



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας, Πληροφορικής & Υπολογιστών

Αποτελεσματική Ανάλυση 3-D Εικόνων για Ανίχνευση Covid-19

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΕΥΑΓΓΕΛΙΑ ΡΕΝΤΙΦΗ

Επιβλέπων : Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2022



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας, Πληροφορικής & Υπολογιστών

Αποτελεσματική Ανάλυση 3-D Εικόνων για Ανίχνευση Covid-19

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΕΥΑΓΓΕΛΙΑ ΡΕΝΤΙΦΗ

Επιβλέπων : Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 21η Φεβρουαρίου 2022.

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

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

.....
Γεώργιος Στάμιου
Καθηγητής Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2022

.....
Ευαγγελία Ρεντίφη

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Ευαγγελία Ρεντίφη, 2022.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

Το πρώτο κρούσμα της νόσου COVID-19 παρουσιάστηκε στα τέλη του 2019 στην πόλη Γουχάν της Κίνας. Εκ τότε ακολούθησε η ραγδαία και σε πολλές περιπτώσεις μη ελεγχόμενη εξάπλωση του ιού σε όλον τον πλανήτη με τον Παγκόσμιο Οργανισμό Υγείας να χαρακτηρίζει εντέλει την νόσο ως πανδημία, ενώ μέχρι σήμερα ο συνολικός αριθμός των κρουσμάτων ανέρχεται στα 415 εκατομμύρια.

Η ανάπτυξη αξιόπιστων διαγνωστικών μεθόδων κρίθηκε καθοριστικής σημασίας τόσο για την φροντίδα των ίδιων των ασθενών που πάσχουν, όσο και τον περιορισμό της διασποράς της νόσου μέσω εφαρμογής μέτρων κοινωνικής αποστασιοποίησης. Τα μοριακά τεστ RT-PCR καθώς και τα τεστ ταχείας ανίχνευσης αντιγόνου, γνωστά ως rapid test, αποτελούν τα πιο διαδεδομένα διαγνωστικά εργαλεία. Ωστόσο ο μεγάλος όγκος των διαγνωστικών ελέγχων που διενεργούνται καθημερινά, ειδικότερα σε περιόδους έξαρσης του ιού, καθώς και ο χαμηλός δείκτης ειδικότητας στην περίπτωση των rapid tests, επιβάλλουν την ανεύρεση νέων εναλλακτικών μεθόδων που θα τεθούν στην διάθεση του παγκόσμιου συστήματος υγείας.

Σκοπός της παρούσας εργασίας είναι η ανάπτυξη ενός ευφυούς συστήματος διάγνωσης της νόσου COVID-19, με την εφαρμογή μεθόδων βαθιάς μάθησης σε ακολουθίες αξονικών τομογραφιών θώρακος. Οι αξονικές τομογραφίες θώρακος χρησιμοποιούνταν μέχρι σήμερα για την διάγνωση πνευμονικών λοιμώξεων, καρκίνου του πνεύμονα, πνευμονικής εμβολής καθώς και άλλων πνευμονικών προβλημάτων ωστόσο οι μεταβολές που προκαλεί ο ιός στην εικόνα των πνευμόνων με την ανάπτυξη κυρίως θολών υάλων συνηγορούν στην χρήση τους και για την διάγνωση του COVID-19. Το αναπτυσσόμενο μοντέλο καταφέρνει να εκμεταλλευτεί τα επαναλαμβανόμενα μοτίβα στις εικόνες των ασθενών με COVID-19 και να παράγει έτσι την τελική διάγνωση.

Λέξεις κλειδιά

Βαθιά μάθηση, Συνελικτικά Νευρωνικά Δίκτυα, Ανατροφοδοτούμενα Νευρωνικά Δίκτυα, Κορονοϊός (Covid-19), Αξονική Τομογραφία

Abstract

The first case of COVID-19 disease occurred in late 2019 in Wuhan, China. Since then, the rapid and in many cases uncontrolled spread of the virus around the world has followed, leading the World Health Organization to characterize the disease as an ongoing pandemic, while the total number of reported cases is 415 million at the time of writing.

The development of reliable diagnostic methods was considered crucial both for the care of the patients themselves and for reducing the spread of the disease through measures of social distancing. RT-PCR molecular tests as well as rapid antigen tests, are the most common diagnostic tools. However, the large volume of diagnostic tests performed daily, especially during periods of virus outbreak, as well as the low specificity in the case of rapid tests, require new alternative methods that will be made available to the global health system.

The aim of the current study is the development of an intelligent diagnostic system for COVID-19 disease, by applying deep learning methods to chest CT scans. Chest CT scans have been used to diagnose lung infections, lung cancer, pulmonary embolism and other lung problems. However, the changes caused by the virus in the the lungs area, which mainly include ground glass opacities, leads to the use of chest CT scans for diagnosis of COVID-19 .The developing model manages to take advantage of the repetitive patterns in the images of COVID-19 patients and thus produce the final diagnosis.

Key words

Deep Learning, Convolutional Neural Networks (CNNs), Recursive Neural Networks (RNNs), Covid-19, Computerized Tomography (CT).

Ευχαριστίες

Η ολοκλήρωση της παρούσας διπλωματικής σηματοδοτεί το τέλος των προπτυχιακών σπουδών μου. Αισθάνομαι λοιπόν την ανάγκη να ευχαριστήσω όλους όσους στάθηκαν δίπλα μου αυτό το διάστημα.

Αρχικά οφείλω ένα μεγάλο ευχαριστώ στον κ. Στέφανο Κόλλια που μου έδωσε την ευκαιρία να εργαστώ πάνω σε ένα τόσο επίκαιρο και συνδυαστικό θέμα. Ακόμη οφείλω να ευχαριστήσω τον Δημήτρη Κόλλια και τον Αναστάση Αρσένο για την σημαντική καθοδήγηση και τις πολύτιμες παρατηρήσεις τους σε όλα τα στάδια εκπόνησης της διπλωματικής.

Παράλληλα πρέπει να ευχαριστήσω την οικογένεια μου και κυρίως την αδερφή μου (αλλά και συγκάτοικο μου) Χριστίνα για την αγάπη, την στήριξη και την κατανόηση τους σε όλες τις εύκολες αλλά και δύσκολες στιγμές. Τέλος ευχαριστώ από καρδιάς όλους τους φίλους μου τόσο για τις αξέχαστες αναμνήσεις (εντός και εκτός σχολής) στην διάρκεια αυτών των πέντε χρόνων όσο και για τις στιγμές που στάθηκαν δίπλα μου και με βοήθησαν να ξεπεράσω κάθε εμπόδιο.

Ευαγγελία Ρεντίφη,
Αθήνα, 21η Φεβρουαρίου 2022

Περιεχόμενα

Περίληψη	5
Abstract	7
Ευχαριστίες	9
Περιεχόμενα	11
Κατάλογος πινάκων	13
Κατάλογος σχημάτων	14
1. Εισαγωγή	17
Εισαγωγή	17
1.1 COVID-19	17
1.1.1 Προέλευση της νόσου COVID-19	17
1.1.2 Δομή και προσβολή ξενιστή	17
1.1.3 Μετάδοση και συμπτώματα	18
1.1.4 Διάγνωση	19
1.1.5 Πρόληψη	21
1.1.6 Θεραπεία	22
1.2 Σκοπός και δομή της εργασίας	22
2. Θεωρητικό Υπόβαθρο	23
Θεωρητικό Υπόβαθρο	23
2.1 Νευρωνικά Δίκτυα	23
2.1.1 Βιολογικά νευρωνικά δίκτυα	23
2.1.2 Τεχνητά Νευρωνικά Δίκτυα	24
2.1.3 Γιατί Νευρωνικά Δίκτυα ;	25
2.1.4 Συναρτήσεις ενεργοποίησης	26
2.1.5 Αλγόριθμοι Βελτιστοποίησης	28
2.1.6 Οπισθοδιάδοση (Back-Propagation)	32
2.2 Συνελκτικά Νευρωνικά Δίκτυα	34
2.3 Ανατροφοδοτούμενα Νευρωνικά Δίκτυα (RNN)	37
2.3.1 Δομή	37
2.3.2 Οπισθοδιάδοση μέσα στον χρόνο (Backpropagation through time)	38
2.4 Επεκτάσεις Ανατροφοδοτούμενων Νευρωνικών Δικτύων	40
2.4.1 Δίκτυα Μακράς Βραχέας Μνήμης (Long Short-Term Memory Network)	40
2.4.2 Μονάδα Ανάδρασης με Δρομολόγηση (Gated Recurrent Unit)	44
2.4.3 Αμφίδρομα Επαναληπτικά Νευρωνικά Δίκτυα (Bidirectional Recurrent Neural Networks)	45
2.5 Μεταφορά Μάθησης (Transfer Learning)	47

3. Ανάλυση Μεθοδολογίας	50
Ανάλυση Μεθοδολογίας	50
3.1 Αρχιτεκτονική Δικτύου	51
3.2 BCDU-Net	51
3.2.1 Κωδικοποιητής (encoding path)	52
3.2.2 Αποκωδικοποιητής (decoding path)	53
3.2.3 Κανονικοποίηση παρτίδας (batch normalization)	53
3.2.4 Bi-Directional ConvLSTM	55
3.3 ResNet	56
3.3.1 Residual blocks	57
3.3.2 Εφαρμογές	58
3.4 EfficientNet	58
3.4.1 EfficientNet-B0	59
3.4.2 Inverted Bottleneck MBConv	59
3.4.3 Squeeze-and-Excitation	61
3.4.4 Εφαρμογές	61
3.5 Inception	62
3.5.1 Εφαρμογές	64
3.6 CNN-RNN	65
4. Πειραματική διαδικασία	67
Πειραματική διαδικασία	67
4.1 Μετρικές Αξιολόγησης	67
4.2 Μείωση ct scan volume	68
4.3 Κατηγοριοποίηση ct scans σε COVID-19/non-COVID-19	69
5. Συμπεράσματα και Μελλοντική Εργασία	73
Συμπεράσματα και Μελλοντική Εργασία	73
Βιβλιογραφία	74

Κατάλογος πινάκων

4.1	Μετρικές αξιολόγησης του μοντέλου ResNet-50 για τις κλάσεις LUNGS/NO-LUNGS	69
4.2	Μετρικές αξιολόγησης του μοντέλου EfficientNet-b0 για τις κλάσεις LUNGS/NO-LUNGS	69
4.3	Μετρικές αξιολόγησης του μοντέλου Inception-V3 για τις κλάσεις LUNGS/NO-LUNGS	69
4.4	Μετρικές αξιολόγησης του μοντέλου ResNet-50-RNN για τις κλάσεις non-COVID-19/COVID-19	72
4.5	Μετρικές αξιολόγησης του μοντέλου EfficientNet-b0-RNN για τις κλάσεις non-COVID-19/COVID-19	72
4.6	Μετρικές αξιολόγησης του μοντέλου Inception-V3-RNN για τις κλάσεις non-COVID-19/COVID-19	72

Κατάλογος σχημάτων

1.1	Αναπαράσταση της δομής του ιού SARS-CoV-2	17
1.2	Αποστάσεις εκπνοής μικροσωματιδίων αερολύματος και μεγάλων σταγονιδίων . .	18
1.3	Εξέλιξη της νόσου COVID-19	19
1.4	A. Αξονική τομογραφία άνδρα 28 ετών με θετικό τεστ RT-PCR, όπου φαίνεται αμφοτερόπλευρη, πολυεστιακή, στρογγυλεμένη, περιφερειακή θολή ύαλος(μπλε βέλος) καθώς και αδιαφάνειες λόγω πύκνωσης(κόκκινο βέλος) B. Αξονική τομογραφία άνδρα 71 ετών με θετικό τεστ RT-PCR όπου εντοπίζεται εκτεταμένη, αμφοτερόπλευρη GGO καθώς και παχύνση μεσολοβίδων και ενδολοβίδων διαφραγμάτων C. Αξονική τομογραφία άνδρα 40 ετών με θετικό τεστ RT-PCR όπου φαίνεται περιφερειακή αμφοτερόπλευρη GGO. Ένα αντίστροφο σημάδι φωτοστέφανου είναι ορατό στον κάτω δεξι λοβό(βέλος)	21
2.1	Δομή βιολογικού (αριστερά) και τεχνητού (δεξιά) νευρώνα	25
2.2	Εμπρόσθια (feed-forward ή εν συντομία FNN) και αναδρομική (recurrent ή εν συντομία RNN) τοπολογία ενός τεχνητού νευρωνικού δικτύου	25
2.3	Οπτική αναπαράσταση σημείου σέλας (saddle point)	29
2.4	SGD χωρίς momentum (αριστερά) : Παρατηρείται ταλάντωση στις πλαγιές της χαράδρας και διστακτική πρόοδος προς το τοπικό βέλτιστο , SGD με momentum (δεξιά) : επιτάχυνση του SGD και μείωση των ταλαντώσεων	31
2.5	Οπτικό παράδειγμα εφαρμογής του αλγορίθμου Nesterov	31
2.6	Νευρώνες εισόδου	35
2.7	Αντιστοίχιση νευρώνα κρυφού επιπέδου(hidden neuron) με το αντίστοιχο πεδίο υποδοχής του (receptive field)	35
2.8	Αντιστοίχιση των 28×28 νευρώνων εισόδου σε 3 χάρτες χαρακτηριστικών	36
2.9	Αναδρομική αναπαράσταση δικτύου RNN (αριστερά), «ξεδιπλωμένο» δίκτυο RNN σε διαδοχικές χρονικές στιγμές (δεξιά)	38
2.10	«Ξεδιπλωμένο» RNN, στο οποίο κάθε δομική μονάδα χρησιμοποιεί την συνάρτηση της υπερεφαπτομένης	41
2.11	«Ξεδιπλωμένο» δίκτυο LSTM στο οποίο κάθε δομική μονάδα περιλαμβάνει 4 επίπεδα αλληλεπίδρασης	41
2.12	Κατάσταση κελιού C_t	42
2.13	Πύλη λήθης f_t	42
2.14	Πύλη εισόδου i_t και πύλη \tilde{C}_t	43
2.15	Δημιουργία ενημέρωσης για την νέα κατάσταση	43
2.16	Καθορισμός της νέας εσωτερικής κατάστασης του κελιού μέσω της πύλης εξόδου o_t	44
2.17	Αρχιτεκτονική δομικής μονάδας δικτύου GRU	45
2.18	Μοντέλο κρυφής αλυσίδας Markov (Hidden Markov Model-HMM)	45
2.19	Γενική δομή αμφίδρομου επαναληπτικού νευρωνικού δικτύου (BRNN), «ξεδιπλωμένου» για τρία χρονικά βήματα	47
2.20	Τρεις τρόποι με τους οποίους η μεταφορά μάθησης μπορεί να βελτιώσει την εκπαίδευση	48

3.1	Αρχιτεκτονική δικτύου BCDU-Net που περιλαμβάνει αμφίδρομες μονάδες ConvLSTM στις παρακαμπτήριες συνδέσεις και πυκνά συνδεδεμένα συνελκτικά στρώματα (densely connected convolutions)	52
3.2	Αμφίδρομες μονάδες ConvLSTM in CUA-Net	53
3.3	Υπολογισμοί σε ένα επίπεδο κανονικοποίησης παρτίδας (batch normalization layer)	54
3.4	Μετατροπή 2D εικόνας σε 3D τένσορα	56
3.5	Εσωτερική δομή ConvLSTM	56
3.6	Τυπική δομή ενός residual block	57
3.7	Διαδρομές μετάδοσης κλίσης κατά την εκπαίδευση σε ένα residual block	57
3.8	Αρχιτεκτονική των μοντέλων της οικογένειας ResNet	58
3.9	Αρχιτεκτονική δικτύου EfficientNet-B0: κάθε γραμμή περιγράφει ένα στάδιο i με \hat{L}_i επίπεδα, με είσοδο διαστάσεων $\langle \hat{H}_i, \hat{W}_i \rangle$ και έξοδο \hat{C}_i κανάλια	59
3.10	Δομή bottleneck block	60
3.11	Residual block	60
3.12	Inverted residual block	61
3.13	Squeeze-and-Excitation blocks	62
3.14	Αρχική δομική μονάδα δικτύου Inception	62
3.15	Δομική μονάδα δικτύου Inception με μείωση διαστάσεων	63
3.16	Αντικατάσταση φίλτρου 5×5 από δυο φίλτρα 3×3	64
3.17	Αντικατάσταση φίλτρων $n \times n$ από αλληλουχία φίλτρων $1 \times n, n \times 1$	64
3.18	Αρχιτεκτονική δικτύου Inception-V3	65
4.1	Αρχικός όγκος CT scan slices	70
4.2	Τελικός όγκος CT scan slices	71

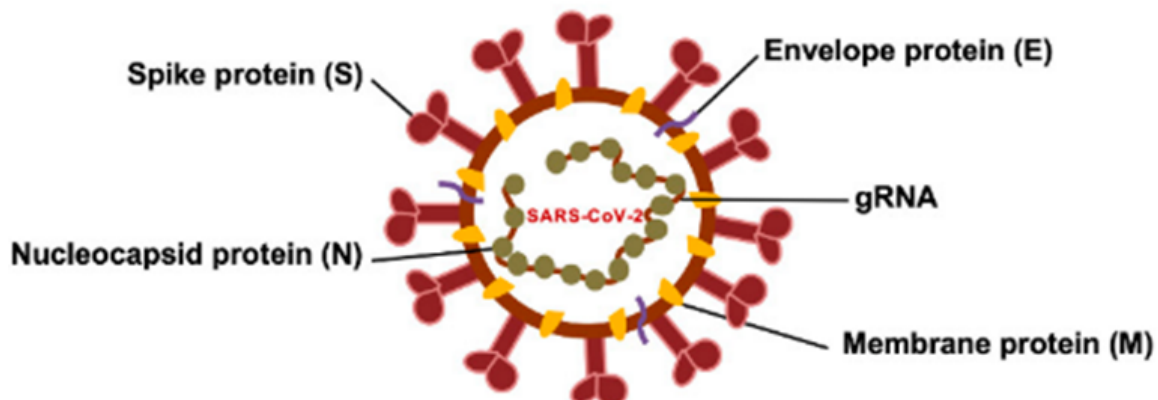
Κεφάλαιο 1

Εισαγωγή

1.1 COVID-19

1.1.1 Προέλευση της νόσου COVID-19

Η οξεία αναπνευστική νόσος 2019-nCoV (ή SARS-CoV-2) ανήκει στην οικογένεια Coronaviridae, που περιλαμβάνει ιούς που φέρουν ως γονιδίωμα μονόκλωνο RNA θετικής πολικότητας. Οι κορονοϊοί διακρίνονται σε τέσσερα γένη, εκ των οποίων το άλφα και το βήτα μπορούν να μεταδοθούν σε ανθρώπους. Πρόκειται για ζωνοσογόνους ιούς που μπορεί να μεταδοθούν από ζώο σε άνθρωπο και το συμβάν της πρώτης μετάδοσης αναφέρεται ως συμβάν διάχυσης. Ο SARS-CoV-2 βρέθηκε ότι σχετίζεται στενά με κορονοϊούς που απαντώνται στον πληθυσμό των νυχτερίδων καθώς και με τον κορονοϊό σοβαρού οξέος αναπνευστικού συνδρόμου(SARS-CoV). Ειδικότερα οι κορονοϊοί RaTG13 και RmYN02 του πληθυσμού των νυχτερίδων φαίνεται να έχουν ομολογία αλληλουχίας 96,2% και 93,3% αντίστοιχα με τον SARS-CoV-2. Επιπλέον, ομοιότητες φαίνεται να υπάρχουν και με κορονοϊούς των Μαλαισιανών πανγκολίνο. [1]



Σχήμα 1.1: Αναπαράσταση της δομής του ιού SARS-CoV-2 [1]

1.1.2 Δομή και προσβολή ξενιστή

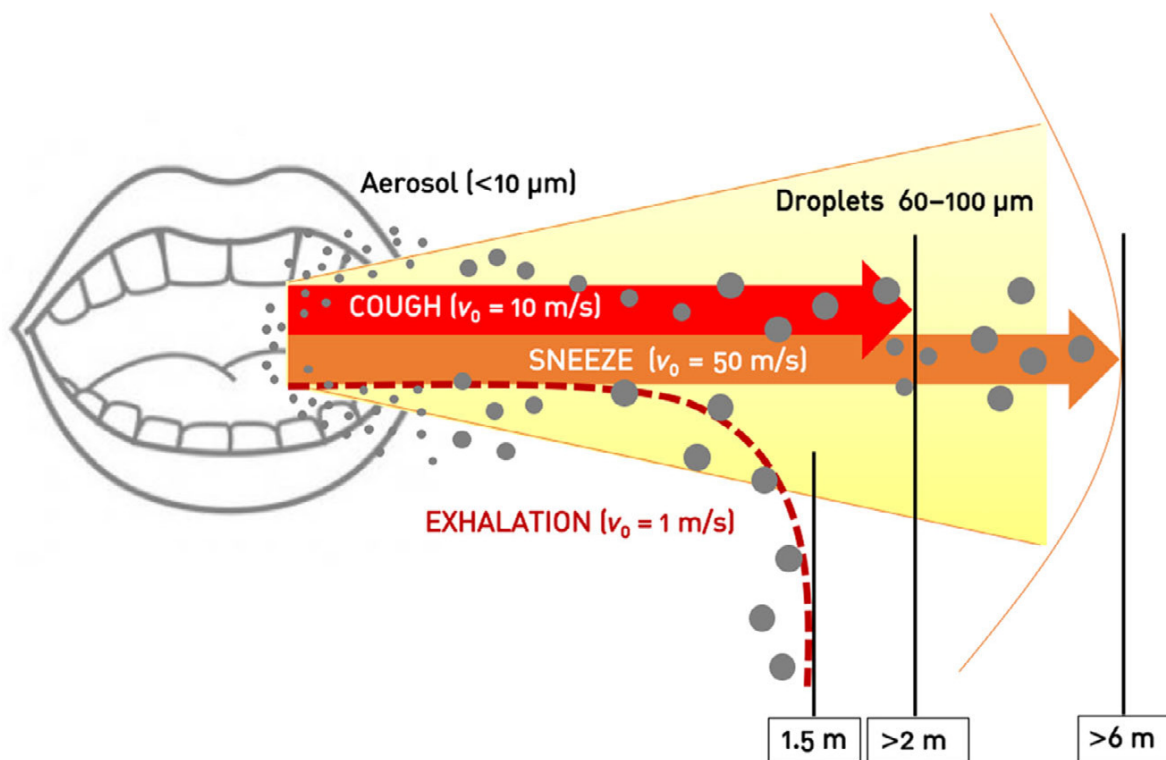
Ο SARS-CoV-2 είναι ένας ιός με περίβλημα σφαιρικού σχήματος. Έχει τέσσερις δομικές και δεκαέξι μη δομικές πρωτεΐνες. Οι δομικές πρωτεΐνες είναι η πρωτεΐνη νουκλεοκαψιδίου (N), η μεμβρανική πρωτεΐνη (M), η αγκαθωτή πρωτεΐνη(S) και η πρωτεΐνη μεμβρανοειδούς φακέλου (E). Το RNA του ιού είναι προσανατολισμένο με κατεύθυνση 5'-3', γεγονός που το καθιστά θετικής πολι-

κότητας και έτσι μπορεί να διαβαστεί απευθείας ως μήνυμα RNA. Η μη δομική πρωτεΐνη 14 (nsp14) επιτελεί διορθωτικό έργο και κρατά τον ρυθμό των μεταλλάξεων σε χαμηλά επίπεδα.

Η πρωτεΐνη S συμβάλει στην προσκόλληση του ιού στον υποδοχέα του μετατρεπτικού ενζύμου της αγγειοτενσίνης 2 (ACE2), που υπάρχει στη μεμβράνη των κυψελιδικών κύτταρων των πνευμόνων του ξενιστή. Έπειτα προκαλείται σύντηξη της ιικής λιπιδικής μεμβράνης με την κυτταρική μεμβράνη του ξενιστή εσωτερικεύοντας έτσι τον ιό. Ακολούθως το κύτταρο ξενιστής μεταφράζει το ιικό RNA και παράγεται ρεπλικάση και δομικές πρωτεΐνες του ιού. Η ρεπλικάση διασπάται σε μη δομικές πρωτεΐνες μια εκ των οποίων είναι η RNA πολυμεράση (RdRp). Ο ιός μεταφέρεται έξω από το κύτταρο μέσω του συμπλέγματος Golgi. Πρόκειται για ένα οργανίδιο το οποίο συσκευάζει μόρια σε μεμβράνες και τα στέλνει σε άλλα μέρη του κυττάρου. Εκεί, ο ιός σχηματίζει ένα περιβλημα λιπιδίων από τη μεμβράνη του συμπλέγματος Golgi. Τα νεοσυσταθέντα σωματίδια του ιού μεταφέρονται μέσα στα κυστίδια Golgi στην κυτταρική επιφάνεια, όπου οδηγούνται έξω από το κύτταρο. [1]

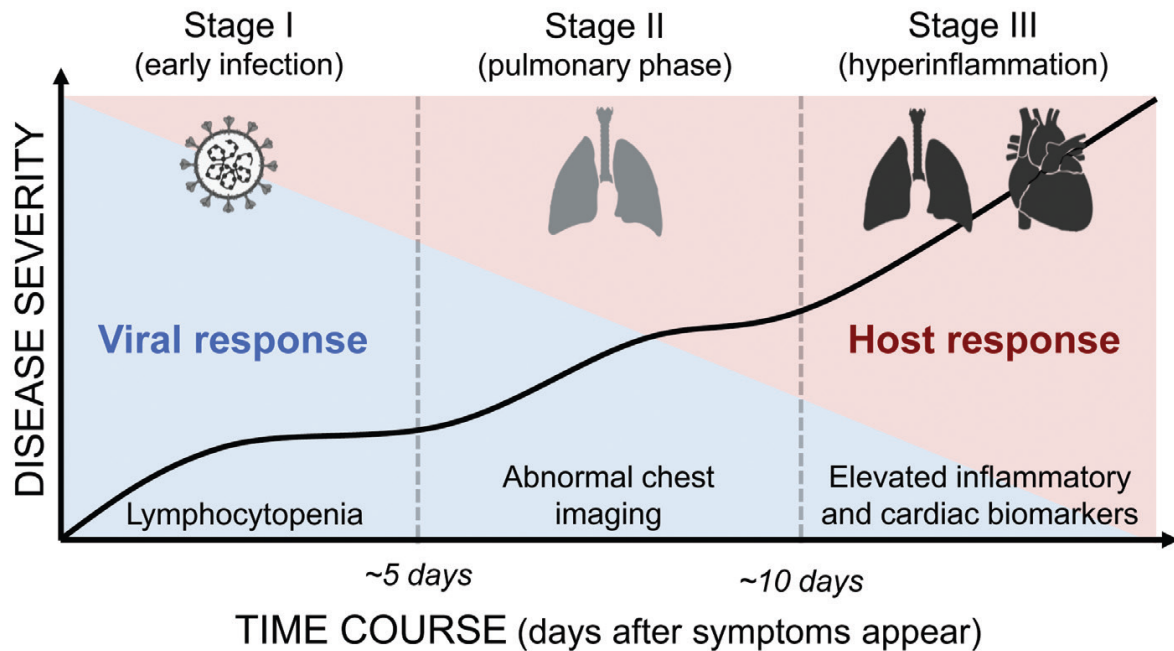
1.1.3 Μετάδοση και συμπτώματα

Η μετάδοση του ιού είναι δυνατή τόσο μέσω του αέρα όσο και μέσω της άμεσης επαφής. Στην πρώτη περίπτωση μεταφέρονται σταγονίδια που απελευθερώνονται με τον βήχα, το φτέρνισμα, την εκπνοή ή την ομιλία ενώ στην δεύτερη περίπτωση έχουμε μόλυνση μέσω της επαφής με μολυσμένες επιφάνειες και του επακόλουθου αγγίγματος των ματιών, της μύτης ή του στόματος. Η μετάδοση μέσω του αέρα αφορά κυρίως εσωτερικούς, πολυσύχναστους και ανεπαρκώς αεριζόμενους χώρους.



Σχήμα 1.2: Αποστάσεις εκπνοής μικροσωματιδίων αερολύματος και μεγάλων σταγονιδίων [2]

Η περίοδος επώασης του ιού κυμαίνεται μεταξύ 3 έως 14 ημερών. Ωστόσο έχει παρατηρηθεί και περίπτωση επώασης του ιού για διάστημα 24 ημερών. Στις περισσότερες περιπτώσεις, οι ασθενείς που προσβάλλονται από τον ιό SARS-CoV-2 είναι ασυμπτωματικοί ή έχουν λίγα ήπια συμπτώματα. Τα κυριότερα εξ αυτών είναι νυχτερινός πυρετός, ξηρός βήχας, πονόλαιμος, καταρροή, ρινική συμφόρηση, μυαλγίες, εξασθένιση, πονοκέφαλος, ζάλη καθώς και ξαφνική απώλεια όσφρησης και γεύσης. Σε σοβαρότερες περιπτώσεις παρουσιάζεται δύσπνοια. Τα πιο σοβαρά-έντονα συμπτώματα εμφανίζονται στο 15% –25% των μολυσμένων ασθενών, δυσχεραίνοντας την αναπνευστική λειτουργία και οδηγώντας σε νοσηλεία με υποβοηθούμενο αερισμό. [2]



Σχήμα 1.3: Εξέλιξη της νόσου COVID-19 [3]

Η εξέλιξη της νόσου με την πάροδο του χρόνου χωρίζεται σε τρεις φάσεις όπως φαίνεται και στο διάγραμμα 1.3: την πρώιμη φάση μόλυνσης, την πνευμονική φάση και την φάση υπερφλεγμονής. Κατά την πρώιμη φάση, που χαρακτηρίζεται από ήπια ή και καθόλου συμπτώματα, ο ιός διεισδύει στους πνεύμονες και αρχίζει να πολλαπλασιάζεται. Η λεμφοκυτταροπενία είναι βασικό εργαστηριακό εύρημα σε αυτό το στάδιο και παρουσιάζεται στην πλειοψηφία των ασθενών. Η νόσος εξελίσσεται εν συνεχεία στην πνευμονική φάση, που χαρακτηρίζεται από αναπνευστική δυσλειτουργία και ανωμαλίες στην απεικόνιση του θώρακα. Στην τρίτη και πιο επικίνδυνη φάση οι δείκτες φλεγμονής είναι αυξημένοι και η δευτερογενής βλάβη οργάνων γίνεται εμφανής. [3]

Οι επιπλοκές που οδηγούν στον θάνατο μπορεί να είναι η αναπνευστική ανεπάρκεια, το σύνδρομο οξείας αναπνευστικής δυσχέρειας, η σηψαιμία και το σηπτικό σοκ, ο θρομβοεμβολισμός και η πολυοργανική ανεπάρκεια.

