμβατικό τρόπο.

2.3.7 Αναδρομικά Νευρωνικά Δίκτυα

Τα αναδρομικά νευρωνικά δίκτυα (Recursive Neural Networks) ανήκουν σε μια κατηγορία νευρωνικών δικτύων που εξειδικεύεται στην επεξεργασία ακολουθιακών δεδομένων όπως για παράδειγμα τα καρτέ σε ένα βίντεο ή οι λέξεις σε ένα κείμενο. Προκειμένου να επεξεργαστούν και να αναλύσουν την ακολουθιακή πληροφορία των δεδομένων τα αναδρομικά δίκτυα μιμούνται τον άνθρωπο εισάγοντας την έννοια της μνήμης. Έτσι είναι σε θέση να θυμούνται εισόδους που χρησιμοποιήθηκαν για την παραγωγή προηγούμενων εξόδων. Η έξοδος του δικτύου δεν εξαρτάται μόνο από δεδομένα της εισόδου εκείνη την χρονική στιγμή και τις παραμέτρους του δικτύου αλλά και από ένα διάνυσμα “κρυφής” κατάστασης (hidden state) που αναπαριστά όλες τις παρελθοντικές εισόδους στο δίκτυο. Επομένως, ένα RNN για δύο περιπτώσεις με την ίδια είσοδο μπορεί να αποδώσει διαφορετικές εξόδους όταν τα προηγούμενα δεδομένα εισόδου είναι διαφορετικά σε κάθε περίπτωση.

Για το σχήμα 2.12 ο υπολογισμός της επόμενης κρυφής κατάστασης εξαρτάται από την είσοδο και την προηγούμενη κρυφή κατάσταση. Παρατηρείται ότι δεν έχει προσδιοριστεί η επιλογή της συνάρτησης ενεργοποίησης για τους νευρώνες της κρυφής κατάστασης, για αυτό θα επιλεγεί αυθαίρετα η συνάρτηση υπερβολικής εφαιπτομένης. Επιπλέον θεωρείται ότι η έξοδος o είναι διακριτή. Στην συνέχεια μπορεί να εφαρμοστεί συνάρτηση ενεργοποίησης Softmax. Η πρόσθια διάδοση στο δίκτυο του σχήματος 2.12 ξεκινά με την

αρχικοποίηση της κρυφής μνήμης $h^{(0)}$. Έπειτα σε κάθε βήμα από την χρονική στιγμή $t = 1$ ως την $t = \tau$ εφαρμόζονται οι εξής τύποι:

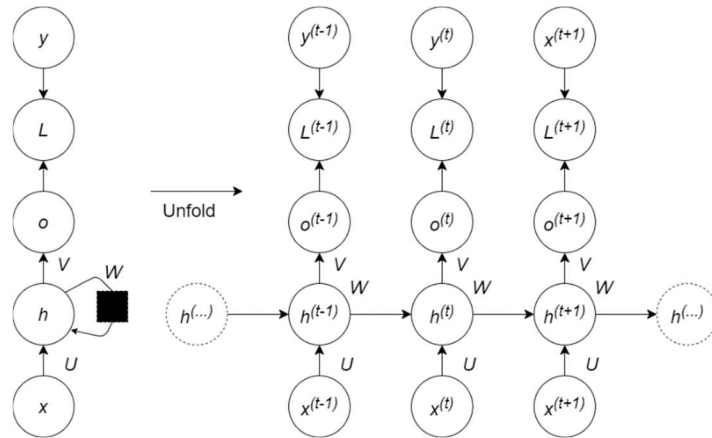
$$\mathbf{a}^{(t)} = \mathbf{b} + \mathbf{W}\mathbf{h}^{(t-1)} + \mathbf{U}\mathbf{x}^{(t)} \quad (2.17)$$

$$\mathbf{h}^{(t)} = \tanh(\mathbf{a}^{(t)}) \quad (2.18)$$

$$\mathbf{o}^{(t)} = \mathbf{c} + \mathbf{V}\mathbf{h}^{(t)} \quad (2.19)$$

$$\hat{\mathbf{y}} = \text{softmax}(\mathbf{o}^{(t)}) \quad (2.20)$$

Όπου $\mathbf{c}$ και $\mathbf{b}$ τα βάρη πόλωσης και $\mathbf{U}, \mathbf{V}, \mathbf{W}$ οι πίνακες παραμέτρων για την είσοδο του νευρώνα κρυφής κατάστασης, την έξοδο του νευρώνα κατάστασης και την μετάβαση στον επόμενο νευρώνα κατάστασης αντίστοιχα.



Σχήμα 2.12 Ακολουθιακό Δίκτυο

2.3.8 LSTM

Τα δίκτυα μακράς βραχυπρόθεσμης μνήμης (Long Short-Term Memory) ([28]) είναι αναδρομικά νευρωνικά δίκτυα τα οποία ανήκουν σε μια κατηγορία δικτύων που ονομάζεται αναδρομικά δίκτυα με πύλες (gated recurrent networks). Τα δίκτυα αυτά σε σχέση με τα απλά RNN επιτρέπουν την διαχείριση πληροφορίας για μεγάλη διάρκεια ενώ αν κρίνεται σκόπιμο η πληροφορία αυτή μπορεί να διατηρηθεί, να ανανεωθεί ή να διαγραφεί. Αυτό είναι ιδιαίτερα χρήσιμο σε περιπτώσεις όπου η ακολουθία που εξετάζεται αποτελείται

από υπακολουθίες στις οποίες αναζητείται πληροφορία, οπότε είναι απαραίτητος ένας μηχανισμός που να ανανεώνει την κατάσταση του δικτύου όσο εξετάζει μια υπακολουθία ενώ να την διαγράφει καθώς μεταβαίνει σε μια νέα. Επιπλέον τα LSTM αντιμετωπίζουν το πρόβλημα της εξαφανιζόμενης κλίσης που συναντάται στα απλά RNNs.

Τα LSTM διαφέρουν από τα κλασσικά RNNs στην δομή (βλέπε σχήμα 2.13) των μονάδων τους οποίες ονομάζονται κύτταρα (cell). Συγκεκριμένα εκτός του βρόχου τροφοδότησης προστίθεται εντός της μονάδας ένας εσωτερικός βρόχος. Ο βρόχος αυτός διατηρεί την κατάσταση κυττάρου (cell state) ώστε το δίκτυο να είναι ικανό να διαχειριστεί πληροφορία για μεγάλα διαστήματα. Ακόμη τα κύτταρα των LSTM εκτός της εισόδου και εξόδου διαθέτουν τρεις πύλες, την πύλη άγνοιας (forget gate), την πύλη εισόδου (input gate) και την πύλη εξόδου (output gate). Η λειτουργία των πυλών και κατά συνέπεια της δομικής μονάδας των LSTM θα γίνει κατανοητή μέσω της διαδικασίας της πρόσθιας διάδοσης, η οποία θα χωριστεί σε τέσσερα στάδια. Αναλυτικά για δεδομένη χρονική στιγμή t σε ένα κύτταρο i :

Στο πρώτο στάδιο η είσοδος $\mathbf{x}^{(t)}$ και η προηγούμενη κρυφή κατάσταση $\mathbf{h}^{(t-1)}$ του συστήματος εισάγονται στην πύλη άγνοιας η οποία είναι ένας νευρώνας με βάρη $\mathbf{U}^f, \mathbf{W}^f$, βάρη πόλωσης $\mathbf{b}^f$ και συνάρτηση ενεργοποίησης την σιγμοειδή. Η πύλη αυτή υπολογίζει μια τιμή $f_i^{(t)}$ μεταξύ του 0 και 1 που καθορίζει την πληροφορία που θα διατηρηθεί ή θα διαγραφεί στην συνέχεια από την νέα κατάσταση κυττάρου $\mathbf{s}^{(t)}$.

$$f_i^{(t)} = \sigma(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)}) \quad (2.21)$$

Το δεύτερο στάδιο περιλαμβάνει δύο βήματα. Στο πρώτο βήμα η είσοδος $\mathbf{x}^{(t)}$ και εισάγεται σε έναν νευρώνα με βάρη $\mathbf{U}, \mathbf{W}$, βάρη πόλωσης $\mathbf{b}$ και συνάρτηση ενεργοποίησης την σιγμοειδή προκειμένου να παράγει μια υποψήφια τιμή που μπορεί να προστεθεί στην συνέχεια στην τρέχουσα κατάσταση. Στο δεύτερο βήμα η είσοδος $\mathbf{x}^{(t)}$ και η προηγούμενη κρυφή κατάσταση $\mathbf{h}^{(t-1)}$ του συστήματος εισάγονται στην πύλη εισόδου η οποία είναι ένας νευρώνας με βάρη $\mathbf{U}^g, \mathbf{W}^g$, βάρη πόλωσης $\mathbf{b}^g$ και συνάρτηση ενεργοποίησης την σιγμοειδή. Η πύλη αυτή υπολογίζει μια τιμή $g_i^{(t)}$ μεταξύ του 0 και 1 που καθορίζει κατά πόσο θα υπολογιστεί η νέα υποψήφια τιμή στην νέα κατάσταση κυττάρου $\mathbf{s}^{(t)}$.

$$i_i^{(t)} = \sigma(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)}) \quad (2.22)$$

$$g_i^{(t)} = \sigma(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)}) \quad (2.23)$$

