

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ κ' ΠΛΗΡΟΦΟΡΙΚΗΣ



ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τεχνητά Νευρωνικά Δίκτυα για την διάκριση τύπων
πνευμονίας

Αλέξιος Γεωργαντόπουλος Γιαννούτσος

ΑΜ:1062268

ΕΠΙΤΡΟΠΗ ΕΞΕΤΑΣΗΣ:

Γ. - Π. Κ. Οικονόμου, Ε.Δ.Ι.Π. (Επιβλέπων)

Χρήστος Μακρής, Δ.Ε.Π (Μέλος)

Δημήτριος Κοσμόπουλος, Δ.Ε.Π (Μέλος)

Ευχαριστίες

Με την ολοκλήρωση της παρούσας διπλωματικής εργασίας, θα ήθελα πρώτα να ευχαριστήσω θερμά τον επιβλέποντα κ. Οικονόμου για την καθοδήγηση και την υποστήριξή του. Η υπομονή του, η διαθεσιμότητά του και οι εύστοχες παρατηρήσεις του αποτέλεσαν πολύτιμη βοήθεια σε κάθε στάδιο της ερευνητικής πορείας.

Αντίστοιχα, εκφράζω τις ευχαριστίες μου και σε όλα τα μέλη της Τριμελούς Εξεταστικής Επιτροπής για τη βοήθειά τους.

Θέλω επίσης να ευχαριστήσω τον κ. Χατζηδούκα και τους υπεύθυνους στο Υπολογιστικό Κέντρο του Τμήματος Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής, η συμβολή των οποίων υπήρξε καθοριστική, καθώς με την υποστήριξή τους μπόρεσα να υλοποιήσω τα υπολογιστικά απαιτητικά μοντέλα της εργασίας.

Ακόμα, οφείλω να ευχαριστήσω όλους τους διδάσκοντες του Τμήματος Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής, οι οποίοι με εφοδίασαν με τις απαραίτητες γνώσεις και εργαλεία, συμβάλλοντας ουσιαστικά στη διαμόρφωση των ακαδημαϊκών και επαγγελματικών μου δεξιοτήτων.

Τέλος, θέλω να εκφράσω την απέραντη ευγνωμοσύνη μου στην οικογένειά μου. Η στήριξη, η ενθάρρυνση και οι θυσίες τους υπήρξαν καταλυτικές για να μπορέσω να φτάσω μέχρι εδώ και να ολοκληρώσω αυτό το ταξίδι.

Περίληψη

Η παρούσα Διπλωματική Εργασία επικεντρώνεται στην ανίχνευση και διάκριση τύπων πνευμονίας μέσω της αξιοποίησης Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ) και πιο συγκεκριμένα Συνελικτικών Νευρωνικών Δικτύων (CNNs). Η πνευμονία αποτελεί μία από τις πιο συχνές και επικίνδυνες λοιμώξεις του αναπνευστικού, με σημαντική επίπτωση τόσο σε παγκόσμιο επίπεδο όσο και στην Ελλάδα. Η διάγνωσή της βασίζεται κυρίως σε ακτινογραφίες θώρακος, οι οποίες χρησιμοποιούνται ως πρώτη απεικονιστική μέθοδος, αν και παρουσιάζουν περιορισμούς λόγω της ομοιότητας των απεικονίσεων με άλλες παθήσεις ή της υποκειμενικότητας στην εκτίμηση από τον ιατρό. Σε πιο πολύπλοκες περιπτώσεις, όπως όταν απαιτείται διαφοροποίηση μάζας από υγρό, αξιοποιούνται επιπλέον εξετάσεις όπως η αξονική ή η μαγνητική τομογραφία, ώστε να δοθεί πιο ακριβής διάγνωση. Η εισαγωγή των ΤΝΔ στην ιατρική διάγνωση έχει αναδείξει νέες δυνατότητες, καθώς διαθέτουν την ικανότητα να αναγνωρίζουν μοτίβα και συσχετίσεις σε μεγάλες ποσότητες δεδομένων, παρέχοντας πιο ακριβείς και γρήγορες προβλέψεις. Στόχος της εργασίας είναι η μελέτη και σύγκριση δύο σύγχρονων αρχιτεκτονικών Νευρωνικών Δικτύων, της VGG16 και της Xception, ως προς την απόδοσή τους στην ταξινόμηση ακτινογραφιών θώρακος σε τρεις κατηγορίες: υγιείς περιπτώσεις, πνευμονία ιογενούς αιτιολογίας και πνευμονία βακτηριδιακής αιτιολογίας. Η υλοποίηση πραγματοποιήθηκε σε περιβάλλον Python με κατάλληλη προεπεξεργασία των δεδομένων και εφαρμογή τεχνικών βελτιστοποίησης, με στόχο την επίτευξη αποτελεσμάτων μεγαλύτερης ακρίβειας. Τα πειραματικά ευρήματα επιβεβαιώνουν ότι τα CNNs μπορούν να αποτελέσουν πολύτιμο εργαλείο υποστήριξης της κλινικής διάγνωσης, αποδεικνύοντας τη δυναμική των ΤΝΔ ως εργαλείο υποστήριξης της κλινικής διάγνωσης.

Abstract

This Thesis focuses on the detection and classification of pneumonia types through the use of Artificial Neural Networks (ANNs) and more specifically Convolutional Neural Networks (CNNs). Pneumonia is one of the most common and dangerous respiratory infections, with significant impact both globally and Greece-wide. Its diagnosis is mainly based on chest X-rays, a process that can be challenging due to the similarity of imaging features with other conditions or the subjectivity in medical interpretation. The introduction of ANNs in medical diagnosis has revealed new possibilities, as they have the ability to recognize patterns and correlations in large volumes of data, providing more accurate and faster predictions. The aim of this work is the study and comparison of two modern Neural Network architectures, VGG16 and Xception, regarding their performance in classifying chest X-rays into three categories: healthy cases, viral pneumonia, and bacterial pneumonia. The implementation was carried out by employing the Python environment along with appropriate data preprocessing and optimization techniques, since the aim was to achieve higher accuracy results. Our experimental findings confirm that CNNs can be a valuable tool in supporting clinical diagnosis, demonstrating the potential of ANNs as a means of assisting medical decision-making.