1.1.4 Διάγνωση

Στις περισσότερες περιπτώσεις η ανίχνευση του ιού SARS-CoV-2 γίνεται εργαστηριακά μέσω της αλυσιδωτής αντίδραση πολυμεράσης με χρήση αντίστροφης μεταγραφάσης σε πραγματικό χρόνο (RT-PCR), με το δείγμα της δοκιμής να συλλέγεται μέσω στείλεου από την στοματοφαρυγ-

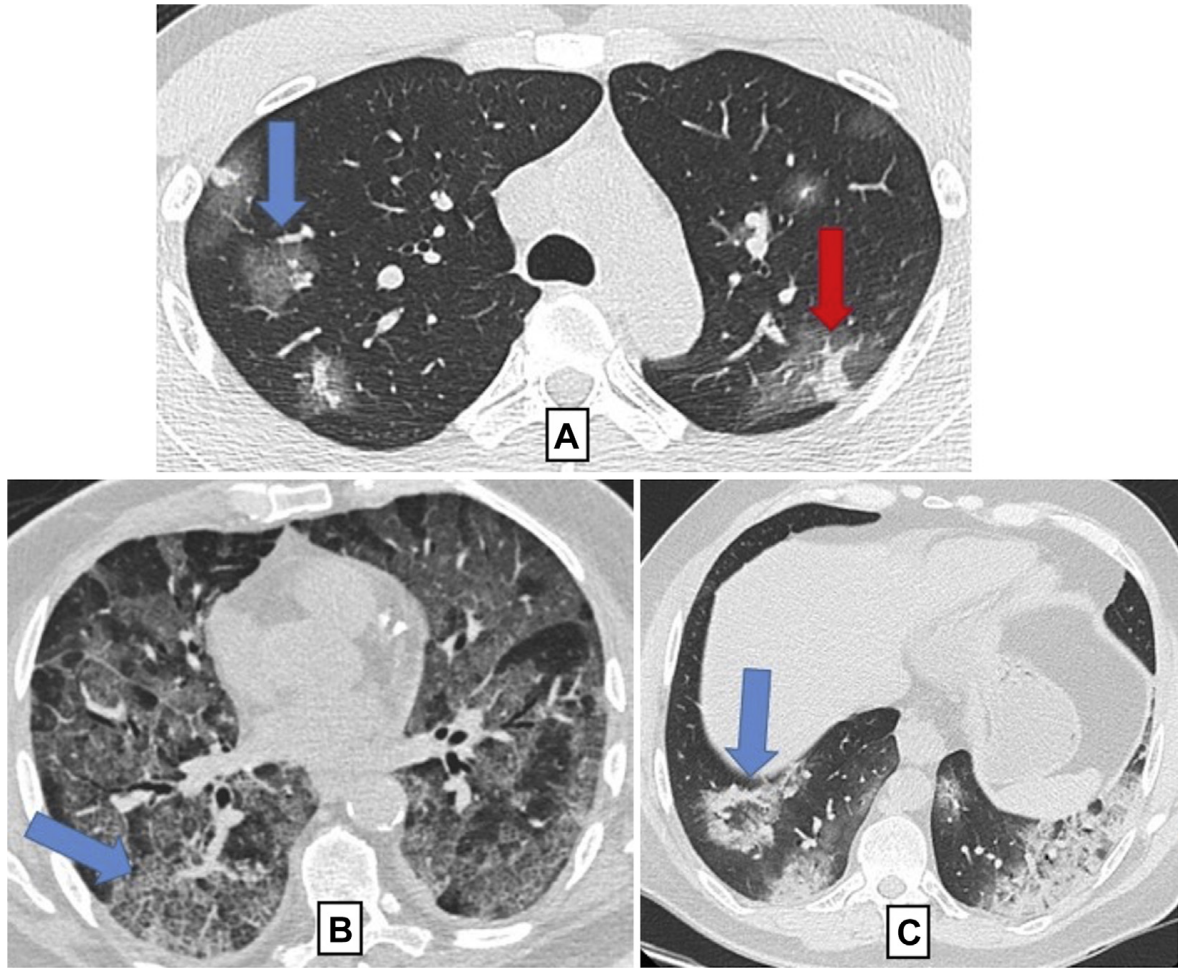
γική ή την ρινοφαρυγγική κοιλότητα. Αν και το τεστ RT-PCR αποτελεί το πρότυπο αναφοράς για την οριστική διάγνωση της λοίμωξης από COVID-19, έχει αυστηρές εργαστηριακές προδιαγραφές και χρειάζεται κάποιος χρόνος για την έκδοση των αποτελεσμάτων. Παράλληλα αρκετές μελέτες διαπίστωσαν ότι τα αρχικά αποτελέσματα RT-PCR για ασθενείς που είχαν προσβληθεί από τον ιό ήταν ψευδώς αρνητικά. Αυτά τα ψευδώς αρνητικά ευρήματα δεν μπορούν να αγνοηθούν, ειδικά για συμπτωματικά άτομα που υποπτεύονται ότι έχουν μολυνθεί. [4]

Ευρέως διαδομένα είναι επίσης και τα τεστ ταχείας ανίχνευσης αντιγόνου (rapid antigen test) που ανιχνεύουν αντιγόνα του ιού, δηλαδή μικρές ιικές πρωτεΐνες και γίνονται κυρίως με επίχρισμα από τον ρινοφάρυγγα. Ωστόσο και με αυτή την μέθοδο παρατηρούνται αρκετά αποτελέσματα ψευδώς αρνητικά.

Οι αξονικές τομογραφίες θώρακα χρησιμοποιούνταν μέχρι τώρα ως εξέταση ρουτίνας για τη διάγνωση της πνευμονίας. Ως εκ τούτου, μπορεί να φανούν χρήσιμες και για τη διάγνωση του COVID-19 σε άτομα με υψηλή κλινική υποψία λοίμωξης. Ωστόσο, παρότι η διάγνωση μέσω αξονικών τομογραφιών παρουσιάζει υψηλά ποσοστά ευαισθησίας (sensitivity) στερείται αξιολογών ποσοστών ειδικότητας (specificity). Επισημαίνεται ότι ως ευαισθησία ενός διαγνωστικού ελέγχου αναφέρεται η πιθανότητα σωστής θετικής διάγνωσης ενώ ως ειδικότητα η πιθανότητα σωστής αρνητικής διάγνωσης. Αυτό μπορεί να οφείλεται στις δυσκολίες που αντιμετωπίζουν οι ακτινολόγοι στην διάκριση μεταξύ του COVID-19 και άλλων ασθενειών σε αξονικές τομογραφίες θώρακα καθώς δεν είναι εξοικειωμένοι με τα μέχρι στιγμής ευρήματα για τα χαρακτηριστικά του ιού σε αυτές.

Την λύση σε αυτό το πρόβλημα έρχεται να δώσει η τεχνητή νοημοσύνη και συγκεκριμένα τα βαθιά νευρωνικά δίκτυα, που έχουν προταθεί πολλάκις για την ανάλυση και επεξεργασία δεδομένων ιατρικής απεικόνισης όπως είναι οι αξονικές τομογραφίες, προκειμένου να βοηθήσουν τους ακτινολόγους αλλά και το σύνολο των γιατρών να βελτιώσουν την ακρίβεια της διάγνωσης τους ([5], [6], [7], [8], [9], [10], [11], [12], [13]).

Κατά τα πρώτα στάδια της νόσου η ακτινογραφία θώρακος (chest x-rays) δεν είναι αρκετά ευαίσθητη ώστε να ανιχνεύσει τις αλλαγές στους πνεύμονες και γι' αυτό χρησιμοποιούνται αξονικές τομογραφίες. Το κύριο χαρακτηριστικό των αξονικών τομογραφιών μολυσμένων ασθενών είναι η εμφάνιση θολών υάλων (ground glass opacity ή GGO). Ένα ακόμη τυπικό χαρακτηριστικό αποτελεί η πύκνωση, που είναι μια περιοχή θολότητας που συσκοτίζει τα όρια των αγγείων και τα τοιχώματα των αεραγωγών, ενώ συχνά παρατηρούνται και GGO με δικτυωτή ή μεσολοβίδια πάχυνση του διαφράγματος. Αποτελέσματα πρόσφατης μελέτης κατέδειξαν ότι 73/80 περιπτώσεις, ποσοστό δηλαδή 91%, εμφάνισε GGO, 50/80 περιπτώσεις, δηλαδή το 63%, πύκνωση και 47/80, δηλαδή το 59%, πάχυνση μεσολοβιακού διαφράγματος. [4] Αξιοσημείωτο είναι επίσης ότι ακόμη και σε τομογραφίες ασυμπτωματικών ασθενών εντοπίζονται πολυεστιακές, μονόπλευρες θολές ύαλοι, γεγονός που επιτρέπει την καθολική εφαρμογή του συγκεκριμένου τύπου διάγνωσης δίχως να απαιτείται η εκδήλωση συμπτωμάτων. [14] Τονίζεται ότι τα χαρακτηριστικά του ιού, όπως απεικονίζονται στις αξονικές τομογραφίες, εμφανίζουν διαφοροποιήσεις σχετικές με το στάδιο της νόσου (πρώιμο ή προχωρημένο) και τον ασθενή. [4]



Σχήμα 1.4: Α. Αξονική τομογραφία άνδρα 28 ετών με θετικό τεστ RT-PCR ,όπου φαίνεται αμφοτερόπλευρη, πολυεστιακή, στρογγυλεμένη ,περιφερειακή θολή ύαλος(μπλε βέλος) καθώς και αδιαφάνειες λόγω πύκνωσης(κόκκινο βέλος) Β. Αξονική τομογραφία άνδρα 71 ετών με θετικό τεστ RT-PCR όπου εντοπίζεται εκτεταμένη, αμφοτερόπλευρη GGO καθώς και παχύνση μεσολοβίδιων και ενδολοβίδιων διαφραγμάτων C. Αξονική τομογραφία άνδρα 40 ετών με θετικό τεστ RT-PCR όπου φαίνεται περιφερειακή αμφοτερόπλευρη GGO. Ένα αντίστροφο σημάδι φωτοστέφανου είναι ορατό στον κάτω δεξί λοβό(βέλος)

[15]

1.1.5 Πρόληψη

Το κύριο μέτρο πρόληψης που διαθέτουμε σήμερα κατά του COVID-19 είναι ο εμβολιασμός. Κατά την ανάπτυξη των εμβολίων έγινε χρήση πολλών διαφορετικών τεχνολογιών και έτσι προέκυψαν εμβόλια RNA (το εμβόλιο Pfizer-BioNTech και το εμβόλιο Moderna), συμβατικά αδρανοποιημένα εμβόλια (BBIBP-CorV, Covaxin και CoronaVac), εμβόλια αδενοϊού (AstraZeneca/Oxford, Johnson Johnson) καθώς και εμβόλια πεπτιδίου (Novavax).

Ωστόσο, παρότι πρόκειται για ένα αρκετά αποτελεσματικό εργαλείο κανένα από αυτά δεν παρέχει 100% προστασία από την νόσο COVID-19. Ένα σημαντικό ποσοστό του πλήρους εμβολιασμένου πληθυσμού προσβάλλεται εντέλει από τον ιό. Παρόλα αυτά οι ασθενείς εμφανίζουν ήπια συμπτώματα και χαμηλότερη μεταδοτικότητα σε σχέση με τον ανεμβολίαστο πληθυσμό. [16]

Η κοινωνική απομόνωση, η αυτοαπομόνωση, η χρήση χειρουργικών масκών, το τακτικό πλύσιμο των χεριών, ο αερισμός εσωτερικών χώρων και ο καθαρισμός επιφανειών συστήνονται σαν επιπλέον μέτρα για την πρόληψη και την ανάσχεση της μετάδοσης του ιού.

1.1.6 Θεραπεία

Μέχρι σήμερα ο Οργανισμός Υγείας και Φαρμάκων (Food and Drug Administration-FDA) των ΗΠΑ έχει εγκρίνει δυο αντικά χάπια ενάντια της νόσου COVID-19.

Το Paxlovid εγκρίθηκε στις 22 Δεκεμβρίου 2021 για την θεραπεία ήπιου έως μέτριου βαθμού συμπτωμάτων σε άτομα ηλικίας 12 ετών και άνω, που διατρέχουν κίνδυνο σοβαρής νόσησης. Η θεραπεία διατίθεται μόνο με ιατρική συνταγή σε άτομα με θετικό τεστ COVID-19 και εντός πέντε ημερών από την έναρξη των συμπτωμάτων. Το Paxlovid περιέχει νιρματρελβίρη και ριτοναβίρη(που χρησιμοποιείται για την καταπολέμηση του HIV). Η νιρματρελβίρη, που αναπτύχθηκε από την Pfizer, παρεμβαίνει στην ικανότητα αναπαραγωγής του ιού ενώ η ριτοναβίρη επιβραδύνει τη διάσπαση της νιρματρελβίρης και έτσι δρα για μεγαλύτερα χρονικά διαστήματα.

Το Molnupiravir εγκρίθηκε στις 23 Δεκεμβρίου 2021 για την θεραπεία ήπιου έως μέτριου βαθμού συμπτωμάτων σε άτομα ηλικίας 18 ετών και άνω, που διατρέχουν κίνδυνο σοβαρής νόσησης. Η θεραπεία διατίθεται μόνο με ιατρική συνταγή, μετά από θετικό τεστ COVID-19 και εντός πέντε ημερών από την έναρξη των συμπτωμάτων. Ωστόσο, ο Οργανισμός Υγείας και Φαρμάκων ανέφερε ότι η χρήση του Molnupiravir θα πρέπει να περιορίζεται σε καταστάσεις στις οποίες άλλες θεραπείες για τον COVID-19 «δεν είναι προσβάσιμες ή κλινικά κατάλληλες». [17]

1.2 Σκοπός και δομή της εργασίας

Σκοπός της παρούσας διπλωματικής είναι η ανάπτυξη συστημάτων βαθιάς μάθησης με σκοπό την έγκυρη και αξιόπιστη διάγνωση της νόσου COVID-19, μέσω της χρήσης αξονικών τομογραφιών θώρακα. Η εργασία διαρθρώνεται ως εξής:

- Στο **κεφάλαιο 2** αναλύονται βασικές θεωρητικές έννοιες που κρίνονται απαραίτητες για την πλήρη κατανόηση του αντικειμένου.
- Στο **κεφάλαιο 3** παρουσιάζεται εκτενώς η αρχιτεκτονική του δικτύου που αναπτύχθηκε.
- Στο **κεφάλαιο 4** αναλύεται η πειραματική διαδικασία που ακολουθήθηκε και παρατίθενται τα αποτελέσματα που προέκυψαν.
- Τέλος στο **κεφάλαιο 5** αναφέρονται τα βασικά συμπεράσματα καθώς και στόχοι για μελλοντικές εργασίες.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

2.1 Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα είναι ένα αφηρημένο αλγοριθμικό κατασκευάσμα το οποίο εμπίπτει στον τομέα της υπολογιστικής νοημοσύνης και στοχεύει στην επίλυση ενός υπολογιστικού προβλήματος. Αυτό επιτυγχάνεται μέσω της εκπαίδευσης του δικτύου πάνω σε ένα σύνολο παραδειγμάτων-δεδομένων σχετικών με το εκάστοτε πρόβλημα. [18]

2.1.1 Βιολογικά νευρωνικά δίκτυα

Η έρευνα σχετικά με τα τεχνητά νευρωνικά δίκτυα είναι εμπνευσμένη από την δομή και την λειτουργία του εγκεφάλου. Ο ανθρώπινος εγκέφαλος είναι μια από τις περίπλοκες βιολογικές δομές και η κατανόηση της λειτουργίας του αποτελεί μία από τις πιο δύσκολες και συναρπαστικές προκλήσεις που αντιμετωπίζει η επιστήμη.

Βασικό δομικό στοιχείο του εγκεφάλου είναι οι νευρώνες ή νευρικά κύτταρα, που δημιουργούν ένα πυκνό δίκτυο επικοινωνίας μεταξύ τους. Μέσα σε αυτόν υπάρχουν περίπου 10^{11} νευρώνες, οι οποίοι παρότι παρουσιάζουν μεγάλη ποικιλία ως προς την μορφή τους, έχουν τα εξής κοινά δομικά χαρακτηριστικά: το σώμα, τους δενδρίτες, τον άξονα και τις συνάψεις. Οι δενδρίτες λειτουργούν ως είσοδος του νευρώνα, ενώ ο άξονας λειτουργεί ως έξοδος. Η επικοινωνία μεταξύ των νευρώνων γίνεται μέσω των συνάψεων, που αποτελούν τα σημεία ένωσης μεταξύ των διακλαδώσεων του άξονα ενός νευρώνα και των δενδριτών από άλλους νευρώνες. Κάθε νευρώνας συνήθως συνδέεται με πολλές χιλιάδες άλλους νευρώνες, με αποτέλεσμα ο συνολικός αριθμός συνάψεων στον εγκέφαλο να υπερβαίνει τις 10^{14} .

Παρόλο που κάθε νευρώνας είναι ένα σχετικά αργό σύστημα επεξεργασίας πληροφοριών (λειτουργεί σε αποτελεσματική χρονική κλίμακα περίπου 1 ms), ο μαζικός παραλληλισμός της επεξεργασίας πληροφοριών σε πολλές συνάψεις οδηγεί ταυτόχρονα σε μια αποτελεσματική επεξεργαστική ισχύ που υπερβαίνει κατά πολύ αυτήν των σημερινών υπολογιστών.

Στους βιολογικούς νευρώνες, φορείς πληροφορίας είναι οι ηλεκτρικοί παλμοί. Όταν οι νευρώνες « πυροβολούν », στέλνουν έναν ηλεκτρικό παλμό (που ονομάζεται δυναμικό δράσης), ο οποίος διαδίδεται κατά μήκος του νευράξονα. Όταν αυτό το σήμα φτάσει σε μια σύναψη ενεργοποιεί την απελευθέρωση χημικών νευροδιαβιβαστών που διασχίζουν τη σύναψη με τον επόμενο νευρώνα. Ανάλογα με τον τύπο της σύναψης, αυτό μπορεί είτε να αυξήσει (διεγερτική σύναψη) είτε να μειώσει (ανασταλτική σύναψη) την πιθανότητα της επακόλουθης πυροδότησης νευρώνων.

Κάθε σύναψη έχει μια σχετική δύναμη (ή βάρος) που καθορίζει το μέγεθος της επίδρασης μιας νευρικής ώσης (δλδ του ηλεκτρικού παλμού) στον μετα-συναπτικό νευρώνα. Κάθε νευρώνας υπο-

λογίζει έτσι ένα σταθμισμένο άθροισμα εισροών από άλλους νευρώνες και αν αυτή η συνολική διέγερση ξεπερνάει κάποιο όριο-κατώφλι, ο νευρώνας «πυροβολεί».

Βασική ιδιότητα τόσο του πραγματικού όσο και του τεχνητού νευρωνικού συστήματος είναι η ικανότητά τους να τροποποιούν τις αποκρίσεις τους ως αποτέλεσμα της έκθεσης σε εξωτερικά σήματα. Αυτό γενικά αναφέρεται ως μάθηση και συμβαίνει κυρίως μέσω αλλαγών στις δυνάμεις των συνάψεων. [18]

2.1.2 Τεχνητά Νευρωνικά Δίκτυα

Βασικό δομικό στοιχείο κάθε τεχνητού νευρωνικού δικτύου είναι ο τεχνητός νευρώνας, που παρουσιάζει σημαντικές ομοιότητες, ως προς την δομή και τις λειτουργίες του με τον βιολογικό νευρώνα.

Υπάρχουν διαφορετικά είδη νευρώνων και η επιλογή για την υλοποίηση του νευρωνικού δικτύου γίνεται με βάση την φύση του προβλήματος που καλούμαστε να επιλύσουμε. Πολλές φορές μάλιστα συνδυάζονται και διαφορετικά είδη νευρώνα.

Στην βασική δομή του, αυτή δηλαδή που χρησιμοποιείται ως επι το πλείστο, ο τεχνητός νευρώνας δέχεται ένα σύνολο εισόδων που πολλαπλασιάζονται με τα αντίστοιχα βάρη, που είναι ανάλογα των συνάψεων του βιολογικού νευρώνα. Ακολουθώντας, το σώμα του νευρώνα αθροίζει τις σταθμισμένες εισόδους και το άθροισμα που προκύπτει τροφοδοτείται σε μια συνάρτηση ενεργοποίησης, η οποία δίνει την έξοδο του νευρώνα. Εκτός από τις εισόδους τους και τα αντίστοιχα βάρη, πολλές φορές ο νευρώνας μπορεί να έχει και κάποιο βάρος b το οποίο ονομάζεται πόλωση ($bias$) και σε αντίθεση με τα υπόλοιπα βάρη επιδρά σε όλες τις τιμές τις εισόδου. Το παραπάνω συνοψίζονται στην εξής μαθηματική σχέση:

$$y(k) = F\left(\sum_{i=0}^m w_i(k) x_i(k) + b\right) \quad (2.1)$$

Όπου:

$x_i(k)$: η τιμή της εισόδου i σε μια διακριτή χρονική στιγμή k , όπου το i παίρνει τιμές από 0 μέχρι m

$w_i(k)$: το συναπτικό βάρος i σε μια διακριτή χρονική στιγμή k , όπου το i παίρνει τιμές από 0 μέχρι m

b : πόλωση ή παράγοντας προδιάθεσης του νευρώνα

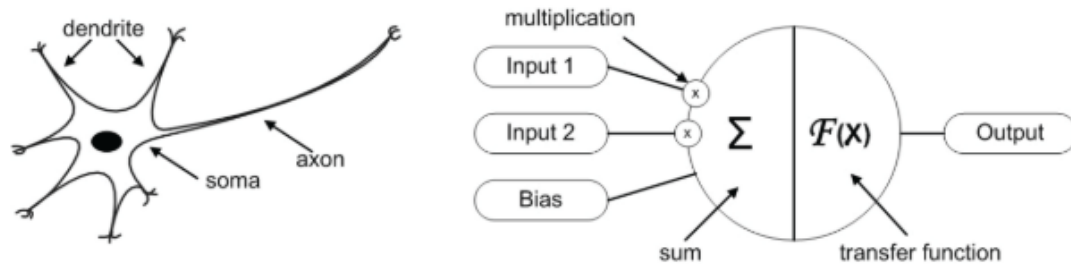
F : η συνάρτηση ενεργοποίησης

$y_i(k)$: η έξοδος του νευρώνα σε μια διακριτή χρονική στιγμή k

Στην εικόνα 2.1 φαίνεται αριστερά ένας βιολογικός νευρώνας (όπου διακρίνονται οι δενδρίτες, το σώμα, οι συνάψεις και ο άξονας), ενώ δεξιά ένας τεχνητός νευρώνας (όπου διακρίνονται οι είσοδοι, τα συναπτικά βάρη, η συνάρτηση ενεργοποίησης, η πόλωση και η έξοδος του).

Ο συνδυασμός δυο ή περισσότερων τεχνητών νευρώνων, δημιουργεί ένα τεχνητό νευρωνικό δίκτυο. Ο τρόπος με τον οποίο διασυνδέονται οι νευρώνες μεταξύ τους καλείται τοπολογία, αρχιτεκτονική ή γράφος. [19]

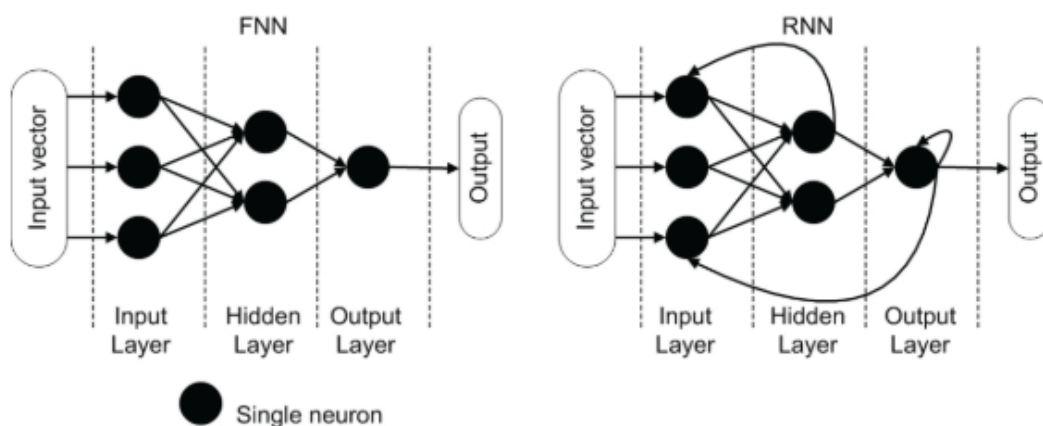
Οι νευρώνες του δικτύου οργανώνονται σε μια σειρά από στρώματα ή επίπεδα (layers). Το πιο αριστερό στρώμα του δικτύου ονομάζεται στρώμα εισόδου (input layer) και δέχεται τα διανύσματα



Σχήμα 2.1: Δομή βιολογικού (αριστερά) και τεχνητού (δεξιά) νευρώνα [19]

εισόδου του δικτύου. Το στρώμα αυτό δεν λογίζεται ως στρώμα νευρώνων καθώς οι κόμβοι της εισόδου δε λειτουργούν ως νευρώνες αλλά απλώς μεταδίδουν τις τιμές της εισόδου στο επόμενο στρώμα. Το πιο δεξί στρώμα καλείται στρώμα εξόδου (output layer) και αποτελείται από τον νευρώνα ή τους νευρώνες που δίνουν την έξοδο του δικτύου. Οποιοδήποτε στρώμα παρεμβάλλεται μεταξύ του στρώματος εισόδου και του στρώματος εξόδου καλείται ενδιάμεσο ή κρυφό επίπεδο (hidden layer). Στην περίπτωση που το δίκτυο περιλαμβάνει περισσότερα του ενός κρυφά επίπεδα, τότε μιλάμε για βαθύ νευρωνικό δίκτυο (deep neural network).

Αξίζει να σημειωθεί ότι ανάλογα με τον τρόπο που συνδέονται τα επίπεδα του νευρώνων μεταξύ τους τα νευρωνικά δίκτυα διακρίνονται σε διάφορες κατηγορίες. Οι βασικότερες από αυτές είναι τα δίκτυα πρόσθιας τροφοδότησης (feedforward networks) [20] και τα ανατροφοδοτούμενα νευρωνικά δίκτυα (recurrent neural networks). Στα δίκτυα πρόσθιας τροφοδότησης η έξοδος ενός επιπέδου χρησιμοποιείται σαν είσοδος στο επόμενο επίπεδο. Αν η έξοδος ενός επιπέδου χρησιμοποιείται σαν είσοδος στο ίδιο ή σε προηγούμενο επίπεδο τότε το δίκτυο είναι ανατροφοδοτούμενο. [21]



Σχήμα 2.2: Εμπρόσθια (feed-forward ή εν συντομία FNN) και αναδρομική (recurrent ή εν συντομία RNN) τοπολογία ενός τεχνητού νευρωνικού δικτύου [19]

2.1.3 Γιατί Νευρωνικά Δίκτυα ;

Ιστορικά το ενδιαφέρον για τα νευρωνικά δίκτυα προέκυψε από την επιθυμία του ανθρώπου να κατανοήσει τον τρόπο που λειτουργεί ο ανθρώπινος εγκέφαλος και να δημιουργήσει μηχανές

ικανές για την επίλυση προβλημάτων, τα οποία οι σειριακών λειτουργιών, προγραμματιζόμενοι υπολογιστές, που στηρίχθηκαν στις θεωρίες των Babbage και von Neumann, δεν είναι κατάλληλοι να επιλύσουν.

Έτσι τα νευρωνικά δίκτυα δίνουν την δυνατότητα παράλληλης κατανεμημένης επεξεργασίας, επιταχύνοντας κατά αυτό τον τρόπο την διαδικασία επίλυσης προβλημάτων. Επίσης έχουν την πολύ σημαντική ικανότητα να μαθαίνουν την γενική λύση ενός προβλήματος από ένα σύνολο πολύ συγκεκριμένων παραδειγμάτων χωρίς να έχουν γνώση συγκεκριμένων λογικών κανόνων, όπως συμβαίνει με τους περισσότερους αλγόριθμους τεχνητής νοημοσύνης. Ένα ακόμη σημαντικό πλεονέκτημα των νευρωνικών δικτύων είναι ο υψηλός βαθμός ανοχής σε σφάλματα. Ένας συμβατικός υπολογιστής μπορεί να αποτύχει εντελώς σε μια λειτουργία του εάν ένα μόνο bit αποθηκευμένης πληροφορίας ή εάν η κατάσταση ενός προγράμματος αποθηκευτεί λάθος. Αυτό μεταφράζεται σε πολύ υψηλό ρίσκο αποτυχίας αν το σύστημα δεν μπορεί να εξυπηρετηθεί συνεχώς. Από την άλλη πλευρά, η λειτουργία του νευρωνικού δικτύου παραμένει ανεπηρέαστη αν ένας νευρώνας ή μια συναπτική σύνδεση δεν λειτουργήσουν όπως πρέπει. [22]

Το κύριο μειονέκτημα των νευρωνικών δικτύων έγκειται στο γεγονός ότι πρέπει να παρέχουμε ένα κατάλληλο σύνολο δεδομένων για την εκπαίδευση του δικτύου, γεγονός που το καθιστά αρκετά προσανατολισμένο στην επίλυση ενός συγκεκριμένου προβλήματος και αποτρέπει την εξερεύνηση νέων περιοχών προβλημάτων. Για παράδειγμα ένα νευρωνικό δίκτυο που έχει εκπαιδευτεί πάνω σε ένα σύνολο εικόνων με γάτες και σκύλους, με σκοπό την ταξινόμηση των εικόνων που θα του δώσουμε σε αυτές τις δυο κατηγορίες, δεν θα είναι αποτελεσματικό στο να αναγνωρίσει ένα τραπέζι σε μια εικόνα.

2.1.4 Συναρτήσεις ενεργοποίησης

Η συνάρτηση ενεργοποίησης F είναι μια μαθηματική συνάρτηση που τροφοδοτείται από την διέγερση του νευρώνα και δίνει την έξοδο του. Οι βασικότερες συναρτήσεις ενεργοποίησης που χρησιμοποιούνται είναι οι παρακάτω:

1. Γραμμική συνάρτηση (Linear Function)

$$F(z) = z \quad (2.2)$$

Η τιμή του $F(z)$ αυξάνεται αναλογικά με την τιμή του z , όπου το z είναι το σταθμισμένο άθροισμα των εισόδων του νευρώνα, όπως περιγράφεται παραπάνω. Επομένως η έξοδος της συνάρτησης δεν περιορίζεται εντός κάποιου συγκεκριμένου εύρους, κάτι το οποίο δεν είναι απαραίτητα πρόβλημα. Ωστόσο η χρήση αυτής της συνάρτησης ενεργοποίησης έχει κάποια σημαντικά μειονεκτήματα:

- Η παράγωγος της συνάρτησης είναι μια σταθερά. Αυτό σημαίνει ότι υπάρχει σταθερή κλίση. Συνεπώς δεν είναι δυνατό να εφαρμοστεί ο αλγόριθμος οπίσθιας διάδοσης, προκειμένου να καθοριστούν οι μεταβολές στα βάρη του νευρωνικού δικτύου με βάση το σφάλμα που προέκυψε. Άρα το νευρωνικό δίκτυο στην ουσία δεν εκπαιδεύεται.
- Καθώς η συνάρτηση ενεργοποίησης είναι γραμμική δεν έχει νόημα να παρατεθούν 2 ή περισσότερα κρυφά επίπεδα (hidden layers) με την ίδια συνάρτηση ενεργοποίησης. Τα N επίπεδα μπορούν να «συμπυχθούν» σε ένα.