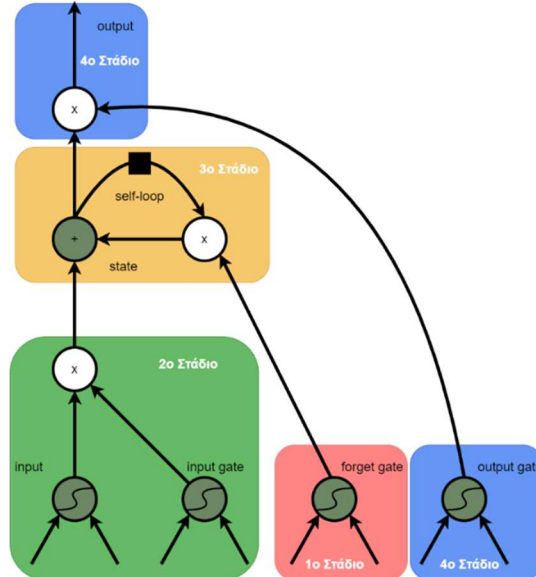
Στο τρίτο στάδιο θα γίνει ανανέωση της κατάστασης κυττάρου. Συγκεκριμένα η τιμή $f_i^{(t)}$ πολλαπλασιάζεται με την τρέχουσα κατάσταση κυττάρου $s^{(t-1)}$ έτσι ώστε να διατηρηθεί η πληροφορία που αποφασίστηκε στο πρώτο στάδιο. Στην συνέχεια προστίθεται το γινόμενο της νέας υποψήφιας τιμής με την τιμή $g_i^{(t)}$. Από το άθροισμα των δυο γινομένων προκύπτει η νέα κατάσταση κυττάρου $s^{(t)}$.

$$s^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} i_i^{(t)} \quad (2.24)$$

Τέλος στο τέταρτο στάδιο θα υπολογιστεί η έξοδος της πύλης εξόδου $q_i^{(t)}$ και η νέα κρυφή κατάσταση $h_i^{(t)}$. Αρχικά η είσοδος $x^{(t)}$ και η προηγούμενη κρυφή κατάσταση $h^{(t-1)}$ του συστήματος εισάγονται στην πύλη εισόδου η οποία είναι ένας νευρώνας με βάρη U^o, W^o , βάρη πόλωσης b^o και συνάρτηση ενεργοποίησης την σιγμοειδή. Η πύλη αυτή υπολογίζει μια τιμή $o_i^{(t)}$ μεταξύ του 0 και 1 που καθορίζει κατά πόσο θα υπολογιστεί η προηγούμενη κρυφή κατάσταση τιμή στον υπολογισμό της νέας κρυφής κατάστασης. Έπειτα η νέα κατάσταση κυττάρου $s^{(t)}$ εισάγεται σε μια σιγμοειδή συνάρτηση ενεργοποίησης και πολλαπλασιάζεται με την έξοδο $q_i^{(t)}$. Το αποτέλεσμα του γινομένου είναι η νέα κρυφή κατάσταση $h_i^{(t)}$.

$$q_i^{(t)} = \sigma(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)}) \quad (2.25)$$

$$h_i^{(t)} = \tanh(s_i^{(t)}) q_i^{(t)} \quad (2.26)$$



Σχήμα 2.13 Δομή κυττάρου LSTM

2.3.9 GRU

Το δίκτυο GRU (Gated Recurrent Unit) ([29]) είναι ένα αναδρομικό νευρωνικό δίκτυο με πύλες όπως και το LSTM. Η βασική διαφορά με το LSTM είναι ότι χρησιμοποιεί δυο πύλες αντί για τρεις αντικαθιστώντας την πύλη εισόδου και την πύλη άγνοιας με μια πύλη που ονομάζεται πύλη ανανέωσης, ενώ ταυτόχρονα οι κατάσταση κυττάρου και η κρυφή κατάσταση συγχωνεύονται. Έτσι το GRU χρησιμοποιεί λιγότερη μνήμη και εκπαιδεύεται γρηγορότερα από το LSTM, υστερεί όμως σε ακρίβεια σε ακολουθιακά δεδομένα με μεγάλο μήκος. Η συνάρτηση ανανέωσης δίνεται από την εξίσωση που ακολουθεί:

$$h_i^{(t)} = u_i^{(t-1)} h_i^{(t-1)} + (1 - u_i^{(t-1)}) \sigma(b_i + \sum_j U_{i,j} x_j^{(t-1)} + \sum_j W_{i,j} r_j^{(t-1)} h_j^{(t-1)}) \quad (2.27)$$

Όπου $\mathbf{u}$ είναι η έξοδος της πύλης ανανέωσης και $\mathbf{r}$ η έξοδος της πύλης επαναφοράς.

$$u_i^{(t)} = \sigma(b_i^u + \sum_j U_{i,j}^u x_j^{(t)} + \sum_j W_{i,j}^u h_j^{(t-1)}) \quad (2.28)$$

$$r_i^{(t)} = \sigma(b_i^r + \sum_j U_{i,j}^r x_j^{(t)} + \sum_j W_{i,j}^r h_j^{(t-1)}) \quad (2.29)$$

2.3.10 Bidirectional LSTM

Τα δίκτυα μακράς βραχυπρόθεσμης μνήμης διπλής κατεύθυνσης ([30], [31]) είναι μια επέκταση των τυπικών LSTM τα οποία έχουν σημαντικά καλύτερη απόδοση σε προβλήματα κατηγοριοποίησης ακολουθιών. Αυτά τα δίκτυα εκπαιδεύουν συγχρόνως δυο LSTM δίκτυα στην ακολουθία εισόδου, το ένα στην ακολουθία στην αρχική της μορφή και το δεύτερο σε ένα αντίγραφο της ακολουθίας που έχει αντίστροφη φορά. Αυτό το καινούργιο περιεχόμενο εισάγεται στο δίκτυο δημιουργώντας καλύτερα και γρηγορότερα αποτελέσματα.

2.2.11 Γενίκευση

Μια σημαντική πρόκληση που υπάρχει στο τομέα της μηχανικής μάθησης είναι το πόσο καλά αποδίδουν οι αλγόριθμοι σε νέα δεδομένα. Η ικανότητα ενός αλγορίθμου να αποδίδει καλά σε νέα δεδομένα ονομάζεται γενίκευση (generalization). Ένα σύστημα μηχανικής μάθησης κατά την εκπαίδευση στοχεύει στην ελαχιστοποίηση του σφάλματος εκπαίδευσης (training error). Πολλές φορές όμως αποτυγχάνει να διατηρήσει ταυτόχρονα χαμηλό και το σφάλμα γενίκευσης δηλαδή το σφάλμα στο σύνολο ελέγχου. Ανάλογα με το σφάλμα εκπαίδευσης και το σφάλμα γενίκευσης να μοντέλο μπορεί να χαρακτηριστεί ως υπερεκπαιδευμένο (overfitted) και υποεκπαιδευμένο (underfitted). Το μοντέλο χαρακτηρίζεται ως underfitted όταν έχει υψηλό σφάλμα εκπαίδευσης. Ενώ στην περίπτωση που

το μοντέλο διατηρεί χαμηλό σφάλμα εκπαίδευσης και έχει υψηλό σφάλμα γενίκευσης, τότε το μοντέλο έχει μάθει καλά το σύνολο εκπαίδευσης αλλά δεν μπορεί να γενικεύσει την γνώση του σε νέα δεδομένα, τέτοια μοντέλα ονομάζονται *overfitted*. Για την αντιμετώπιση αυτού του προβλήματος χρησιμοποιούνται τεχνικές γενίκευσης.

Εμπλουτισμός Δεδομένων (Data Augmentation). Η τεχνική αυτή αφορά την αύξηση του συνόλου δεδομένων εκπαίδευσης με προσθήκη μετασχηματισμών των υπαρχόντων δεδομένων. Σε εικόνες τέτοιοι μετασχηματισμοί μπορεί να είναι ολισθήσεις, περιστροφές, αλλαγή φωτεινότητας και προσθήκη θορύβου. Κατά αυτό τον τρόπο προστίθενται νέες “ψεύτικες” εικόνες που μπορούν να βοηθήσουν στην εκπαίδευση του συστήματος.

Dropout. Κατά την τεχνική του dropout επιλέγονται τυχαία νευρώνες ενός επιπέδου του δικτύου οι οποίοι δεν χρησιμοποιούνται στην εκπαίδευση. Συγκεκριμένα αυτοί οι κόμβοι δεν συμπεριλαμβάνονται στον υπολογισμό στην πρόσθια καθώς και στην οπίσθια διάδοση. Αυτό βοηθάει στην αποφυγή της προσαρμογής των νευρώνων στα χαρακτηριστικά του συνόλου εκπαίδευσης.

2.3.12 Transfer Learning

Η μεταφορική μάθηση (Transfer Learning) ([31]) είναι μια τεχνική μηχανικής μάθησης που επικεντρώνεται στην αποθήκευση της γνώσης που αποκτήθηκε, επιλύοντας ένα πρόβλημα και στη συνέχεια εφαρμόζοντας τη γνώση αυτή για την επίλυση ενός διαφορετικού αλλά σχετικού προβλήματος. Ωστόσο, στη γενική περίπτωση δεν είναι από πριν προφανές ότι η μεταφορική μάθηση θα είναι ωφέλιμη αυτό διαπιστώνεται μετά την ανάπτυξη και την αξιολόγηση του μοντέλου. Για το λόγο αυτό η χρήση των μηχανισμών μεταφορικής μάθησης δεν πρέπει να γίνεται αυθαίρετα αλλά μόνο μετά από προσεκτική μελέτη. Παράδειγμα μεταφορικής μάθησης είναι η χρήση προεκπαιδευμένων νευρωνικών δικτύων στο σύνολο δεδομένων ImageNet.

Κεφάλαιο 3

Επεξεργασία Δεδομένων

3.1 Εισαγωγή

Για την μελέτη ενός συστήματος μηχανικής μάθησης συχνά απαιτείται η επεξεργασία των σχετικών δεδομένων ώστε να είναι σε κατάλληλη μορφή για να ενσωματωθούν στο σύστημα. Στο κεφάλαιο αυτό θα γίνει αρχικά αναφορά στις αξονικές τομογραφίες θώρακος και στον τύπο των αρχείων που αποθηκεύονται και στην συνέχεια θα αναλυθούν οι μέθοδοι που χρησιμοποιήθηκαν για την εξαγωγή της απαραίτητης πληροφορίας για την διάγνωση του COVID-19, καθώς και για την προσαρμογή αυτής στο σύστημα. Τέλος θα παρουσιαστεί το σύνολο δεδομένων που χρησιμοποιήθηκε.