Περιεχόμενα

1. Εισαγωγή	1
1.1 Μια αναδρομή στην ιστορία της Τεχνητής Νοημοσύνης	1
1.2 Η συμβολή της Τεχνητής Νοημοσύνης στον ιατρικό χώρο	3
1.2.1 Ιατρική Απεικόνιση και Αναγνώριση Προτύπων.....	3
1.2.2 Προγνωστική Αναλυτική, Παρακολούθηση και Εξατομικευμένη Ιατρική	4
1.2.3 Τεχνητή Νοημοσύνη στην Ανακάλυψη Φαρμάκων και στη Ρομποτική Χειρουργική	4
1.2.4 Ηθικά Ζητήματα και Προκλήσεις	5
1.3 Βασικά Πλεονεκτήματα των Τεχνητών Νευρωνικών Δικτύων	6
1.3.1 Μάθηση, Γενίκευση και Μη Γραμμικότητα	6
1.3.2 Ανθεκτικότητα, Προσαρμοστικότητα και Μάθηση από Άκρο σε Άκρο	6
1.3.3 Υπολογιστική Αποδοτικότητα και Διαχείριση Δεδομένων	7
1.3.4 Ευελιξία σε Πολλούς Τομείς.....	8
1.4 Τα Νευρωνικά Δίκτυα Γύρω Μας.....	8
1.4.1 Μέσα Κοινωνικής Δικτύωσης και Χρηματοοικονομικά	8
1.4.2 Μεταφορές και Συσκευές IoT	9
1.5 Προκλήσεις των Τεχνητών Νευρωνικών Δικτύων.....	9
1.5.1 Υψηλό Υπολογιστικό Κόστος και Απαιτήσεις Πόρων.....	9
1.5.2 Έλλειψη Ερμηνευσιμότητας και Διαφάνειας.....	10
1.5.3 Αξιοπιστία και Γενίκευση	10
1.5.4 Ευπάθεια σε Εχθρικές Επιθέσεις	10
1.6 Η Λειτουργία ενός Νευρωνικού Δικτύου	11
1.6.1 Συναρτήσεις.....	11
1.6.2 Νευρώνες	12
1.7 Εκπαίδευση Νευρωνικών Δικτύων.....	13
1.7.1 Εποπτευόμενη Μάθηση (Supervised Learning)	13
1.7.2 Μη Εποπτευόμενη Μάθηση (Unsupervised Learning)	13
1.7.3 Ενισχυτική Μάθηση (Reinforcement Learning)	14
1.8 Αλγόριθμοι Εκπαίδευσης και Βελτιστοποίηση	15
1.8.1 Οπισθοδιάδοση Σφάλματος (Backpropagation)	15
1.8.2 Καθοδική Κλίση (Gradient Descent).....	16
1.8.3 Adam (Adaptive Moment Estimation).....	17
2 Συστήματα Υποστήριξης αποφάσεων.....	18
2.1 Εισαγωγή	18
2.2 Ιστορική Εξέλιξη των Συστημάτων Υποστήριξης Αποφάσεων	19

2.3 Κατηγορίες Συστημάτων Υποστήριξης Αποφάσεων	20
2.4 Εφαρμογές των Συστημάτων Υποστήριξης Αποφάσεων στην Ιατρική	21
2.4.1 Δομή ενός Κλινικού Συστήματος Υποστήριξης Αποφάσεων.....	22
2.4.2 Περιορισμοί και Προκλήσεις Εφαρμογής	22
3. Πνευμονία	24
3.1 Γενικές πληροφορίες.....	24
3.2 Κατηγορίες Πνευμονίας	26
3.3 Συμπτώματα και Διάγνωση	27
3.4 Μέθοδοι Ανίχνευσης και Ιατρική Απεικόνιση.....	28
3.5 Προκλήσεις στη Διάγνωση Πνευμονίας.....	30
4. Επεξεργασία Δεδομένων.....	31
4.1 Συλλογή Δεδομένων.....	31
4.2 Καθαρισμός Δεδομένων και Αντιμετώπιση Ελλιπών Τιμών.....	32
4.3 Κανονικοποίηση και Μετασχηματισμός Εικόνων.....	34
4.4 Τεχνικές Αύξησης Δεδομένων (Data Augmentation)	35
4.5 Ετεροσυσχέτιση Δεδομένων	37
4.6 Εργαλεία που χρησιμοποιήθηκαν.....	37
5. Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN)	39
5.1 Εισαγωγή	39
5.2 Αρχιτεκτονικές CNNs – Εξέλιξη και μοντέλα αναφοράς	39
5.3 Δομή και λειτουργία Συνελικτικών Τεχνητών Νευρωνικών Δικτυών	40
5.3.1 Συνελικτικό επίπεδο (Convolutional Layer)	40
5.3.2 Στρώμα Pooling	42
5.3.3 Πλήρως Συνδεδεμένα Στρώματα (Fully Connected Layers).....	43
5.4 Επιλογή Αρχιτεκτονικής CNN	44
5.5 VGG16: Δομή και χαρακτηριστικά	45
5.6 Xception: Δομή και χαρακτηριστικά	47
5.7 Τεχνικές βελτίωσης απόδοσης (Dropout, Regularization, Batch Normalization)	50
5.8 Transfer Learning και Fine-tuning	51
6. Υλοποίηση και εκπαίδευση των δικτυών.....	52
6.1 Περιβάλλον Ανάπτυξης (Hardware & Software).....	52
6.2 Εκπαίδευση μοντέλων.....	54
6.2.1 Υπερπαράμετροι Εκπαίδευσης: Learning Rate και Adam Optimizer.....	54
6.2.2 Επιλογή Batch Size και Επίδρασή του στην Εκπαίδευση	55
6.3 Προσαρμογή Τελικής Κεφαλής μέσω Μεταφοράς Μάθησης	56
6.4 Callbacks στην Εκπαίδευση Νευρωνικών Δικτύων	58

7. Αποτελέσματα και Μετρικές αξιολόγησης	61
7.1 Εισαγωγή στα αποτελέσματα	61
7.2 Μετρικές Αξιολόγησης	62
7.2.1 Accuracy και Loss.....	62
7.2.2 Precision, Recall, F1-score και Support	63
7.2.3 ROC Curves και AUC-ROC Score	64
7.2.4 Confusion Matrix	65
7.3 Ακρίβεια και απώλεια (Accuracy/Loss)	66
7.3.1 Πίνακες ακρίβειας και απώλειας	66
7.3.2 Διαγράμματα ακριβείας και απώλειας	68
7.3.3 Συγκριτική Ανάλυση Διαγραμμάτων VGG16 και Xception	70
7.4 Σύγκριση χρόνου Εκπαίδευσης και εποχών.....	71
7.4.1 Αριθμοί εποχών και Πίνακες	71
7.4.2 Χρόνος εκπαίδευσης και Πίνακες	72
7.4.3 Σχολιασμός Αποτελεσμάτων	74
7.5 Καμπύλες ROC και AUC-ROC score	75
7.5.1 Καμπύλες ROC.....	75
7.5.2 Πίνακες AUC-ROC score	77
7.6 Μετρικές και Confusion Matrix	78
7.6.1 Precision (Ακρίβεια Θετικών Προβλέψεων).....	78
7.6.2 Recall (Ανάκληση).....	79
7.6.3 F1-score	80
7.6.4 Specificity (Ειδικότητα).....	81
7.6.5 Confusion Matrix	82
8. Οπτικοποίηση Αποτελεσμάτων και Ερμηνευσιμότητα Προβλέψεων	85
8.1 Εισαγωγή στην Οπτικοποίηση με Heatmaps	85
8.2 Μεθοδολογία Grad-CAM	85
8.3 Παρουσίαση Αποτελεσμάτων	86
8.3.1 Ενδεικτικά Παραδείγματα Heatmaps	86
8.3.2 Ανάλυση Ποσοστών Πρόβλεψης Heatmaps	90
9. Συμπεράσματα και μελλοντικές προοπτικές	92
9.1 Συμπεράσματα	92
9.2 Μελλοντικές προοπτικές.....	93
Βιβλιογραφία	95