- Δεν έχει καλή απόδοση σε μη γραμμικά δεδομένα.

2. Σιγμοειδής συνάρτηση (Sigmoid function)

Η σιγμοειδής συνάρτηση είναι μία πραγματική, συνεχής και φραγμένη συνάρτηση με θετική παράγωγο. Το πεδίο ορισμού της είναι το σύνολο των πραγματικών αριθμών ενώ το σύνολο τιμών της είναι συνήθως το διάστημα $[0, 1]$ ή $[-1, 1]$ ανάλογα με το ποια σιγμοειδή χρησιμοποιείται. Οι τυπικές σιγμοειδής είναι δυο:

- **Λογιστική σιγμοειδής** (με σύνολο τιμών $[0, 1]$)

$$F(z) = \frac{1}{1 + e^{-z}} \quad (2.3)$$

- **Υπερβολική εφασπτομένη** (με σύνολο τιμών $[-1, 1]$)

$$F(z) = \tanh(z) \quad (2.4)$$

Η λογιστική σιγμοειδής όπως και η υπερβολική συνάρτηση έχουν ομαλή κλίση, αποτρέποντας τα μεγάλα «άλματα» στις τιμές της εξόδου. Επιπλέον στην περίπτωση της λογιστικής σιγμοειδούς συνάρτησης οι τιμές της εξόδου κυμαίνονται μεταξύ 0 και 1 οπότε στην ουσία έχουμε μια κανονικοποιημένη έξοδο. Το κύριο μειονέκτημα αυτών των συναρτήσεων ενεργοποίησης είναι ότι υποφέρουν από το πρόβλημα της εξαφάνισης κλίσεων (vanishing gradient). Για πολύ υψηλές τιμές ή πολύ χαμηλές τιμές της εισόδου, δεν υπάρχει σχεδόν καμία αλλαγή στην πρόβλεψη. Αυτό μπορεί να έχει ως αποτέλεσμα το δίκτυο να αρνείται να μάθει περισσότερα ή να είναι πολύ αργό για να φτάσει σε μια ακριβή πρόβλεψη. Επίσης πρόκειται για υπολογιστικά ακριβή συνάρτηση.

3. Συνάρτηση ReLU

Πρόκειται για την πιο ευρέως χρησιμοποιούμενη συνάρτηση ενεργοποίησης στα νευρωνικά δίκτυα. Η συνάρτηση αυτή δίνει σαν έξοδο 0 παίρνοντας αρνητική είσοδο αλλά για κάθε θετική είσοδο δίνει ως έξοδο ακριβώς την τιμή της εισόδου, όπως θα έκανε και μια γραμμική συνάρτηση ενεργοποίησης. Μαθηματικά η συνάρτηση αυτή περιγράφεται ως εξής :

$$F(z) = \max(0, z) \quad (2.5)$$

Η συνάρτηση αυτή είναι υπολογιστικά αποδοτική και επιτρέπει στο δίκτυο να συγκλίνει πολύ γρήγορα. Επίσης είναι μη γραμμική παρόλο που μοιάζει με γραμμική και επιτρέπει την εφαρμογή του αλγορίθμου οπίσθια διάδοσης και επομένως γίνεται εκπαίδευση του δικτύου. Αυτό βέβαια ισχύει μόνο για θετικές εισόδους καθώς όταν οι εισοδοί πλησιάζουν στο μηδέν ή είναι αρνητικές, η κλίση της συνάρτησης γίνεται μηδέν, το δίκτυο δεν μπορεί να εφαρμόσει οπίσθια διάδοση και άρα δεν μπορεί να μάθει.

4. Συνάρτηση Softmax

Η συνάρτηση Softmax αποτελεί μία γενίκευση της σιγμοειδούς συνάρτησης για μετασχηματισμό τιμών σε πιθανότητες. Συγκεκριμένα η συνάρτηση αυτή δέχεται σαν είσοδο ένα διάνυσμα

και μετασχηματίζει τις τιμές σε πιθανότητες οι οποίες αθροίζουν στη μονάδα. Η Softmax χρησιμοποιείται ευρέως σε προβλήματα κατηγοριοποίησης, όπου μόνο μία κατηγορία είναι η σωστή.

$$F(z_i) = \frac{e^{z_i}}{\sum_{i=0}^N e^{z_i}} \quad (2.6)$$

Η συνάρτηση αυτή χρησιμοποιείται στο στρώμα εξόδου του νευρωνικού δικτύου με σκοπό την κατηγοριοποίηση σε κλάσεις.

2.1.5 Αλγόριθμοι Βελτιστοποίησης

Οι αλγόριθμοι βελτιστοποίησης [23] έχουν ως στόχο την ελαχιστοποίηση μια συνάρτησης κόστους $J(\theta)$ όπου θ οι παράμετροι-βάρη του δικτύου. Η συνάρτηση αυτή μοντελοποιεί στην ουσία πόσο απέχουν οι προβλέψεις του δικτύου μας από τις πραγματικές τιμές-στόχους. Μεταβάλλοντας τις τιμές των βαρών, ώστε να ελαχιστοποιείται η συνάρτηση αυτή, επιτυγχάνεται αποδοτικότερη εκπαίδευση του δικτύου που οδηγεί σε ακριβέστερα αποτελέσματα μετέπειτα. Καθοριστικής σημασίας λοιπόν είναι η επιλογή του κατάλληλου αλγορίθμου βελτιστοποίησης. Οι πιο διαδεδομένοι από αυτούς παρουσιάζονται παρακάτω:

Αλγόριθμος Επικλινούς Καθόδου (Gradient Descent Algorithm)

Πρόκειται για έναν από τους πιο συνηθισμένους αλγόριθμους βελτιστοποίησης που χρησιμοποιείται στην εκπαίδευση νευρωνικών δικτύων. Ο αλγόριθμος αυτός ανανεώνει τις παραμέτρους του δικτύου προς την αντίθετη κατεύθυνση από αυτή της κλίσης (gradient) της συνάρτησης κόστους, προκειμένου να επιτύχει την ελαχιστοποίηση της. Ως κλίση μιας συνάρτησης σε ένα σημείο ορίζεται ως το διάνυσμα που δείχνει πάντοτε προς την διεύθυνση του μεγαλύτερου ρυθμού μεταβολής της συνάρτησης στο σημείο αυτό. Με άλλα λόγια, ακολουθείται η κατεύθυνση της καμπύλης της επιφάνειας που δημιουργεί η συνάρτηση κόστους μέχρι μια «κοιλιάδα». Ο ρυθμός εκμάθησης (learning rate), είναι αυτός που καθορίζει το μέγεθος των βημάτων που θα γίνουν προκειμένου να φτάσουμε σε ένα (τοπικό) ελάχιστο της συνάρτησης. Υπάρχουν τρεις παραλλαγές του αλγορίθμου, οι οποίες διαφέρουν ως προς τον όγκο των δεδομένων που χρησιμοποιούνται για κάθε ανανέωση των παραμέτρων του δικτύου. Αναλογικά με τον όγκο αυτό παρατηρείται μια αντιστάθμιση μεταξύ της ακρίβειας ενημέρωσης των παραμέτρων του δικτύου και του χρόνου που απαιτείται για την εκτέλεση της ενημέρωσης.

- **Batch gradient descent**

Ο αλγόριθμος αυτός υπολογίζει τη κλίση της συνάρτησης κόστους ως προς τις παραμέτρους θ του δικτύου για ολόκληρο το σύνολο δεδομένων εκπαίδευσης:

$$\vartheta_{new} = \vartheta_{old} - \eta \nabla_{\vartheta} J(\vartheta_{old}) \quad (2.7)$$

Έτσι καθώς πρέπει να υπολογίσουμε τις κλίσεις για όλα τα στοιχεία του συνόλου για να εκτελεστεί μόνο μία ενημέρωση, ο αλγόριθμος αυτός μπορεί να είναι πολύ αργός και δεν μπορεί να χρησιμοποιηθεί για σύνολα δεδομένων που δεν χωρούν στη μνήμη.

- **Stochastic gradient descent**

Ο αλγόριθμος αυτός πραγματοποιεί ανανέωση παραμέτρων για καθένα από τα δεδομένα εκπαίδευσης, δηλαδή για κάθε ζεύγος διανυσμάτων εισόδων-στόχων $x^{(i)}, y^{(i)}$ ξεχωριστά, σύμφωνα με την παρακάτω εξίσωση:

$$\vartheta_{new} = \vartheta_{old} - \eta \nabla_{\vartheta} J(\vartheta_{old}; x^{(i)}; y^{(i)}) \quad (2.8)$$

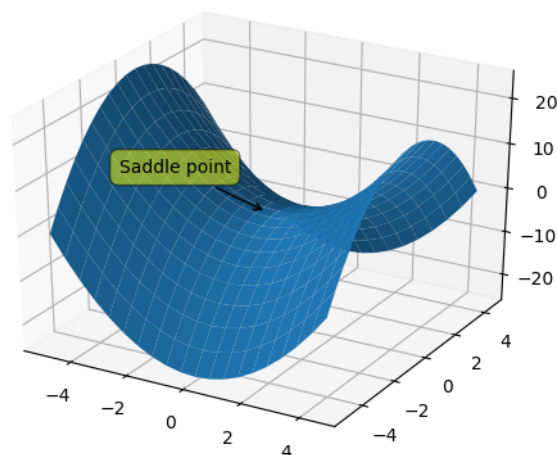
Έτσι ενώ αλγόριθμος Batch gradient descent πραγματοποιεί περιττούς υπολογισμούς για μεγάλο σύνολο δεδομένων καθώς είναι πιθανό να υπολογίζει κλίσεις για παρόμοια παραδείγματα εκπαίδευσης πριν από κάθε ενημέρωση, ο αλγόριθμος Stochastic gradient descent καταργεί αυτόν τον πλεονασμό εκτελώντας μια ενημέρωση για κάθε παράδειγμα εκπαίδευσης. Ως αποτέλεσμα, από την μια αυτές οι συχνές ενημερώσεις μπορούν να προσφέρουν περισσότερη λεπτομέρεια και ταχύτητα, αλλά από την άλλη μπορούν επίσης να οδηγήσουν σε απώλειες υπολογιστικής απόδοσης σε σύγκριση με τον Batch Gradient Descent. Παράλληλα, οι συχνές ενημερώσεις μπορούν να οδηγήσουν σε θορυβώδεις κλίσεις κάτι το οποίο ωστόσο μπορεί εντέλει να φανεί χρήσιμο στην απόδραση από κάποιο τοπικό ελάχιστο και στην εύρεση του ολικού ελάχιστου που είναι και το ζητούμενο.

- **Mini-batch gradient descent**

Ο αλγόριθμος αυτός αποτελεί έναν συνδυασμό των παραπάνω δυο αλγορίθμων καθώς πραγματοποιεί ανανέωση των παραμέτρων του δικτύου για ένα υποσύνολο, n παραδειγμάτων εκπαίδευσης, του συνόλου εκπαίδευσης όπως περιγράφεται από την παρακάτω εξίσωση:

$$\vartheta_{new} = \vartheta_{old} - \eta \nabla_{\vartheta} J(\vartheta_{old}; x^{(i:i+n)}; y^{(i:i+n)}) \quad (2.9)$$

Αυτή η προσέγγιση επιτυγχάνει μια ισορροπία μεταξύ της υπολογιστικής απόδοσης του Batch gradient descent και της ταχύτητας του Stochastic gradient descent.



Σχήμα 2.3: Οπτική αναπαράσταση σημείου σέλας (saddle point)

Ωστόσο και οι τρεις παραπάνω παραλλαγές εγείρουν προκλήσεις που σχετίζονται πρωτίστως με την επιλογή του κατάλληλου ρυθμού μάθησης. Ένας πολύ μικρός ρυθμός μάθησης μπορεί να οδηγήσει σε οδυνηρά αργή σύγκλιση, ενώ ένας πολύ μεγάλος ρυθμός μάθησης μπορεί να εμποδίσει την σύγκλιση μέσα από μεγάλες διακυμάνσεις, με πιθανότητα να προκαλέσει εντέλει και απόκλιση. Παράλληλα, μια ακόμη σημαντική πρόκληση που σχετίζεται με την ελαχιστοποίηση των μη κυρτών συναρτήσεων κόστους, είναι η αποφυγή παγίδευσης στα πολυάριθμα μη βέλτιστα τοπικά τους ελάχιστα αλλά και σε σημεία σέλας (ένα σημείο στην επιφάνεια της γραφικής παράστασης μιας συνάρτησης όπου οι κλίσεις σε ορθογώνιες κατευθύνσεις είναι όλες μηδενικές, αλλά δεν είναι τοπικό άκρο της συνάρτησης, όπως φαίνεται στην εικόνα 2.3).

Ακολουθώς παρουσιάζουμε μια σειρά αλγορίθμων βελτιστοποίησης που χρησιμοποιούνται ευρέως για την αντιμετώπιση των παραπάνω προκλήσεων.

Momentum

Με τον αλγόριθμο αυτό στην ουσία προσπαθούμε να συλλέξουμε κάποιες πληροφορίες σχετικά με τις προηγούμενες ενημερώσεις που έχει υποστεί ένα βάρος, πριν από την εκτέλεση της τρέχουσας ενημέρωσης. Αναλυτικότερα, εάν ένα βάρος «κινείται» μόνιμα προς μια κατεύθυνση, δηλαδή αυξάνεται μόνιμα ή μειώνεται μόνιμα, θα συσσωρευτεί αργά κάποια «ορμή» προς αυτή την κατεύθυνση. Επομένως όταν συναντήσει κάποια αντίσταση και πρέπει να κινηθεί προς την αντίθετη κατεύθυνση, θα συνεχίσει να πηγαίνει για ένα μικρό διάστημα προς την αρχική του κατεύθυνση λόγω αυτής της συσσωρευμένης ορμής.

Η παραπάνω θεωρία είναι σε πλήρη αντιστοιχία με την έννοια της ορμής στην επιστήμη της φυσικής. Έστω λοιπόν ότι έχουμε μια μικρή κοιλάδα και μια μπάλα. Εάν αφήσουμε τη μπάλα στη μία πλευρά της κοιλάδας, θα κυλάει συνεχώς προς τα κάτω με αυξανόμενη ορμή προς αυτήν την κατεύθυνση. Όταν η μπάλα φτάσει στον πάτο, δεν θα σταματήσει εκεί. Αντ' αυτού, θα ανέλθει για λίγο και στην άλλη πλευρά της κοιλάδας, αφού έχει αποκτήσει κάποια ορμή, παρότι η δύναμη του βάρους το τραβάει προς τον πάτο της κοιλάδας.

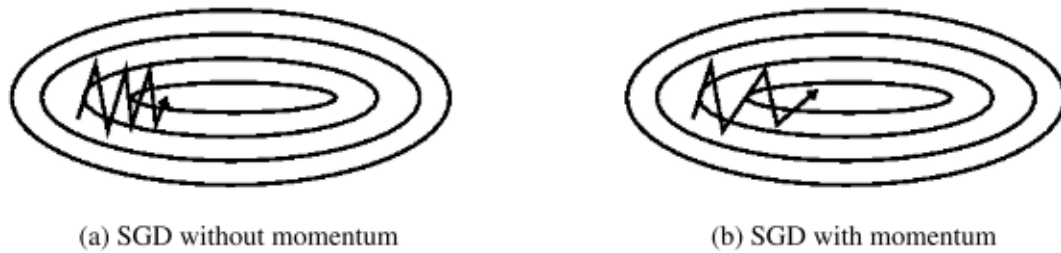
Ουσιαστικά προσπαθούμε να αναδημιουργήσουμε αυτό το σενάριο μαθηματικά, έτσι ώστε ο αλγόριθμος gradient descent να προσπαθήσει να παρακάμψει τα τοπικά ελάχιστα ώστε να φτάσει στα ολικά ελάχιστα, που είναι και το ζητούμενο. Ας δούμε τον τύπο:

$$u_{new} = \alpha * u_{old} - \eta \nabla_{\theta} J(\vartheta_{old}) \quad (2.10)$$

$$\vartheta_{new} = \vartheta_{old} + u_{new} \quad (2.11)$$

Εδώ, το u είναι ο παράγοντας ορμής και $\alpha \in [0, 1]$ μια παράμετρος η τιμή της οποίας καθορίζει κατά πόσο συνυπολογίζεται η τιμή της ορμής στα βάρη. Καθώς το βάρος ενημερώνεται, ουσιαστικά θα αποθηκεύει ένα μέρος όλων των προηγούμενων κλίσεων στον συντελεστή ορμής. Επομένως, μόλις συναντήσει για παράδειγμα ένα τοπικό ελάχιστο, θα συνεχίσει αρχικά να κινείται προς την ίδια κατεύθυνση που κινούνταν έως ότου το μέγεθος του συντελεστή ορμής μειωθεί σταδιακά και αρχίσει να δείχνει προς την αντίθετη κατεύθυνση. Στις περισσότερες περιπτώσεις, ο παράγοντας ορμής είναι συνήθως αρκετός για να κάνει τα βάρη να ξεπεράσουν τα τοπικά ελάχι-

στα. Με την χρήση του αλγορίθμου αυτού λοιπόν πετυχαίνουμε ταχύτερη σύγκλιση και μειωμένη ταλάντωση στις περιοχές των τοπικών ελαχίστων.



Σχήμα 2.4: SGD χωρίς momentum (αριστερά) : Παρατηρείται ταλάντωση στις πλαγιές της χαράδρας και διστακτική πρόοδος προς το τοπικό βέλτιστο , SGD με momentum (δεξιά) : επιτάχυνση του SGD και μείωση των ταλαντώσεων [23]

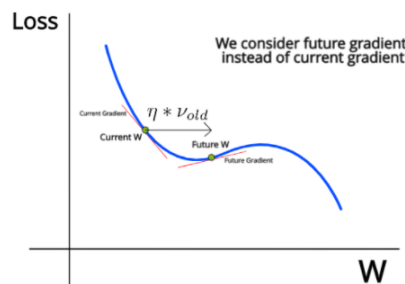
Nesterov accelerated gradient

Στηριζόμενοι πάλι στο παράδειγμα με την μπάλα και την κοιλάδα, διαπιστώνουμε ότι η ιδέα της μπάλας που ακολουθεί τυφλά την πλαγιά, είναι εξαιρετικά μη ικανοποιητική. Θα θέλαμε να έχουμε μια πιο έξυπνη μπάλα, μια μπάλα που έχει μια ιδέα για το πού πηγαίνει, έτσι ώστε να ξέρει να επιβραδύνει πριν ο λόφος ανηφορίσει ξανά. Σε αυτή την ιδέα στηρίζεται και ο αλγόριθμος Nesterov accelerated gradient.

Εδώ αντί να υπολογιστεί η κλίση στην τρέχουσα θέση, υπολογίζεται σε μια κατά προσέγγιση μελλοντική θέση. Έτσι πριν φτάσουμε σε ένα ελάχιστο, η ορμή θα αρχίσει να μειώνεται πριν φτάσει σε αυτό, καθώς χρησιμοποιείται η κλίση από ένα μελλοντικό σημείο. Αυτό οδηγεί σε βελτιωμένη σταθερότητα ,μικρότερες ταλαντώσεις ενώ συγκλίνει και αποδίδει καλύτερα από την προηγούμενη, πιο απλοποιημένη μορφή του αλγορίθμου. Οι αντίστοιχες μαθηματικές σχέσεις φαίνονται παρακάτω :

$$v_{future} = a * v_{old} + \eta \nabla_{\theta} J(u_{old}) \quad (2.12)$$

$$u_{new} = a * u_{old} - \eta \nabla_{\theta} J(v_{future}) \quad (2.13)$$



Σχήμα 2.5: Οπτικό παράδειγμα εφαρμογής του αλγορίθμου Nesterov [24]

Adam

Ο αλγόριθμος Adam (Adaptive Moment Estimation) συνδυάζει την προσαρμογή του ρυθμού μάθησης με έναν όρο Ορμής (Momentum). Συγκεκριμένα, διατηρεί έναν κινούμενο μέσο όρο m_t των κλίσεων u_t καθώς και έναν κινούμενο μέσο όρο u_t των τετραγώνων g_t^2 , και μετά χρησιμοποιεί αυτούς τους μέσους όρους για να αναπροσαρμόσει τα βάρη με ξεχωριστούς ρυθμούς εκπαίδευσης για κάθε βάρη του δικτύου. Με άλλα λόγια, η μεταβλητή m_t προσεγγίζει αθροιστικά τη ροπή πρώτης τάξης, δηλαδή τη μέση τιμή των εξομαλυμένων παραγώγων του κάθε βάρους και μεταβλητή u_t προσεγγίζει αθροιστικά τη ροπή δεύτερης τάξης, δηλαδή τη διασπορά των παραγώγων. Οι εξισώσεις που περιγράφουν τη μεταβολή των παραμέτρων στο βήμα t είναι:

$$m_t = \beta_1 m_t + (1 - \beta_1) g_t \quad (2.14)$$

$$u_t = \beta_2 u_{t-1} + (1 - \beta_2) g_t^2 \quad (2.15)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (2.16)$$

$$\hat{u}_t = \frac{u_t}{1 - \beta_2^t} \quad (2.17)$$

$$\theta_{new} = \theta_{old} - \frac{n * \hat{m}_t}{\sqrt{\hat{u}_t} + \epsilon} \quad (2.18)$$

Η νέα τιμή για κάθε παράμετρο του δικτύου προκύπτει εντέλει με αφαίρεση από την προηγούμενη τιμή, της ποσότητας που φαίνεται στην τελευταία εξίσωση και αποτελεί συνολικά το μεμονωμένο ρυθμό απόδοσης για το κάθε βάρη.

Ο Adam είναι ένας ευρέως χρησιμοποιούμενος αλγόριθμος για εκπαίδευση νευρωνικών δικτύων, που αποδίδει καλύτερα από οποιοδήποτε άλλον βελτιστοποιητή προσφέροντας ταχύτητα σύγκλισης και αντιμετώπιση των προκλήσεων που αναφέρθηκαν παραπάνω. [23], [24]

2.1.6 Οπισθοδιάδοση (Back-Propagation)

Ο αλγόριθμος οπισθοδιάδοσης σφάλματος αποτελεί τον πιο γνωστό αλγόριθμο που χρησιμοποιείται κατά την εκπαίδευση ενός τεχνητού νευρωνικού δικτύου για την εύρεση των κατάλληλων τιμών παραμέτρων. Σκοπός του αλγορίθμου είναι ο υπολογισμός της κλίσης της συνάρτησης κόστους, η οποία προσδιορίζει το πόσο καλά έχει εκπαιδευτεί το δίκτυο μας. Έπειτα έχοντας αυτή την γνώση, με την χρήση του κατάλληλου αλγορίθμου βελτιστοποίησης γίνεται η ανανέωση των παραμέτρων του δικτύου, έτσι ώστε να ελαχιστοποιηθεί το σφάλμα-κόστος.

Έστω λοιπόν νευρωνικό δίκτυο στρωμάτων με N_n νευρώνες στο επίπεδο n . Έστω επίσης W^n και b^n τα βάρη και οι πολώσεις του επιπέδου n αντιστοίχως και a^n η έξοδος του. Έτσι ορίζουμε το διάνυσμα:

$$z^n = W^n a^{n-1} + b^n \quad (2.19)$$

που προσδιορίζει την έξοδο του επιπέδου n πριν περάσει από την συνάρτηση ενεργοποίησης F .
Οπότε:

$$\alpha^n = F(z^n) \quad (2.20)$$

(όπου η συνάρτηση ενεργοποίησης εφαρμόζεται σε κάθε έξοδο του νευρώνα χωριστά).

Το επίπεδο 0 λογίζεται σαν επίπεδο εισόδου. Οπότε ορίζεται η έξοδος α^0 αλλά όχι τα z^0, W^0, b^0 . Έτσι λοιπόν οι έξοδοι των επιπέδων του δικτύου είναι από α^1 μέχρι α^N .

Προκειμένου λοιπόν να ανανεώσουμε τις παραμέτρους του δικτύου πρέπει να γνωρίζουμε την μερική παράγωγο της συνάρτησης κόστους ως προς τα βάρη W^n και τις πολώσεις του δικτύου b^n . Ατομικά για τον νευρώνα i του επιπέδου n η παραπάνω εξίσωση μετασχηματίζεται ως εξής:

$$z_i^n = \sum_j W_{i,j}^n \alpha_j^{n-1} + b_i^n \quad (2.21)$$

Οπότε $\frac{\partial z_i^n}{\partial b_i^n} = 1$ και σύμφωνα με τον κανόνα της αλυσίδας έχουμε:

$$\frac{\partial J}{\partial b_i^n} = \frac{\partial J}{\partial z_i^n} \frac{\partial z_i^n}{\partial b_i^n} = \frac{\partial J}{\partial z_i^n} \quad (2.22)$$

Οπότε $\nabla_{b^n} J = \nabla_{z^n} J$

Ομοίως κάνοντας πάλι χρήση του κανόνα της αλυσίδας έχουμε :

$$\frac{\partial J}{\partial W_{i,j}^n} = \frac{\partial J}{\partial z_i^n} \frac{\partial z_i^n}{\partial W_{i,j}^n} = \frac{\partial J}{\partial z_i^n} \alpha_j^{n-1} \quad (2.23)$$

Οπότε $\nabla_{W^n} J = (\nabla_{z^n} J)(\alpha^{n-1})^T$

Επίσης ισχύει :

$$\frac{\partial J}{\partial \alpha_j^{n-1}} = \sum_i \frac{\partial J}{\partial z_i^n} \frac{\partial z_i^n}{\partial \alpha_j^{n-1}} = \sum_i \frac{\partial J}{\partial z_i^n} W_{i,j}^n \quad (2.24)$$

Άρα $\nabla_{\alpha^{n-1}} J = (W^n)^T (\nabla_{z^n} J)$

Επίσης αφού $\alpha_i^{n-1} = F(z_i^{n-1})$ έχουμε:

$$\frac{\partial J}{\partial z_i^{n-1}} = \frac{\partial J}{\partial \alpha_i^{n-1}} \frac{\partial \alpha_i^{n-1}}{\partial z_i^{n-1}} = \frac{\partial J}{\partial \alpha_i^{n-1}} F'(z_i^{n-1}) \quad (2.25)$$

Οπότε έχουμε:

$$\nabla_{z^{n-1}} J = (W^n)^T (\nabla_{z^n} J) \odot (\nabla_{z^{n-1}} F) \quad (2.26)$$

Εδώ το $\nabla_{z^{n-1}} F$ αναπαριστά στην ουσία την παράγωγο κάθε στοιχείου z^{n-1} , οπότε πρόκειται κυρίως περισσότερο για ιακωβιανή και όχι για παράγωγο ενώ με \odot αναπαριστούμε το προϊόν Hadamard το οποίο παίρνει δυο πίνακες των ίδιων διαστάσεων και παράγει έναν άλλον πίνακα της ίδιας διάστασης με τους παραπάνω, όπου κάθε στοιχείο (i, j) είναι το γινόμενο των στοιχείων (i, j) των δύο αρχικών πινάκων.

Οπότε ξεκινώντας από το επίπεδο εξόδου του δικτύου έχουμε:

$$\nabla_{z^n} J = (\nabla_{\alpha^n} J) \odot (\nabla_{z^n} F) \quad (2.27)$$

και εφαρμόζουμε την εξίσωση επαναληπτικά προκειμένου να υπολογίσουμε τις κλίσεις: $\nabla_{z^{n-1}} J$, $\nabla_{z^{n-2}} J$ κ.ο.κ. Για να υλοποιήσουμε τον αλγόριθμο αποδοτικά δεν χρειάζεται να αποθηκεύουμε όλες τις κλίσεις που έχουμε υπολογίσει μέχρι τώρα. Πρέπει να κρατήσουμε μόνο την πιο πρόσφατη και να έχουμε ένα χώρο μνήμης ώστε να υπολογίζουμε την επόμενη. Οπότε αν δεσμεύσουμε δυο πίνακες με τόσα στοιχεία όσα οι νευρώνες του επιπέδου με τους περισσότερους νευρώνες, μπορούμε να επαναχρησιμοποιούμε αυτούς τους πίνακες για τους υπολογισμούς μας. Για τον αλγόριθμο stochastic gradient descent ωστόσο η ανανέωση των βαρών και των πολώσεων γίνεται επιτόπου (in place) οπότε δεν απαιτείται αυτός ο επιπλέον αποθηκευτικός χώρος.

Συνοψίζοντας λοιπόν διαπιστώνουμε ότι ο αλγόριθμος οπισθοδιάδοσης αποτελείται από τρεις φάσεις:

1. **Πρόσθια διάδοση (forward propagation)**: οι είσοδοι τροφοδοτούνται στα διαδοχικά στρώματα του νευρωνικού δικτύου και στο τέλος επιστρέφεται μια έξοδος. Έπειτα αυτή συγκρίνεται με την επιθυμητή έξοδο και υπολογίζεται το σφάλμα με βάση την συνάρτηση κόστους.

2. **Οπίσθια διάδοση (backward propagation)**: σε αυτή την φάση η πληροφορία του σφάλματος μεταδίδεται από το τελευταίο επίπεδο του νευρωνικού προς τα πίσω μέχρι να φτάσει στο επίπεδο εισόδου, ώστε να υπολογιστεί η κλίση της συνάρτησης κόστους με βάση τον κανόνα της αλυσίδας όπως φαίνεται από τις παραπάνω εξισώσεις.

3. **Ενημέρωση των παραμέτρων (update phase)**: σε αυτή την φάση μπορούμε πλέον να μεταβάλλουμε τις παραμέτρους του δικτύου με βάση τον αλγόριθμο βελτιστοποίησης που έχουμε επιλέξει.