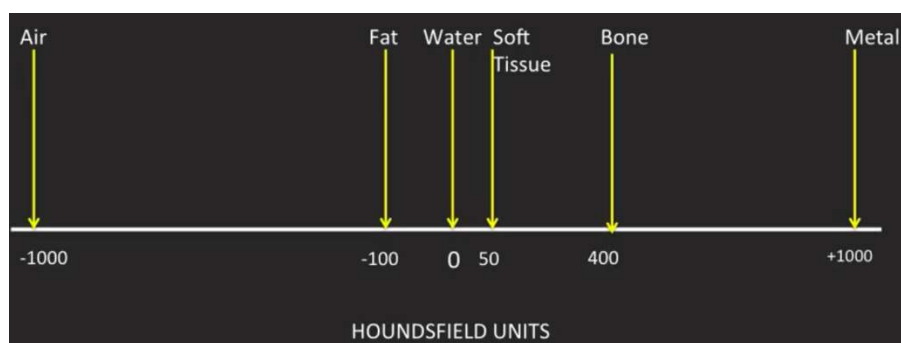
3.2 Δεδομένα

3.2.1 Αξονικές Τομογραφίες

Η αξονική τομογραφία είναι μία μη επεμβατική διαγνωστική εξέταση, η οποία χρησιμοποιεί ένα συνδυασμό των ακτίνων Χ και της τεχνολογίας των υπολογιστών, για να παράγει οριζόντιες αξονικές εικόνες (τομές) του σώματος που ονομάζονται φέτες (slices). Η αξονική τομογραφία μπορεί να απεικονίσει λεπτομερείς εικόνες από οποιοδήποτε μέρος του σώματος, συμπεριλαμβανομένων των οστών, των μυών και του λίπους. Οι εικόνες αυτές παρέχουν πολύ μεγαλύτερη δυνατότητα λεπτομερούς διάγνωσης από εκείνες της κλασικής ακτινογραφίας.

Κάθε εικόνα που παράγεται από έναν αξονικό τομογράφο είναι ένας χάρτης διαστάσεων 512x512 αποτελούμενος από τιμές που ονομάζονται συντελεστές εξασθένησης $\mu(x, y)$ και αφορούν την υπό απεικόνιση τομή. Για την ευκολότερη διαχείριση του χάρτη γίνεται μετατροπή κάθε συντελεστή εξασθένησης σε αριθμούς CT με μονάδα μέτρησης το HU (Hounsfield Unit). Είναι εύκολο να καταλάβει κανείς πως η απεικόνιση των εικόνων αυτών είναι μονοδιάστατη καθώς κάθε τιμή CT του χάρτη απεικονίζεται σε ένα pixel. Συγκεκριμένα για την αναπαράσταση της τιμής CT σε ένα pixel γίνεται αντιστοίχιση της κλίμακας των αριθμών CT σε κλίμακα αποχρώσεων του γκρι. Όσο μεγαλύτερος είναι ο αριθμός CT τόσο πιο ανοιχτό (λευκό) φαίνεται το pixel στην εικόνα (π.χ. οστό) ενώ όσο μικραίνει η τιμή του τόσο πιο σκούρο (μαύρο) φαίνεται. Αυτό έχει να κάνει με την πυκνότητα

των απεικονιζόμενων στοιχείων με την τιμή CT να αυξάνεται με την αύξηση της πυκνότητας. Στο σχήμα 3.1 παρουσιάζονται οι τιμές CT ορισμένων στοιχείων που παρατηρούνται σε μια αξονική τομογραφία. Είναι σημαντικό να αναφερθεί εδώ πως το ανθρώπινο μάτι μπορεί να ξεχωρίσει μόνο ένα συγκεκριμένο αριθμό αποχρώσεων του γκρι. Συνεπώς δεν έχει νόημα να γίνει αντιστοίχιση της πλήρους κλίμακας των αριθμών CT σε πλήρη κλίμακα αποχρώσεων του γκρι γιατί οι μικρές διαφορές στον αριθμό CT δεν θα είναι ορατές. Έτσι απεικονίζεται τμήμα της κλίμακας CT στην πλήρη κλίμακα αποχρώσεων του γκρι ενισχύοντας οπτικά μικρές διαφορές στον αριθμό CT. Το τμήμα αυτό της κλίμακας ονομάζεται παραθύρο (window) και χαρακτηρίζεται από το εύρος ή πλάτος (width) και από το κέντρο (center) ή επίπεδο (level).



Σχήμα 3.1 Κλίμακα αριθμών CT

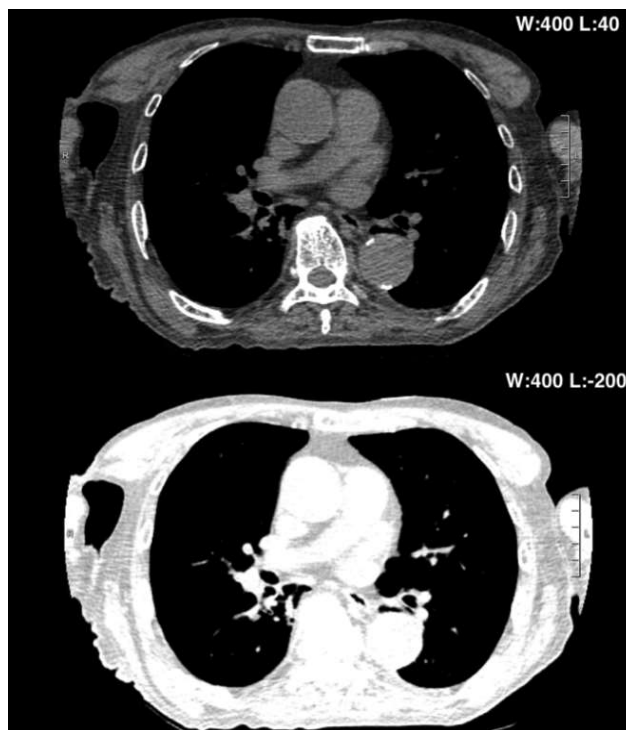
Το εύρος του παραθύρου καθορίζει την έκταση της διακύμανσης δηλαδή τον αριθμό διαφορετικών αριθμών CT που θα αντιπροσωπεύονται από διαφορετικές αποχρώσεις του γκρι. Οποιαδήποτε τιμή CT είναι μικρότερη από το κατώτατο όριο του πλάτους του παραθύρου θα εμφανίζεται ως μαύρο, ενώ οποιαδήποτε τιμή είναι μεγαλύτερη από το ανώτερο όριο θα εμφανίζεται ως άσπρο, κατά αυτόν τον τρόπο επηρεάζεται η αντίθεση της εικόνας. Αναλυτικότερα η αύξηση του πλάτους του παραθύρου (βλέπε σχήμα 3.2(α)) μειώνει την αντίθεση καθώς απαιτείται μεγαλύτερη αλλαγή στην πυκνότητα για να αλλάξει η απόχρωση του γκρι που αντιπροσωπεύει μια συγκεκριμένη τιμή CT. Αυτό έχει ως αποτέλεσμα στην απεικόνιση μιας εικόνας τα περισσότερα στοιχεία που αναπαρίστανται να φαίνονται παρόμοια παρά τις διάφορες στις πυκνότητες τους. Αντίθετα η μείωση του πλάτους του παραθύρου θα αυξήσει την αντίθεση της εικόνας καθώς μια μικρή αλλαγή στην πυκνότητα θα οδηγήσει σε μεγάλη αλλαγή του χρώματος. Έτσι, στοιχεία που διαφέρουν λίγο στην πυκνότητα θα αντιστοιχούν σε διαφορετικές αποχρώσεις του γκρι.

Το κέντρο του παραθύρου καθορίζει το κεντρικό αριθμό της απεικονιζόμενης εικόνας και ταυτόχρονα την θέση του παραθύρου πάνω στην κλίμακα CT. Με την αύξηση του κέντρου του παραθύρου μειώνεται η φωτεινότητα της εικόνας. Αυτό συμβαίνει γιατί καθώς αυξάνεται το κέντρο του παραθύρου, θα απαιτούνται υψηλότερες τιμές CT, προκειμένου η πυκνότητα να αναπαρασταθεί ως λευκή. Αντίθετα η μείωση του κέντρου του

παραθύρου (βλέπε σχήμα 3.2(β)) θα αυξήσει τη φωτεινότητα της εικόνας αφού πλέον για την εμφάνιση μιας πυκνότητας ως λευκή απαιτούνται χαμηλότερες τιμές CT.



Σχήμα 3.2(α) Μείωση της αντίθεσης με την αύξηση του πλάτους παραθύρου.



Σχήμα 3.2(β) Αύξηση της φωτεινότητας με μείωση του κέντρου του παραθύρου.

Η επιλογή του κατάλληλου παραθύρου για την απεικόνιση των τομογραφιών που συνοδεύουν μια μελέτη συμβάλλει καθοριστικά στην σωστή ερμηνεία της παθογένειας και της ανατομίας του ασθενούς. Αυτό συμβαίνει διότι ανάλογα με το παράθυρο που επιλέγεται διακρίνονται συγκεκριμένα στοιχεία της αξονικής. Για αυτό τον λόγο οι αξονικές τομογραφίες πρέπει να απεικονίζονται επιλέγοντας το κατάλληλο εύρος και το κατάλληλο κέντρο παραθύρου πριν την ερμηνεία τους από έναν ακτινολόγο. Επομένως μια εικόνα μπορεί να έχει πολλές αναπαραστάσεις. Στην μελέτη των αξονικών στήθους ανάλογα με την πληροφορία που απαιτείται χρησιμοποιούνται τα εξής παράθυρα:

Παράθυρο Πνεύμονα. Για μήκος παραθύρου 1500 και κέντρο -500 απεικονίζεται (σχήμα 3.3(α)) το παράθυρο του πνεύμονα (lung window) το οποίο μπορεί να οπτικοποιήσει μια μεγάλη ποικιλία πυκνότητας στο στήθος όπως το παρέγχυμα του πνεύμονα καθώς και γειτονικά αιμοφόρα αγγεία. Το παράθυρο του πνεύμονα είναι χρήσιμο για την αξιολόγηση του παρεγχύματος του πνεύμονα. Ορισμένες παθογένειες στο στήθος (όπως συγκεκριμένοι τύποι οζιδίων) μπορούν να εκτιμηθούν μόνο σε αυτό το παράθυρο.