Πίνακας εικόνων

Εικόνα 1: Δομή ενός νευρώνα και ενός τεχνητού νευρωνικού δικτύου	2
Εικόνα 2: Σχηματική απεικόνιση της διαδικασίας υποστήριξης κλινικής απόφασης	18
Εικόνα 3: Γραφική απεικόνιση των κατηγοριών DSS	21
Εικόνα 4: Σχηματική απεικόνιση κυψελίδας σε πνευμονία	26
Εικόνα 5: Ενδεικτικές ακτινογραφίες θώρακος	29
Εικόνα 6: Ποσοστά διαχωρισμού του συνόλου δεδομένων	32
Εικόνα 7: Οπτική αναπαράσταση Convolutional Layer	41
Εικόνα 8: Απεικόνιση της αρχιτεκτονικής ενός CNN	43
Εικόνα 9: Γραφική απεικόνιση της αρχιτεκτονικής του αλγορίθμου VGG-16	45
Εικόνα 10: Η φιλοσοφία πίσω από την αρχιτεκτονική του Xception	48
Εικόνα 11: Γραφική απεικόνιση της αρχιτεκτονικής του αλγορίθμου Xception	49
Εικόνα 12: Εξέλιξη της ακριβείας και της απώλειας κατά την εκπαίδευση και την επικύρωση	60
Εικόνα 13: Στιγμιότυπο αποτελεσμάτων στο terminal	61
Εικόνα 14: Σύγκριση ακριβείας εκπαίδευσης/επικύρωσης και απώλειας εκπαίδευσης/επικύρωσης (για το μοντέλο VGG16)	68
Εικόνα 15: Σύγκριση ακριβείας εκπαίδευσης/επικύρωσης και απώλειας εκπαίδευσης/επικύρωσης για το μοντέλο Xception	68
Εικόνα 16: Απεικόνιση μιας τυχαίας τυπικής εκπαίδευσης στο μοντέλο VGG16	72
Εικόνα 17: Απεικόνιση μιας τυχαίας τυπικής εκπαίδευσης στο μοντέλο Xception	73
Εικόνα 18: Γραφική απεικόνιση καμπυλών ROC για το μοντέλο VGG16	75
Εικόνα 19: Γραφική απεικόνιση καμπυλών ROC για το μοντέλο Xception	76
Εικόνα 20: Διάγραμμα Confusion Matrix για το μοντέλο Xception	83
Εικόνα 21: Παράδειγμα Grad-CAM heatmap για την κλάση normal του μοντέλου VGG16	86
Εικόνα 22: Παράδειγμα Grad-CAM heatmap για την κλάση normal του μοντέλου Xception	87
Εικόνα 23: Παράδειγμα Grad-CAM heatmap για την κλάση bacterial του μοντέλου VGG16	87
Εικόνα 24: Παράδειγμα Grad-CAM heatmap για την κλάση bacterial του μοντέλου Xception	88
Εικόνα 25: Παράδειγμα Grad-CAM heatmap για την κλάση viral του μοντέλου VGG16	88
Εικόνα 26: Παράδειγμα Grad-CAM heatmap για την κλάση viral του μοντέλου Xception	89

Κατάλογος πινάκων

Πίνακας 1: Συνοπτική παρουσίαση των κύριων διάφορων μεταξύ εποπτευόμενης, μη εποπτευόμενης και ενισχυτικής μάθησης	15
Πίνακας 2: Σύνοψη βασικών τεχνικών αύξησης δεδομένων (data augmentation) και της επίδρασή τους στη γενίκευση και σε ανισόρροπα σύνολα δεδομένων.....	36
Πίνακας 3: Σύνοψη εργαλείων και της χρήσης τους στην παρούσα εργασία	38
Πίνακας 4: Χαρακτηριστικά Υλικού και Λογισμικού	52
Πίνακας 5: Αποτελέσματα ακριβείας για το μοντέλο VGG16.....	66
Πίνακας 6: Αποτελέσματα ακριβείας για το μοντέλο Xception.....	66
Πίνακας 7: Αποτελέσματα απώλειας για το μοντέλο VGG16.	67
Πίνακας 8: Αποτελέσματα απώλειας για το μοντέλο Xception.	67
Πίνακας 9: Αριθμός εποχών οπού το μοντέλο σταμάτησε να εκπαιδεύεται για το μοντέλο VGG16.....	71
Πίνακας 10: Αριθμός εποχών οπού το μοντέλο σταμάτησε να εκπαιδεύεται για το μοντέλο Xception.....	71
Πίνακας 11: Μέσος ορός χρόνου εκπαίδευσης μιας εποχής για το μοντέλο VGG16.....	73
Πίνακας 12: Μέσος ορός χρόνου εκπαίδευσης μιας εποχής για το μοντέλο Xception.....	73
Πίνακας 13: Αποτελέσματα για το AUC-ROC score του μοντέλου VGG16.....	77
Πίνακας 14: Αποτελέσματα για το AUC-ROC score του μοντέλου Xception.....	77
Πίνακας 15: Αποτελέσματα Precision ανά κλάση για το μοντέλο VGG16.	78
Πίνακας 16: Αποτελέσματα Precision ανά κλάση για το μοντέλο Xception.	79
Πίνακας 17: Αποτελέσματα Recall ανά κλάση για το μοντέλο VGG16.	79
Πίνακας 18: Αποτελέσματα Recall ανά κλάση για το μοντέλο Xception.	80
Πίνακας 19: Αποτελέσματα F1-score ανά κλάση για το μοντέλο VGG16.	80
Πίνακας 20: Αποτελέσματα F1-score ανά κλάση για το μοντέλο Xception.	81
Πίνακας 21: Αποτελέσματα Specificity ανά κλάση για το μοντέλο VGG16.....	81
Πίνακας 22: Αποτελέσματα Specificity ανά κλάση για το μοντέλο Xception.....	82
Πίνακας 23: Μέσο Confidence προβλέψεων για το μοντέλο VGG16.....	90
Πίνακας 24: Μέσο Confidence προβλέψεων για το μοντέλο Xception	90

1. Εισαγωγή

1.1 Μια αναδρομή στην ιστορία της Τεχνητής Νοημοσύνης

Η έννοια της Τεχνητής Νοημοσύνης (TN) δεν είναι καινούργια, ως επιστημονικός όρος καθιερώθηκε στα μέσα του 20ού αιώνα. Η πρώτη ουσιαστική απόπειρα επιστημονικής θεμελίωσης της TN έγινε το 1950, όταν ο Άγγλος μαθηματικός Alan Turing δημοσίευσε το πρωτοποριακό άρθρο του με τίτλο “*Computing Machinery and Intelligence*” [1]. Εκεί, έθεσε για πρώτη φορά το ερώτημα «Μπορούν οι μηχανές να σκεφτούν;» και ανέπτυξε τη γνωστή πλέον δοκιμή Turing, ως κριτήριο για τον προσδιορισμό της τεχνητής νοημοσύνης. Σύμφωνα με αυτή, αν ένας παρατηρητής δεν μπορεί να διακρίνει αν επικοινωνεί με άνθρωπο ή μηχανή, τότε η μηχανή μπορεί να θεωρηθεί νοήμων.