2.2 Συνελικτικά Νευρωνικά Δίκτυα

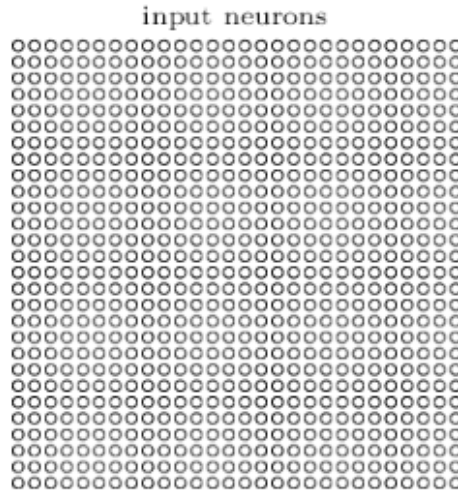
Τα συνελικτικά νευρωνικά δίκτυα [21] είναι μια ειδική κατηγορία τεχνητών νευρωνικών δικτύων που εξειδικεύονται στην επεξεργασία δεδομένων που είναι σε μορφή πινάκων όπως εικόνες και βίντεο [25]. Η αρχιτεκτονική τους στηρίζεται σε τρεις βασικές ιδέες: τα τοπικά πεδία υποδοχής (local receptive fields), τον διαμοιρασμό παραμέτρων (parameter sharing) και την χωρική υποδειγματοληψία (pooling). Αναλυτικότερα έχουμε:

Τοπικά πεδία υποδοχής (local receptive fields)

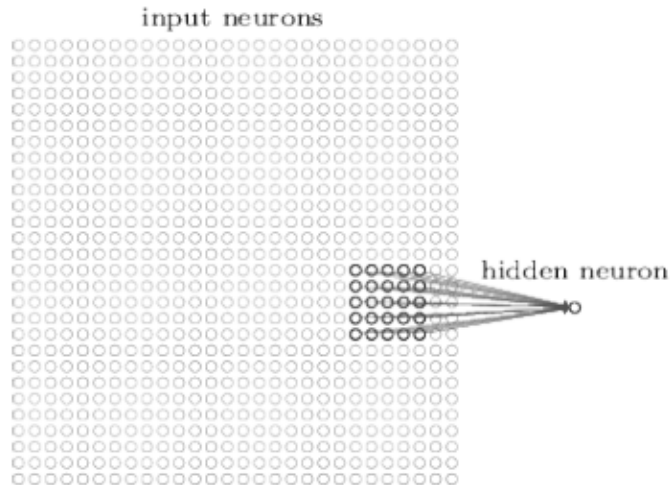
Σε ένα συνελικτικό δίκτυο θα πρέπει να θεωρήσουμε τις εισόδους σαν ένα τετράγωνο νευρώνων και όχι σαν ένα διάνυσμα, όπως μέχρι τώρα, που μπορεί να έχει διαστάσεις 28×28 , των οποίων οι τιμές αντιστοιχούν στις εντάσεις 28×28 εικονοστοιχείων όπως φαίνεται στην εικόνα 2.6.

Ως συνήθως, θα αντιστοιχίσουμε τους νευρώνες εισόδου σε ένα επίπεδο κρυφών νευρώνων. Ωστόσο δεν θα συνδέσουμε κάθε νευρώνα εισόδου με κάθε νευρώνα του κρυφού επιπέδου. Αντ' αυτού, θα συνδέσουμε κάθε νευρώνα του κρυφού επιπέδου μόνο με μια μικρή περιοχή της εικόνας εισόδου, που μπορεί για παράδειγμα να έχει διαστάσεις 5×5 και άρα να αντιστοιχεί σε 25 εικονοστοιχεία, όπως αποτυπώνεται στην εικόνα 2.7.

Αυτή η μικρή περιοχή λοιπόν ονομάζεται τοπικό πεδίο υποδοχής και κάθε νευρώνας του κρυφού επιπέδου μαθαίνει να «αναλύει» το πεδίο που του αντιστοιχεί. Κάθε μια από τις παραπάνω



Σχήμα 2.6: Νευρώνες εισόδου [21]



Σχήμα 2.7: Αντιστοίχιση νευρώνα κρυφού επιπέδου(hidden neuron) με το αντίστοιχο πεδίο υποδοχής του (receptive field)

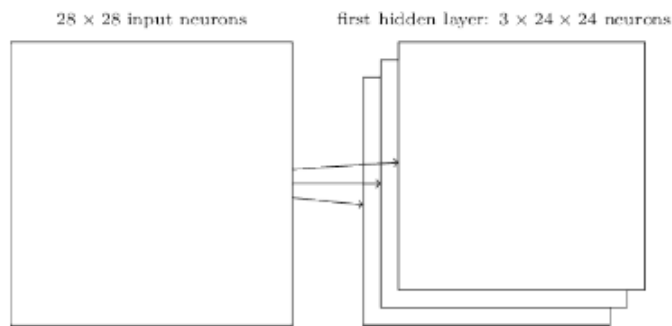
[21]

συνδέσεις αντιστοιχεί σε ένα βάρος. Ας σημειωθεί ότι αν έχουμε εικόνα εισόδου διαστάσεων 28×28 και τοπικό πεδίο υποδοχής διαστάσεων 5×5 τότε θα έχουμε 24×24 νευρώνες στο κρυφό επίπεδο και αυτό γιατί μπορούμε να μετακινήσουμε το αρχικό πεδίο υποδοχής μόνο κατά 23 νευρώνες σε κάθε διάσταση ώστε να εξακολουθεί να βρίσκεται εντός των ορίων της εικόνας. Βέβαια αυτό ισχύει μόνο για πεδία που μετακινούμε κατά ένα εικονοστοιχείο κάθε φορά. Πολλές φορές το πεδίο υποδοχής μπορεί να μετακινείται ανά περισσότερα εικονοστοιχεία. Αυτό προσδιορίζεται από την τιμή του βήματος (stride) του πεδίου. Ταυτόχρονα προκειμένου το επίπεδο των κρυφών νευρώνων να είναι ίδιων διαστάσεων με την εικόνα εισόδου υπάρχει η δυνατότητα συμπλήρωσης της εικόνας με μηδενικά (zero-padding).

Διαμοιρασμός παραμέτρων(parameter sharing)

Όπως αναφέρθηκε παραπάνω κάθε νευρώνας του κρυφού επιπέδου συνδέεται μέσω μιας ομάδας 5×5 βαρών με ένα τοπικό πεδίο υποδοχής. Αυτό που δεν αναφέρθηκε είναι ότι θα χρησιμοποιήσουμε την ίδια ομάδα βαρών για καθέναν από τους 24×24 νευρώνες του κρυφού επιπέδου. Τα διαμοιραζόμενα αυτά βάρη αναφέρονται συχνά ως φίλτρα (filters) ή πυρήνες (kernels). Σημειώνεται ότι ένα φίλτρο έχει το ίδιο βάθος με την εικόνα εισόδου.

Ο διαμοιρασμός παραμέτρων είναι ένας από τους βασικούς παράγοντες που καθιστά τα συνελκτικά νευρωνικά δίκτυα αρκετά αποδοτικά όσον αφορά την επεξεργασία εικόνων και βίντεο καθώς αν ένα φίλτρο εκπαιδευτεί να ανιχνεύει ακμές σε ένα κομμάτι της εικόνας τότε πολύ πιθανό να χρησιμοποιηθεί για την ανίχνευση ακμών και σε άλλα σημεία της εικόνας. Κάθε φορά που εφαρμόζουμε ένα φίλτρο σε μια εικόνα εξάγεται ένας χάρτης χαρακτηριστικών (feature map). Έτσι είναι δυνατόν να έχουμε πολλά διαφορετικά φίλτρα, καθένα από τα εφαρμόζεται στην ίδια εικόνα και εξάγει διαφορετικό είδος χαρακτηριστικών.



Σχήμα 2.8: Αντιστοίχιση των 28×28 νευρώνων εισόδου σε 3 χάρτες χαρακτηριστικών [21]

Στην παραπάνω εικόνα φαίνεται ότι έχουμε τρεις χάρτες χαρακτηριστικών. Κάθε χάρτης ορίζεται από μια ομάδα 5×5 διαμοιραζόμενων βαρών. Ως αποτέλεσμα το δίκτυο μπορεί να ανιχνεύσει τρία διαφορετικά είδη χαρακτηριστικών σε όλη την εικόνα που δίνεται σαν είσοδος στο δίκτυο μας. Στην πράξη συνήθως χρησιμοποιούνται πολλά παραπάνω από τρία φίλτρα.

Κάθε χάρτης χαρακτηριστικών προκύπτει με την χρήση της παρακάτω εξίσωσης:

$$a^1 = \sigma(b + w * a^0) \quad (2.28)$$

όπου ως a^1 λογίζονται τα στοιχεία του χάρτη χαρακτηριστικών που προκύπτει, ως a^0 τα στοιχεία εισόδου, w το φίλτρο, b η πόλωση και με $*$ η πράξη της συνέλιξης. Έτσι το επίπεδο του δικτύου που θεμελιώνεται στην παραπάνω εξίσωση, ονομάζεται συνελκτικό και κατ' επέκταση τα δίκτυα που περιέχουν τέτοια επίπεδα ονομάζονται συνελκτικά.

Όσον αφορά τις διαστάσεις εξόδου ενός συνελκτικού επιπέδου ισχύουν τα εξής:

- έστω ότι δέχεται είσοδο διαστάσεων $W_1 \times H_1 \times D_1$
- έστω επίσης
 - αριθμός φίλτρων K
 - διάσταση φίλτρου F

- βήμα (stride) μετακίνησης του φίλτρου S
- αριθμός μηδενικών που προστίθενται στα άκρα της εικόνας P (stride)
- οπότε έχουμε σαν έξοδο ένα χάρτη χαρακτηριστικών με διαστάσεις $W_2 \times H_2 \times D_2$:
 - $W_2 = \frac{W_1 - F + 2P}{S} + 1$
 - $H_2 = \frac{H_1 - F + 2P}{S} + 1$
 - $D_2 = K$

Ένα μεγάλο πλεονέκτημα του διαμοιρασμού βάρων είναι ότι μειώνεται αισθητά ο αριθμός των παραμέτρων που εμπλέκονται στο συνελκτικό δίκτυο. Στο παράδειγμα μας για κάθε χάρτη χαρακτηριστικών χρειαζόμαστε $25 = 5 \times 5$ διαμοιραζόμενα βάρη και ενδεχομένως και μια πόλωση (bias). Οπότε κάθε χάρτης χαρακτηριστικών απαιτεί 26 παραμέτρους. Έτσι αν έχουμε 20 χάρτες χαρακτηριστικών συνολικά θα έχουμε $20 \times 26 = 520$ παραμέτρους να ορίζουν το επίπεδο του δικτύου. Στον αντίποδα, αν είχαμε ένα πλήρως συνδεδεμένο επίπεδο με 30 νευρώνες, όπου δηλαδή κάθε νευρώνας θα συνδεόταν με κάθε νευρώνα του αμέσως προηγούμενου επιπέδου, για μια εικόνα 28×28 θα είχαμε $30 \times 28 \times 28$ παραμέτρους συν 30 πολώσεις που υπολογίζεται σε σύνολο 23.550 παραμέτρων.

Χωρική υποδειγματοληψία (pooling)

Πέρα από τα συνελκτικά επίπεδα, τα συνελκτικά δίκτυα περιέχουν και επίπεδα υποδειγματοληψίας, τα οποία συνήθως τοποθετούνται αμέσως μετά τα συνελκτικά.

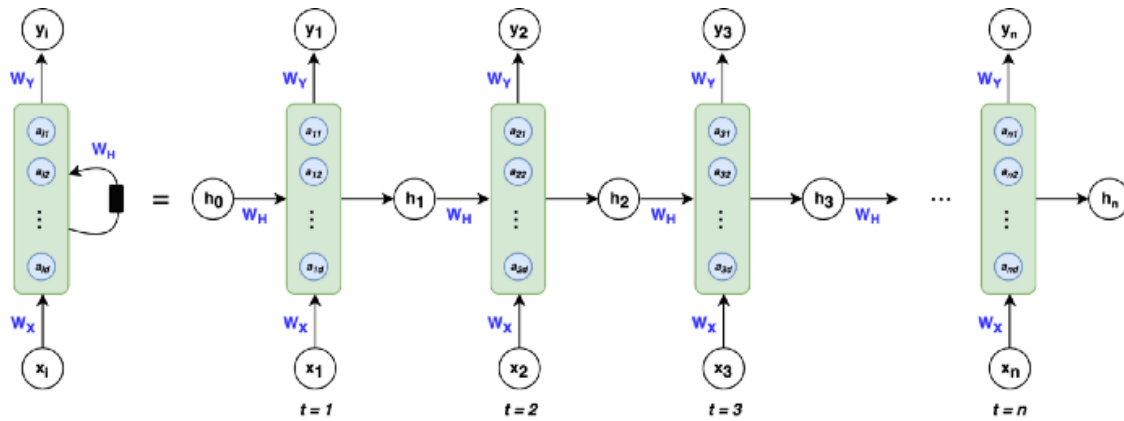
Πιο αναλυτικά ένα επίπεδο υποδειγματοληψίας παίρνει έναν χάρτη χαρακτηριστικών που έχει παραχθεί από ένα συνελκτικό επίπεδο και τον συμπύσσει. Έτσι λοιπόν οι νευρώνες σε αυτό το επίπεδο υπολογίζουν το μέγιστο (max) ή το μέσο (average) από περιοχές 2×2 του χάρτη χαρακτηριστικών. Έτσι μειώνονται οι διαστάσεις του χάρτη χαρακτηριστικών. Στην ουσία μέσω αυτού του επιπέδου το δίκτυο μαθαίνει εάν ένα δεδομένο χαρακτηριστικό βρίσκεται σε οποιοδήποτε περιοχή της εικόνας χωρίς να μας ενδιαφέρει επακριβώς η θέση του που είναι ιδιαίτερα χρήσιμο πχ σε προβλήματα κατηγοριοποίησης εικόνων (image classification).

2.3 Ανατροφοδοτούμενα Νευρωνικά Δίκτυα (RNN)

2.3.1 Δομή

Τα Ανατροφοδοτούμενα Νευρωνικά Δίκτυα (Recursive Neural Networks) είναι μια ειδική κατηγορία νευρωνικών δικτύων που πέρα από τις πρόσθιες συνδέσεις μεταξύ των στρωμάτων υπάρχουν επιπλέον συνδέσεις προς τα πίσω, δηλαδή νευρώνες μεγαλύτερων στρωμάτων τροφοδοτούν είτε νευρώνες προηγούμενων στρωμάτων είτε νευρώνες του ίδιου στρώματος. Τα ANΔ χρησιμοποιούνται ευρέως για δεδομένα που έχουν ακολουθιακή μορφή και έτσι βρίσκουν εφαρμογή σε πεδία όπως η αναγνώριση φωνής, η αναγνώριση συναισθημάτων, η μετάφραση κειμένων και η αναγνώριση δραστηριοτήτων σε ένα βίντεο. Η γενική μορφή ενός ANΔ φαίνεται στην εικόνα 2.9.

Έτσι λοιπόν τα ορθογώνια που απεικονίζονται αντιπροσωπεύουν τις κρυφές καταστάσεις (hidden states) του δικτύου και οι νευρώνες εντός των ορθογωνίων καλούνται κρυφοί κόμβοι (hidden nodes) ή κρυφές μονάδες (hidden units). Το συγκεκριμένο επίπεδο επιτελεί την λειτουργία της ενεργοποίησης μέσω της συνάρτησης που έχουμε επιλέξει. Η έξοδος κάθε επιπέδου συμβολίζεται με h , ενώ



Σχήμα 2.9: Αναδρομική αναπαράσταση δικτύου RNN (αριστερά), «ξεδιπλωμένο» δίκτυο RNN σε διαδοχικές χρονικές στιγμές (δεξιά) [26]

τα W_X, W_Y, W_Z αποτελούν τα βάρη του ΑΝΔ που διαμοιράζονται σε όλο το δίκτυο. Το διάνυσμα x_i αναπαριστά την είσοδο του δικτύου, όπου ο δείκτης i αναφέρεται στο i -οστό στοιχείο του διανύσματος που τροφοδοτείται στο i -οστό επίπεδο.

Η λειτουργία του ΑΝΔ στηρίζεται στις παρακάτω εξισώσεις:

$$\alpha_t = W_H h_{t-1} + W_X X_t \quad (2.29)$$

$$h_t = f(\alpha_t) \quad (2.30)$$

$$\hat{y}_t = g(W_Y h_t) \quad (2.31)$$

Η επιλογή των συναρτήσεων f και g εξαρτάται από το είδος του ΑΝΔ αλλά στην πλειοψηφία των περιπτώσεων θέτουμε $f = \tanh$ και $g = \text{softmax}$.

2.3.2 Οπισθοδιάδοση μέσα στον χρόνο (Backpropagation through time)

Η εκπαίδευση ενός ΑΝΔ γίνεται με την χρήση του αλγορίθμου οπισθοδιάδοσης στο χρόνο [27], που αποτελεί μια παραλλαγή του κλασικού αλγορίθμου οπισθοδιάδοσης, ο οποίος περιεγράφηκε παραπάνω και χρησιμοποιείται για την εκπαίδευση δικτύων, όπως τα MLP.

Για λόγους απλοποίησης και καλύτερης κατανόησης, κατά την περιγραφή του αλγορίθμου χρησιμοποιούμε ΑΝΔ χωρίς πολώσεις και θέτουμε ως συνάρτηση κόστους $J_t(y_t, \hat{y}_t) = -y_t \log(\hat{y}_t)$, γνωστή και ως διασταυρωμένη εντροπία (cross entropy loss function).

Η βασική ιδέα και αυτού του αλγορίθμου έγκειται στον υπολογισμό της παραγώγου της συνάρτησης κόστους ως προς τις παραμέτρους του δικτύου, δηλ ως προς τα βάρη και τις πολώσεις, η οποία προκύπτει ως το άθροισμα των επιμέρους παραγώγων για κάθε χρονική στιγμή :

$$J_{total}(y, \hat{y}) = - \sum_{t=1}^n y_t \log(\hat{y}_t) \quad (2.32)$$

Πρέπει λοιπόν να υπολογιστεί η παράγωγος της συνάρτησης κόστους ως προς τα βάρη W_X, W_Y, W_Z του δικτύου, δηλ οι τιμές των $\frac{\partial J}{\partial W_X}, \frac{\partial J}{\partial W_Y}, \frac{\partial J}{\partial W_Z}$.

Αρχικά υπολογίζουμε την παράγωγο $\frac{\partial J}{\partial W_Y}$. Ας εξετάσουμε τις εξαρτήσεις μεταξύ των μεγεθών. Έχουμε:

$$J_t = -y_t \log(\hat{y}_t) \rightarrow \hat{y}_t = g(z_t) \rightarrow z_t = W_Y h_t \quad (2.33)$$

Οπότε με βάση τον κανόνα της αλυσίδας προκύπτει:

$$\frac{\partial J_t}{\partial W_Y} = \frac{\partial J}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial z_t} \frac{\partial z_t}{\partial W_Y} \quad (2.34)$$

και άρα έχουμε:

$$\frac{\partial J_{total}}{\partial W_Y} = - \sum_{t=1}^n \frac{\partial J_t}{\partial W_Y} \quad (2.35)$$

Ο υπολογισμός των παραγώγων $\frac{\partial J}{\partial W_X}, \frac{\partial J}{\partial W_Z}$ ακολουθεί παρόμοια διαδικασία. Οπότε για τον υπολογισμό της παραγώγου $\frac{\partial J}{\partial W_X}$ εξετάζουμε τις εξαρτήσεις μεταξύ των μεγεθών. Έτσι έχουμε:

$$J_t = -y_t \log(\hat{y}_t) \rightarrow \hat{y}_t = g(z_t) \rightarrow z_t = W_Y h_t \rightarrow h_t = f(W_H h_{t-1} + W_X X_t) \quad (2.36)$$

Επομένως με βάση τον κανόνα της αλυσίδας έχουμε :

$$\begin{aligned} \frac{\partial J_t}{\partial W_X} &= \frac{\partial J_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial z_t} \frac{\partial z_t}{\partial h_t} \frac{\partial h_t}{\partial W_X} + \frac{\partial J_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial z_t} \frac{\partial z_t}{\partial h_t} \frac{\partial h_t}{\partial h_{t-1}} \frac{\partial h_{t-1}}{\partial W_X} + \dots + \\ & \frac{\partial J_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial z_t} \frac{\partial z_t}{\partial h_t} \frac{\partial h_t}{\partial h_{t-n}} \frac{\partial h_{t-n}}{\partial W_X} = \sum_{k=0}^n \frac{\partial J_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial z_t} \frac{\partial z_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W_X} \end{aligned} \quad (2.37)$$

και άρα:

$$\frac{\partial J_{total}}{\partial W_X} = - \sum_{t=1}^n \sum_{k=0}^n \frac{\partial J_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial z_t} \frac{\partial z_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W_X} \quad (2.38)$$

Σημειώνεται επίσης ότι:

$$\frac{\partial h_t}{\partial h_k} = \frac{\partial h_t}{\partial h_{t-1}} \frac{\partial h_{t-1}}{\partial h_{t-2}} \dots \frac{\partial h_{k+1}}{\partial h_k} \quad (2.39)$$

Έτσι λοιπόν έχοντας υπολογίσει τις παραγώγους $\frac{\partial J}{\partial W_X}, \frac{\partial J}{\partial W_Y}, \frac{\partial J}{\partial W_Z}$ ανανεώνουμε τα βάρη του δικτύου με βάση τον αλγόριθμο βελτιστοποίησης που έχουμε επιλέξει.

Δυο βασικά προβλήματα μπορούν να προκύψουν κατά την εφαρμογή του παραπάνω αλγορίθμου λόγω της αύξησης των παραγόντων των γινομένων που μετέχουν στον υπολογισμό των παραγώγων. Αυτό της εκραγείσας κλίσης (exploding gradients) και αυτό της εξαφάνισης κλίσης (vanishing gradients). Στην πρώτη περίπτωση οι τιμές των κλίσεων είναι υπερβολικά μεγάλες, προκαλώντας συχνά υπερχειλίση ενώ στην δεύτερη περίπτωση τείνουν εκθετικά προς το μηδέν καθιστώντας

αδύνατο για το μοντέλο να συσχετίσει κατά την μάθηση χρονικά μακρινά γεγονότα. Η εύρεση λύσεων για αυτούς τους περιορισμούς είναι καθοριστικής σημασίας για την ανάπτυξη μοντέλων που πρέπει να συσχετίσουν γεγονότα που απέχουν λεπτά, ώρες ακόμη και μέρες.

Υπάρχουν λοιπόν δυο διαφορετικές προσεγγίσεις. Από την μια έχουμε λύσεις που στηρίζονται στην αρχιτεκτονική του δικτύου και από την άλλη στην βελτιστοποίηση του αλγορίθμου μάθησης. [28]

Στην πρώτη κατηγορία συγκαταλέγονται τα echo state δίκτυα τα οποία έχουν προκαθορισμένα βάρη για όλες τις συνδέσεις (πέρα από αυτή με την έξοδο του δικτύου) τα οποία δεν υφίστανται κάποια μεταβολή κατά την διαδικασία της μάθησης και έχουμε φροντίσει να αρχικοποιηθούν καταλλήλως ώστε να μην προκύψουν τα προαναφερθέντα προβλήματα. Παράλληλα η ανάπτυξη δικτύων Μακράς Βραχέας Μνήμης (Long Short-Term Memory Network-LSTM) αλλά και Φραγμένων Επαναληπτικών δικτύων (Gated Recurrent Unit – GRU) που χρησιμοποιούν πύλες-ειδικούς νευρώνες που ελέγχουν την ροή της πληροφορίας αποτελούν εναλλακτικές που θα αναλυθούν εκτενώς παρακάτω.

Από την άλλη πλευρά, μια βελτιστοποίηση του αλγορίθμου μάθησης ενδέχεται να περιλαμβάνει τον προσδιορισμό ενός κατωφλίου με σκοπό τον περιορισμό των τιμών της παραγωγού και την αποφυγή έκρηξης. Έτσι λοιπόν συμβολίζοντας ως g την τιμή της παραγωγού και ως u την τιμή του κατωφλίου, αν $\|g\| > u$ τότε πριν από την ανανέωση των παραμέτρων του δικτύου θέτουμε εκ νέου την τιμή της παραγωγού $g \leftarrow \frac{gu}{\|g\|}$. Μια ακόμη εναλλακτική αποτελεί ο αλγόριθμος περικομμένης οπισθοδιάδοσης μέσα στον χρόνο (Truncated BPTT), ο οποίος εκτελεί ουσιαστικά τον αλγόριθμο της οπισθοδιάδοσης μέσα στον χρόνο όπως αυτός παρουσιάστηκε μέχρι τώρα, αλλά σε μικρά χρονικά παράθυρα ώστε να περιοριστεί ο αριθμός των παραγόντων των γινομένων που χρησιμοποιούνται κατά τον υπολογισμό των παραγώγων. Έτσι επιλύεται τόσο το πρόβλημα της έκρηξης όσο και της εξαφάνισης κλίσης, αποδυναμώνοντας ωστόσο την αποτελεσματικότητα του δικτύου αφού δεν μπορούν πλέον να συσχετιστούν γεγονότα που βρίσκονται εκτός αυτού του χρονικού παραθύρου, ακόμη και στις περιπτώσεις που δεν έχουμε εξαφάνιση ή έκρηξη κλίσης.

2.4 Επεκτάσεις Ανατροφοδοτούμενων Νευρωνικών Δικτύων

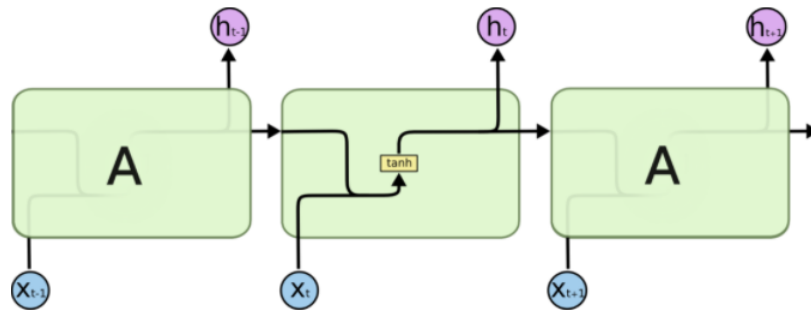
2.4.1 Δίκτυα Μακράς Βραχέας Μνήμης (Long Short-Term Memory Network)

Τα δίκτυα Μακράς Βραχέας Μνήμης (Long Short-Term Memory Network – LSTM) [29] αποτελούν μια ειδική κατηγορία ΑΝΔ, που έχουν την ικανότητα να συσχετίζουν χρονικά μακρινά γεγονότα επιλύοντας έτσι το πρόβλημα της απόσβεσης κλίσης που αναφέρθηκε παραπάνω.

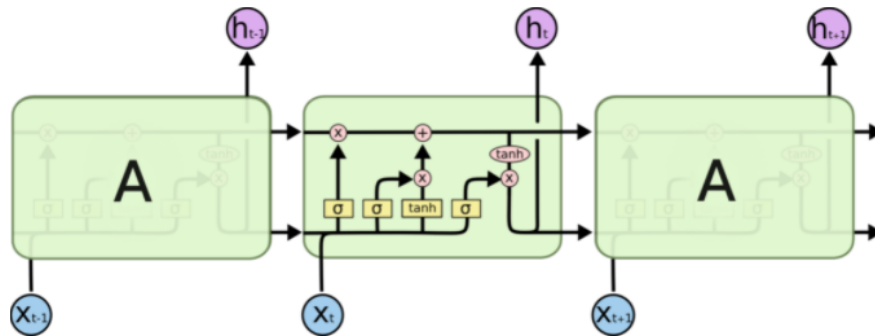
Όλα τα ΑΝΔ έχουν την μορφή αλυσίδας επαναλαμβανόμενων δομών. Στην περίπτωση των απλών ΑΝΔ οι δομές αυτές έχουν πολύ απλή αρχιτεκτονική, όπως ένα νευρωνικό επίπεδο με συνάρτηση ενεργοποίησης την υπερβολική εφαπτομένη.

Ομοίως και τα δίκτυα LSTM έχουν αυτή την αλυσιδωτή δομή, ωστόσο τώρα οι δομικές μονάδες έχουν μια πιο σύνθετη αρχιτεκτονική όπως φαίνεται και παρακάτω.

Στο διάγραμμα 2.10, κάθε μονή γραμμή μεταφέρει ένα διάνυσμα από την έξοδο ενός κόμβου στις εισόδους των επόμενων. Οι ροζ κύκλοι αντιπροσωπεύουν λειτουργίες όπως η πρόσθεση διανυσμάτων ενώ τα κίτρινα πλαίσια αποτελούν επίπεδα του νευρωνικού δικτύου. Γραμμές που συγ-



Σχήμα 2.10: «Ξεδιπλωμένο» RNN, στο οποίο κάθε δομική μονάδα χρησιμοποιεί την συνάρτηση της υπερεφαπτομένης [29]



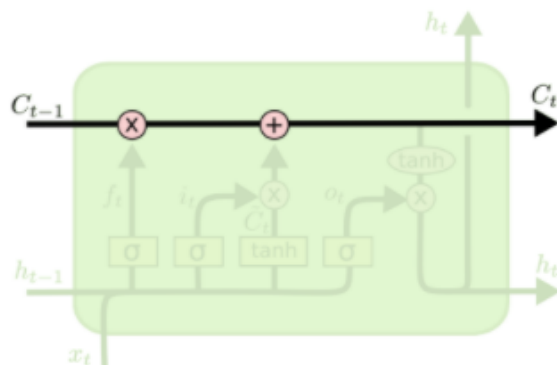
Σχήμα 2.11: «Ξεδιπλωμένο» δίκτυο LSTM στο οποίο κάθε δομική μονάδα περιλαμβάνει 4 επίπεδα αλληλεπίδρασης [29]

χωνεύονται υποδηλώνουν συνένωση, ενώ μια γραμμή που διασπάται σε δυο κλάδους υποδηλώνει ότι το περιεχόμενο της αντιγράφεται και κάθε αντίγραφο ακολουθεί διαφορετική διαδρομή.

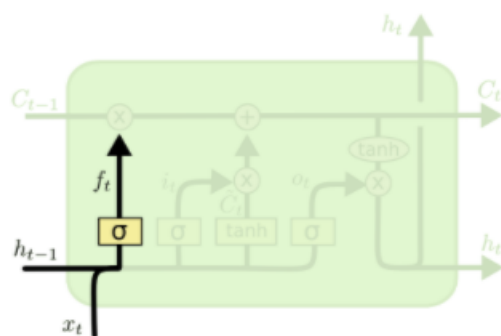
Η κεντρική ιδέα στην οποία στηρίζεται η λειτουργία των δικτύων LSTM είναι η κατάσταση κελιού (Cell State). Πρόκειται για την οριζόντια γραμμή που βρίσκεται στην κορυφή του διαγράμματος και μέσω αυτής δημιουργείται ένας διάδρομος ροής πληροφορίας.

Τα LSTM έχουν την δυνατότητα να εγγράψουν και να διαγράψουν πληροφορίες από την κατάσταση κελιού μέσω του μηχανισμού των πυλών. Κάθε δίκτυο LSTM περιέχει τις ακόλουθες προκειμένου να μεταβάλει την κατάσταση του κελιού:

- Η f_t ονομάζεται πύλη λήθης (Forget Gate) και καθορίζει ποιες πληροφορίες θα διαγραφούν από την κατάσταση κελιού του LSTM. Η λειτουργία αυτή επιτελείται μέσω του νευρωνικού επιπέδου, με την σιγμοειδή συνάρτηση, το οποίο τροφοδοτείται με την έξοδο h_{t-1} της προηγούμενης μονάδας του δικτύου και με την είσοδο x_t . Αν η έξοδος του επιπέδου είναι κοντά στην μονάδα η πληροφορία διατηρείται, ενώ αν είναι κοντά στο 0 διαγράφεται.
- Η i_t ονομάζεται πύλη εισόδου (Input Gate) και καθορίζει ποιες πληροφορίες στην κατάσταση κελιού πρέπει να ανανεώσουμε. Έτσι αν έχουμε έξοδο κοντά στην μονάδα γίνεται η εγγραφή της πληροφορίας, ενώ με έξοδο κοντά στο μηδέν η πληροφορία αγνοείται.