Παράθυρο Οστού. Για μήκος παραθύρου 2000 και κέντρο 500 απεικονίζεται (σχήμα 3.3(β)) το παράθυρο οστού (bone window) όπου διακρίνονται τα οστά.

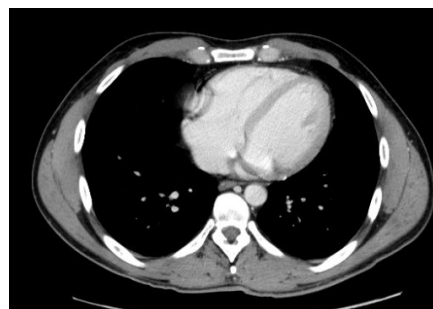
Παράθυρο Μαλακού Ιστού. Για μήκος παραθύρου 350 και κέντρο 50 απεικονίζεται (σχήμα 3.3(γ)) το παράθυρο μαλακού ιστού (soft tissue window) όπου διακρίνονται οι μαλακοί ιστοί.



Σχήμα 3.3(α) Παράθυρο Πνεύμονα



Σχήμα 3.3(β) Παράθυρο Οστού



Σχήμα 3.3(γ) Παράθυρο Μαλακού Ιστού

Προκειμένου οι αξονικές τομογραφίες να μετασχηματιστούν σε μορφή κατάλληλη για επεξεργασία κρίνεται αναγκαία η εξοικείωση με τον τύπο των αρχείων που αποθηκεύονται.

3.2.2 DICOM

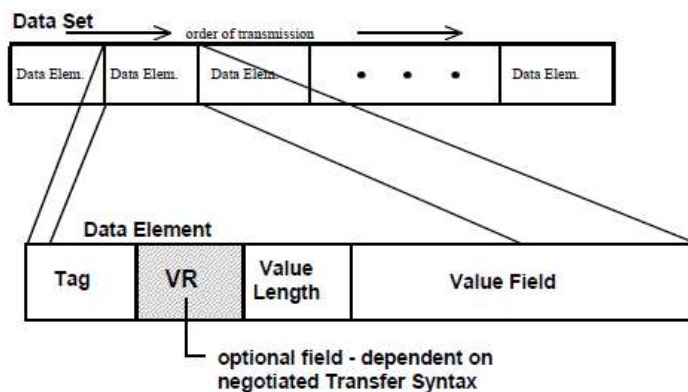
Το DICOM (Digital Imaging and Communications in Medicine) είναι το παγκόσμιο πρότυπο για την αποθήκευση και την μεταφορά ψηφιακών εικόνων μεταξύ ιατρικών μηχανημάτων καθώς και συστημάτων αρχειοθέτησης και επεξεργασίας ιατρικών εικόνων (PACS, Process Picture archiving and communication system) διαφορετικών κατασκευαστών. Αυτό το πρωτόκολλο χρησιμοποιείται και για την εξαγωγή των αξονικών τομογραφιών σε αρχεία.

Ειδικότερα κάθε slice της αξονικής τομογραφίας είναι αποθηκευμένο σε ένα αρχείο DICOM. Το αρχείο αυτό αποτελείται από την επικεφαλίδα (header) και το σύνολο δεδομένων (data set). Η επικεφαλίδα περιέχει το προοίμιο (preamble), το πρόθεμα (prefix) και ένα σύνολο μεταδεδομένων που περιέχουν πληροφορίες σχετικά με την κωδικοποίηση που έχει χρησιμοποιηθεί στο αρχείο. Με την επικεφαλίδα εξασφαλίζεται η συμβατότητα του αρχείου DICOM με πλήθος εφαρμογών. Το υπόλοιπο κομμάτι του αρχείου καταλαμβάνεται από το σύνολο δεδομένων και απαρτίζεται από τα στοιχεία δεδομένων (data elements). Είναι σημαντικό να αναφερθεί εδώ ότι κάθε στοιχείο δεδομένων όπως και τα μεταδεδομένα περιλαμβάνουν μέχρι τέσσερα πεδία (βλέπε σχήμα 3.4). Τα τρία από αυτά υπάρχουν σε όλα τα στοιχεία και είναι η ετικέτα του στοιχείου (tag), το μήκος της τιμής (value length) και το πεδίο τιμής (value field), το προαιρετικό πεδίο είναι η αναπαράσταση της τιμής (value representation).

Ετικέτα (tag). Η ετικέτα αποτελείται από δυο τιμές, την τιμή της ομάδας και την τιμή του στοιχείου στην ομάδα, ο συνδυασμός αυτός είναι μοναδικός για κάθε στοιχείο σε ένα αρχείο DICOM. Οι ομάδες αυτές μπορεί να σχετίζονται με τον ασθενή (π.χ. ηλικία, όνομα, ύψος), την εξέταση καθώς και με την εικόνα που περιέχεται στο αρχείο (π.χ. αριθμός γραμμών, αριθμός στηλών). Για παράδειγμα η ετικέτα (0028, 0010) προσδιορίζει το στοιχείο των γραμμών της εικόνας και περιέχει πληροφορία σχετικά με το ύψος της, ενώ η (0028, 0011) είναι η ετικέτα του στοιχείου των στηλών και περιέχει πληροφορία σχετικά με το πλάτος της εικόνας.

Αναπαράσταση Τιμής (Value Representation). Η αναπαράσταση τιμής περιέχει δυο χαρακτήρες που προσδιορίζουν το τύπο δεδομένων που περιέχονται σε ένα στοιχείο. Για παράδειγμα για το στοιχείο των γραμμών της εικόνας η αναπαράσταση τιμής είναι US (unsigned integer) που σημαίνει μη προσημασμένος αριθμός.

Μήκος Τιμής (Value Length) και Πεδίο Τιμής (Value Field). Επειδή το DICOM είναι δυαδικό πρωτόκολλο τα στοιχεία πρέπει να περιλαμβάνουν το μήκος της τιμής τους. Όλες οι τιμές των στοιχείων έχουν άρτιο μήκος. Για τα στοιχεία που έχουν αναπαράσταση τιμής αποτελούμενη από χαρακτήρες, όταν το πλήθος των χαρακτήρων τους είναι περιττό προστίθεται στο τέλος της συμβολοσειράς ένας κενός χαρακτήρας (Space character 20H) προκειμένου το στοιχείο να έχει τιμή περιττού μήκους. Η προσθήκη αυτή ισχύει για όλα τα στοιχεία με αναπαράσταση χαρακτήρων εκτός των στοιχείων με αναπαράσταση χαρακτήρων UI (Unique Identifier) και OB (other byte) όπου γίνεται προσθήκη μηδενικού χαρακτήρα (null character 00H). Για παράδειγμα το στοιχείο που προσδιορίζει το φύλο του ασθενούς έχει αναπαράσταση UI μπορεί να έχει τις εξής τιμές “M” (Male), “F” (Female) και “O” (Other), όταν χρησιμοποιείται μια από αυτές στο τέλος της προστίθεται ο μηδενικός χαρακτήρας ώστε να έχει άρτιο μήκος.



Σχήμα 3.4 Δομή Στοιχείου Δεδομένων

3.3 Προεπεξεργασία Αξονικών

Στην διάγνωση της νόσου του COVID-19 χρησιμοποιούνται αξονικές τομογραφίες με αναπαράσταση παραθύρου πνεύμονα. Από εκεί ένας έμπειρος ακτινολόγος μπορεί να διακρίνει αλλοιώσεις σχετικές με την νόσο. Επομένως για τα συστήματα που θα αναλυθούν στο επόμενο κεφάλαιο θα χρησιμοποιηθούν αξονικές τομογραφίες θώρακος όπου θα έχει εφαρμοστεί το παράθυρο του πνεύμονα.

3.3.1 Εφαρμογή παραθύρου πνεύμονα

Για να αναπαρασταθεί το παράθυρο πνεύμονα σε μια φέτα της αξονικής τομογραφίας είναι απαραίτητη η λήψη της απαραίτητης πληροφορίας από το σύνολο δεδομένων του αρχείου DICOM προκειμένου να μετασχηματιστεί κατάλληλα η εικόνα. Αρχικά πριν εφαρμοστεί το παράθυρο του πνεύμονα θα πρέπει να εξακριβωθεί ότι ο χάρτης της εικόνας να είναι εκφρασμένος σε μονάδες HU και στην περίπτωση που δεν είναι να γίνει μετατροπή.

Αναλυτικότερα θα αναζητηθούν τα στοιχεία Pixel Data, Rescale intercept, Rescale slope. Στο πρώτο στοιχείο ενθυλακώνεται ο χάρτης της εικόνας ο οποίος στις αξονικές τομογραφίες είναι συνήθως εκφρασμένος σε μονάδες HU. Αυτό μπορεί κανείς να διαπιστώσει ελέγχοντας τα στοιχεία Rescale intercept (σταθερά διαμόρφωσης) και Rescale slope (κλίση διαμόρφωσης). Σε περίπτωση που το πρώτο είναι 0 και το δεύτερο 1 ο χάρτης εκφράζεται σε μονάδες HU. Διαφορετικά, με την εφαρμογή (σχέση 3.1) των σταθερών αυτών σε κάθε στοιχείο του χάρτη της εικόνας γίνεται μετατροπή της εικόνας σε μονάδες HU.

$$HU(i) = rescale_slope * OV(i) + rescale_intercept \quad (3.1)$$

Όπου $HU(i)$ είναι η αναπαράσταση σε Hounsfield Units και $OV(i)$ οι αρχικές τιμές του χάρτη.

Αφού εξασφαλιστεί ότι ο πίνακας είναι εκφρασμένος σε τιμές HU είναι δυνατή η εφαρμογή του παραθύρου του πνεύμονα. Όπως αναφέρθηκε προηγουμένως στην ενότητα 3.2.1 για να γίνει απεικόνιση του παραθύρου του πνεύμονα πρέπει να εφαρμοστεί επίπεδο παράθυρου -500 και εύρος παραθύρου 1500. Έτσι τα στοιχεία του χάρτη που έχουν τιμή CT άνω των 250 HU θα γίνουν ίσα με την μέγιστη τιμή που αναπαρίσταται σε αυτό το παράθυρο δηλαδή 250 ώστε στην συνέχεια στην αντιστοίχιση της κλίμακας του παραθύρου με την κλίμακα αποχρώσεων του γκρι να γίνουν λευκά, ενώ εκείνα που έχουν τιμή κατώτερη

του κάτω ορίου να γίνουν ίσα με -1250 προκειμένου να γίνουν μαύρα. Κατά αυτό το τρόπο γίνεται διακριτή η επιθυμητή πληροφορία.