Το 1956, στο συνέδριο του Dartmouth College, επιστήμονες όπως οι John McCarthy, Marvin Minsky, Nathaniel Rochester και Claude Shannon, διατύπωσαν το φιλόδοξο όραμα της TN ως ανεξάρτητου πεδίου έρευνας. Στο πλαίσιο εκείνου του συνεδρίου εισήχθη επίσημα ο όρος “*Artificial Intelligence*” και διατυπώθηκε η υπόθεση ότι «κάθε πτυχή της μάθησης ή κάθε άλλη εκδήλωση της ανθρώπινης νοημοσύνης μπορεί να περιγραφεί τόσο ακριβώς, ώστε να μπορεί να αναπαραχθεί από μια μηχανή» [2]. Η δήλωση αυτή αποτέλεσε τη βάση για δεκαετίες ερευνητικής προσπάθειας και τεχνολογικής ανάπτυξης.

Τα πρώτα συστήματα που αναπτύχθηκαν, όπως το Logic Theorist των Newell και Simon (1956), στόχευαν στην αναπαραγωγή της ανθρώπινης λογικής σκέψης. Το πρόγραμμα κατάφερε να αποδείξει μαθηματικά θεωρήματα και θεωρείται το πρώτο λειτουργικό πρόγραμμα TN [3]. Ακολούθησε το General Problem Solver, ένα γενικής χρήσης μοντέλο επίλυσης προβλημάτων.

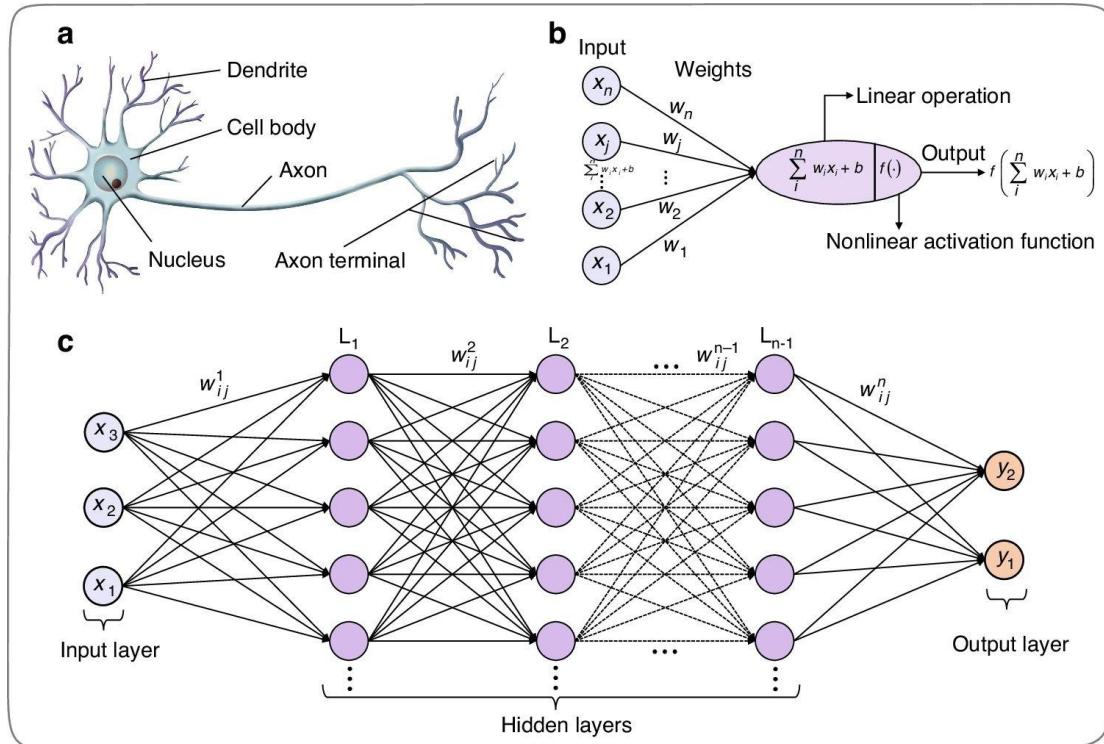
Τη δεκαετία του 1960, εμφανίστηκαν τα πρώτα προγράμματα που επιχειρούσαν την κατανόηση φυσικής γλώσσας. Το ELIZA του Joseph Weizenbaum [4] προσομοίωνε έναν ψυχοθεραπευτή, ενώ το SHRDLU του Terry Winograd [5] αλληλεπιδρούσε σε ένα απλοποιημένο περιβάλλον μέσω γλώσσας. Αργότερα, αναπτύχθηκαν και άλλα συστήματα, όπως το LUCY, το οποίο συνέβαλε στην περαιτέρω έρευνα στον τομέα της γλωσσικής κατανόησης και καταδεικνύει ότι, ακόμα και σε περιορισμένα περιβάλλοντα, μια μηχανή μπορεί να αλληλεπιδράσει με έναν άνθρωπο.

Ωστόσο, η αρχική αισιοδοξία σύντομα μετριάστηκε. Κατά τη δεκαετία του 1970, η TN γνώρισε την πρώτη της περίοδο παρακμής – τον λεγόμενο “χειμώνα της Τεχνητής Νοημοσύνης” (AI Winter) – εξαιτίας της υπερεκτίμησης των δυνατοτήτων της και της αποτυχίας εκπλήρωσης των προσδοκιών. Η περιορισμένη υπολογιστική ισχύς και η αδυναμία επίλυσης σύνθετων προβλημάτων, όπως η μάθηση και η όραση, οδήγησαν στη διακοπή πολλών ερευνητικών προγραμμάτων και στη μείωση της χρηματοδότησης.

Η κατάσταση άρχισε να αλλάζει τη δεκαετία του 1980 με την επανεμφάνιση του ενδιαφέροντος για τα Τεχνητά Νευρωνικά Δίκτυα (TNΔ). Παρόλο που τα TNΔ είχαν προταθεί ήδη από τη δεκαετία του 1940 με το μοντέλο των McCulloch και Pitts (1943), καθώς και με το perceptron του Rosenblatt τη δεκαετία του 1950, η πρακτική τους εφαρμογή ήταν περιορισμένη λόγω ανεπαρκούς υπολογιστικής ισχύος και θεωρητικών αδυναμιών. Ο αλγόριθμος backpropagation, ο οποίος επέτρεπε την αποτελεσματική εκπαίδευση πολυεπίπεδων δικτύων, αποτέλεσε σημείο καμπής. Η εποχή αυτή έφερε στο προσκήνιο

νέες εφαρμογές και δυνατότητες, αν και οι περιορισμοί που κρατούσαν την TN πίσω δεν είχαν ακόμα ξεπεραστεί.

Παρακάτω ακολουθεί μια σχηματική απεικόνιση της δομής ενός νευρώνα και ενός Τεχνητού Νευρωνικού Δικτύου.