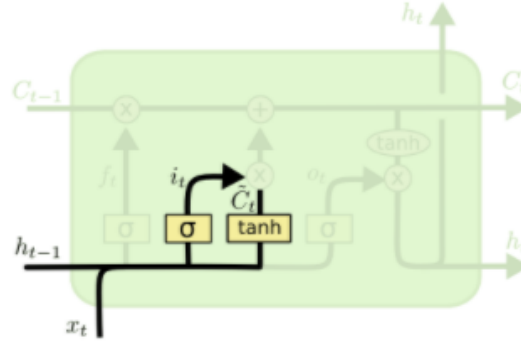


Σχήμα 2.12: Κατάσταση κελιού C_t
[29]

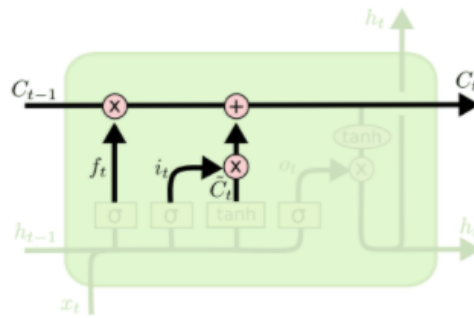


Σχήμα 2.13: Πύλη λήθης f_t
[29]

- Η πύλη \tilde{C}_t αποτελείται από ένα νευρωνικό επίπεδο που χρησιμοποιεί ως συνάρτηση ενεργοποίησης την υπερβολική εφαπτομένη και παράγει τις νέες υποψήφιας τιμές που θέλουμε να εγγράψουμε στην κατάσταση του κελιού.
- Ακολούθως οι έξοδοι των πυλών i_t και \tilde{C}_t συνδυάζονται προκειμένου να παραχθεί η ενημέρωση για την νέα κατάσταση. Έτσι πολλαπλασιάζουμε την παλιά κατάσταση με την έξοδο της f_t , ώστε να διαγραφούν οι κατάλληλες πληροφορίες και ακολούθως προσθέτουμε το γινόμενο $i_t * \tilde{C}_t$, που αντιστοιχεί στις νέες υποψήφιας τιμές οι οποίες είναι κλιμακωμένες ανάλογα με την ενημέρωση που πρέπει να γίνει.
- Τέλος η πύλη o_t καλείται πύλη εξόδου (Output Gate) και καθορίζει την νέα εσωτερική κατάσταση του κελιού ως εξής: η τρέχουσα εσωτερική κατάσταση του κελιού, C_t , περνάει από ένα νευρωνικό επίπεδο, που χρησιμοποιεί ως συνάρτηση ενεργοποίησης την υπερβολική εφαπτομένη ώστε οι τιμές της εξόδου να είναι μεταξύ -1 και 1 και έπειτα η έξοδος αυτή πολλαπλασιάζεται με την έξοδο της πύλης εξόδου.



Σχήμα 2.14: Πύλη εισόδου i_t και πύλη \tilde{C}_t [29]



Σχήμα 2.15: Δημιουργία ενημέρωσης για την νέα κατάσταση [29]

Οι εξισώσεις που περιγράφουν αναλυτικά την λειτουργία ενός δικτύου LSTM είναι οι παρακάτω:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (2.40)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2.41)$$

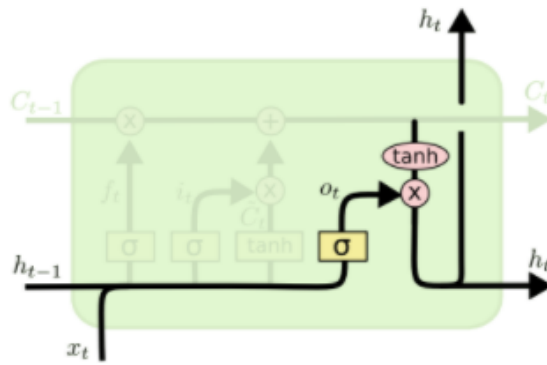
$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \quad (2.42)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (2.43)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (2.44)$$

$$h_t = o_t \odot \tanh(C_t) \quad (2.45)$$

Πρέπει να σημειωθεί ότι ένα σημαντικό μειονέκτημα των δικτύων LSTM ότι απαιτούν μεγαλύτερο χρόνο εκπαίδευσης σε σχέση με τα συμβατικά ANΔ, καθώς κάθε μονάδα τους έχει τετραπλάσιο αριθμό παραμέτρων σε σχέση με μια μονάδα ενός ANΔ, λόγω των πυλών που υπάρχουν.



Σχήμα 2.16: Καθορισμός της νέας εσωτερικής κατάστασης του κελιού μέσω της πύλης εξόδου o_t [29]

2.4.2 Μονάδα Ανάδρασης με Δρομολόγηση (Gated Recurrent Unit)

Η Μονάδα Ανάδρασης με Δρομολόγηση (GRU) αποτελεί μια μεταγενέστερη εκδοχή των LSTM, όπως γίνεται εμφανές και από τις ομοιότητες στην αρχιτεκτονική των δυο δικτύων. Ωστόσο εδώ δεν χρησιμοποιείται πλέον η κατάσταση κελιού (cell state) αλλά μόνο η εσωτερική- κρυφή κατάσταση (hidden state) ώστε να μεταδοθεί η πληροφορία. Παράλληλα οι πύλες λήθης (Forget Gate) και εισόδου (Input Gate) των LSTM συγχωνεύονται σε μια ενιαία πύλη ενημέρωσης (Update Gate). Αναλυτικότερα, λοιπόν η λειτουργία των GRU στηρίζεται στην λειτουργία δύο πυλών:

- Η πύλη ενημέρωσης z_t (Update Gate) καθορίζει ποιες από τις πληροφορίες της προηγούμενης εσωτερικής κατάστασης θα διατηρηθούν αναλλοίωτες και θα προωθηθούν στο δίκτυο αλλά και ποιες θα αντικατασταθούν από νέες εγγραφές.
- Η πύλη επαναφοράς r_t (Reset Gate) καθορίζει τι μέρος της προηγούμενης εσωτερικής κατάστασης θα διατηρηθεί σε μια προσωρινή εσωτερική κατάσταση \tilde{h}_t , η οποία περιλαμβάνει τις νέες υποψήφιες προς εγγραφή πληροφορίες.

Οι εξισώσεις που περιγράφουν αναλυτικά την λειτουργία των GRU φαίνονται παρακάτω:

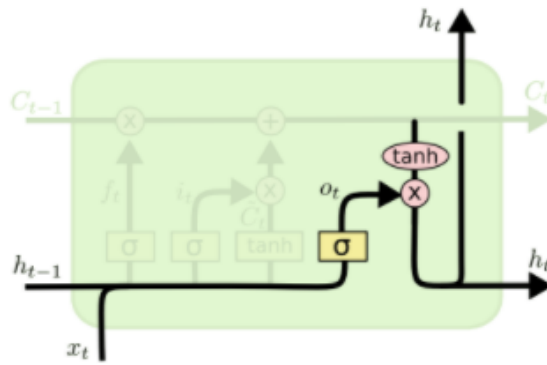
$$z_t = \sigma(W_z[h_{t-1}, x_t]) \quad (2.46)$$

$$r_t = \sigma(W_r[h_{t-1}, x_t]) \quad (2.47)$$

$$\tilde{h}_t = \tanh(W[r_t \odot h_{t-1}, x_t]) \quad (2.48)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (2.49)$$

Συμπερασματικά θα λέγαμε ότι τόσο τα δίκτυα που χρησιμοποιούν μονάδες GRU όσο και LSTM συμβάλλουν αποτελεσματικά στην διατήρηση των μακροχρόνιων εξαρτήσεων μεταξύ των δεδομένων δια μέσου του δικτύου. Η δομή των GRU είναι πιο απλοϊκή από αυτή των LSTM, γεγονός



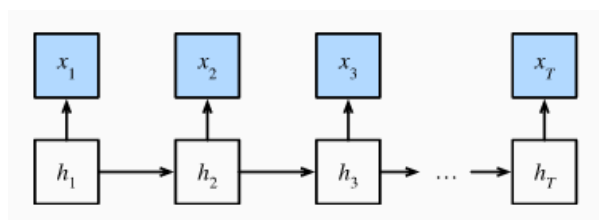
Σχήμα 2.17: Αρχιτεκτονική δομικής μονάδας δικτύου GRU [29]

που προσφέρει μεγαλύτερη ταχύτητα χωρίς σημαντικές απώλειες στην απόδοση. Ωστόσο εμπειρικές μελέτες καταδεικνύουν ότι αυτό το πλεονέκτημα περιορίζεται μόνο σε περιπτώσεις κειμένου ή μικρών συνόλων δεδομένων εισόδου. [30]

2.4.3 Αμφίδρομα Επαναληπτικά Νευρωνικά Δίκτυα (Bidirectional Recurrent Neural Networks)

Ιδιαίτερα χρήσιμη για την βελτιστοποίηση και επέκταση των νευρωνικών μοντέλων που αναφέρθηκαν μέχρι τώρα, κρίνεται η μελέτη του μοντέλου κρυφής αλυσίδας Markov (Hidden Markov Model-HMM). Ως αλυσίδα Markov θεωρείται μια ακολουθία ενδεχομένων – καταστάσεων (states) καθένα από τα οποία εξαρτάται μόνο από το αμέσως προηγούμενο του ή διαφορετικά το κάθε ενδεχόμενο καθορίζει με κάποια πιθανότητα το αμέσως επόμενο του. Έτσι οι αλυσίδες Markov χρησιμοποιούνται για τον υπολογισμό μιας ακολουθίας γεγονότων. Ωστόσο, πολλές φορές τα γεγονότα αυτά δεν είναι άμεσα αλλά έμμεσα παρατηρήσιμα (για παράδειγμα διαβάζοντας ένα κείμενο έχουμε σαν άμεση παρατήρηση τις λέξεις που το αποτελούν και σαν έμμεση το μέρος του λόγου στο οποίο ανήκουν). Έτσι ένα μοντέλο Markov μας επιτρέπει να μιλάμε για ένα σύνολο παρατηρούμενων γεγονότων ,ένα σύνολο κρυφών καταστάσεων και δυο σύνολα πιθανοτήτων, τις πιθανότητες μετάβασης και τις πιθανότητες εκπομπής ή εμφάνισης συμβόλων. [31]

Έστω λοιπόν ότι κάθε χρονική στιγμή t , η κρυφή κατάσταση h_t καθορίζει την αντίστοιχη παρατήρηση x_t με πιθανότητα $P(x_t|h_t)$ και επιπλέον κάθε μετάβαση από $h_t \rightarrow h_{t+1}$ γίνεται με πιθανότητα $P(h_{t+1}|h_t)$. Το πιθανοτικό αυτό μοντέλο αποτυπώνεται παρακάτω γραφικά.



Σχήμα 2.18: Μοντέλο κρυφής αλυσίδας Markov (Hidden Markov Model-HMM) [32]

Έτσι για μια ακολουθία T παρατηρήσεων έχουμε την εξής από κοινού συνάρτηση κατανομής ως προς τις παρατηρούμενες και κρυφές καταστάσεις:

$$P(x_1, x_2, \dots, x_T, h_1, h_2, \dots, h_T) = \prod_{t=1}^T P(h_t|h_{t-1})P(x_t|h_t), P(h_1|h_0) = P(h_1) \quad (2.50)$$

Έστω τώρα ότι παρατηρούμε όλα τα γεγονότα x_i εκτός από ένα γεγονός x_j και θέλουμε να υπολογίσουμε την πιθανότητα $P(x_j|x_{-j})$ όπου $x_{-j} = (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_T)$. Μέσω του δυναμικού προγραμματισμού καταλήγουμε στην παρακάτω εξίσωση που μας δίνει την πιθανότητα η αλυσίδα Markov να βρεθεί στην κατάσταση h_{t+1} την χρονική στιγμή $t + 1$:

$$\pi_{t+1}(h_{t+1}) = \sum_{h_t} \pi_t(h_t)P(x_t|h_t)P(h_{t+1}|h_t) \quad (2.51)$$

θέτοντας αρχικά $\pi_1(h_1) = P(h_1)$. Σε πιο γενικευμένη μορφή η παραπάνω εξίσωση είναι ισοδύναμη με $\pi_{t+1} = f(\pi_t, x_t)$. Η εξίσωση αυτή καλείται εμπρόσθια αναδρομή και μοιάζει αρκετά με την εξίσωση ενημέρωσης της κρυφής κατάστασης στα τυπικά αναδρομικά δίκτυα.

Αντιστοίχως προκύπτει και η εξίσωση οπίσθιας αναδρομής που ορίζεται ως εξής:

$$\varrho_{t-1}(h_{t-1}) = \sum_{h_t} P(h_t|h_{t-1})P(x_t|h_t)\varrho_t(h_t) \quad (2.52)$$

έχοντας θέσει αρχικά $\varrho_T(h_T) = 1$, η οποία σε πιο γενικευμένη μορφή είναι ισοδύναμη με $\varrho_{t-1} = g(\varrho_t, x_t)$. Έτσι και αυτή θυμίζει την εξίσωση ενημέρωσης της κρυφής κατάστασης στα RNN με την διαφορά ότι τρέχει προς τα πίσω.

Συνδυάζοντας τις εξισώσεις εμπρόσθιας και οπίσθιας αναδρομής καταλήγουμε στην σχέση:

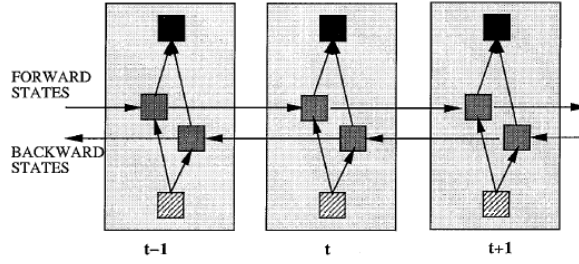
$$P(x_j|x_{-j}) \propto \sum_{h_j} \pi_j(h_j)\varrho_j(h_j)P(x_j|h_j) \quad (2.53)$$

Έναν αντίστοιχο μηχανισμό λοιπόν αναζητούμε προκειμένου να βελτιστοποιήσουμε τα μοντέλα επεξεργασίας ακολουθιακών δεδομένων, όπως τα RNNs και τα LSTMs, τα οποία πλέον δεν θα έχουν γνώση μόνο του παρελθόντος της ακολουθίας αλλά και του μέλλοντος.

Ένα αμφίδρομο RNN προκύπτει από ένα μονόδρομο με την προσθήκη ενός κρυφού επιπέδου που μεταδίδει την πληροφορία προς αντίθετη κατεύθυνση σε σχέση με το πρώτο κρυφό επίπεδο (δλδ προς την αρνητική κατεύθυνση του χρόνου). Έτσι οδηγούμαστε σε μια γενικότερη δομή αντίστοιχη με αυτή που φαίνεται στην εικόνα 2.19, όπου έχουμε ένα «ξεδιπλωμένο» αμφίδρομο επαναληπτικό δίκτυο για τρεις χρονικές στιγμές.

Παρατηρούμε ότι χωρίς τις οπίσθιες καταστάσεις (backwards states) η παραπάνω δομή προσομοιάζει την δομή ενός τυπικού αναδρομικού δικτύου, ενώ χωρίς τις πρόσθιες καταστάσεις την δομή ενός τυπικού αναδρομικού δικτύου με ανεστραμμένο χρονικό άξονα.

Έστω ότι για κάποια χρονική στιγμή t έχουμε μια δέσμη εισόδου $X_t \in R^{n \times d}$ (όπου n ο αριθμός των δειγμάτων εισόδου σε κάθε παράδειγμα d) και συνάρτηση ενεργοποίησης του κρυφού επιπέδου ϕ . Έστω επίσης $\vec{H}_t \in R^{n \times h}$ και $\overleftarrow{H}_t \in R^{n \times h}$ οι πρόσθιες και οι οπίσθιες καταστάσεις αντιστοίχως



Σχήμα 2.19: Γενική δομή αμφίδρομου επαναληπτικού νευρωνικού δικτύου (BRNN), «ξεδιπλωμέ- νου» για τρία χρονικά βήματα

[33]

για την δεδομένη χρονική στιγμή (όπου h ο αριθμός των κρυφών νευρώνων) Η ενημέρωση αυτών των κρυφών καταστάσεων γίνεται σύμφωνα με τις παρακάτω εξισώσεις:

$$\vec{H}_t = \phi(X_t W_{xh}^{(f)} + \vec{H}_{t-1} W_{hh}^{(f)} + b_h^{(f)}) \quad (2.54)$$

$$\overleftarrow{H}_t = \phi(X_t W_{xh}^{(b)} + \overleftarrow{H}_{t+1} W_{hh}^{(b)} + b_h^{(b)}) \quad (2.55)$$

όπου για τα βάρη έχουμε $W_{xh}^{(f)} \in R^{d \times h}$, $W_{hh}^{(f)} \in R^{h \times h}$, $W_{xh}^{(b)} \in R^{d \times h}$ και $W_{hh}^{(b)} \in R^{h \times h}$ και για τις πολώσεις $b_h^{(f)} \in R^{1 \times h}$, $b_h^{(b)} \in R^{1 \times h}$.

Ακολουθώς συνενώνουμε τις πρόσθιες και τις οπίσθιες κρυφές καταστάσεις ώστε να πάρουμε την συνολική κρυφή κατάσταση $H_t \in R^{n \times 2h}$ η οποία θα προωθηθεί στο στρώμα εξόδου. Στα βαθιά αμφίδρομα επαναληπτικά δίκτυα με πολλά επίπεδα αυτή η πληροφορία μεταβιβάζεται ως είσοδος στο επόμενο αμφίδρομο κρυφό επίπεδο. Τελικά στο επίπεδο εξόδου έχουμε:

$$O_t = H_t W_{hq} + b_q \quad (2.56)$$

όπου $O_t \in R^{n \times q}$ (q : αριθμός εξόδων), $W_{hq} \in R^{2h \times q}$ και $b_q \in R^{1 \times q}$. Σημειώνεται ότι τα δύο κρυφά επίπεδα, εμπρόσθιο και οπίσθιο, μπορούν να έχουν διαφορετικό πλήθος νευρώνων. [32]

2.5 Μεταφορά Μάθησης (Transfer Learning)

Πολλές τεχνικές μηχανικής μάθησης εφαρμόζονται επιτυχώς κάτω από μια βασική προϋπόθεση: τα σύνολα εκπαίδευσης (training set) και αξιολόγησης (test set) βρίσκονται στον ίδιο χώρο (feature space) και έχουν την ίδια κατανομή. Όταν η κατανομή διαφοροποιείται πολλά στατιστικά μοντέλα εκπαιδεύονται και πάλι εκ του μηδενός (from scratch) χρησιμοποιώντας ένα νέο σύνολο εκπαίδευσης. Ωστόσο σε πολλές περιπτώσεις, κρίνεται ιδιαίτερα ασύμφορο οικονομικά αλλά και αδύνατο να επανασυλλέξουμε τα απαιτούμενα δεδομένα. Έτσι λοιπόν η μεταφορά γνώσης ή μεταφορά μάθησης μεταξύ διαφορετικών προβλημάτων (domain tasks) θα ήταν ιδιαίτερα χρήσιμη.

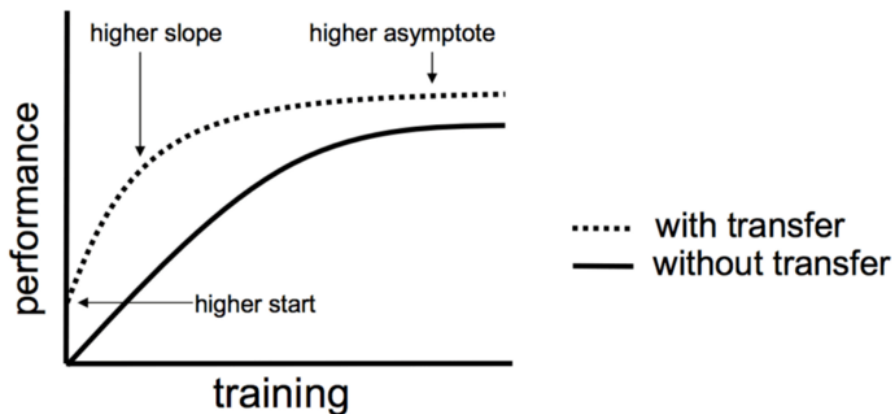
Η ανάπτυξη αυτού του πεδίου στηρίχτηκε στην ίδια την ικανότητα του ανθρώπου να χρησιμοποιεί μια προϋπάρχουσα γνώση για την επίλυση ενός νέου προβλήματος. Για παράδειγμα η γνώση κιθάρας μπορεί να βοηθήσει κάποιον να μάθει μετέπειτα πιάνο, λόγω του μουσικού υπόβαθρου που

ήδη υπάρχει. Ομοίως λοιπόν, κατά την μεταφορά μάθησης, ένα μοντέλο εκπαιδεύεται σε ένα συγκεκριμένο περιβάλλον, που καλείται περιβάλλον προέλευσης (source domain), αποκτά γνώση πάνω σε αυτό και μπορεί με επιτυχία να εκτελέσει μια εργασία προέλευσης (source task). Έπειτα το μοντέλο αυτό μπορεί να χρησιμοποιηθεί και σε ένα άλλο περιβάλλον, που καλείται περιβάλλον στόχος (target domain) και να εκτελέσει επιτυχώς μια εργασία στόχο (target task).

Αξίζει να δώσουμε τους εξής πιο σαφείς ορισμούς για τους όρους περιβάλλον(domain) και εργασία (task) όσον αφορά την μεταφορά μάθησης:

- Ένα περιβάλλον (domain) D ορίζεται από δύο στοιχεία: τον χώρο χαρακτηριστικών (feature space) και την οριακή κατανομή πιθανότητας $P(X)$ όπου $X = \{x_1, \dots, x_n\} \in X$
- Δεδομένου ενός περιβάλλοντος $D = \{X, P(X)\}$ μια εργασία (task) ορίζεται από δύο στοιχεία: τον χώρο ετικετών Y (label space) και μια συνάρτηση πρόβλεψης $f(\bullet)$, η οποία εκπαιδεύεται με βάση το σύνολο των δεδομένων εκπαίδευσης, που αποτελείται από ζεύγη διανυσμάτων χαρακτηριστικών και ετικετών $\{x_i, y_i\}$, όπου $x_i \in X$ και $y_i \in Y$. Από πλευράς πιθανοτήτων, η συνάρτηση αυτή είναι ισοδύναμη με $P(y|x)$. [34]

Ιδιαίτερα στην σημερινή εποχή όπου υπάρχουν μεγάλα αποθετήρια δεδομένων δίνεται η δυνατότητα να μεταφέρουμε γνώση ήδη καλά εκπαιδευμένων μοντέλων σε νέα περιβάλλοντα (target domain) για τα οποία διαθέτουμε περιορισμένη πληροφορία, που δεν θα ήταν ικανή για την επαρκή εκπαίδευση ενός μοντέλου εκ του μηδενός. Έτσι λοιπόν η μεταφορά μάθησης παρότι απλή στον ορισμό και την περιγραφή της συνιστά ένα πολύ ισχυρό εργαλείο για προβλήματα βαθιάς μάθησης, όταν το πλήθος των δεδομένων που διαθέτουμε δεν είναι ικανοποιητικό ώστε να εξασφαλίσει την σωστή εκπαίδευση του μοντέλου. Χρησιμοποιώντας λοιπόν ένα κατάλληλο προεκπαιδευμένο μοντέλο μπορούμε να εξασφαλίσουμε τόσο καλύτερες επιδόσεις όσον αφορά την ακρίβεια των προβλέψεων που επιτυγχάνεται όσο και υψηλότερη ταχύτητα σύγκλισης, όπως φαίνεται και πιο παραστατικά στο παρακάτω διάγραμμα:



Σχήμα 2.20: Τρεις τρόποι με τους οποίους η μεταφορά μάθησης μπορεί να βελτιώσει την εκπαίδευση

[35]

Ωστόσο, οι προκλήσεις και οι προβληματισμοί γύρω από την μεταφορά μάθησης δεν είναι ανύπαρκτοι. Το φαινόμενο της αρνητικής μεταφοράς δεν είναι τόσο σπάνιο και είναι άμεσα συνδεδεμένο με την μεγάλη ανομοιότητα που μπορεί να υπάρχει μεταξύ του περιβάλλοντος προέλευσης

(source domain) και του περιβάλλοντος στόχου(target domain) ,με αποτέλεσμα η γνώση που διαθέτει το προεκπαιδευμένο μοντέλο να μην λειτουργεί θετικά στην εκπαίδευση του στο περιβάλλον στόχου μειώνοντας εντέλει την προβλεπτική του ακρίβεια. Στην βιβλιογραφία [36], [37]αναφέρονται μερικές από τις προτεινόμενες λύσεις για τον περιορισμό ή και την αποφυγή του συγκεκριμένου προβλήματος.

Κεφάλαιο 3

Ανάλυση Μεθοδολογίας

Στην παρούσα εργασία, αναπτύσσεται ένα ολοκληρωμένο μοντέλο διάγνωσης της νόσου COVID-19 μέσω αξονικών τομογραφιών με χρήση μεθόδων βαθιάς μάθησης. Όπως επισημαίνεται και στο εισαγωγικό κεφάλαιο, παρότι οι ακτινογραφίες θώρακα έχουν χαμηλότερο κόστος και είναι διαθέσιμες άμεσα (η διαδικασία λήψης διαρκεί περίπου 10 λεπτά) [38], για την ανάπτυξη του παρόντος μοντέλου προτιμήθηκε η χρήση αξονικών τομογραφιών καθώς τα μειονεκτήματα του υψηλότερου κόστους και της μη άμεσης διαθεσιμότητας των αποτελεσμάτων (η διαδικασία λήψης διαρκεί μεταξύ 10 και 30 λεπτών ενώ τα αποτελέσματα είναι διαθέσιμα συνήθως μετά από 24 ώρες) [39], αντισταθμίζονται από το γεγονός ότι παρέχουν μια αρκετά λεπτομερή εικόνα των πνευμονικών κυψελίδων, οδηγώντας σε καλύτερα διαγνωστικά αποτελέσματα. Όπως καταδεικνύεται στα [4], [40] οι ακτινογραφίες θώρακα δεν χαρακτηρίζονται από υψηλή ευαισθησία και έτσι παρατηρούνται περιπτώσεις μη ανίχνευσης του ιού σε ασθενείς είτε ασυμπτωματικούς είτε στα πρώτα στάδια της νόσου.

Ως **πρώτο στάδιο** κατά την ανάπτυξη του μοντέλου, προτείνεται η κατάτμηση(segmentation) [41], [42] των αξονικών τομογραφιών με σκοπό την ανίχνευση της περιοχής των πνευμόνων. Πρόκειται για ένα σημαντικό στάδιο προεπεξεργασίας πριν την εφαρμογή του μοντέλου κατηγοριοποίησης των τομογραφιών σε COVID-19/non-COVID-19, καθώς σε περίπτωση που παραληφθεί, το μοντέλο ενδέχεται να χρησιμοποιεί άλλα χαρακτηριστικά της εικόνας, πέρα των θολών υάλων και παρόμοιων ενδείξεων συσχετιζόμενων με την εμφάνιση της νόσου, που δεν ανήκουν καν στην περιοχή των πνευμόνων. Έτσι είναι αρκετά πιθανό παρότι επιτυγχάνει σημαντικές αποδόσεις, το μοντέλο να μην είναι ιδιαίτερα αξιόπιστο. Με την εφαρμογή μοντέλου κατάτμησης των πνευμόνων, αφαιρείται ένα σημαντικό μέρος του θορύβου καθώς και της περιοχής γύρω από τους πνεύμονες, αναγκάζοντας το μοντέλο κατηγοριοποίησης που θα εφαρμοστεί μετέπειτα, να λαμβάνει πληροφορίες μόνο από την περιοχή των πνευμόνων, ενώ παράλληλα εξαλείφεται η περιττή πληροφορία που μπορεί να το οδηγήσει σε overfitting. [43], [44]

Το **δεύτερο στάδιο** του διαγνωστικού μοντέλου, περιλαμβάνει την μείωση του όγκου των εικόνων μιας δεδομένης αξονικής τομογραφίας με την χρήση ενός μοντέλου κατηγοριοποίησης των εικόνων που προέκυψαν από το στάδιο της κατάτμησης, σε εικόνες που απεικονίζονται ή δεν απεικονίζονται πνεύμονες. Και αυτό το στάδιο κρίνεται πολύ σημαντικό για την μείωση της περιττής πληροφορίας, καθώς ένας σημαντικός αριθμός των πρώτων και των τελευταίων εικόνων που ανήκουν σε μια δεδομένη αξονική τομογραφία δεν αποτυπώνουν καμία περιοχή των πνευμόνων και συνεπώς δεν ενδείκνυνται προς τροφοδότηση στο μοντέλο κατηγοριοποίησης σε COVID-19/non-COVID-19.

Το **τρίτο και τελευταίο στάδιο**, περιλαμβάνει ένα μοντέλο κατηγοριοποίησης των τομογραφιών σε COVID-19/non-COVID-19. Ενώ σε πολλές περιπτώσεις γίνεται χρήση μοντέλων κατηγοριοποίησης ανά εικόνα (slice), εντούτοις είναι αρκετά χρονοβόρα η διαδικασία δημιουργίας ενός μεγάλου συνόλου δεδομένων, ικανού να εξασφαλίσει αξιόλογες επιδόσεις κατά την εκπαίδευση του δικτύου, καθώς απαιτείται επισημείωση (annotation) κάθε ξεχωριστής εικόνας από ακτινολόγους. [43] Έτσι στην παρούσα εργασία το μοντέλο κατηγοριοποίησης τροφοδοτείται με το σύνολο των εικόνων κάθε τομογραφίας και όπως γίνεται αντιληπτό το labeling έγινε ανά scan και όχι ανά slice.

3.1 Αρχιτεκτονική Δικτύου

Για την υλοποίηση του διαγνωστικού μοντέλου, όπως αυτό περιγράφεται περιληπτικά παραπάνω, έγινε χρήση γνωστών μοντέλων που επιλέχθηκαν προσεχτικά για τους σκοπούς της παρούσας εργασίας.