3.3.2 Προσαρμογή του μήκους της ακολουθίας

Οι αξονικές τομογραφίες αποτελούνται από έναν αριθμό φετών ο οποίος ποικίλει ανάλογα με τον αξονικό τομογράφο που χρησιμοποιείται καθώς και με την ρύθμιση που έχει επιλεγεί από τον χειριστή του. Για την ενσωμάτωσή τους στο σύστημα θα πρέπει όλες οι ακολουθίες να έχουν το ίδιο μήκος l . Προκειμένου να επιτευχθεί η προσαρμογή κάθε ακολουθίας σε ένα ενιαίο μήκος υλοποιήθηκαν δυο συναρτήσεις που δέχονται ως είσοδο μια ακολουθία $X = \{x_1, x_2, \dots, x_n\}$ και την επιθυμητή τιμή μήκους l και επιστρέφουν μια νέα ακολουθία $X' = \{x'_1, x'_2, \dots, x'_l\}$ βάσει της X με μήκος l . Αυτές είναι οι `remove_slices` στην περίπτωση όπου το μήκος της ακολουθίας είναι μεγαλύτερο από το l και η `add_slices` όταν το μήκος της είναι μικρότερο του l . Αναλυτικότερα για κάθε συνάρτηση:

Add_slices.

1. Υπολόγισε την ακέραια διαίρεση του επιθυμητού μήκους l με το μήκος ($|X|$) της ακολουθίας X . Έστω π το πηλίκο και v το υπόλοιπο.
2. Δημιούργησε μια νέα ακολουθία $X' = \{\underbrace{x_1 \dots x_1}_{\pi}, \underbrace{x_2 \dots x_2}_{\pi}, \dots, \underbrace{x_n \dots x_n}_{\pi}\}$ η οποία έχει ως στοιχεία υπακολουθίες από στοιχεία της αρχικής ακολουθίας X και μήκος $|\{\underbrace{x_1 \dots x_1}_{\pi}, \underbrace{x_2 \dots x_2}_{\pi}, \dots, \underbrace{x_n \dots x_n}_{\pi}\}| = \pi|X| = |X'|$.

3. Βρες την υπακολουθία στην οποία ανήκει ο **median** της νέας ακολουθίας και πήγαινε $\frac{v}{2}$ υπακολουθίες πίσω. Στις επόμενες v υπακολουθίες πρόσθεσε ένα ίδιο στοιχείο με αυτά που περιέχουν

$$|\{\underbrace{x_1 \dots x_1}_{\pi}, \underbrace{x_2 \dots x_2}_{\pi}, \dots, \underbrace{x_s \dots x_s}_{\pi+1}^{\text{median}-\frac{v}{2}}, \dots, \underbrace{x_k \dots x_k}_{\pi+1}^{\text{median}}, \dots, \underbrace{x_f \dots x_f}_{\pi+1}^{\text{median}+\frac{v}{2}}, \dots, \underbrace{x_n \dots x_n}_{\pi}\}| = \pi|X| + v = |X'| = l.$$

4. Επέστρεψε την ακολουθία X' .

Remove_slices.

1. Σύγκρινε το επιθυμητό μήκος l με το $|X| = n$, αν ισχύει η ανίσωση $n > l + 1$ συνέχισε στο βήμα 2, διαφορετικά προχώρα στο βήμα 5.
2. Υπολόγισε την διαφορά του μήκους ($|X| = n$) της ακολουθίας X με το επιθυμητό μήκος l . $d = |X| - l$
3. Υπολόγισε το πηλίκο της ακέραιας διαίρεσης του μήκους ($|X|$) της ακολουθίας X με την διαφορά που υπολογίστηκε στο βήμα 1. $\frac{|X|}{d} = \pi d + v$.
4. Αφαίρεσε τα στοιχεία που έχουν δείκτη πολλαπλάσιο του $(\pi + 1)$,

$$X = \{x_1, x_2, \dots, \overbrace{x_{\pi+1}}^{1^o}, \dots, \overbrace{x_{2\pi+2}}^{2^o}, \dots, \overbrace{x_{k\pi+k}}^{k^o \text{στο}}, \dots, x_n\}$$

$$X = \{x'_1, x'_2, \dots, x'_m\}$$

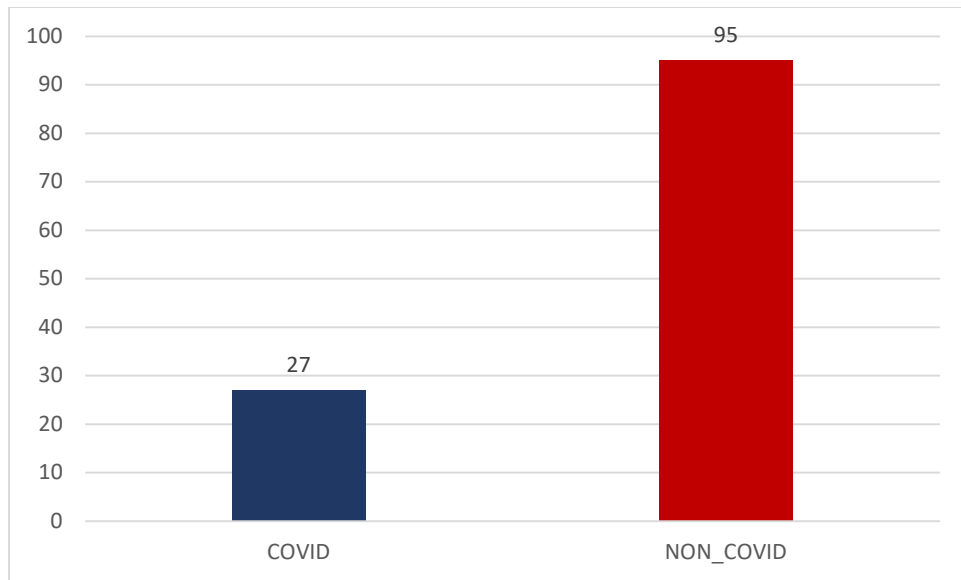
Θα αφαιρεθούν $k = \frac{|X|}{\pi+1}$ στοιχεία της ακολουθίας, οπότε το μήκος της ακολουθίας που προκύπτει είναι $m = n - k$.

5. Βρες τον median του της ακολουθίας X και αφάιρεσε τον.
6. Επέστρεψε την ακολουθία X' .

Συνεπώς με τις συναρτήσεις `add_slices` και `remove_slices`, κάθε ακολουθία μετασχηματίζεται σε μια νέα με μήκος l διατηρώντας την σειρά των εικόνων τους.

3.4 Σύνολο Δεδομένων

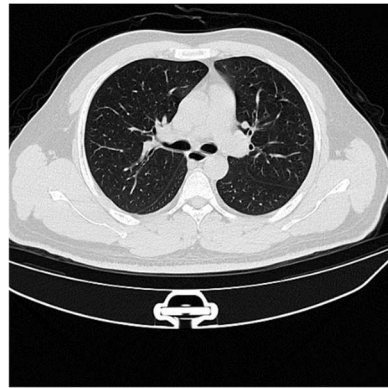
Το σύνολο δεδομένων αποτελείται από σύνολα δεδομένων 3 πηγών ([1], [2], [3]) Τα δεδομένα αυτά ήταν αρχικά αποθηκευμένα σε αρχεία DICOM και NIFTI (Neuroimaging Informatics Technology Initiative) και στην συνέχεια μετατράπηκαν σε μορφή PNG. Στα αρχεία DICOM πριν την μετατροπή τους σε PNG χρησιμοποιήθηκαν οι μέθοδοι που αναφέρθηκαν στο 3.3.1 ώστε να είναι εμφανής η απαραίτητη πληροφορία. Οι ακολουθίες που περιείχαν τα αρχεία NIFTI δεν απαιτούσαν επεξεργασία καθώς ήταν ήδη εκφρασμένες σε παράθυρο πνεύμονα. Συνολικά συγκεντρώθηκαν 27 σειρές αξονικών με διάγνωση COVID-19 και 95 στις οποίες δεν έχει διαγνωστεί η νόσος.



Σχήμα 3.5 Σύνολο Δεδομένων



Σχήμα 3.6(α) Αξονική Ασθενούς με COVID-19



Σχήμα 3.6(β) Αξονική Non_COVID Ασθενούς

Κεφάλαιο 4

Αρχιτεκτονικές Μοντέλων Διάγνωσης COVID-19

4.1 Εισαγωγή

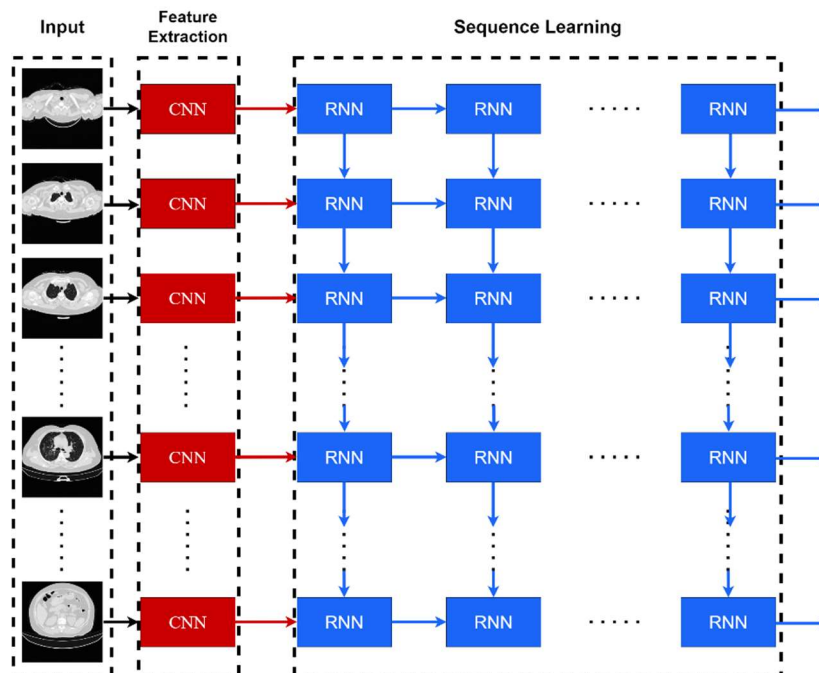
Στο κεφάλαιο αυτό θα γίνει παρουσίαση των μοντέλων που χρησιμοποιήθηκαν για την διάγνωση ασθενών COVID-19. Αρχικά θα γίνει ανάλυση της αρχιτεκτονικής CNN-RNN που χρησιμοποιείται στα νευρωνικά που θα μελετηθούν εκθέτοντας του λόγους που επιλέχθηκε για το πρόβλημα. Στην συνέχεια θα παρουσιαστούν οι παραλλαγές αυτής της αρχιτεκτονικής και το πως ενσωματώθηκαν τα δεδομένα σε αυτές.