Εικόνα 1: Δομή ενός νευρώνα και ενός τεχνητού νευρωνικού δικτύου. a) Δομή ενός βιολογικού νευρώνα. b) Μαθηματική διαδικασία εξαγωγής συμπερασμάτων στους τεχνητούς νευρώνες, που περιλαμβάνει τα δεδομένα εισόδου, τα βάρη, τη συνάθροιση, τη συνάρτηση ενεργοποίησης και την έξοδο. c) Τεχνητό νευρωνικό δίκτυο τύπου MLP(multi-layer perceptron). ([Optical neural networks: progress and challenges | Light: Science & Applications](#))

Τα TNΔ αποτελούν το επίκεντρο της παρούσας διπλωματικής εργασίας, καθώς η δομή, η λειτουργία και οι αλγόριθμοι εκπαίδευσής τους συνδέονται άμεσα με το αντικείμενο της ανάλυσης ιατρικών εικόνων και της αναγνώρισης τύπων πνευμονίας. Περισσότερο, δε, θα αναλυθούν εκτενέστερα στα επόμενα κεφάλαια.

Η δεκαετία του 1990 έφερε στο προσκήνιο την εξόρυξη δεδομένων (data mining) και τις πρώτες απόπειρες εφαρμογής της TN στη βιομηχανία. Παράλληλα, οι μηχανές υποστήριξης διανυσμάτων (SVMs) και οι αλγόριθμοι μάθησης προσέφεραν λύσεις σε πρακτικά προβλήματα [18].

Η πραγματική επανάσταση όμως ήρθε μετά το 2010. Η ανάπτυξη των βαθέων Νευρωνικών Δικτύων (Deep Neural Networks) και η επεξεργαστική δύναμη των GPU έφεραν εντυπωσιακές επιδόσεις σε τομείς όπως η αναγνώριση εικόνας, η φυσική γλώσσα και η

αυτόνομη πλοήγηση. Το AlexNet (2012), που κατέκτησε το διαγωνισμό ImageNet με διαφορά, άνοιξε τον δρόμο για τη σημερινή κυριαρχία της βαθιάς μάθησης [7].

Σήμερα, η Τεχνητή Νοημοσύνη βρίσκεται παντού: από τους εικονικούς βοηθούς (Siri, Google Assistant), τα έξυπνα οικιακά συστήματα, τα αυτόνομα οχήματα και την ιατρική διάγνωση, μέχρι τους αλγορίθμους συστάσεων στο Netflix το YouTube και τις αναζητήσεις στο Google. Η πιο πρόσφατη εξέλιξη είναι τα Μεγάλα Γλωσσικά Μοντέλα (LLMs) όπως το ChatGPT, το οποίο εκπαιδεύτηκε σε δισεκατομμύρια λέξεις και μπορεί να παράγει ανθρώπινο κείμενο με υψηλή συνοχή [6].

Το μέλλον της TN προδιαγράφεται εξίσου συναρπαστικό και σύνθετο. Η πρόοδος των υπολογιστικών συστημάτων, η ενοποίηση TN με το διαδίκτυο των πραγμάτων (IoT), την υπολογιστική όραση, τη ρομποτική και τη βιοπληροφορική, καθιστούν την Τεχνητή Νοημοσύνη ένα από τα σημαντικότερα επιστημονικά και κοινωνικά πεδία του 21ου αιώνα [14].

1.2 Η συμβολή της Τεχνητής Νοημοσύνης στον Ιατρικό χώρο

«Η Τεχνητή Νοημοσύνη (TN) έχει εδραιωθεί ως μία από τις πιο σημαντικές και ραγδαία αναπτυσσόμενες τεχνολογίες στον τομέα της ιατρικής, επιφέροντας θεμελιώδεις αλλαγές στη διάγνωση, τη θεραπεία, την έρευνα και τη διαχείριση ασθενών. Χρησιμοποιώντας μεθόδους μηχανικής μάθησης (ML), που κυμαίνονται από απλούστερους αλγορίθμους, όπως η λογιστική παλινδρόμηση, τα δέντρα αποφάσεων και οι υποστηρικτικοί διανυσματικοί μηχανισμοί (SVMs), έως πιο σύνθετες προσεγγίσεις όπως η επεξεργασία φυσικής γλώσσας (NLP) και η βαθιά μάθηση (DL), η TN προσφέρει τη δυνατότητα ενίσχυσης της κλινικής λήψης αποφάσεων, μείωσης των σφαλμάτων και εξατομίκευσης της φροντίδας. Από την απεικονιστική ιατρική και τη γονιδιωματική, μέχρι την ανάπτυξη φαρμάκων και τη ρομποτική χειρουργική, η TN μετασχηματίζει με γρήγορο ρυθμό τη σύγχρονη ιατρική πρακτική [8].»

1.2.1 Ιατρική Απεικόνιση και Αναγνώριση Προτύπων

Μία από τις πιο σημαντικές συνεισφορές της TN είναι στον τομέα της Ιατρικής Απεικόνισης. Τα μοντέλα TN—και ειδικά τα συνελικτικά νευρωνικά δίκτυα (CNNs)—έχουν αποδείξει εξαιρετικές δυνατότητες στην ανίχνευση ανωμαλιών σε ακτινογραφίες, αξονικές τομογραφίες (CT), μαγνητικές τομογραφίες (MRI) και υπερήχους. Σε μια σημαντική μελέτη, ερευνητές του Stanford ανέπτυξαν ένα μοντέλο TN ικανό να διαγνώσει πνευμονία από ακτινογραφίες θώρακος με απόδοση που ξεπερνά εκείνη των έμπειρων ακτινολόγων [8]. Αντίστοιχα, η DeepMind της Google επέδειξε τη δυναμική της TN ανιχνεύοντας πάνω από 50 οφθαλμολογικές παθήσεις με υψηλή ακρίβεια μέσω απεικόνισης συνεκτικής οπτικής τομογραφίας [9].

Στη δερματολογία, ερευνητές από το Stanford[10], με επικεφαλής τον Andre Esteva εκπαίδευσαν έναν αλγόριθμο βαθιάς μάθησης σε πάνω από 129.000 κλινικές εικόνες για την ταξινόμηση του καρκίνου του δέρματος, με απόδοση συγκρίσιμη με αυτή των δερματολόγων. Αυτές οι εξελίξεις δείχνουν ότι η TN μπορεί να λειτουργήσει όχι μόνο ως εργαλείο υποστήριξης αλλά και ως μηχανισμός πρώιμου ελέγχου, ιδιαίτερα σε περιοχές με περιορισμένους ανθρωπίνους και οικονομικούς πόρους.

1.2.2 Προγνωστική Αναλυτική, Παρακολούθηση και Εξατομικευμένη Ιατρική

Η τεχνητή νοημοσύνη μεταμορφώνει επίσης την προγνωστική αναλυτική εντοπίζοντας ασθενείς υψηλού κινδύνου πριν εμφανιστούν κλινικά συμπτώματα. Για παράδειγμα, μοντέλα βασισμένα σε ηλεκτρονικούς φακέλους υγείας (EHR) μπορούν να προβλέψουν την πιθανότητα επανεισαγωγής στο νοσοκομείο, εμφάνισης σηψαμίας ή επιδείνωσης σε ΜΕΘ. Η AI ομάδα της Google ανέπτυξε ένα μοντέλο που προέβλεπε τη θνησιμότητα ασθενών 24–48 ώρες νωρίτερα με ακρίβεια άνω του 90% αναλύοντας EHR δεδομένα [11].