Το δίκτυο που χρησιμοποιήθηκε για την κατάτμηση των αξονικών τομογραφιών με σκοπό την ανίχνευση πνευμόνων είναι το BCDU-Net [45], που αποτελεί μια προέκταση του δημοφιλούς δικτύου U-Net [46], εισάγοντας μονάδες BConvLSTM στις παρακαμπτήριες συνδέσεις (skip connections) και επαναχρησιμοποιώντας του χάρτες χαρακτηριστικών μέσω πυκνών μπλοκ (dense blocks). Δεν εκπαιδεύτηκε εκ νέου αλλά χρησιμοποιήθηκε ήδη προεκπαιδευμένο μοντέλο, που παρατίθεται από τους δημιουργούς στο repository [BCDU-Net pretrained models](#)

Τα δίκτυα ResNet-50 [47], EfficientNet-B0 [48] και Inception-V3 [49] χρησιμοποιήθηκαν τόσο για την μείωση του όγκου των εικόνων κάθε αξονικής τομογραφίας, επιτελώντας κατηγοριοποίηση σε εικόνες με ή χωρίς πνεύμονες, όσο και σε υβριδικά δίκτυα CNN-RNN για την έκδοση της τελικής διάγνωσης, χαρακτηρίζοντας κάθε τομογραφία ως COVID-19/non-COVID-19.

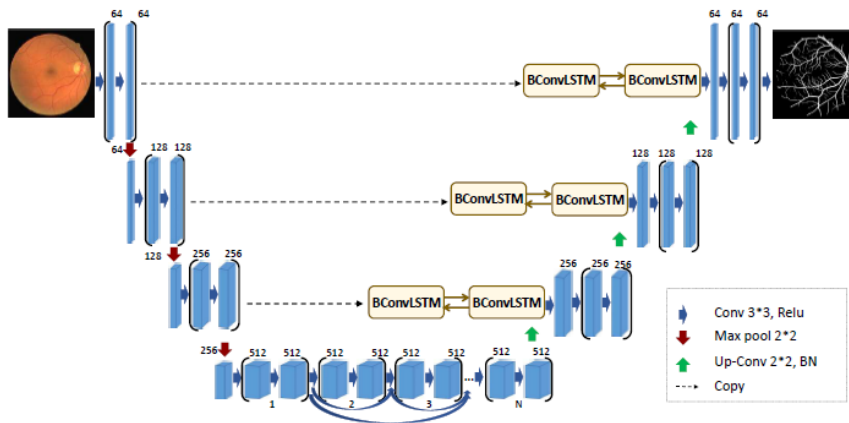
Σημειώνεται ότι κατά την χρήση των μοντέλων ResNet-50, EfficientNet-B0, Inception-V3 για την κατηγοριοποίηση των slices κάθε αξονικής τομογραφίας ως προς το αν απεικονίζονται πνεύμονες ή όχι, αφαιρέθηκαν τα πλήρως διασυνδεδεμένα επίπεδα που επιτελούν το στάδιο της κατηγοριοποίησης και προστέθηκαν ένα επίπεδο global average pooling και ένα νέο πλήρως διασυνδεδεμένο επίπεδο.

Ακολουθώς παρουσιάζονται λεπτομερώς τα παραπάνω μοντέλα.

3.2 BCDU-Net

Όσον αφορά την γενική ιδέα της αρχιτεκτονικής του δικτύου BCDU-Net (εικόνα 3.1), οι χάρτες χαρακτηριστικών από ένα δεδομένο επίπεδο κωδικοποίησης (encoding layer) περιέχουν χαρακτηριστικά υψηλής ανάλυσης (χαμηλού επιπέδου) ενώ οι χάρτες χαρακτηριστικών που εξάγονται από το αντίστοιχο επίπεδο υπερδειγματοληψίας περιέχουν κυρίως σημασιολογική πληροφορία και έχουν χαμηλότερη ανάλυση. Αντί για μια απλή συνένωση λοιπόν, συνδυάζοντας αυτά τα δυο επίπεδα πληροφοριών με την χρήση μη γραμμικών συναρτήσεων οδηγούμαστε σε πιο ακριβή αποτελέσματα κατάτμησης. Γι' αυτό το λόγο στο δίκτυο BCDU-Net γίνεται εισαγωγή μονάδων BConvLSTM στις παρακαμπτήριες συνδέσεις (skip connections).

Έχοντας μια ακολουθία συνεκτικών επιπέδων ένα δίκτυο μπορεί να μάθει πολλά διαφορετικά είδη χαρακτηριστικών, εκ των οποίων κάποια μπορεί να είναι περιττά. Προκειμένου να με-



Σχήμα 3.1: Αρχιτεκτονική δικτύου BCDU-Net που περιλαμβάνει αμφίδρομες μονάδες ConvLSTM στις παρακαμπτήριες συνδέσεις και πυκνά συνδεδεμένα συνελκτικά στρώματα (densely connected convolutions)

[45]

τριάσουμε αυτό το πρόβλημα στο τελευταίο επίπεδο της συμβιβαστικής διαδρομής του δικτύου (contracting path), κάθε συνελκτικό μπλοκ συνδέεται με κάθε επόμενο συνελκτικό μπλοκ του επιπέδου κατά τον άξονα των καναλιών (channel-wise concatenation). Έτσι τα χαρακτηριστικά που προκύπτουν από ένα συνελκτικό μπλοκ προωθούνται σε όλα τα επόμενα μπλοκ. Κατά αυτόν τον τρόπο το δίκτυο ωθείται να μάθει διαφορετικά χαρακτηριστικά, στηριζόμενο όμως άμεσα στην γνώση προηγούμενων επιπέδων και έτσι αποφεύγεται ο πλεονασμός πληροφορίας.

Επιπλέον εφαρμόζοντας την μέθοδο της κανονικοποίησης παρτίδας (batch normalization) μετά τα φίλτρα αποκωδικοποίησης αυξάνεται η ταχύτητα σύγκλισης. Τα παραπάνω αποτελούν τις κύριες διαφοροποιήσεις-προσθήκες του δικτύου BCDU-Net σε σχέση με U-Net. Αναλυτικότερα έχουμε:

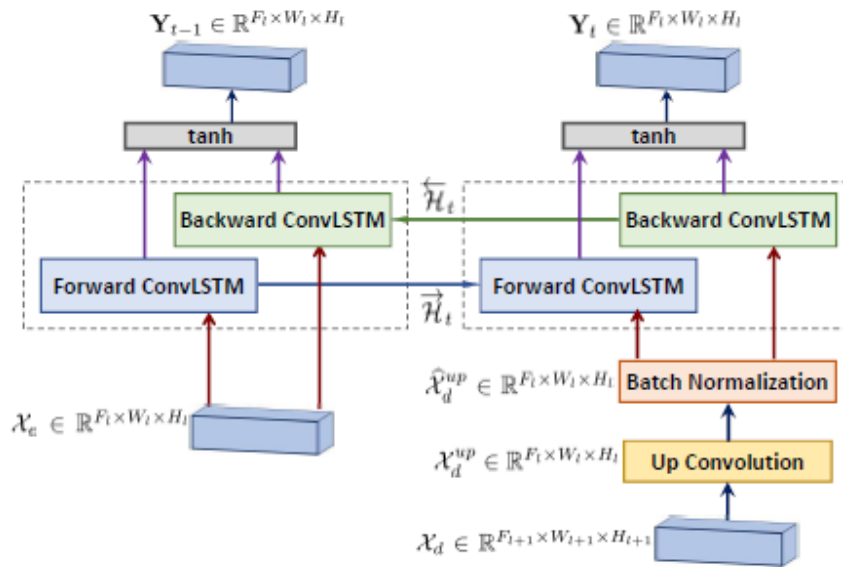
3.2.1 Κωδικοποιητής (encoding path)

Η συμβιβαστική διαδρομή του δικτύου αποτελείται από τέσσερα βήματα. Το κάθε βήμα αποτελείται από δυο συνελκτικά φίλτρα 3×3 ,μετά από τα οποία εφαρμόζεται υποδειγματοληψία μεγίστου σε περιοχές 2×2 και η συνάρτηση ReLU. Ο αριθμός των χαρτών χαρακτηριστικών διπλασιάζεται σε κάθε βήμα. Το τελευταίο βήμα του κωδικοποιητή παράγει μια υψηλής διάστασης αναπαράσταση με σημαντική σημασιολογική πληροφορία. Στο σημείο αυτό υλοποιείται και η ιδέα των πυκνά συνδεδεμένων συνελκτικών στρωμάτων (densely connected convolutions). Λογίζοντας ως μπλοκ δυο συνεχόμενα συνελκτικά στρώματα, στο τελευταίο βήμα της συμβιβαστικής διαδρομής παρουσιάζεται μια ακολουθία από μπλοκ τα οποία είναι πυκνά συνδεδεμένα. Ειδικότερα: έστω X_e^i η έξοδος του i -οστού συνελκτικού μπλοκ. Η είσοδος αυτού του μπλοκ λαμβάνει μια αλληλουχία από χάρτες χαρακτηριστικών από όλα τα προηγούμενα μπλοκ, $[X_e^1, X_e^2, \dots, X_e^{i-1}] \in R^{(i-1)F_i \times W_i \times H_i}$ και τελικά προκύπτει η έξοδος $X_e^i \in R^{F_i \times W_i \times H_i}$.

Η μέθοδος αυτή ευνοεί την επαναχρησιμοποίηση χαρακτηριστικών καθώς όλα τα μπλοκ έχουν εύκολη πρόσβαση στα προηγούμενα βοηθώντας και το δίκτυο να μάθει διαφορετικά χαρακτηριστικά στηριζόμενο στην ήδη υπάρχουσα γνώση. Ακόμη εξαιτίας αυτού το δίκτυο προστατεύεται από φαινόμενα εξαφάνισης ή έκρηξης παραγώγου.

3.2.2 Αποκωδικοποιητής (decoding path)

Κάθε βήμα της επεκτατικής διαδρομής ξεκινά με μια υπερδειγματοληψία επί της εξόδου του προηγούμενου στρώματος. Στο αρχικό δίκτυο U-Net, οι χάρτες ενεργοποίησης που παράγονται σε κάθε βήμα της συμβιβαστικής διαδρομής περικόπτονται και συνδέονται σειριακά με τους χάρτες χαρακτηριστικών που προκύπτουν ως έξοδος κάθε βήματος υπερδειγματοληψίας (upsampling). Αντ' αυτού στο δίκτυο BCDU-Net χρησιμοποιούνται μονάδες BConvLSTM για την συνένωση των δυο ειδών χαρτών χαρακτηριστικών κατά ένα πιο περίπλοκο τρόπο. Έστω λοιπόν $X_e \in \mathbb{R}^{F_l \times W_l \times H_l}$ ο σύνολο των χαρτών που παράγονται από ένα δεδομένο βήμα της συμβιβαστικής διαδρομής και $X_d \in \mathbb{R}^{F_{l+1} \times W_{l+1} \times H_{l+1}}$ το σύνολο των χαρτών χαρακτηριστικών του προηγούμενου συνελκτικού επιπέδου, όπου F_l ο αριθμός των χαρτών χαρακτηριστικών στο επίπεδο l και $W_l \times H_l$ οι διαστάσεις καθενός από αυτούς τους χάρτες. Επίσης σημειώνεται ότι $F_{l+1} = 2 * F_l$, $W_{l+1} = 2 * W_l$ και $H_{l+1} = 2 * H_l$. Όπως φαίνεται και στην παρακάτω εικόνα, το σύνολο χαρτών X_d προωθείται στον αποκωδικοποιητή όπου υφίσταται μια διαδικασία υπερδειγματοληψίας ακολουθούμενη από δυο διαδοχικά συνελκτικά φίλτρα 2×2 με αποτέλεσμα των διπλασιασμό του μεγέθους κάθε χάρτη χαρακτηριστικών και τον υποδιπλασιασμό του αριθμού των καναλιών, παράγοντας εντέλει το σύνολο $X_d^{up} \in \mathbb{R}^{F_l \times W_l \times H_l}$. Με άλλα λόγια η επεκτατική διαδρομή αυξάνει το μέγεθος των χαρτών χαρακτηριστικών σε κάθε επίπεδο με σκοπό να φτάσει το αρχικό μέγεθος της εικόνας εισόδου μετά και το τελευταίο επίπεδο.



Σχήμα 3.2: Αμφίδρομες μονάδες ConvLSTM in CUA-Net [45]

3.2.3 Κανονικοποίηση παρτίδας (batch normalization)

Ένα σύνηθες πρόβλημα που συναντάται κατά την διαδικασία εκπαίδευσης ενός νευρωνικού δικτύου είναι ότι η κατανομή των εισόδων κάθε επιπέδου μεταβάλλεται σε κάθε βήμα εκπαίδευσης προς αντιστοιχία με την μεταβολή των παραμέτρων του προηγούμενου επιπέδου. Το γεγονός αυτό επιβραδύνει την εκπαίδευση του δικτύου καθώς σε κάθε βήμα πρέπει να υιοθετηθεί μια νέα κατανομή από κάθε επίπεδο. Μια λύση για την αντιμετώπιση του συγκεκριμένου προβλήματος αποτελεί

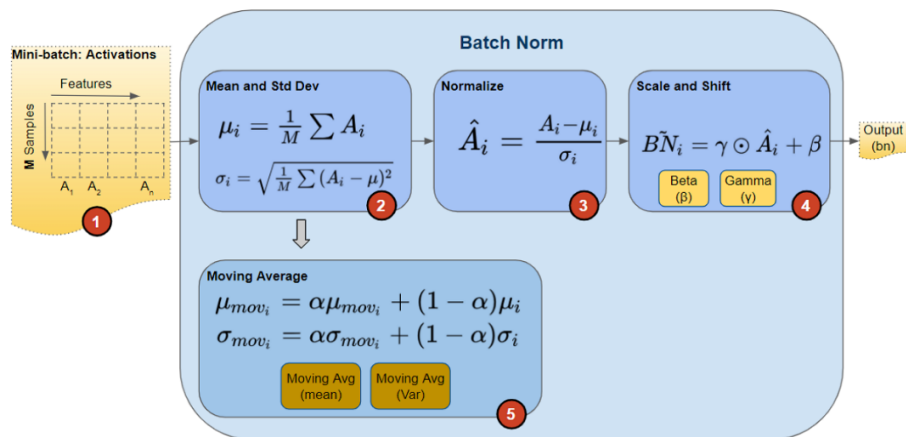
η κανονικοποίηση των εισόδων κάθε επιπέδου, [50] επιτρέποντας μας τελικά να αυξήσουμε τον ρυθμό εκπαίδευσης και να κάνουμε μια «λιγότερο προσεκτική» αρχικοποίηση των παραμέτρων του δικτύου, βελτιώνοντας σε πολλές περιπτώσεις και την απόδοση.

Κάθε στρώμα κανονικοποίησης παρτίδων χαρακτηρίζεται από:

- τις εκπαιδευόμενες (learnable) παραμέτρους β (beta) και γ (gamma)
- και τις μη εκπαιδευόμενες (non learnable) παραμέτρους Μέσος Κινούμενος Μέσος όρος (Mean Moving Average) και Διακύμανση Κινούμενου Μέσο όρου (Variance Moving Average)

Κατά την διαδικασία πρόσθιας τροφοδότησης, τα δεδομένα που διέρχονται του στρώματος κανονικοποίησης ακολουθούν την εξής πορεία:

Αρχικά υπολογίζεται ο μέσος όρος και η διακύμανση των δεδομένων της παρτίδας. Ακολούθως, γίνεται η κανονικοποίηση των διανυσμάτων ενεργοποίησης (activation vector) όπως φαίνεται στο πλαίσιο 2 του 3.3. Έτσι πλέον έχουμε μηδενικό μέσο όρο και μοναδιαία διακύμανση. Έπειτα οι κανονικοποιημένες τιμές πολλαπλασιάζονται κατά Hadamard με τον παράγοντα γ (gamma) και στο γινόμενο προστίθεται ο παράγοντας β (beta). Παράλληλα υπολογίζονται και οι τιμές των παραμέτρων Μέσος Κινούμενος Μέσος όρος (Mean Moving Average) και Διακύμανση Κινούμενου Μέσο όρου (Variance Moving Average) που δεν χρησιμοποιούνται κατά την φάση της εκπαίδευσης (training phase) αλλά κατά την φάση των συμπερασμάτων (inference phase) όπου το δίκτυο δεν τροφοδοτείται πλέον με παρτίδες δεδομένων αλλά με ένα μεμονωμένο δείγμα (sample).



Σχήμα 3.3: Υπολογισμοί σε ένα επίπεδο κανονικοποίησης παρτίδας (batch normalization layer) [51]

Η καινοτομία του στρώματος κανονικοποίησης παρτίδων έγκειται στο γεγονός ότι δεν επιτελεί απλά μια κανονικοποίηση των διανυσμάτων εισόδου κάθε επιπέδου αλλά εκπαιδεύεται στο να βρει τις καλύτερες τιμές των παραμέτρων β (beta) και γ (gamma) ώστε να μετατοπίσει και να κλιμακώσει τις κανονικοποιημένες τιμές εισόδου και να πετύχει μεγαλύτερη προβλεπτική ακρίβεια για το δίκτυο μας.

Στην περίπτωση λοιπόν του δικτύου BCDU-Net μετά την υπερδειγματοληψία, το σύνολο χαρτών X_d^{up} διέρχεται ενός στρώματος κανονικοποίησης παρτίδων (batch normalization layer), που τελικά παράγει το σύνολο χαρτών \hat{X}_d^{up} .

3.2.4 Bi-Directional ConvLSTM

Το βασικότερο μειονέκτημα των δικτύων LSTM όσων αφορά την χρήση τους για κατάτμηση εικόνων εντοπίζεται στο ότι λόγω των πλήρων συνδέσεων (full connections) που χρησιμοποιούν για τις μεταβάσεις μεταξύ εισόδου-κατάστασης (input-state) και κατάστασης-κατάστασης (state-state) δεν λαμβάνουν υπόψη χωρικές συσχετίσεις. Προς επίλυση αυτού του προβλήματος προτείνεται η χρήση μονάδων ConvLSTM [52], όπου όλες οι εισοδοί X_1, \dots, X_t , οι έξοδοι των κελιών C_1, \dots, C_t , οι κρυφές καταστάσεις H_1, \dots, H_t και οι πύλες i_t, f_t, o_t αποτελούν 3D ταυυστές (tensors) των οποίων οι δύο τελευταίες διαστάσεις είναι χωρικές (γραμμές και στήλες). Προκειμένου να αποκτήσουμε μια καλύτερη εικόνα των εισόδων και των καταστάσεων του ConvLSTM, μπορούμε να τα φανταστούμε ως διανύσματα ενός χωρικού πλέγματος (spatial grid). Το ConvLSTM ορίζει την μελλοντική κατάσταση ενός συγκεκριμένου κελιού στο πλέγμα με βάση τις εισόδους αλλά και παρελθοντικές καταστάσεις των γειτόνων του. Αυτό μπορεί εύκολα να επιτευχθεί χρησιμοποιώντας τον τελεστή της συνέλιξης (εικόνα 3.5) για τις μεταβάσεις μεταξύ εισόδου-κατάστασης (input-state) και κατάστασης-κατάστασης (state-state). Η λειτουργία της μονάδας ConvLSTM περιγράφεται μαθηματικά από τις παρακάτω εξισώσεις:

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \odot C_{t-1} + b_i) \quad (3.1)$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \odot C_{t-1} + b_f) \quad (3.2)$$

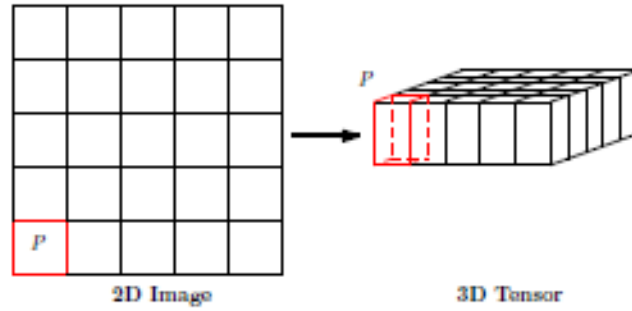
$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (3.3)$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \odot C_t + b_o) \quad (3.4)$$

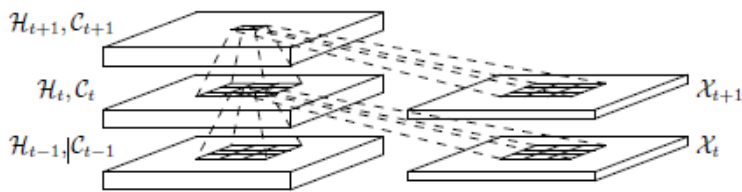
$$H_t = o_t \odot \tanh(C_t) \quad (3.5)$$

Προκειμένου να διασφαλίσουμε ότι τα διανύσματα καταστάσεων έχουν τον ίδιο αριθμό στηλών και γραμμών με τα διανύσματα εισόδου γίνεται padding πριν από την υλοποίηση της συνέλιξης. Συνήθως πριν τροφοδοτήσουμε την πρώτη είσοδο στο δίκτυο, αρχικοποιούμε όλες τις καταστάσεις του LSTM στο μηδέν το οποία συνεπάγεται «πλήρη άγνοια» για τις μελλοντικές καταστάσεις. Έτσι ομοίως, αν εφαρμόσουμε zero-padding στις κρυφές καταστάσεις θεωρούμε ότι δεν έχουμε προηγούμενη γνώση στην περιοχή που συμπληρώνεται με μηδενικά.

Στο BCDU-Net λοιπόν γίνεται χρήση αμφίδρομων ConvLSTM ή διαφορετικά BConvLSTM [53] προκειμένου να κωδικοποιήσουμε τα σύνολα χαρτών X_e και X_d^{up} . Μια μονάδα BConvLSTM χρησιμοποιεί δυο ConvLSTM προκειμένου να επεξεργαστεί τα δεδομένα που δέχεται και προς τις δυο κατευθύνσεις και ακολούθως να παρθεί μια απόφαση για την τρέχουσα είσοδο. Σε ένα συμβατικό ConvLSTM λαμβάνονται υπόψη μόνο οι εξαρτήσεις που προκύπτουν από την πρόσθια πέρασμα των δεδομένων. Ωστόσο πρέπει να λάβουμε υπόψη το σύνολο της πληροφορίας σε μια ακολουθία εισόδου και γι' αυτό θα ήταν ωφέλιμο να λάβουμε υπόψη και τις εξαρτήσεις που προκύπτουν από το οπίσθιο πέρασμα καθώς φαίνεται να βελτιώνει την απόδοση του μοντέλου. Σε ένα BConvLSTM



Σχήμα 3.4: Μετατροπή 2D εικόνας σε 3D τένσορα [52]



Σχήμα 3.5: Εσωτερική δομή ConvLSTM [52]

έχουμε δυο σύνολα παραμέτρων που αναφέρονται στις καταστάσεις πρόσθιας και οπίσθιας κατεύθυνσης. Η έξοδος του BConvLSTM υπολογίζεται ως εξής:

$$Y_t = \tanh(W_y \vec{H}_t * \vec{H}_t + W_y \overleftarrow{H}_t + b) \quad (3.6)$$

Όπου \vec{H}_t και \overleftarrow{H}_t αποτελούν τους ταχυστές για τις πρόσθιες και οπίσθιες κρυφές καταστάσεις αντίστοιχα, το b αντιστοιχεί στην πόλωση και το $Y_t \in R^{F_l \times W_l \times H_l}$ στην τελική έξοδο λαμβάνοντας υπόψη την αμφίδρομη χωρο-χρονική πληροφορία. Επιπλέον η συνάρτηση \tanh χρησιμοποιείται προκειμένου να συνδυαστούν οι έξοδοι και των δυο κατευθύνσεων με έναν μη-γραμμικό τρόπο.

3.3 ResNet

Το βάθος ενός νευρωνικού δικτύου φαίνεται να αναδεικνύεται σε σημαντικό παράγοντα βελτίωσης της απόδοσης του [49], [54]. Ωστόσο καθώς αυξάνεται ο αριθμός των στρωμάτων εντείνεται ο κίνδυνος εξαφάνισης/έκρηξης κλίσης (vanishing/exploding gradient) οδηγώντας εντέλει σε μη αποτελεσματική εκπαίδευση του. Το πρόβλημα αυτό αντιμετωπίζεται μέσω κανονικοποιημένης αρχικοποίησης των βαρών του δικτύου (normalized initialization) [55] καθώς και με την προσθήκη ενδιάμεσων επιπέδων κανονικοποίησης (intermediate normalization layers)

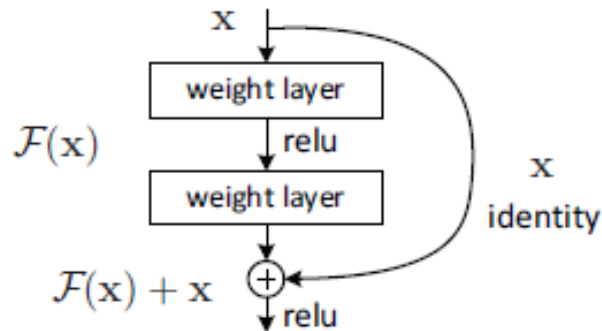
Παρ' όλα αυτά σε πολλές περιπτώσεις η αύξηση του βάθους εξακολουθεί να μην συνοδεύεται από αντίστοιχη αύξηση των επιδόσεων του δικτύου. Με την εισαγωγή των residual blocks οι σχεδιαστές της οικογένειας των δικτύων ResNet [47] κατέστησαν εφικτή την αύξηση του βάθους του δικτύου με ταυτόχρονη βελτίωση των επιδόσεων.

3.3.1 Residual blocks

Η λειτουργία κάθε residual block περιγράφεται από την παρακάτω μαθηματική σχέση:

$$y_l = x_l + F(x_l, W_l) \quad (3.7)$$

όπου x_l, y_l η είσοδος και η έξοδος του l residual block, W_l το σύνολο των βαρών(και των πολώσεων) του και F η συνάρτηση που εφαρμόζεται στην είσοδο x_l του block.

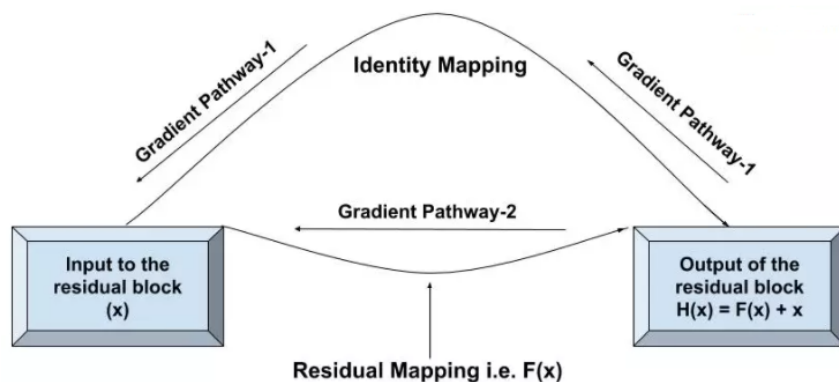


Σχήμα 3.6: Τυπική δομή ενός residual block [47]

Όπως φαίνεται στην εικόνα 3.6, κάθε block διαθέτει μια σύνδεση που παρακάμπτει όλα τα ενδιαμέσα επίπεδα διαμορφώνοντας μέσω πρόσθεσης την τελική έξοδο $F + x$. Η σύνδεση αυτή δεν εισάγει επιπλέον παραμέτρους και δεν αυξάνει την υπολογιστική πολυπλοκότητα του δικτύου. Σημειώνεται ότι η είσοδος x του δικτύου καθώς και η συνάρτηση που χρησιμοποιείται πρέπει να έχουν ίδιες διαστάσεις. Σε περίπτωση που η συνθήκη αυτή δεν ικανοποιείται εφαρμόζεται γραμμική προβολή W_s μέσω της παρακαμπτήριας σύνδεσης ως εξής:

$$y_l = W_s x_l + F(x_l, W_l) \quad (3.8)$$

Κατά την διάρκεια της οπισθοδιάδοσης, στην φάση εκπαίδευσης του δικτύου, η μετάδοση των κλίσεων μπορεί να ακολουθήσει δυο διαδρομές καθώς διασχίζει το residual block όπως φαίνεται στην εικόνα 3.7.



Σχήμα 3.7: Διαδρομές μετάδοσης κλίσης κατά την εκπαίδευση σε ένα residual block [56]

Κατά την μετάδοση μέσω της διαδρομής 2 (Gradient Pathway-2) ακολουθείται ο αλγόριθμος οπισθοδιάδοσης, όπως περιγράφεται σε προηγούμενη ενότητα, με αποτέλεσμα την μεταβολή στην τιμή της κλίσης αλλά και των βαρών του residual block. Για την αποφυγή προβλημάτων έκρηξης/εξαφάνισης κλίσης γίνεται εισαγωγή της παρακαμπτήριας σύνδεσης (Gradient Pathway 1) που φροντίζει για την αυτούσια μετάδοση του σφάλματος προς την είσοδο, δίχως μεταβολές στις τιμές των βαρών του block. Έτσι το βάθος του δικτύου αυξάνεται χωρίς επιπλοκές στην διαδικασία εκπαίδευσης και το δίκτυο συγκλίνει ομαλά. [57]

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Σχήμα 3.8: Αρχιτεκτονική των μοντέλων της οικογένειας ResNet [47]

3.3.2 Εφαρμογές

Μεταξύ των μοντέλων της οικογένειας ResNet, επιλέξαμε να χρησιμοποιήσουμε το ResNet-50 λόγω της επιτυχούς εφαρμογής του σε tasks που απαιτείται επεξεργασία ιατρικών εικόνων όπως διάγνωσης καρκίνου του μαστού με ιστοπαθολογικές εικόνες [58], πρόβλεψης της ανταπόκρισης του διατηρητικού χημειοβολισμού στο ηπατοκυτταρικό καρκίνωμα μέσω αξονικών τομογραφιών [59] καθώς και διάγνωσης COVID-19 με ακτινογραφίες θώρακα. [60]

3.4 EfficientNet

Στην δημοσίευση [48] προτείνεται η ταυτόχρονη κλιμάκωση του βάθους d , του πλάτους w καθώς και της ανάλυσης r ενός νευρωνικού δικτύου μέσω ενός σύνθετου παράγοντα ϕ . Όπως καταδεικνύεται και στο [61], τόσο το βάθος όσο και το πλάτος του δικτύου είναι καθοριστικής σημασίας της εκφραστικότητας του, ενώ αύξηση της ανάλυσης σε δίκτυα για αναγνώριση αντικειμένων [62] και χαρακτήρων [63] οδήγησε σε αύξηση της ακρίβειας που επιτυγχάνουν. Έτσι αν και είναι δυνατή η αυθαίρετη κλιμάκωση δυο ή τριών εκ των παραμέτρων του δικτύου (d, w, r), στην πλειοψηφία των περιπτώσεων δεν επιτυγχάνεται βέλτιστη ακρίβεια.