4.2 CNN-RNN

Η αρχιτεκτονική CNN-RNN περιλαμβάνει ένα συνελικτικό δίκτυο (CNN) ακολουθούμενο από ένα αναδρομικό δίκτυο (RNN). Συναντάται στην μοντελοποίηση προβλημάτων όπου η είσοδος τους είναι μια ακολουθία ήχου ή εικόνας, όπως η όραση υπολογιστών (computer vision), η κατηγοριοποίηση (classification) και ο εντοπισμός αντικειμένων σε βίντεο (object detection). Το πρόβλημα της διάγνωσης COVID-19 από αξονικές τομογραφίες μπορεί να αναχθεί σε πρόβλημα κατηγοριοποίησης βίντεο. Συγκεκριμένα μια φέτα (slice) μια αξονικής μπορεί να αντιστοιχιστεί σε ένα καρέ (frame) ενός βίντεο καθώς και στις δυο περιπτώσεις υπάρχει εξάρτηση στην σειρά εμφάνισης. Η αρχιτεκτονική αυτή δέχεται ως είσοδο ακολουθίες αξονικών ιδίου μήκους και τις κατηγοριοποιεί σε δυο στις κατηγορίες Covid, non_Covid.

Στο πρώτο στάδιο της αρχιτεκτονικής για κάθε slice από τα l στο πλήθος της ακολουθίας εισόδου υπάρχει ένα CNN που υπολογίζει τους χάρτες των χαρακτηριστικών του. Η εξαγωγή των χαρακτηριστικών γίνεται ταυτόχρονα ενώ σημαντικό είναι να αναφερθεί ότι τα CNN έχουν την ίδια αρχιτεκτονική και μοιράζονται τα ίδια βάρη. Έπειτα οι χάρτες κάθε slice απλοποιούνται σε έναν με την χρήση ενός επιπέδου global pooling. Η διαφορά του global από το κλασσικό pooling είναι ότι επιλέγονται και απλοποιούνται τμήματα από όλους τους χάρτες χαρακτηριστικών με αποτέλεσμα να επιστρέφεται στην έξοδο του επιπέδου ένας χάρτης. Στο σημείο αυτό η αξονική που δόθηκε ως είσοδος έχει μετατραπεί σε μια ακολουθία από l χάρτες έναν για κάθε slice.

Στην συνέχεια τα χαρακτηριστικά αυτά τροφοδοτούνται στο δεύτερο στάδιο το οποίο περιλαμβάνει ένα αναδρομικό νευρωνικό δίκτυο. Ο ρόλος του σταδίου είναι η εύρεση πληροφορίας μεταξύ των slices, έτσι ώστε στην συνέχεια με την χρήση ενός επιπέδου εξόδου ή πλήρως συνδεδεμένων επιπέδων να γίνει η κατηγοριοποίηση των αξονικών.



Σχήμα 4.1 Αρχιτεκτονική CNN-RNN

4.3 Ενσωμάτωση των Δεδομένων

Η χρήση ενός μοντέλου κατηγοριοποίησης απαιτεί την ενσωμάτωση των δεδομένων έτσι ώστε να είναι σε κατάλληλη μορφή για επεξεργασία. Τα δεδομένα που θα τροφοδοτηθούν στα μοντέλα είναι αξονικές τομογραφίες οι οποίες μπορούν να θεωρηθούν αφαιρετικά ως ακολουθίες εικόνων. Οι ακολουθίες αυτές πρέπει να πληρούν δυο κριτήρια. Το πρώτο είναι να περιέχουν εικόνες στις οποίες είναι διακριτή η απαραίτητη πληροφορία για την κατηγοριοποίησή τους. Αυτό το κριτήριο έχει ικανοποιηθεί κατά την δημιουργία των συνόλων δεδομένων με την εφαρμογή του παραθύρου πνεύμονα σε κάθε εικόνα της αξονικής (βλέπε υποενότητα 3.3.1). Το δεύτερο κριτήριο αφορά τις διαστάσεις της ακολουθίας, δηλαδή το μήκος της και την διάσταση των εικόνων που περιλαμβάνει. Ειδικότερα οι ακολουθίες που εξετάζονται πρέπει να έχουν το ίδιο μήκος δηλαδή να περιέχουν το ίδιο πλήθος εικόνων ενώ όλες οι εικόνες της ακολουθίας πρέπει να έχουν τις ίδιες διαστάσεις. Η εξασφάλιση ενιαίου μήκους για όλες τις ακολουθίες πραγματοποιείται με την χρήση των

συναρτήσεων `add_slices` και `remove_slices` (βλέπε υποενότητα 3.3.2). Είναι σημαντικό να αναφερθεί ότι η επιλογή του μήκους ακολουθίας για όλα τα μοντέλα θα είναι ίση με το μήκος της μέγιστης ακολουθίας του συνόλου δεδομένων. Επιπλέον οι εικόνες γίνονται `re-scale` από 512×512 σε 224×224 προκειμένου να έχουν το ίδιο μέγεθος και να μειωθεί η υπολογιστική και χωρική πολυπλοκότητα του μοντέλου. Ακόμα θα προστεθούν δυο αντίγραφα της εικόνας προκειμένου να συμβαδίζει με τις διαστάσεις εισόδου των δικτύων καθώς δέχονται εικόνες με τρία κανάλια.

4.3 Παρουσίαση Αρχιτεκτονικών

Οι αρχιτεκτονικές που θα αναλυθούν στην συνέχεια στηρίζονται την αρχιτεκτονική CNN-RNN που παρουσιάστηκε στην ενότητα 4.2 συγκεκριμένα ως CNN θα χρησιμοποιηθεί το VGG-16. Αυτό ακολουθείται από ένα GRU. Η έξοδος του RNN θα δίνεται σε μια σιγμοειδή συνάρτηση ή σε μια σειρά πλήρως συνδεδεμένων επιπέδων. Τα μοντέλα χωρίζονται σε κατηγορίες ανάλογα με το στάδιο που ακολουθεί το RNN και σχετίζεται με την κατηγοριοποίηση των αξονικών.

4.3.1 Αρχιτεκτονική με επίπεδο εξόδου

Στο μοντέλο αυτό εισάγονται ακολουθίες μήκους n με εικόνες μεγέθους 224×224 . Τηρώντας την σειρά των εικόνων της ακολουθίας κάθε εικόνα δίνεται σε ένα προεκπαιδευμένο συνελικτικό VGG-16 (βλέπε 2.3.6) από το οποίο έχει αφαιρεθεί το στάδιο της κατηγοριοποίησης που περιλαμβάνει τα πλήρως συνδεδεμένα επίπεδα και έχει προστεθεί ένα επίπεδο `global max pooling`. Για να προσαρμοστεί η εικόνα στις διαστάσεις εισόδου του δικτύου την εισάγουμε με δυο αντίγραφα της προκειμένου να έχει διάσταση $224 \times 224 \times 3$. Η έξοδος του CNN έπειτα από το επίπεδο `global max pooling` είναι για κάθε εικόνα της ακολουθίας ένα διάνυσμα 512 χαρακτηριστικών το οποίο δίνεται σε ένα RNN.

Το RNN που επιλέγεται είναι ένα GRU που περιλαμβάνει n νευρώνες όσους και το μήκος της ακολουθίας. Οι νευρώνες δέχονται από έναν χάρτη χαρακτηριστικών τηρώντας ταυτόχρονα την σειρά στην ακολουθία. Το RNN προσπαθεί να εντοπίσει συσχετίσεις στην ακολουθία με κάθε νευρώνα να δίνει ως έξοδο ένα διάνυσμα 128 στοιχείων στο οποίο περιέχονται τα χαρακτηριστικά της ακολουθίας από την αρχή της μέχρι και την εικόνα που αντιστοιχεί σε αυτόν τον νευρώνα. Η έξοδος του τελευταίου νευρώνα περιέχει πληροφορία για ολόκληρη την ακολουθία και για αυτό δίνεται στην συνέχεια σε μια σειρά από πλήρως συνδεδεμένα επίπεδα που λειτουργούν ως ταξιμομητής. Συγκεκριμένα για την υλοποίηση της κατηγοριοποίησης η έξοδος του τελευταίου επιπέδου δίνεται σε μια σιγμοειδή συνάρτηση.

4.3.2 Αρχιτεκτονική με πλήρως συνδεδεμένα επίπεδα

Στο μοντέλο αυτό εισάγονται ακολουθίες μήκους n με εικόνες μεγέθους 224×224 . Τηρώντας την σειρά των εικόνων της ακολουθίας κάθε εικόνα δίνεται σε ένα προεκπαιδευμένο συνελικτικό VGG-16 (βλέπε 2.3.6) από το οποίο έχει αφαιρεθεί το στάδιο της κατηγοριοποίησης που περιλαμβάνει τα πλήρως συνδεδεμένα επίπεδα και έχει προστεθεί ένα επίπεδο global max pooling. Για να προσαρμοστεί η εικόνα στις διαστάσεις εισόδου του δικτύου την εισάγουμε με δυο αντίγραφα της προκειμένου να έχει διάσταση $224 \times 224 \times 3$. Η έξοδος του CNN έπειτα από το επίπεδο global max pooling είναι για κάθε εικόνα της ακολουθίας ένα διάνυσμα 512 χαρακτηριστικών το οποίο δίνεται σε ένα RNN.