Φορητές συσκευές που λειτουργούν με τεχνητή νοημοσύνη (όπως wearables) συμβάλλουν στην παρακολούθηση ασθενών σε πραγματικό χρόνο και ενισχύουν την προληπτική ιατρική μέσω συνεχόμενα δεδομένα (όπως καρδιακό ρυθμό, επίπεδα οξυγόνου στο αίμα, και πρότυπα κίνησης) εντοπίζοντας εγκαίρως μεταβολές σε βιοσημεία, ειδοποιώντας γιατρούς/παρόχους υγείας, βελτιώνοντας έτσι σημαντικά τη διαχείριση χρόνιων παθήσεων.

Μία ακόμη σημαντική συμβολή της τεχνητής νοημοσύνης βρίσκεται στον τομέα της εξατομικευμένης ιατρικής. Τα συστήματα τεχνητής νοημοσύνης μπορούν να αναλύσουν γονιδιακά προφίλ, περιβαλλοντικούς παράγοντες και ιατρικά ιστορικά ώστε να προσαρμόσουν τα θεραπευτικά πλάνα. Το IBM Watson for Genomics, για παράδειγμα, ερμηνεύει γονιδιωματικά δεδομένα και προτείνει στοχευμένες θεραπείες για τον καρκίνο βάσει γνωστών μεταλλάξεων [12].

Τα μοντέλα μηχανικής μάθησης χρησιμοποιούνται επίσης στη φαρμακογονιδιωματική, προβλέποντας πώς θα ανταποκριθεί ένας ασθενής σε συγκεκριμένα φάρμακα. Με αυτόν τον τρόπο, η τεχνητή νοημοσύνης μειώνει τον κίνδυνο παρενεργειών και ενισχύει την αποτελεσματικότητα της θεραπείας. Αυτές οι δυνατότητες ανοίγουν τον δρόμο για πιο εξατομικευμένες και ακριβείς θεραπείες σε πολλές ιατρικές ειδικότητες.

1.2.3 Τεχνητή Νοημοσύνη στην Ανακάλυψη Φαρμάκων και στη Ρομποτική Χειρουργική

Η τεχνητή νοημοσύνη επιταχύνει σημαντικά τη συνήθως αργή και δαπανηρή διαδικασία ανακάλυψης φαρμάκων. Μέσω βαθιάς μάθησης και γενετικών μοντέλων, μπορεί να εντοπίσει υποσχόμενες μοριακές δομές, να προβλέψει τη βιολογική τους δραστηριότητα και να τις βελτιστοποιήσει για ασφάλεια και αποτελεσματικότητα.

Ένα αξιοσημείωτο παράδειγμα είναι το μόριο για τη θεραπεία της ίνωσης που σχεδιάστηκε από την Insilico Medicine με τεχνητή νοημοσύνη και ανακαλύφθηκε σε μόλις 46 ημέρες—μια διαδικασία που συνήθως διαρκεί χρόνια[13]. Επιπλέον, η τεχνητή νοημοσύνη έπαιξε κρίσιμο ρόλο στην επιτάχυνση της έρευνας για τα εμβόλια και τις θεραπείες κατά της COVID-19, αναλύοντας υικά πρωτεΐνικά μόρια, αλληλεπιδράσεις με τον ξενιστή και υποψήφιες ενώσεις.

Στη ρομποτικά υποβοηθούμενη χειρουργική, η τεχνητή νοημοσύνη βελτιώνει την ακρίβεια, μειώνει τα ποσοστά επιπλοκών και επιτρέπει ελάχιστα επεμβατικές τεχνικές. Συστήματα όπως το Da Vinci Surgical System ενσωματώνουν αλγόριθμους μηχανικής μάθησης για να ενισχύσουν την επιδεξιότητα του χειρουργού και να προσφέρουν ανάδραση απτικής

αίσθησης (haptic feedback). Έρευνες έχουν δείξει ότι οι λαπαροσκοπικές επεμβάσεις με υποστήριξη τεχνητής νοημοσύνης έχουν μικρότερους χρόνους ανάρρωσης και λιγότερες μετεγχειρητικές επιπλοκές[14].

Επιπλέον, αναπτύσσονται αυτόνομα ρομποτικά συστήματα για την εκτέλεση χειρουργείων „ρουτίνας“. Για παράδειγμα, το Smart Tissue Autonomous Robot (STAR) εκτέλεσε επιτυχώς επέμβαση σε μαλακό ιστό σε γουρούνι, παρουσιάζοντας μεγαλύτερη συνέπεια ακόμα και από ανθρώπινους χειρουργούς[15].

1.2.4 Ηθικά Ζητήματα και Προκλήσεις

Η ενσωμάτωση της Τεχνητής Νοημοσύνης (TN) στην υγειονομική περίθαλψη, παρά τα σημαντικά οφέλη που υπόσχεται, συνοδεύεται από ένα πλήθος ηθικών και πρακτικών προκλήσεων που δεν μπορούν να αγνοηθούν. Ένα από τα κρισιμότερα ζητήματα είναι η μεροληψία στα δεδομένα εκπαίδευσης. Όταν τα μοντέλα TN εκπαιδεύονται σε μη αντιροσωπευτικά σύνολα δεδομένων, οι προβλέψεις τους τείνουν να είναι αναξιόπιστες ή λιγότερο ακριβείς για συγκεκριμένες υποομάδες ασθενών. Για παράδειγμα, ένα σύστημα που έχει εκπαιδευτεί κυρίως σε δεδομένα από Καυκάσιους ασθενείς μπορεί να παρουσιάζει μειωμένη ακρίβεια σε άτομα διαφορετικής εθνοτικής καταγωγής, ενισχύοντας ανισότητες στην περίθαλψη [17].

Επιπλέον, τίθεται το ζήτημα της διαφάνειας και της εξηγησιμότητας. Τα περισσότερα σύγχρονα μοντέλα, όπως τα συνελικτικά νευρωνικά δίκτυα (CNNs), λειτουργούν ως «μαύρα κουτιά», με αποτέλεσμα οι γιατροί να μην έχουν πάντα σαφή εικόνα για τον τρόπο λήψης μιας απόφασης. Αυτό γεννά ανησυχίες σχετικά με την αξιοπιστία και την αποδοχή τέτοιων συστημάτων στην κλινική πράξη, όπου η τεκμηρίωση και η υπευθυνότητα είναι κρίσιμες.

Η προστασία προσωπικών δεδομένων είναι επίσης πρωταρχικής σημασίας, καθώς η συλλογή και ανάλυση μεγάλου όγκου ιατρικών εικόνων ενέχει κινδύνους παραβίασης απορρήτου. Η συμμόρφωση με κανονισμούς όπως ο GDPR στην Ευρώπη αποτελεί βασική προϋπόθεση για τη νόμιμη και δεοντολογική χρήση της TN σε ιατρικά περιβάλλοντα.