Η προτεινόμενη μέθοδος σχετίζει τον σύνθετο παράγοντα ϕ με τους επιπλέον υπολογιστικούς πόρους που μπορούν να διατεθούν για το δίκτυο. Έτσι σε περίπτωση διάθεσης 2^N επιπλέον πόρων, το βάθος αυξάνεται κατά α^N , το πλάτος κατά β^N και η ανάλυση κατά γ^N όπου α, β, γ

σταθεροί παράγοντας που καθορίζονται μέσω grid search. Σημειώνεται ότι ο αριθμός των FLOPS (Floating-point Operations per Second) μιας κανονικής συνελικτικής πράξης είναι ανάλογος προς d, w^2, r^2 που συνεπάγεται ότι ο διπλασιασμός του βάθους του δικτύου θα διπλασιάσει τα FLOPS αλλά ο διπλασιασμός του πλάτους ή της ανάλυσης θα τετραπλασιάσει τα FLOPS. Έτσι λοιπόν επισημαίνεται ότι καθώς οι συνελικτικές πράξεις καθορίζουν την πολυπλοκότητα των συνελικτικών δικτύων, η κλιμάκωση αυτών δια της προτεινόμενης μεθόδου θα αυξήσει τον συνολικό αριθμό των FLOPS κατά $(\alpha \cdot \beta^2 \cdot \gamma^2)^\phi$.

Διαισθητικά κατανοώντας την συγκεκριμένη μέθοδο, αύξηση στην ανάλυση της εικόνας εισόδου είναι λογικό να συνεπάγεται με αύξηση των επιπέδων, δηλαδή του βάθους του δικτύου, ώστε να αυξηθεί το πεδίο υποδοχής (receptive fields) καθώς και με αύξηση του πλάτους, δηλαδή των καναλιών, προκειμένου να καταγραφούν πιο λεπτομερή μοτίβα.

3.4.1 EfficientNet-B0

Στο [48] γίνεται εισαγωγή μιας νέας οικογένειας δικτύων, που καλούνται EfficientNets. Το EfficientNet-B0 αποτελεί το μοντέλο-βάση για την ανάπτυξη των υπολοίπων μοντέλων της οικογένειας μέσω σύνθετης κλιμάκωσης και η διάρθρωση του φαίνεται στον παρακάτω πίνακα.

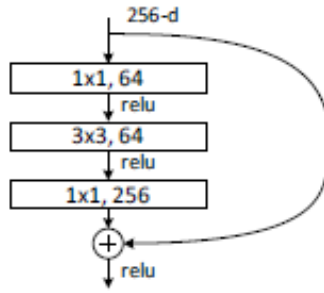
Stage i	Operator \mathcal{F}_i	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels \hat{C}_i	#Layers \hat{L}_i
1	Conv3x3	224×224	32	1
2	MBCConv1, k3x3	112×112	16	1
3	MBCConv6, k3x3	112×112	24	2
4	MBCConv6, k5x5	56×56	40	2
5	MBCConv6, k3x3	28×28	80	3
6	MBCConv6, k5x5	14×14	112	3
7	MBCConv6, k5x5	14×14	192	4
8	MBCConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

Σχήμα 3.9: Αρχιτεκτονική δικτύου EfficientNet-B0: κάθε γραμμή περιγράφει ένα στάδιο i με \hat{L}_i επίπεδα, με είσοδο διαστάσεων $\langle \hat{H}_i, \hat{W}_i \rangle$ και έξοδο \hat{C}_i κανάλια [48]

Βασική δομική μονάδα του δικτύου αποτελεί το mobile inverted bottleneck MBCConv [64] στο οποίο προστέθηκε τεχνολογία συμπίεσης-διέγερσης (squeeze-and-excitation) [65] προς βελτιστοποίηση του.

3.4.2 Inverted Bottleneck MBCConv

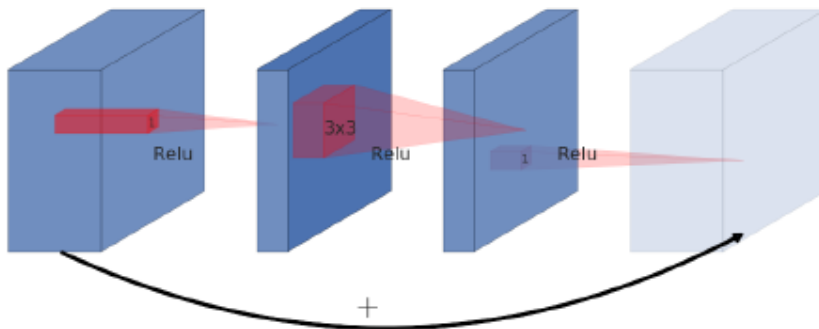
Η αρχική ιδέα του Bottleneck block εισάγεται στο [47]. Η δομή του περιλαμβάνει ένα συνελικτικό φίλτρο 1×1 που μειώνει το πλήθος των καναλιών της εισόδου η είσοδος του του φίλτρου ορίζεται ως $B \times C \times H \times W$ και η έξοδος που προκύπτει είναι διαστάσεων $B \times \frac{C}{r} \times H \times W$. Έπειτα εφαρμόζεται ένα συνελικτικό φίλτρο 3×3 και ακολούθως και πάλι ένα συνελικτικό φίλτρο 1×1 για να επαναφέρει την έξοδο στον ίδιο χώρο χαρακτηριστικών (δηλαδή στις ίδιες διαστάσεις) με την είσοδο του block. Στην έξοδο κάθε συνελικτικού επιπέδου εφαρμόζεται η συνάρτηση ενεργοποίησης ReLU. Η χαρακτηριστική ονομασία Bottleneck προκύπτει λόγω της αρχικής μείωσης του βάθους της εισόδου.



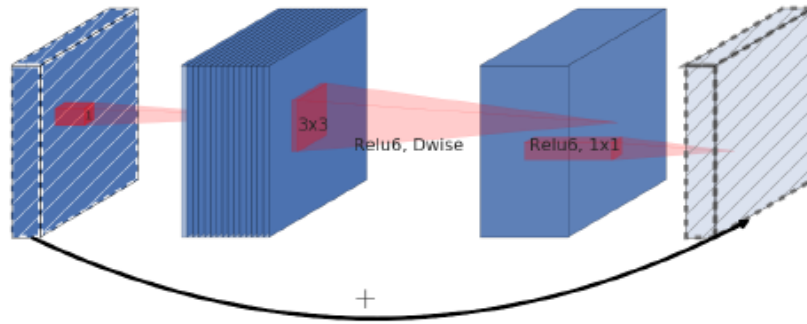
Σχήμα 3.10: Δομή bottleneck block [47]

Κατά την ανάπτυξη του MobileNetV2 [64] γίνεται εισαγωγή της ιδέας των γραμμικών Bottleneck blocks, τα οποία προκύπτουν από ένα συμβατικό Bottleneck block όπως αυτό αναφέρθηκε παραπάνω χωρίς την τελευταία συνάρτηση ενεργοποίησης προκειμένου να αποτραπεί η μεγάλη απώλεια πληροφορίας που προκαλείται από μη γραμμικές συναρτήσεις ενεργοποίησης όπως η ReLU.

Εν συνεχεία, τα Inverted Residual blocks προκύπτουν ως ανεστραμμένα Bottleneck layers. Έτσι αρχικά αυξάνεται ο αριθμός των καναλιών του διανύσματος εισόδου αντί να μειώνεται και τελικά μέσω του block η είσοδος $B \times C \times H \times W$ γίνεται $B \times eC \times H \times W$, έπειτα $B \times eC \times H \times W$ και τελικά $B \times C \times H \times W$, όπου e η αναλογία επέκτασης (expansion ratio).



Σχήμα 3.11: Residual block [64]



Σχήμα 3.12: Inverted residual block
[64]

3.4.3 Squeeze-and-Excitation

Στο [65] αναπτύσσεται η ιδέα των Squeeze-and-Excitation blocks που στοχεύουν στην βελτίωση της ποιότητας των αναπαραστάσεων που παράγονται από ένα δίκτυο, μέσα από την μοντελοποίηση των αλληλεξαρτήσεων μεταξύ των καναλιών των συνελκτικών επιπέδων. Προτείνεται δηλαδή ένας μηχανισμός που αναπροσαρμόζει τις τιμές ενός διανύσματος χαρακτηριστικών (feature recalibration) ανά επίπεδο βάθους, μαθαίνοντας να χρησιμοποιεί καθολική (global) πληροφορία που εξάγεται ανά κανάλι ώστε να αξιοποιούνται τα σημαντικά χαρακτηριστικά και να μην δίνεται τόση έμφαση στα λιγότερο χρήσιμα.

Έτσι λοιπόν αρχικά τα διανύσματα χαρακτηριστικών U διέρχονται ενός Global Average Pooling επιπέδου και η είσοδος διαστάσεων $C \times H \times W$ δίνει έξοδο $C \times 1 \times 1$ σύμφωνα με την παρακάτω εξίσωση:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (3.9)$$

Έτσι επιτελείται συμπίεση (squeeze) καθολικής χωρικής πληροφορίας ανά κανάλι. Ακολούθως εφαρμόζεται η εξίσωση και κάθε είσοδος z αντιστοιχίζεται σε ένα σύνολο βαρών για κάθε κανάλι.

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (3.10)$$

όπου:

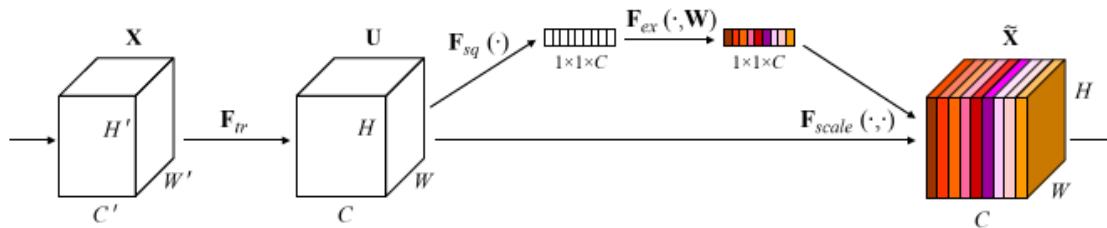
δ : η συνάρτηση ReLU

σ : η σιγμοειδής συνάρτηση

W_1, W_2 : τα διανύσματα βαρών δυο Fully Connected επιπέδων

3.4.4 Εφαρμογές

Τα μοντέλα της οικογένειας EfficientNet έχουν χρησιμοποιηθεί για κατηγοριοποίηση όγκων του εγκεφάλου με χρήση μαγνητικών τομογραφιών [66], περιστατικών COVID-19 και πνευμονίας με χρήση ακτινογραφιών θώρακα [67] καθώς και διάφορων τύπων καρκίνων του δέρματος [68]



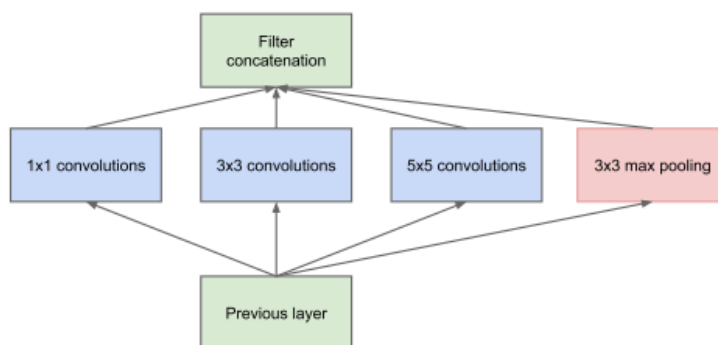
Σχήμα 3.13: Squeeze-and-Excitation blocks [65]

επιτυγχάνοντας σημαντικές επιδόσεις με υψηλά ποσοστά ακρίβειας. Αυτό αποτέλεσε σημαντικό επιχείρημα υπέρ της χρήσης του μοντέλου EfficientNet-B0 για τους σκοπούς της παρούσας εργασίας.

3.5 Inception

Η πληροφορία σε μια εικόνα μπορεί να εντοπίζεται άλλοτε σε μια μικρή περιοχή της και άλλοτε να είναι διάχυτη σε όλη την απεικόνιση. Το γεγονός αυτό καθιστά ιδιαίτερα δύσκολη την επιλογή μεγέθους για το συνελκτικό φίλτρο που εφαρμόζεται. Φίλτρα διαστάσεων 1×1 , 3×3 φαίνονται καταλληλότερα για τον εντοπισμό πληροφορίας που κατανέμεται σε μικρή περιοχή της απεικόνισης ενώ ένα φίλτρο 5×5 είναι προτιμότερο σε περίπτωση που η πληροφορία είναι διάχυτη στην εικόνα. Επιπλέον η τεχνική της «αφελούς» στοίβαξης συνελκτικών επιπέδων προσθέτει σημαντικό υπολογιστικό κόστος στο υπό κατασκευή δίκτυο.

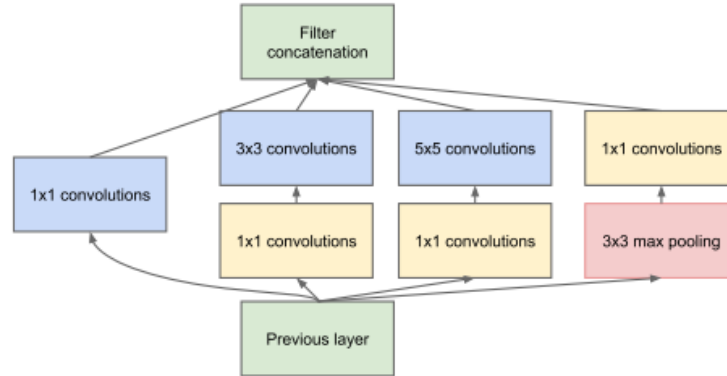
Οι παραπάνω διαπιστώσεις υποκίνησαν την δημιουργία της οικογένειας μοντέλων Inception. [49] Κύριος άξονας της αρχιτεκτονικής τους είναι η εφαρμογή φίλτρων διαφόρων μεγεθών στο ίδιο επίπεδο. Έτσι το δίκτυο δεν επεκτείνεται σε βάθος αλλά σε πλάτος. Η αρχική δομική μονάδα ενός δικτύου Inception φαίνεται παρακάτω:



Σχήμα 3.14: Αρχική δομική μονάδα δικτύου Inception [49]

Προκειμένου να μειωθεί το υπολογιστικό κόστος που ανακύπτει από την εφαρμογή των συνελκτικών φίλτρων 3×3 , 5×5 προστίθεται πριν από αυτά ένα επιπλέον συνελκτικό φίλτρο 1×1 . Παρότι αρχικά μπορεί να φαίνεται πιο «ακριβή» αυτή η αλληλουχία συνελκτικών φίλτρων ένα-

ντι ενός φίλτρου, εντούτοις το φίλτρο 1×1 μειώνει το πλήθος των καναλιών του διανύσματος χαρακτηριστικών που διέρχονται των φίλτρων 3×3 , 5×5 και έτσι το συνολικό υπολογιστικό κόστος μειώνεται. Παράλληλα, ένα ακόμη συνελκτικό φίλτρο 1×1 προστίθεται μετά το max pooling layer. Η νέα δομική μονάδα φαίνεται παρακάτω:

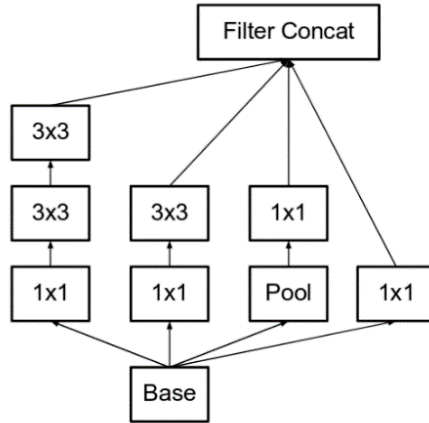


Σχήμα 3.15: Δομική μονάδα δικτύου Inception με μείωση διαστάσεων [49]

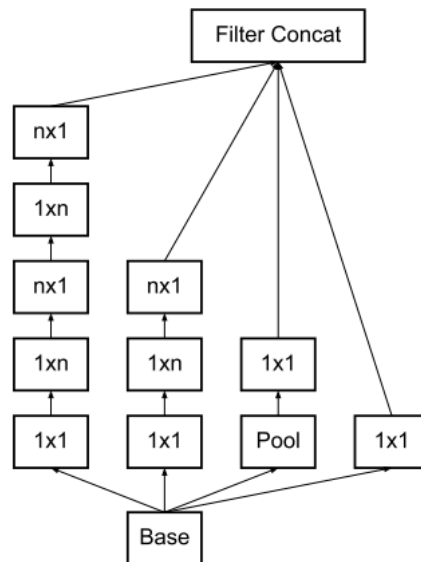
Το GoogleNet που αποτέλεσε το αρχικό δίκτυο που προτάθηκε στο [49] περιέχει 9 inception modules , ενώ το συνολικό βάθος του είναι 22 επίπεδα, συνυπολογίζοντας μόνο τα επίπεδα με παραμέτρους (με την προσθήκη των max pooling layers έχουμε 27 επίπεδα). Δεδομένου του βάθους του δικτύου, προκειμένου να αποφευχθεί το πρόβλημα της εξαφάνισης κλίσης, οι συγγραφείς εισήγαγαν 2 βοηθητικούς ταξινομητές που τοποθετήθηκαν αντίστοιχα μετά τα 2 inception modules που βρίσκονται στην μέση του δικτύου. Κατά την διάρκεια της εκπαίδευσης η απώλεια αυτών των ταξινομητών σταθμίζεται βάσει ενός παράγοντα (στην δημοσίευση τίθεται 0.3) και προστίθεται στην συνολική απώλεια του δικτύου.

Ωστόσο παρατηρήθηκε ότι η υπερβολική μείωση των διαστάσεων του διανύσματος χαρακτηριστικών , ειδικά όταν αυτή πραγματοποιείται στα πρώτα επίπεδα του δικτύου οδηγεί σε σημαντική απώλεια πληροφορίας, ενώ υπάρχουν και περιθώρια για περαιτέρω βελτίωση του υπολογιστικού κόστους του δικτύου. Έτσι στο [69] προτείνονται ορισμένες βελτιώσεις της παραπάνω αρχιτεκτονικής.

Αρχικά το φίλτρο 5×5 αντικαθίσταται από δυο φίλτρα 3×3 με σκοπό την βελτίωση της υπολογιστικής ταχύτητας του δικτύου. Έτσι δεδομένου ότι το φίλτρο 5×5 είναι 2.78 φορές ακριβότερο υπολογιστικά ($\frac{25}{9} = 2.78$) από το 3×3 ,σημειώνεται βελτίωση της απόδοσης .



Σχήμα 3.16: Αντικατάσταση φίλτρου 5×5 από δυο φίλτρα 3×3 [69]

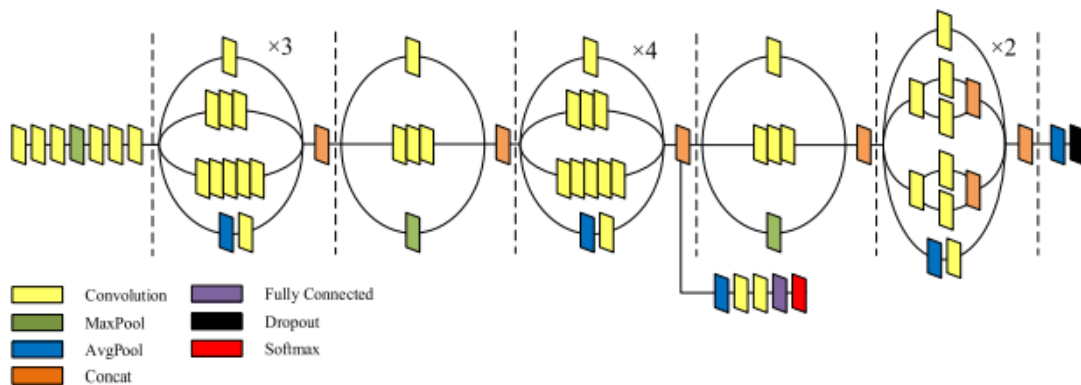


Σχήμα 3.17: Αντικατάσταση φίλτρων $n \times n$ από αλληλουχία φίλτρων $1 \times n, n \times 1$ [69]

Ακολούθως φίλτρα διαστάσεων $n \times n$ αντικαθίστανται με την αλληλουχία φίλτρων $1 \times n$ και $n \times 1$ όπου $n = 3, 5, 7$. Αντικατάσταση του φίλτρου 3×3 με την αλληλουχία 3×1 και 1×3 οδηγεί σε μείωση του υπολογιστικού κόστους κατά 33.3%. Παράλληλα το δίκτυο επεκτείνεται επιπλέον σε πλάτος, γίνεται χρήση RMSProp optimizer και Batch Normalization στους βοηθητικούς ταξινομητές ενώ κάποια φίλτρα 3×3 αντικαθίστανται με φίλτρα 7×7 και τελικά διαμορφώνεται η αρχιτεκτονική του Inception-V3 που χρησιμοποιείται στην παρούσα εργασία.

3.5.1 Εφαρμογές

Η επιτυχής χρήση του μοντέλου Inception-V3 για διάγνωση διαταραχών αυτιστικού φάσματος μέσω λειτουργικών μαγνητικών τομογραφιών [71], πρόβλεψη μετάστασης στους λεμφαδένες μέσω υπέρηχων πρωτοπαθών καρκίνων του μαστού [72] καθώς και για διάγνωση COVID-19 με



Σχήμα 3.18: Αρχιτεκτονική δικτύου Inception-V3 [70]

χρήση ακτινογραφιών θώρακα [73] κατέδειξε την δυνατότητα χρήσης του μοντέλου όπου απαιτείται επεξεργασία ιατρικών εικόνων.

3.6 CNN-RNN

Τα βαθιά συνελκτικά δίκτυα (DCNN) ενδείκνυνται για την επεξεργασία εικόνων ([74], [49], [47]), ενώ τα ανατροφοδοτούμενα νευρωνικά δίκτυα για δεδομένα που έχουν ακολουθιακή μορφή, βρίσκοντας έτσι εφαρμογή μεταξύ άλλων στα πεδία αναγνώρισης φωνής [75] και μετάφρασης κειμένου [76]. Ένα υβριδικό δίκτυο DCNN-RNN, που προτάθηκε αρχικά στο [77] για αναγνώριση ακολουθίας βάσει εικόνας βρίσκοντας εφαρμογή στην αναγνώριση κειμένου σε εικόνες φυσικού περιβάλλοντος, εκμεταλλεύεται τόσο την ικανότητα των συνελκτικών δικτύων να εξάγουν τα χωρικά χαρακτηριστικά της εισόδου όσο και των ανατροφοδοτούμενων δικτύων να εκμεταλλεύονται την χρονική σειρά των δεδομένων.

Στην παρούσα εργασία το δίκτυο CNN χρησιμοποιείται για να εξάγει χάρτες χαρακτηριστικών από τις εικόνες αξονικών τομογραφιών (CT slices) και οι μονάδες GRU, που το ακολουθούν, για να αξιοποιήσουν τις μεταξύ τους χρονικές εξαρτήσεις.

Έτσι λοιπόν οι εικόνες της αξονικής τομογραφίας τροφοδοτούνται αρχικά στο συνελκτικό δίκτυο για την εξαγωγή των απαραίτητων χωρικών χαρακτηριστικών, η οποία γίνεται ταυτόχρονα για όλα τα slices που απαρτίζουν μια δεδομένη αξονική τομογραφία. Υπογραμμίζεται ότι από τα συνελκτικά μοντέλα που χρησιμοποιήθηκαν (ResNet-50, EfficientNet-b0 και Inception-V3) αφαιρέθηκε το στάδιο της κατηγοριοποίησης που περιλαμβάνει τα πλήρως συνδεδεμένα επίπεδα και προστέθηκε ένα επίπεδο global average pooling.

Εν συνεχεία οι χάρτες χαρακτηριστικών τροφοδοτούνται στις αντίστοιχες μονάδες GRU, που αξιοποιούν την αξονική (axial) εξάρτηση των εικόνων της τομογραφίας. Μετά το επίπεδο GRU ακολουθεί ένα Flatten επίπεδο, ώστε να επιτευχθεί η μετατροπή από το 2D διάνυσμα που παράγεται από το επίπεδο GRU σε 1D και να καταστεί εφικτή η επεξεργασία των δεδομένων από τα δυο πλήρως διασυνδεδεμένα (Dense) επίπεδα που βρίσκονται στην κορυφή του μοντέλου και επιτελούν την επιθυμητή κατηγοριοποίηση.

Καθώς ο όγκος των εικόνων (ct slices) που περιέχονται σε μια αξονική τομογραφία δεν είναι σταθερός και δεδομένου ότι πρέπει να καθοριστούν οι συγκεκριμένες διαστάσεις εισόδου για το

προς εκπαίδευση δίκτυο, ορίστηκε 70 ο μέγιστος όγκος εικόνων ανά τομογραφία. Σε περίπτωση που υπολείπονται εικόνες, επαναλαμβάνεται η τελευταία εικόνα ως ότου συμπληρωθεί ο επιθυμητός αριθμός ενώ στην περίπτωση που το πλήθος των εικόνων υπερβαίνει το 70, εφαρμόζεται υποδειγματοληψία βάση ενός καθορισμένου βήματος.

Κεφάλαιο 4

Πειραματική διαδικασία

4.1 Μετρικές Αξιολόγησης

Για την αξιολόγηση των μοντέλων κατηγοριοποίησης θα γίνει χρήση ορισμένων δεικτών που ονομάζονται μετρικές αξιολόγησης. Στην παρούσα εργασία θα χρησιμοποιήσουμε τις μετρικές accuracy, precision, recall και F1-score. Για τον ορισμό αυτών των μετρικών αρχικά ορίζουμε ως:

- **True Positive:** ο αριθμός των δειγμάτων της πρώτης κλάσης που κατηγοριοποιήθηκαν σωστά
- **True Negative:** ο αριθμός των δειγμάτων της δεύτερης κλάσης που κατηγοριοποιήθηκαν σωστά
- **False Positive:** ο αριθμός των δειγμάτων της δεύτερης κλάσης, τα οποία κατηγοριοποιήθηκαν στην πρώτη κλάση
- **False Negative:** ο αριθμός των δειγμάτων της πρώτης κλάσης, τα οποία κατηγοριοποιήθηκαν στην δεύτερη κλάση

Αναλυτικά λοιπόν προκύπτουν οι ακόλουθοι ορισμοί των μετρικών που χρησιμοποιούνται:

Accuracy: είναι ο λόγος των δειγμάτων για τα οποία έγινε σωστή πρόβλεψη προς τον συνολικό αριθμό των προβλέψεων. Η συγκεκριμένη μετρική δεν προτείνεται για μη ισορροπημένα σύνολα δεδομένων (imbalance datasets), όπως επιβεβαιώνεται από το ακόλουθο παράδειγμα. Έστω ότι έχουμε 2 κλάσεις και η πρώτη κλάση αποτελεί το 99% του συνόλου δεδομένων, ενώ η δεύτερη κλάση το 1%. Το μοντέλο μπορεί να πετύχει accuracy 99% προβλέποντας συνεχώς την πρώτη κλάση, αλλά δεν θα μάθει ποτέ να προβλέπει την δεύτερη κλάση, κάτι το οποίο δεν είναι επιθυμητό.

$$Accuracy = \frac{TruePositive + TrueNegative}{TotalNumberofPredictions} \quad (4.1)$$

Precision: Είναι ο λόγος των δειγμάτων μιας κλάσης που σωστά κατηγοριοποιήθηκαν σε αυτήν από το μοντέλο, προς το συνολικό αριθμό των δειγμάτων που κατηγοριοποιήθηκαν σε αυτήν την κλάση. Η συγκεκριμένη μετρική χρησιμοποιείται όταν η απόδοση του μοντέλου εξαρτάται κυρίως από το πλήθος των λανθασμένα προβλεφθέντων θετικών. Έτσι λοιπόν σε ένα μοντέλο που προβλέπει αν μια επένδυση θα είναι επικερδής ή όχι, σε περίπτωση που η επένδυση κατηγοριοποιηθεί

λανθασμένα ως επικερδής προκαλείται ζημιά. Συνεπώς προτιμάται μια επικερδής επένδυση να ταξινομηθεί λανθασμένα ως μη επικερδής, παρά το αντίθετο και η μετρική precision κρίνεται κατάλληλη. Η μετρική precision της κλάσης positive προσδιορίζεται ως εξής:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (4.2)$$

Recall:Είναι ο λόγος των δειγμάτων μιας κλάσης που σωστά κατηγοριοποιήθηκαν σε αυτήν από το μοντέλο προς το συνολικό αριθμό των δειγμάτων της κλάσης. Συγκεκριμένα αναφέρουμε ως:

- **Sensitivity:**τον λόγο των θετικών δειγμάτων που σωστά κατηγοριοποιήθηκαν ως θετικά, προς τον συνολικό αριθμό των θετικών δειγμάτων. Έτσι στην περίπτωση ενός διαγνωστικού μοντέλου προσδιορίζει την περίπτωση ένα θετικό αποτέλεσμα να είναι όντως θετικό.

$$Sensitivity = \frac{TruePositive}{FalseNegative + TruePositive} \quad (4.3)$$

- **Specifity:**τον λόγο των αρνητικών δειγμάτων που σωστά κατηγοριοποιήθηκαν ως αρνητικά, προς τον συνολικό αριθμό των αρνητικών δειγμάτων. Στην περίπτωση ενός διαγνωστικού μοντέλου προσδιορίζει την περίπτωση ένα αρνητικό αποτέλεσμα να είναι όντως αρνητικό.

$$Specifity = \frac{TrueNegative}{TrueNegative + FalsePositive} \quad (4.4)$$

F1-score:Είναι ο αρμονικός μέσος μεταξύ precision και recall και χρησιμοποιείται για μη ισορροπημένα σύνολα δεδομένων.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4.5)$$

4.2 Μείωση ct scan volume

Το dataset που χρησιμοποιήθηκε για την ανάπτυξη του μοντέλου κατηγοριοποίησης των ct scan slices σε LUNGS/NO-LUNGS, δηλαδή σε slices στα οποία απεικονίζονται πνεύμονες και σε slices στα οποία δεν απεικονίζονται πνεύμονες αντίστοιχα, περιλαμβάνει συνολικά 6255 ct scan slices σε μορφή .jpg, στα οποία όποια έχει γίνει segmentation και εκ των οποίων τα 3636 κατατάσσονται στην κατηγορία LUNGS ενώ τα 2619 στην κατηγορία NO-LUNGS. Ο διαχωρισμός του dataset σε train, validation και test set έγινε ως εξής: 65% χρησιμοποιήθηκε για training, 15% για validation και 20% για testing. Όλα τα slices έγιναν resize στις διαστάσεις 224×224 ώστε να έχουν το ίδιο μέγεθος και να μειωθεί η χωρική και η υπολογιστική πολυπλοκότητα των χρησιμοποιούμενων μοντέλων. Επίσης καθώς τα μοντέλα που χρησιμοποιούνται δέχονται σαν είσοδο εικόνες 3 καναλιών, ενώ τα ct scan slices είναι grayscale, εισάγουμε δυο αντίγραφα κάθε slice προκειμένου να λάβουμε διάσταση εισόδου $224 \times 224 \times 3$.