Το RNN που επιλέγεται είναι ένα GRU που περιλαμβάνει n νευρώνες όσους και το μήκος της ακολουθίας. Οι νευρώνες δέχονται από έναν χάρτη χαρακτηριστικών τηρώντας ταυτόχρονα την σειρά στην ακολουθία. Το RNN προσπαθεί να εντοπίσει συσχετίσεις στην ακολουθία με κάθε νευρώνα να δίνει ως έξοδο ένα διάνυσμα 128 στοιχείων στο οποίο περιέχονται τα χαρακτηριστικά της ακολουθίας από την αρχή της μέχρι και την εικόνα που αντιστοιχεί σε αυτόν τον νευρώνα. Οι έξοδοι κάθε νευρώνα δίνονται σε μια σιγμοειδή συνάρτηση ενεργοποίησης. Συγκεκριμένα για την υλοποίηση της ταξινόμησης υπάρχουν δυο πλήρως συνδεδεμένα επίπεδα, το πρώτο επίπεδο συνθέτουν 32 νευρώνες με συνάρτηση ενεργοποίησης ReLU και δέχεται ως είσοδο το διάνυσμα 128 στοιχείων του τελευταίου νευρώνα του GRU. Ενώ το δεύτερο επίπεδο και τελευταίο του μοντέλου περιλαμβάνει 1 νευρώνα με συνάρτηση ενεργοποίησης την σιγμοειδή.

Κεφάλαιο 5

Πειραματική Διαδικασία

Στην ενότητα αυτή θα αναλυθεί η διαδικασία μάθησης των αρχιτεκτονικών με επίπεδο εξόδου και με πλήρως συνδεδεμένα επίπεδα ενώ στην συνέχεια θα γίνει αξιολόγηση της επίδοσης τους. Αναλυτικότερα από το σύνολο δεδομένων (βλέπε 3.4) το 80% χρησιμοποιήθηκε για την εκπαίδευση, ενώ με το υπόλοιπο 20% έγινε η αξιολόγηση των μοντέλων. Για την εκπαίδευση των μοντέλων επιλέχθηκε η συνάρτηση κόστους binary cross-entropy (βλέπε 2.2.3) και ως αλγόριθμος βελτιστοποίησης ο Adam με ρυθμό μάθησης 0.001, τιμή $\beta_1=0.9$ και τιμή $\beta_2=0.999$. Επιπλέον για την εκπαίδευση ορισμένων μοντέλων χρησιμοποιήθηκαν τεχνικές μεταφοράς μάθησης ενώ για όλα τα μοντέλα. Τέλος η αξιολόγηση των μοντέλων γίνεται βάση των μετρικών αξιολόγησης F1 score, Precision και Recall (βλέπε 2.2.4).

Για την επιλογή του μήκους της ακολουθίας θα υπολογίσουμε την ελάχιστη τιμή (min), την μέση τιμή (mean), τον διάμεσο (median) και την μέγιστη τιμή (max) του μήκους των ακολουθιών του συνόλου δεδομένων. Στα πειράματα θα εξετάσουμε το max της ακολουθίας.

Max: 418
Min: 39
Mean: 256.09016393442624
Median 252

Πίνακας 5.1 Στατιστικά Συνόλου Δεδομένων

5.1 Αρχιτεκτονική με επίπεδο εξόδου

Για την εκπαίδευση του μοντέλου στο σύνολο δεδομένων ο αριθμός εποχών επιλέχθηκε ίσος με 10 και το μέγεθος παρτίδας ίσο με 3.

Στο σχήματα απεικονίζονται οι συγκρίσεις των τιμών ακρίβειας εκπαίδευσης και ακρίβειας επαλήθευσης καθώς και οι τιμές σφάλματος εκπαίδευσης και σφάλματος επαλήθευσης. Επιπλέον οι επιδόσεις του μοντέλου παρουσιάζονται στους πίνακες.

- Για μήκος ακολουθίας 418 με χρήση προεκπαιδευμένου VGG-16 στο ImageNet :

	Precision	Recall	F1-score	support
COVID	0.75	0.60	0.67	5
Non_COVID	0.90	0.95	0.92	19
accuracy			0.88	24
Macro average	0.82	0.77	0.79	24
Weighted avg	0.87	0.88	0.87	24

Πίνακας 5.2 Μετρικές χωρίς επαύξηση

Το σύστημα κάνει σωστή πρόβλεψη για τους 3 από τους 5 ασθενείς Covid ενώ ανιχνεύει 18 στους 19 Non_Covid ασθενείς. Θα γίνει εφαρμογή της τεχνικής του εμπλουτισμού δεδομένων προκειμένου να εξεταστεί η δυνατότητα βελτίωσης της απόδοσης του συστήματος.

- Εμπλουτισμός δεδομένων με περιστροφή εύρους 10°.

	Precision	Recall	F1-score	support
COVID	1.00	0.60	0.75	5
Non_COVID	0.90	1.00	0.95	19
accuracy			0.92	24
Macro average	0.95	0.80	0.85	24
Weighted avg	0.92	0.92	0.91	24

Πίνακας 5.3 Μετρικές με επαύξηση εύρους 10°

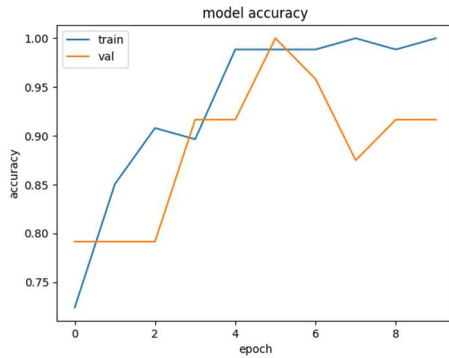
Παρατηρείται βελτίωση της απόδοσης του συστήματος. Παρότι η διάγνωση των COVID ασθενών είναι ίδια το σύστημα καταφέρνει να κάνει σωστή διάγνωση για όλους τους Non_COVID ασθενείς. Στην συνέχεια θα δοκιμαστεί επιπλέον η προσθήκη μετασχηματισμών των αρχικών δεδομένων με ολίσθηση εύρους 5 px.

- Εμπλουτισμός δεδομένων με περιστροφή εύρους 10° και ολίσθηση εύρους 5px.

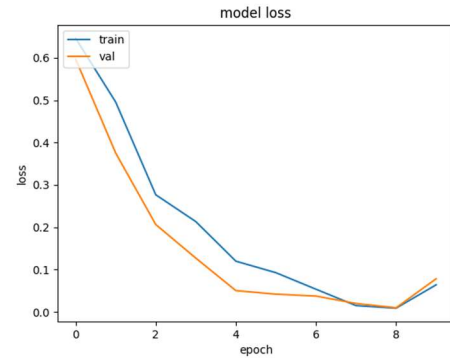
	Precision	Recall	F1-score	support
COVID	1.00	0.60	0.75	5
Non_COVID	0.90	1.00	0.95	19
accuracy			0.92	24
Macro average	0.95	0.80	0.85	24
Weighted avg	0.92	0.92	0.91	24

Πίνακας 5.4 Μετρικές με επαύξηση εύρους 10° και ολίσθηση εύρους 5 px

Η προσθήκη επιπλέον μετασχηματισμών δεν βελτίωσε την απόδοση του συστήματος. Στα σχήματα 5.1 και 5.2 παρουσιάζονται οι καμπύλες ακρίβειας και σφάλματος για την εκπαίδευση του συστήματος με εμπλουτισμό δεδομένων με περιστροφή εύρους 10° και ολίσθηση εύρους 5px.



Σχήμα 5.1 Καμπύλη Ακρίβειας



Σχήμα 5.2 Καμπύλη Σφάλματος

5.2 Αρχιτεκτονική με πλήρως συνδεδεμένα επίπεδα

Για την εκπαίδευση του μοντέλου στο σύνολο δεδομένων ο αριθμός εποχών επιλέχθηκε ίσος με 10 και το μέγεθος παρτίδας ίσο με 3.

Στο σχήματα απεικονίζονται οι συγκρίσεις των τιμών ακρίβειας εκπαίδευσης και ακρίβειας επαλήθευσης καθώς και οι τιμές σφάλματος εκπαίδευσης και σφάλματος επαλήθευσης αντίστοιχα. Επιπλέον οι επιδόσεις του μοντέλου παρουσιάζονται στους πίνακες.

- Για μήκος ακολουθίας 418 με χρήση προεκπαιδευμένου VGG-16 στο ImageNet :

	Precision	Recall	F1-score	support
COVID	1.00	0.80	0.89	5
Non_COVID	0.95	1.00	0.97	19
accuracy			0.96	24
Macro average	0.97	0.90	0.93	24
Weighted avg	0.96	0.96	0.96	24

Πίνακας 5.5 Μετρικές χωρίς επαύξηση

Η αρχιτεκτονική με πλήρως συνδεδεμένα επίπεδα αποδίδει καλύτερα στο σύνολο δεδομένων από την αρχιτεκτονική με το επίπεδο εξόδου χωρίς να γίνει εμπλουτισμός δεδομένων καθώς προβλέπει σωστά 4 στους 5 ασθενείς COVID και ανιχνεύει όλους τους Non_COVID ασθενείς. Θα εξεταστεί αν η εφαρμογή της τεχνικής εμπλουτισμού θα βελτιώσει την απόδοση του συστήματος.

- Εμπλουτισμός δεδομένων με περιστροφή εύρους 10°.