Ένα ακόμη κρίσιμο θέμα είναι η ευθύνη: σε περίπτωση λανθασμένης διάγνωσης από ένα σύστημα τεχνητής νοημοσύνης, ποιος φέρει την τελική ευθύνη; Ο κατασκευαστής του αλγορίθμου, ο οργανισμός που το υιοθέτησε ή ο ιατρός που το χρησιμοποίησε; Η έλλειψη σαφών πλαισίων λογοδοσίας δημιουργεί αβεβαιότητα και ενισχύει τον σκεπτικισμό γύρω από την κλινική εφαρμογή αυτών των συστημάτων.

Τέλος, σημαντικό είναι και το ζήτημα της ισότιμης πρόσβασης. Η υιοθέτηση προηγμένων εργαλείων TN ενδέχεται να είναι πιο εύκολη σε μεγάλους οργανισμούς με επαρκείς πόρους, ενώ μικρότερα ή υποχρηματοδοτούμενα νοσοκομεία μπορεί να μείνουν πίσω, διευρύνοντας περαιτέρω τις ανισότητες στο σύστημα υγείας.

Συνοψίζοντας, η ενσωμάτωση της TN στην ιατρική διάγνωση δεν αποτελεί μόνο τεχνική πρόκληση αλλά και ηθικό στοίχημα. Η αντιμετώπιση θεμάτων όπως η μεροληψία, η διαφάνεια, το απόρρητο, η λογοδοσία και η ισότιμη πρόσβαση είναι απαραίτητη προϋπόθεση ώστε η TN να εξελιχθεί σε ένα εργαλείο που ενδυναμώνει πραγματικά την κλινική πρακτική και υπηρετεί το κοινό καλό. Οι ρυθμιστικοί φορείς, όπως ο FDA στις ΗΠΑ και οι αντίστοιχοι ευρωπαϊκοί οργανισμοί, εργάζονται ήδη προς αυτήν την κατεύθυνση,

αναπτύσσοντας κατευθυντήριες γραμμές και πλαίσια αξιολόγησης για την ασφαλή και υπεύθυνη χρήση της TN στην υγεία.

1.3 Βασικά Πλεονεκτήματα των Τεχνητών Νευρωνικών Δικτύων

Τα Τεχνητά Νευρωνικά Δίκτυα έχουν φέρει επανάσταση στον χώρο της τεχνητής νοημοσύνης χάρη στην ικανότητά τους να μοντελοποιούν πολύπλοκα μοτίβα, να μαθαίνουν από δεδομένα και να προσαρμόζονται σε νέα πληροφορία. Σε αντίθεση με τις παραδοσιακές αλγορίθμικές προσεγγίσεις που απαιτούν ρητά καθορισμένους κανόνες, τα TNΔ εξάγουν συμπεράσματα ανακαλύπτοντας μη φανερές δομές στα δεδομένα μέσω εκπαίδευσης. Η αυξανόμενη σημασία τους σε τομείς όπως η ιατρική και τα χρηματοοικονομικά οφείλεται σε μια σειρά πλεονεκτημάτων που τα καθιστούν ιδιαίτερα κατάλληλα για προβλήματα υψηλής διάστασης και μη γραμμικότητας.

1.3.1 Μάθηση, Γενίκευση και Μη Γραμμικότητα

Ένα από τα σημαντικότερα πλεονεκτήματα των TNΔ είναι η ικανότητά τους να μαθαίνουν από δεδομένα. Στο πλαίσιο της εποπτευόμενης μάθησης όπως θα δούμε και αργότερα (βλέπε 1.7.1), τους παρέχονται ζεύγη εισόδου-εξόδου, δηλαδή παραδείγματα δεδομένων με τις αντίστοιχες ετικέτες κατηγορίας. Μέσα από αυτή τη διαδικασία το δίκτυο αναγνωρίζει πολύπλοκα μοτίβα και αποκτά τη δυνατότητα να προβλέπει σωστά νέα, άγνωστα δεδομένα — μια ιδιότητα γνωστή ως γενίκευση. Αυτό τα καθιστά κατάλληλα για τομείς όπου η χειροκίνητη καταγραφή κανόνων είναι αδύνατη ή ανέφικτη.

Για παράδειγμα, στην αναγνώριση εικόνας, οι παραδοσιακοί αλγόριθμοι δυσκολεύονται με αλλαγές στον φωτισμό, στον προσανατολισμό ή στο φόντο. Τα νευρωνικά δίκτυα και πιο συγκεκριμένα τα συνελικτικά νευρωνικά δίκτυα (CNNs) μπορούν να μάθουν να ανιχνεύουν χαρακτηριστικά όπως άκρα, σχήματα και υφές απευθείας από τα δεδομένα ενώ και άλλες αρχιτεκτονικές, όπως τα επαναλαμβανόμενα δίκτυα ή τα Transformer-based μοντέλα, διαθέτουν αντίστοιχες δυνατότητες σε διαφορετικού τύπου εφαρμογές [23].

Τα TNΔ εισάγουν μη γραμμικότητα μέσω συναρτήσεων ενεργοποίησης όπως η ReLU, η sigmoid και η tanh. Αυτό τους επιτρέπει να μοντελοποιούν εξαιρετικά πολύπλοκες σχέσεις εισόδου-εξόδου, κάτι που δεν μπορούν να επιτύχουν τα γραμμικά μοντέλα.

Η ευελιξία αυτή είναι ζωτικής σημασίας σε εφαρμογές όπως η πρόβλεψη χρηματοοικονομικών τάσεων, όπου οι παράγοντες που επηρεάζουν την αγορά αλληλεπιδρούν με μη προφανείς τρόπους. Τα νευρωνικά δίκτυα μπορούν να συλλάβουν αυτές τις αλληλεπιδράσεις και να παρέχουν ακριβείς προβλέψεις, ακόμη και όταν οι σχέσεις δεν είναι άμεσα εμφανείς [24].

1.3.2 Ανθεκτικότητα, Προσαρμοστικότητα και Μάθηση από Άκρο σε Άκρο

Ένα ακόμη ισχυρό πλεονέκτημα των TNΔ είναι η ανθεκτικότητά τους στον θόρυβο. Κατά τη διάρκεια της εκπαίδευσης, μπορούν να μάθουν να αναγνωρίζουν τα σημαντικά μοτίβα ακόμα και παρουσία άσχετων ή ελλιπών πληροφοριών. Στην ιατρική διάγνωση, για παράδειγμα, τα δεδομένα των ασθενών μπορεί να περιέχουν ελλείψεις ή ανακρίβειες. Τα

ΤΝΔ μπορούν να εντοπίσουν τα πιο σημαντικά χαρακτηριστικά και να αγνοήσουν τον θόρυβο, ιδίως όταν εφαρμόζονται τεχνικές όπως το dropout ή η κανονικοποίηση για την αποφυγή υπερεκπαίδευσης [25].