Κατά την εκπαίδευση χρησιμοποιήθηκε η συνάρτηση κόστους categorical-crossentropy και ο αλγόριθμος βελτιστοποίησης Adam με τις προκαθορισμένες τιμές παραμέτρων για τα μοντέλα ResNet-50 και Inception-V3, ενώ για το μοντέλο EfficientNet-b0 χρησιμοποιήθηκε ο αλγόριθμος

SGD με learning rate 0.001 . Επίσης τα βάρη για κάθε μοντέλο αρχικοποιήθηκαν στις τιμές των βαρών των αντίστοιχων pretrained μοντέλων στο ImageNet. Ο αριθμός εποχών(epochs) ορίστηκε ίσος με 40 ενώ το μέγεθος δέσμης(batch size) ίσο με 100.

Παρακάτω παρατίθενται τα αποτελέσματα που προέκυψαν:

	Accuracy	Precision	Recall	F1-score
LUNGS	0.980	0.978	0.988	0.983
NO-LUNGS		0.984	0.969	0.976

Πίνακας 4.1: Μετρικές αξιολόγησης του μοντέλου ResNet-50 για τις κλάσεις LUNGS/NO-LUNGS

	Accuracy	Precision	Recall	F1-score
LUNGS	0.971	0.966	0.986	0.976
NO-LUNGS		0.980	0.952	0.965

Πίνακας 4.2: Μετρικές αξιολόγησης του μοντέλου EfficientNet-b0 για τις κλάσεις LUNGS/NO-LUNGS

	Accuracy	Precision	Recall	F1-score
LUNGS	0.983	0.974	0.997	0.985
NO-LUNGS		0.996	0.963	0.979

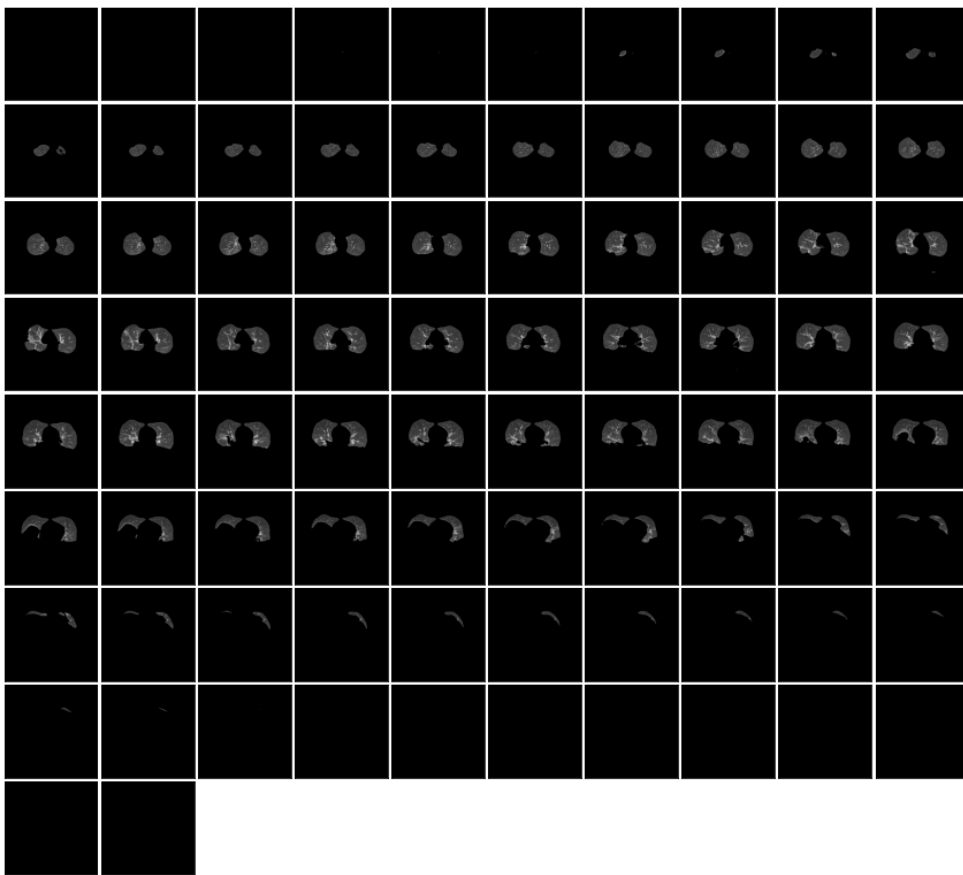
Πίνακας 4.3: Μετρικές αξιολόγησης του μοντέλου Inception-V3 για τις κλάσεις LUNGS/NO-LUNGS

Όπως παρατηρείται όλα τα μοντέλα σημείωσαν αξιόλογες επιδόσεις, ωστόσο αυτό που ξεχωρίζει είναι το **Inception-V3**, το οποίο εντέλει επιλέγεται ως το βασικό μοντέλο για μείωση του όγκου των εικόνων ανά αξονική τομογραφία. Επιτυγχάνει υψηλό ποσοστό accuracy (98.3%) ενώ το υψηλό ποσοστό sensitivity (99.7%) , διασφαλίζει ότι τα slices που έχουν ταξινομηθεί ως LUNGS , σχεδόν βέβαια θα απεικονίζουν περιοχή πνευμόνων και έτσι το μοντέλο που επιτελεί εν συνεχεία την τελική διάγνωση ανά scan σε COVID-19/non-COVID-19 δεν θα τροφοδοτηθεί με περιττό και άχρηστο όγκο εικόνων, που δεν περιέχουν πληροφορία χρήσιμη ώστε να επιτελεστεί επιτυχώς η παραπάνω ταξινόμηση

4.3 Κατηγοριοποίηση ct scans σε COVID-19/non-COVID-19

Το dataset που χρησιμοποιήθηκε για την ανάπτυξη του μοντέλου κατηγοριοποίησης των ct scans σε COVID-19/non-COVID-19 περιλαμβάνει 54 ct scans εκ των οποίων τα 32 ανήκουν σε ασθενείς με COVID-19 και τα 22 σε άτομα που δεν νοσούν από την ασθένεια. Όπως προβλέπεται από την αρχιτεκτονική που έχει δομηθεί στην παρούσα διπλωματική, τα slices των ct scans έχουν διέλθει του segmentation model με σκοπό την ανίχνευση της περιοχής των πνευμόνων και την διατήρηση μόνο αυτής, για το υπόλοιπο του δικτύου. Ακολούθως τα segmented slices κάθε scan διέρχονται του μοντέλου Inception-V3 ,ώστε να διατηρηθούν μόνο όσα από αυτά απεικονίζουν πνεύμονες.

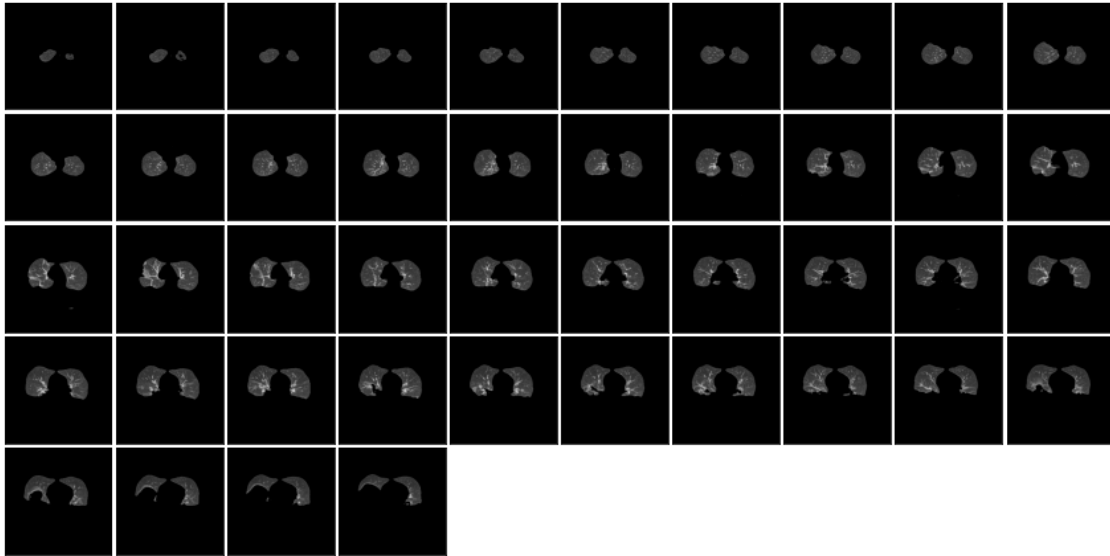
Σημειώνεται ότι χάρη στο Inception-V3 που επιτελεί την κατηγοριοποίηση των slices σε LUNGS/NO-LUNGS επιτυγχάνεται σημαντική μείωση του όγκου των εικόνων ανά αξονική τομογραφία, που κατά μέσο όρο προσδιορίζεται σε ποσοστό 39.63% του αρχικού τους όγκου. Σε πολλές περιπτώσεις μάλιστα το ποσοστό αυτό είναι αρκετά μεγαλύτερο. Ενδεικτικά αναφέρονται μερικά παραδείγματα, όπως αυτά ανέκυψαν από την επεξεργασία του dataset: από 555 slices σε 118 που αντιστοιχεί σε ποσοστό μείωσης 78.74% ,από 313 σε 131 slices που αντιστοιχεί σε ποσοστό μείωσης 58.15% και από 295 σε 149 slices που αντιστοιχεί σε ποσοστό μείωσης 49.49%. Τα υψηλά αυτά ποσοστά καταδεικνύουν ακόμη περισσότερο την χρηστικότητα αυτού του μοντέλου, καθώς φαίνεται ότι ο όγκος της περιττής πληροφορίας ανά scan είναι ιδιαίτερα μεγάλος.



Σχήμα 4.1: Αρχικός όγκος CT scan slices

Στην εικόνα 4.1 παρατίθεται η οπτική αναπαράσταση ενός ct scan που αρχικά αποτελείτο από 82 slices και με την χρήση του μοντέλου κατηγοριοποίησης σε LUNGS/NO-LUNGS παρέμειναν τελικά 44, όπως αποτυπώνεται στην εικόνα 4.2, που αντιστοιχεί σε ποσοστό μείωσης 46.34%. Όπως φαίνεται ο όγκος των «άχρηστων» εικόνων τοποθετείται στην αρχή και στο τέλος του συνολικού όγκου της τομογραφίας και καταφέρνει τελικά να εξαιρεθεί χάρη στο χρησιμοποιούμενο μοντέλο, καθιστώντας και οπτικά την διαφορά εμφανή.

Τελικά ο όγκος των slices ανά scan, που αρχικά κυμαίνονταν από 44 μέχρι 555, μειώνεται στο εύρος 22 με 344. Ωστόσο όπως αναφέρθηκε και παραπάνω καθώς πρέπει να καθοριστούν οι συγκεκριμένες διαστάσεις εισόδου για το προς εκπαίδευση μοντέλο, καθώς το δίκτυο τροφοδοτείται



Σχήμα 4.2: Τελικός όγκος CT scan slices

ταυτόχρονα με όλο τον όγκο των slices ενός δεδομένου ct scan, ο αριθμός των slices ανά τομογραφία ορίστηκε 70.

Ο διαχωρισμός του dataset σε train, validation και test set έγινε όπως και στο προηγούμενο στάδιο, δηλ: 65% χρησιμοποιήθηκε για training, 15% για validation και 20% για testing. Οι διαστάσεις κάθε slice τίθενται και πάλι 224×224 και εν συνεχεία εισάγονται δυο αντίγραφα κάθε slice προκειμένου να λάβουμε τελικά διάσταση εισόδου $224 \times 224 \times 3$, καθώς, όπως έχει ήδη αναφερθεί, τα CNN μοντέλα που χρησιμοποιούνται ως βάση του υβριδικού δικτύου κατηγοριοποίησης σε COVID-19/non-COVID-19 δέχονται σαν είσοδο εικόνες 3 καναλιών. Διευκρινίζεται ότι η συνολική διάσταση εισόδου για το παρών μοντέλο ορίζεται $70 \times 224 \times 224 \times 3$.

Κατά την εκπαίδευση χρησιμοποιήθηκε η συνάρτηση κόστους sparse-categorical-crossentropy και ο αλγόριθμος βελτιστοποίησης Adam για τα υβριδικά μοντέλα που είχαν ως συνελικτική βάση τα ResNet-50 και EfficientNet-b0 αντίστοιχως, ενώ για το μοντέλο με συνελικτική βάση το Inception-V3 χρησιμοποιήθηκε ο αλγόριθμος SGD. Τα συνελικτικά μοντέλα είναι προεκπαιδευμένα στο σύνολο ImageNet και δεν εκπαιδεύτηκαν εκ νέου. Για την ρύθμιση του ρυθμού εκπαίδευσης χρησιμοποιήθηκε η συνάρτηση `tf.keras.optimizers.schedules.ExponentialDecay` για την οποία τέθηκαν οι εξής τιμές παραμέτρων: `initial_learning_rate=0.0001`, `decay_steps=10000`, `decay_rate=0.96` και `staircase=True`. Επιπλέον καθώς το σύνολο δεδομένων που διαθέτουμε είναι imbalanced, προκειμένου να αποφευχθεί κατά το δυνατόν το μοντέλο να μάθει μόνο τα χαρακτηριστικά της κλάσης με τα περισσότερα δείγματα, γίνεται χρήση της παραμέτρου `class_weights` της συνάρτησης `fit` που χρησιμοποιείται κατά την εκπαίδευση του μοντέλου. Ο αριθμός εποχών (epochs) ορίστηκε ίσος με 30 ενώ το μέγεθος δέσμης (batch size) ίσο με 2, δηλαδή 2 ct scans των 70 slices.

Τα αποτελέσματα που προέκυψαν φαίνονται στους πίνακες [4.4-4.6](#).

	Accuracy	Precision	Recall	F1-score
non-COVID-19	0.700	0.667	0.500	0.571
COVID-19		0.714	0.834	0.768

Πίνακας 4.4: Μετρικές αξιολόγησης του μοντέλου ResNet-50-RNN για τις κλάσεις non-COVID-19/COVID-19

	Accuracy	Precision	Recall	F1-score
non-COVID-19	0.600	0.500	0.500	0.500
COVID-19		0.667	0.667	0.667

Πίνακας 4.5: Μετρικές αξιολόγησης του μοντέλου EfficientNet-b0-RNN για τις κλάσεις non-COVID-19/COVID-19

	Accuracy	Precision	Recall	F1-score
non-COVID-19	0.700	0.600	0.750	0.667
COVID-19		0.800	0.667	0.727

Πίνακας 4.6: Μετρικές αξιολόγησης του μοντέλου Inception-V3-RNN για τις κλάσεις non-COVID-19/COVID-19

Όσον αφορά τον σχολιασμό των αποτελεσμάτων, αρχικά πρέπει να διευκρινιστεί ότι η μετρική accuracy δεν είναι πλήρως αντιπροσωπευτική καθώς το dataset που χρησιμοποιήθηκε, όπως υπογραμμίστηκε και παραπάνω δεν είναι πλήρως ισορροπημένο (ειδικότερα 60% των δειγμάτων αντιστοιχεί σε ασθενείς με COVID ενώ 40% σε υγιή άτομα). Σε κάθε περίπτωση δίνει μια πρώτη εικόνα των αποτελεσμάτων και παρατίθεται για λόγους πληρότητας. Δεδομένου του μικρού όγκου των δεδομένων που είχαμε στην διάθεση μας (συνολικά 54 ct scans) θα λέγαμε ότι τα αποτελέσματα είναι ικανοποιητικά. Ειδικότερα το μοντέλο ResNet-50-RNN φαίνεται να πετυχαίνει το υψηλότερο ποσοστό sensitivity (83.4%) μεταξύ των τριών μοντέλων ενώ το Inception-V3-RNN το υψηλότερο ποσοστό specificity (75.0%). Αυτές οι δύο μετρικές όπως αναφέρθηκε και στο εισαγωγικό κεφάλαιο κρίνονται ως οι πιο σημαντικές για ένα διαγνωστικό test, καθώς στην πρώτη περίπτωση διασφαλίζεται η ορθότητα του θετικού αποτελέσματος ενώ στην δεύτερη περίπτωση η ορθότητα του αρνητικού αποτελέσματος, που κρίνεται ιδιαίτερα σημαντική για την ανάσχεση της μετάδοσης του ιού.

Κεφάλαιο 5

Συμπεράσματα και Μελλοντική Εργασία

Στην παρούσα διπλωματική εργασία σημειώθηκε προσπάθεια ανάπτυξης ενός αξιόπιστου διαγνωστικού μοντέλου της νόσου COVID-19 μέσω της χρήσης αξονικών τομογραφιών θώρακα. Ιδιαίτερα σημαντικά κρίθηκαν τα βήματα προεπεξεργασίας που εφαρμόστηκαν προτού γίνει εξαγωγή της τελικής διάγνωσης. Σε αυτά περιλαμβάνονται η χρήση ενός προεκπαιδευμένου μοντέλου για κατάτμηση των εικόνων των αξονικών τομογραφιών καθώς και ενός μοντέλου, το οποίο εκπαιδεύτηκε εκ του μηδενός και μειώνει τον όγκο των εικόνων ανά αξονική τομογραφία, πετώντας στην ουσία τις εικόνες στις οποίες δεν αποτυπώνεται η περιοχή των πνευμόνων. Έτσι το τελικό μοντέλο που επιτελεί την διάγνωση τροφοδοτείται με την κατά το δυνατότερο ελάχιστη ποσότητα περιττής πληροφορίας, γεγονός που κρίνεται ιδιαίτερα σημαντικό ώστε να ανιχνεύσει τα μοτίβα που παρατηρούνται μεταξύ των τομογραφιών.

Τα αποτελέσματα που προέκυψαν ήταν ικανοποιητικά, ωστόσο η συγκέντρωση ενός μεγαλύτερου συνόλου δεδομένων για την εκπαίδευση του μοντέλου που επιτελεί την κατηγοριοποίηση των τομογραφιών σε COVID-19 και non-COVID-19 αναμένεται να οδηγήσει σε σημαντική βελτίωση της απόδοσης. Δεδομένου βέβαια ότι πολλές φορές ο όγκος των ιατρικών δεδομένων που είναι διαθέσιμα, είναι περιορισμένος θα είχε ενδιαφέρον το σύνολο δεδομένων να εμπλουτιστεί και με συνθετικές εικόνες που δημιουργούνται μέσω νευρωνικών δικτύων GANs. Παράλληλα, θα άξιζε να εφαρμοστούν επιπλέον τεχνικές προεπεξεργασίας όπως περικοπή των εικόνων ακριβώς γύρω από την περιοχή των πνευμόνων και μεταβολή της αντίθεσης για βελτίωση της ποιότητας τους. Τέλος, ενδιαφέρον θα παρουσίαζε και η δοκιμή και άλλων γνωστών αρχιτεκτονικών πέραν των όσων χρησιμοποιήθηκαν στα πλαίσια της παρούσας εργασίας.

Βιβλιογραφία

- [1] D. Vasireddy, R. Vanaparthi, G. Mohan, S. V. Malayalad, and P. Atlurie, “Review of covid-19 variants and covid-19 vaccine efficacy: What the clinician should know?”
- [2] V. Checchi, P. Bellini, D. Bencivenni, and U. Consolo, “Covid-19 dentistry-related aspects: a literature overview,” *International dental journal*, vol. 71, no. 1, pp. 21–26, 2021.
- [3] A. Akhmerov and E. Marbán, “Covid-19 and the heart,” *Circulation research*, vol. 126, no. 10, pp. 1443–1455, 2020.
- [4] W. Alsharif and A. Qurashi, “Effectiveness of covid-19 diagnosis and management tools: A review,” *Radiography*, vol. 27, no. 2, pp. 682–687, 2021.
- [5] D. Kollias, A. Arsenos, L. Soukissian, and S. Kollias, “Mia-cov19d: Covid-19 detection through 3-d chest ct image analysis,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 537–544.
- [6] I. Kollia, A.-G. Stafylopatis, and S. Kollias, “Predicting parkinson’s disease using latent information extracted from deep neural networks,” in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.
- [7] 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)*. IEEE, pp. 1–8.
- [8] 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, 2018.
- [9] A. Tagaris, D. 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. 03, p. 1850011, 2018.
- [10] A. Tagaris, D. Kollias, and A. Stafylopatis, “Assessment of parkinson’s disease based on deep neural networks,” in *International Conference on Engineering Applications of Neural Networks*. Springer, 2017, pp. 391–403.
- [11] J. Wingate, I. Kollia, L. Bidaut, and S. Kollias, “Unified deep learning approach for prediction of parkinson’s disease,” *IET Image Processing*, vol. 14, no. 10, pp. 1980–1989, 2020.
- [12] D. Kollias, N. Bouas, Y. Vlaxos, V. Brillakis, M. Seferis, I. Kollia, L. Sukissian, J. Wingate, and S. Kollias, “Deep transparent prediction through latent representation analysis,” *arXiv preprint arXiv:2009.07044*, 2020.

- [13] D. Kollias, Y. Vlaxos, M. Seferis, I. Kollia, L. Sukissian, J. Wingate, and S. Kollias, “Transparent adaptation in deep medical image diagnosis,” in *International Workshop on the Foundations of Trustworthy AI Integrating Learning, Optimization and Reasoning*. Springer, 2020, pp. 251–267.
- [14] H. Shi, X. Han, N. Jiang, Y. Cao, O. Alwalid, J. Gu, Y. Fan, and C. Zheng, “Radiological findings from 81 patients with covid-19 pneumonia in wuhan, china: a descriptive study,” *The Lancet infectious diseases*, vol. 20, no. 4, pp. 425–434, 2020.
- [15] S. Machnicki, D. Patel, A. Singh, A. Talwar, B. Mina, M. Oks, P. Makkar, D. Naidich, A. Mehta, N. S. Hill *et al.*, “The usefulness of chest ct imaging in patients with suspected or diagnosed covid-19: a review of literature,” *Chest*, vol. 160, no. 2, pp. 652–670, 2021.
- [16] A. Pascual-Iglesias, J. Canton, A. M. Ortega-Prieto, J. M. Jimenez-Guardeño, and J. A. Regla-Nava, “An overview of vaccines against sars-cov-2 in the covid-19 pandemic era,” *Pathogens*, vol. 10, no. 8, p. 1030, 2021.
- [17] Harvard-Health, “Treatments for covid-19,” Feb 2022. [Online]. Available: <https://www.health.harvard.edu/diseases-and-conditions/treatments-for-covid-19>
- [18] C. M. Bishop, “Neural networks and their applications,” *Review of scientific instruments*, vol. 65, no. 6, pp. 1803–1832, 1994.
- [19] K. Suzuki, *Artificial neural networks: methodological advances and biomedical applications*. BoD–Books on Demand, 2011.
- [20] M. Wallace, I. Maglogiannis, K. Karpouzis, G. Kormentzas, and S. Kollias, “Intelligent one-stop-shop travel recommendations using an adaptive neural network and clustering of history,” *Information Technology & Tourism*, vol. 6, no. 3, pp. 181–193, 2003.
- [21] K. V. Batista, “Neural networks and deep learning,” Apr 2020. [Online]. Available: https://www.academia.edu/42751314/Neural_Networks_and_Deep_Learning
- [22] B. Muller, J. Reinhardt, and M. Strickland, *Neural networks: an introduction*, 2nd ed. Springer, 1995.
- [23] S. Ruder, “An overview of gradient descent optimization algorithms,” *arXiv preprint arXiv:1609.04747*, 2016.
- [24] H. Bommana, “Understanding optimizers. exploring how the different popular... | by harsha bommana | deep learning demystified | medium,” <https://medium.com/deep-learning-demystified/https-medium-com-deep-learning-demystified-understanding-optimizers-313b787a69fe>, (Accessed on 02/17/2022).
- [25] F. Caliva, F. S. De Ribeiro, A. Mylonakis, C. Demazi’ere, P. Vinai, G. Leontidis, and S. Kollias, “A deep learning approach to anomaly detection in nuclear reactors,” in *2018 International joint conference on neural networks (IJCNN)*. IEEE, 2018, pp. 1–8.

- [26] B. Khuong, “The basics of recurrent neural networks (rnns),” Jun 2020. [Online]. Available: <https://pub.towardsai.net/whirlwind-tour-of-rnns-a11effb7808f>
- [27] G. Chen, “A gentle tutorial of recurrent neural network with error backpropagation,” *arXiv preprint arXiv:1610.02583*, 2016.
- [28] T. P. Lillicrap and A. Santoro, “Backpropagation through time and the brain,” *Current opinion in neurobiology*, vol. 55, pp. 82–89, 2019.
- [29] “Understanding lstm networks.” [Online]. Available: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [30] S. Yang, X. Yu, and Y. Zhou, “Lstm and gru neural network performance comparison study: Taking yelp review dataset as an example,” in *2020 International workshop on electronic communication and artificial intelligence (IWECAI)*. IEEE, 2020, pp. 98–101.
- [31] D. Jurafsky and J. H. Martin, “Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition.”
- [32] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, *Dive into Deep Learning*, 2020, <https://d2l.ai>.
- [33] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
- [34] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
- [35] L. Torrey and J. Shavlik, “Transfer learning,” in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 2010, pp. 242–264.
- [36] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” *Journal of Big data*, vol. 3, no. 1, pp. 1–40, 2016.
- [37] Z. Wang, Z. Dai, B. Póczos, and J. Carbonell, “Characterizing and avoiding negative transfer,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 293–11 302.
- [38] “Differences between x-rays, ct scans amp; mri’s,” Sep 2021. [Online]. Available: <https://www.envrad.com/difference-between-x-ray-ct-scan-and-mri/>
- [39] “Ct scan (computed tomography): What is it, preparation amp; test details.” [Online]. Available: <https://my.clevelandclinic.org/health/diagnostics/4808-ct-computed-tomography-scan>
- [40] M.-Y. Ng, E. Y. Lee, J. Yang, F. Yang, X. Li, H. Wang, M. M.-s. Lui, C. S.-Y. Lo, B. Leung, P.-L. Khong *et al.*, “Imaging profile of the covid-19 infection: radiologic findings and literature review,” *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, p. e200034, 2020.
- [41] N. Tsapatsoulis and S. Kollias, “Face detection in color images and video sequences,” in *2000 10th Mediterranean Electrotechnical Conference. Information Technology and Electrotechnology for the Mediterranean Countries. Proceedings. MeleCon 2000 (Cat. No. 00CH37099)*, vol. 2. IEEE, 2000, pp. 498–502.

- [42] P. Mylonas, E. Spyrou, Y. Avrithis, and S. Kollias, "Using visual context and region semantics for high-level concept detection," *IEEE Transactions on Multimedia*, vol. 11, no. 2, pp. 229–243, 2009.
- [43] M. Kara, Z. Öztürk, S. Akpek, and A. Turupcu, "Covid-19 diagnosis from chest ct scans: A weakly supervised cnn-lstm approach," *AI*, vol. 2, no. 3, pp. 330–341, 2021.
- [44] L. O. Teixeira, R. M. Pereira, D. Bertolini, L. S. Oliveira, L. Nanni, G. D. Cavalcanti, and Y. M. Costa, "Impact of lung segmentation on the diagnosis and explanation of covid-19 in chest x-ray images," *Sensors*, vol. 21, no. 21, p. 7116, 2021.
- [45] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional convlstm u-net with densley connected convolutions," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.
- [46] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [48] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.
- [49] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [50] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.
- [51] K. Doshi, "Batch norm explained visually-how it works, and why neural networks need it," May 2021. [Online]. Available: <https://towardsdatascience.com/batch-norm-explained-visually-how-it-works-and-why-neural-networks-need-it-b18919692739>
- [52] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, vol. 28, 2015.
- [53] H. Song, W. Wang, S. Zhao, J. Shen, and K.-M. Lam, "Pyramid dilated deeper convlstm for video salient object detection," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 715–731.
- [54] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

- [55] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.
- [56] “Detailed guide to understand and implement resnets,” Sep 2019. [Online]. Available: <https://cv-tricks.com/keras/understand-implement-resnets/>
- [57] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in *European conference on computer vision*. Springer, 2016, pp. 630–645.
- [58] Q. A. Al-Haija and A. Adebajo, “Breast cancer diagnosis in histopathological images using resnet-50 convolutional neural network,” in *2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. IEEE, 2020, pp. 1–7.
- [59] J. Peng, S. Kang, Z. Ning, H. Deng, J. Shen, Y. Xu, J. Zhang, W. Zhao, X. Li, W. Gong *et al.*, “Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from ct imaging,” *European radiology*, vol. 30, no. 1, pp. 413–424, 2020.
- [60] S. Sakib, M. A. B. Siddique, M. M. R. Khan, N. Yasmin, A. Aziz, M. Chowdhury, and I. K. Tasawar, “Detection of covid-19 disease from chest x-ray images: a deep transfer learning framework,” *MedRxiv*, 2020.
- [61] M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, and J. Sohl-Dickstein, “On the expressive power of deep neural networks,” in *international conference on machine learning*. PMLR, 2017, pp. 2847–2854.
- [62] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
- [63] B. Liebl and M. Burghardt, “An evaluation of dnn architectures for page segmentation of historical newspapers,” in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 5153–5160.
- [64] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
- [65] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
- [66] A. Zhaputri, M. Hayaty, and A. D. Laksito, “Classification of brain tumour mri images using efficient network,” in *2021 4th International Conference on Information and Communications Technology (ICOIACT)*. IEEE, 2021, pp. 108–113.
- [67] G. Marques, D. Agarwal, and I. de la Torre Díez, “Automated medical diagnosis of covid-19 through efficientnet convolutional neural network,” *Applied soft computing*, vol. 96, p. 106691, 2020.

- [68] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using efficientnets—a first step towards preventing skin cancer," *Neuroscience Informatics*, p. 100034, 2021.
- [69] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
- [70] N. Dong, L. Zhao, C.-H. Wu, and J.-F. Chang, "Inception v3 based cervical cell classification combined with artificially extracted features," *Applied Soft Computing*, vol. 93, p. 106311, 2020.
- [71] L. Herath, D. Meedeniya, M. Marasingha, and V. Weerasinghe, "Autism spectrum disorder diagnosis support model using inception v3," in *2021 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, vol. 4. IEEE, 2021, pp. 1–7.
- [72] L.-Q. Zhou, X.-L. Wu, S.-Y. Huang, G.-G. Wu, H.-R. Ye, Q. Wei, L.-Y. Bao, Y.-B. Deng, X.-R. Li, X.-W. Cui *et al.*, "Lymph node metastasis prediction from primary breast cancer us images using deep learning," *Radiology*, vol. 294, no. 1, pp. 19–28, 2020.
- [73] N. S. Shadin, S. Sanjana, and N. J. Lisa, "Covid-19 diagnosis from chest x-ray images using convolutional neural network (cnn) and inceptionv3," in *2021 International Conference on Information Technology (ICIT)*. IEEE, 2021, pp. 799–804.
- [74] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.
- [75] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*. Ieee, 2013, pp. 6645–6649.
- [76] M. Vathsala and G. Holi, "Rnn based machine translation and transliteration for twitter data," *International Journal of Speech Technology*, vol. 23, no. 3, pp. 499–504, 2020.
- [77] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 11, pp. 2298–2304, 2016.