	Precision	Recall	F1-score	support
COVID	1.00	0.80	0.89	5
Non_COVID	0.95	1.00	0.97	19
accuracy			0.96	24
Macro average	0.97	0.90	0.93	24
Weighted avg	0.96	0.96	0.96	24

Πίνακας 5.6 Μετρικές με επαύξηση εύρους 10°

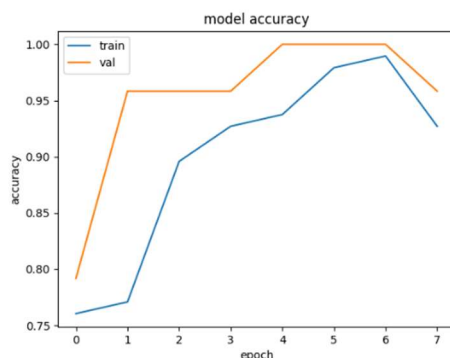
Δεν παρατηρείται βελτίωση της απόδοσης του συστήματος. Θα εξεταστεί αν η προσθήκη επιπλέον μετασχηματισμών βελτιώσει την απόδοση του συστήματος.

- Εμπλουτισμός δεδομένων με περιστροφή εύρους 10° και ολίσθηση εύρους 5px

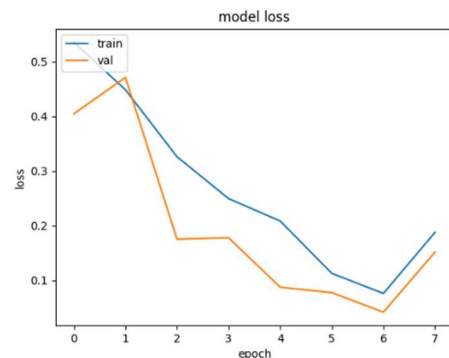
	Precision	Recall	F1-score	support
COVID	1.00	0.80	0.89	5
Non_COVID	0.95	1.00	0.97	19
accuracy			0.96	24
Macro average	0.97	0.90	0.93	24
Weighted avg	0.96	0.96	0.96	24

Πίνακας 5.7 Μετρικές με επαύξηση εύρους 10° και ολίσθηση εύρους 5 px

Η προσθήκη επιπλέον μετασχηματισμών δεν βελτίωσε την απόδοση του συστήματος. Στα σχήματα 5.3 και 5.4 παρουσιάζονται οι καμπύλες ακρίβειας και σφάλματος για την εκπαίδευση του συστήματος με εμπλουτισμό δεδομένων με περιστροφή εύρους 10° και ολίσθηση εύρους 5px.



Σχήμα 5.3 Καμπύλη Ακρίβειας



Σχήμα 5.4 Καμπύλη Σφάλματος

Κεφάλαιο 6

Συμπεράσματα και Μελλοντική Εργασία

Στην παρούσα διπλωματική έγινε ανάλυση ιατρικών δεδομένων που αφορούν αξονικές τομογραφίες προκειμένου να τεθούν σε κατάλληλη μορφή για την επεξεργασία τους από αρχιτεκτονικές βαθιών νευρωνικών δικτύων. Συγκεκριμένα εξετάστηκαν αρχιτεκτονικές CNN-RNN για την δημιουργία μοντέλων ανίχνευσης ασθενών COVID-19, επιτυγχάνοντας υψηλή ακρίβεια.

Σημαντική κρίνεται η εξακρίβωση αυτών των μοντέλων σε μεγαλύτερα σύνολα δεδομένων σε μελλοντική εργασία. Ακόμα ως μελλοντική εργασία μπορεί να θεωρηθεί ο εμπλουτισμός του συνόλου δεδομένων με παραγωγικά αντιπαραθετικά νευρωνικά δίκτυα (GANs) ώστε να δημιουργηθούν νέες συνθετικές εικόνες βάσει των εικόνων του συνόλου δεδομένων. Επιπλέον μπορούν να εξεταστούν τεχνικές κατάτμησης για την απομόνωση της περιοχής ενδιαφέροντος (ROI). Τέλος ως μελλοντική εργασία προτείνεται η εξέταση 3D συνελικτικών νευρωνικών δικτύων και η σύγκριση τους με την CNN-RNN αρχιτεκτονική.

Βιβλιογραφία

- [1] D. Kollias, A. Tagaris, A. Stafylopatis, S. Kollias, and G. Tagaris, “Deep neural architectures for prediction in healthcare,” *Complex & Intelligent Systems*, vol. 4, no. 2, pp. 119–131, Jun. 2018, doi: 10.1007/s40747-017-0064-6.
- [2] A. Tagaris, D. Kollias, and A. Stafylopatis, “Assessment of parkinson’s disease based on deep neural networks,” in *Communications in Computer and Information Science*, 2017, vol. 744, pp. 391–403, doi: 10.1007/978-3-319-65172-9_33.
- [3] A. Tagaris, Di. Kollias, A. Stafylopatis, G. Tagaris, and S. Kollias, “Machine Learning for Neurodegenerative Disorder Diagnosis - Survey of Practices and Launch of Benchmark Dataset,” *International Journal on Artificial Intelligence Tools*, vol. 27, no. 3, May 2018, doi: 10.1142/S0218213018500112.
- [4] J. Wingate, I. Kollia, L. Bidaut, and S. Kollias, “A Unified Deep Learning Approach for Prediction of Parkinson’s Disease,” *IET Image Processing*, vol. 14, no. 10, pp. 1980–1989, Nov. 2019, Available: <http://arxiv.org/abs/1911.10653>.
- [5] D. Kollias and S. P. Zafeiriou, “Exploiting multi-CNN features in CNN-RNN based Dimensional Emotion Recognition on the OMG in-the-wild Dataset,” *IEEE Transactions on Affective Computing*, Oct. 2020, doi: 10.1109/TAFFC.2020.3014171.
- [6] D. Kollias, M. Yu, A. Tagaris, G. Leontidis, A. Stafylopatis, and S. Kollias, “Adaptation and contextualization of deep neural network models,” in *2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017 - Proceedings*, Feb. 2018, vol. 2018-January, pp. 1–8, doi: 10.1109/SSCI.2017.8280975.
- [7] D. Kollias, A. Tagaris, and A. Stafylopatis, “On line emotion detection using retrainable deep neural networks,” Feb. 2017, doi: 10.1109/SSCI.2016.7850049.
- [8] D. Kollias, N. Bouas, Y. Vlaxos, V. Brillakis, M. Seferis, and I. Kollia, “Deep Transparent Prediction through Latent Representation Analysis,” pp. 1–16.
- [9] I. Kollia, A. G. Stafylopatis, and S. Kollias, “Predicting parkinson’s disease using latent information extracted from deep neural networks,” *arXiv*. arXiv, Jan. 23, 2019, Available: <http://arxiv.org/abs/1901.07822>.
- [10] Y. Avrithis, N. Tsapatsoulis, and S. Kollias, “Broadcast news parsing using visual cues: A robust face detection approach,” in *IEEE International Conference on Multi-Media and Expo*, 2000, no. III/WEDNESDAY, pp. 1469–1472, doi: 10.1109/icme.2000.871044.

- [11] K. Rapantzikos, N. Tsapatsoulis, Y. Avrithis, and S. Kollias, “Bottom-up spatiotemporal visual attention model for video analysis,” *IET Image Processing*, vol. 1, no. 2, pp. 237–248, 2007, doi: 10.1049/iet-ipr:20060040.
- [12] D. Kollias, G. Marandianos, A. Raouzaïou, and A. G. Stafylopatis, “Interweaving deep learning and semantic techniques for emotion analysis in human-machine interaction,” in *Proceedings - 10th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2015*, Dec. 2015, pp. 61–65, doi: 10.1109/SMAP.2015.7370086.
- [13] I. Kollia, N. Simou, A. Stafylopatis, and S. Kollias, “Semantic image analysis using a symbolic neural architecture,” *Image Analysis and Stereology*, vol. 29, no. 3, pp. 133–141, 2010, doi: 10.5566/ias.v29.p159-172.
- [14] B. Glimm, Y. Kazakov, I. Kollia, and G. Stamou, “Lower and upper bounds for SPARQL queries over OWL ontologies,” in *CEUR Workshop Proceedings*, 2015, vol. 1350.
- [15] D. Kollias, S. Cheng, M. Pantic, and S. Zafeiriou, “Photorealistic facial synthesis in the dimensional affect space,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Nov. 2019, vol. 11130 LNCS, pp. 475–491, doi: 10.1007/978-3-030-11012-3_36.
- [16] D. Kollias, S. Cheng, E. Ververas, I. Kotsia, and S. Zafeiriou, “Deep Neural Network Augmentation: Generating Faces for Affect Analysis,” *International Journal of Computer Vision*, vol. 128, no. 5, pp. 1455–1484, May 2020, doi: 10.1007/s11263-020-01304-3.
- [17] A. L. Samuel, “Some studies in machine learning using the game of checkers,” *IBM Journal of Research and Development*, vol. 44, no. 1–2, pp. 207–219, 2000, doi: 10.1147/rd.441.0206.
- [18] “Machine Learning textbook.” <http://www.cs.cmu.edu/~tom/mlbook.html>.
- [19] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [20] Y. Lecun *et al.*, “Handwritten digit recognition with a back-propagation network.” Morgan Kaufmann, 1990, Available: <https://nyuscholars.nyu.edu/en/publications/handwritten-digit-recognition-with-a-back-propagation-network-2>.
- [21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.

- [22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, vol. 44, no. 1. MIT Press, 2016.
- [23] B. T. Polyak, “Some methods of speeding up the convergence of iteration methods,” *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964, doi: 10.1016/0041-5553(64)90137-5.
- [24] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” Dec. 2015, Available: <https://arxiv.org/abs/1412.6980v9>.
- [25] “[1409.1556] Very Deep Convolutional Networks for Large-Scale Image Recognition.” <https://arxiv.org/abs/1409.1556>.
- [26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks.”, Available: <http://code.google.com/p/cuda-convnet/>.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2016, vol. 2016-December, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [28] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [29] K. Cho *et al.*, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” in *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, Jun. 2014, pp. 1724–1734, doi: 10.3115/v1/d14-1179.
- [30] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997, doi: 10.1109/78.650093.
- [31] L. Y. Pratt, “Discriminability-Based Transfer between Neural Networks,” in *Advances in neural information processing systems*, 1993, pp. 204–211.