Τα ΤΝΔ έχουν την ικανότητα να βελτιώνονται με τον χρόνο. Μέσω της διαδικασίας του backpropagation και αλγορίθμων βελτιστοποίησης όπως το gradient descent, προσαρμόζουν επαναληπτικά τις εσωτερικές παραμέτρους τους για να μειώσουν τα σφάλματα πρόβλεψης. Επιπλέον, η τεχνική της μεταφοράς μάθησης (transfer learning) επιτρέπει σε ένα προεκπαιδευμένο δίκτυο να προσαρμοστεί σε ένα νέο αλλά συγγενές πρόβλημα με περιορισμένα δεδομένα. Αυτό μειώνει σημαντικά το υπολογιστικό κόστος και τον χρόνο εκπαίδευσης, όπως έχει φανεί σε εφαρμογές όπως η ιατρική απεικόνιση και η επεξεργασία φυσικής γλώσσας [26].

Σε αντίθεση με παραδοσιακά συστήματα που απαιτούν ξεχωριστά στάδια για εξαγωγή χαρακτηριστικών, επεξεργασία και ταξινόμηση, τα ΤΝΔ υποστηρίζουν μάθηση από άκρο σε άκρο. Αυτό σημαίνει ότι, μετά τα βασικά στάδια προεπεξεργασίας (π.χ. κανονικοποίηση ή αλλαγή μεγέθους εικόνων), τα δεδομένα μπορούν να εισαχθούν απευθείας στο μοντέλο και το δίκτυο να μάθει όλες τις απαραίτητες μετατροπές εσωτερικά. Η προσέγγιση αυτή απλοποιεί την αρχιτεκτονική των συστημάτων, αυξάνει την ακρίβεια και μειώνει την ανάγκη για χειροκίνητη μηχανική χαρακτηριστικών — κάτι ιδιαίτερα χρήσιμο σε περίπλοκες εργασίες όπως η μηχανική μετάφραση ή ο αυτόματος υποτιτλισμός βίντεο [30].

1.3.3 Υπολογιστική Αποδοτικότητα και Διαχείριση Δεδομένων

Τα ΤΝΔ είναι κατάλληλα για παράλληλο υπολογισμό, επιτρέποντας την αξιοποίηση σύγχρονου υλικού όπως GPUs και TPUs για ταχύτερη εκπαίδευση και πρόβλεψη. Αυτό τα καθιστά κατάλληλα για πολύ μεγάλα σύνολα δεδομένων και εφαρμογές σε πραγματικό χρόνο όπως η επεξεργασία βίντεο ή η αυτόνομη οδήγηση. Για παράδειγμα, η αρχιτεκτονική TPU v4 της Google υποστηρίζει μεγάλα μοντέλα βαθιάς μάθησης μέσω μαζικών παραλληλισμένων πράξεων, μειώνοντας δραστικά τον χρόνο εκπαίδευσης για μοντέλα αιχμής όπως τα BERT και GPT [27].

Τα ΤΝΔ αποδίδουν εξαιρετικά σε προβλήματα με εισόδους υψηλής διάστασης, όπως γονιδιωματικά δεδομένα, εικονοστοιχεία ή ηχητικά σήματα. Σε αντίθεση με τους παραδοσιακούς αλγόριθμους που υποφέρουν από τα προβλήματα που συνοδεύουν τα δεδομένα με πολλές διαστάσεις. Τα ΤΝΔ μπορούν να ανακαλύψουν λανθάνουσες αναπαραστάσεις που μειώνουν την πολυπλοκότητα διατηρώντας την ουσία της πληροφορίας.

Autoencoders και τεχνικές μείωσης διαστάσεων στη βαθιά μάθηση προσφέρουν συμπίεση χαρακτηριστικών, ιδιαίτερα χρήσιμη σε τομείς όπως η κυβερνοασφάλεια και η βιοπληροφορική, όπου τα δεδομένα είναι άφθονα αλλά οι ετικέτες περιορισμένες [29]. Στα Τεχνητά Νευρωνικά Δίκτυα, ο όρος *ετικέτα* (label) αναφέρεται στην πληροφορία-στόχο που αντιστοιχεί σε κάθε δείγμα εκπαίδευσης· για παράδειγμα, στην ταξινόμηση ιατρικών εικόνων της παρούσας εργασίας η ετικέτα είναι η διάγνωση ("Υγιής", "Βακτηριακή Πνευμονία", "Ιογενής Πνευμονία"). Οι ετικέτες χρησιμοποιούνται για να καθοδηγήσουν το μοντέλο κατά τη διάρκεια της εκπαίδευσης, επιτρέποντάς του να μάθει τη συσχέτιση μεταξύ εισόδων και εξόδων.»

1.3.4 Ευελιξία σε Πολλούς Τομείς

Ένα από τα πιο αξιοσημείωτα χαρακτηριστικά των ΤΝΔ είναι η ευελιξία τους. Μια βασική αρχιτεκτονική, με ελάχιστες προσαρμογές, μπορεί να εφαρμοστεί σε πληθώρα εφαρμογών:

- Αναγνώριση εικόνων (CNNs)
- Αναγνώριση ομιλίας (RNNs, Transformers)
- Μετάφραση γλώσσας (μοντέλα Seq2Seq)
- Πρόβλεψη χρονοσειρών (LSTM, GRU)
- Παραγωγικά μοντέλα (GANs, VAEs)

Αυτή η προσαρμοστικότητα μειώνει την ανάγκη ανάπτυξης εξειδικευμένων αλγορίθμων για κάθε πεδίο και επιτρέπει τη χρήση ενός κοινού εργαλείου σε πολλές προκλήσεις [28].

1.4 Τα Νευρωνικά Δίκτυα Γύρω Μας

Τα Νευρωνικά Δίκτυα, που άλλοτε περιορίζονταν σε ακαδημαϊκή έρευνα, έχουν πλέον καταστεί πανταχού παρόντα στην κοινωνία. Από τα smartphones και τα μέσα κοινωνικής δικτύωσης μέχρι τα αυτόνομα οχήματα και τα χρηματοοικονομικά συστήματα, τα Νευρωνικά Δίκτυα αποτελούν πλέον τον αόρατο κινητήριο μοχλό πολλών τεχνολογιών που χρησιμοποιούμε καθημερινά. Η ικανότητά τους να μαθαίνουν μοτίβα, να προσαρμόζονται με τον χρόνο και να επεξεργάζονται τεράστιες ποσότητες μη δομημένων δεδομένων έχει καταστήσει δυνατή τη χρήση τους σε ένα ευρύ φάσμα εφαρμογών [23]. Αυτό το εδάφιο εξερευνά πώς τα Νευρωνικά Δίκτυα ενσωματώνονται στην καθημερινότητά μας, διαμορφώνοντας τον τρόπο που αλληλεπιδρούμε, επικοινωνούμε και λαμβάνουμε αποφάσεις — συχνά χωρίς καν να το αντιλαμβανόμαστε.

1.4.1 Μέσα Κοινωνικής Δικτύωσης και Χρηματοοικονομικά

Τα Νευρωνικά Δίκτυα αποτελούν τη ραχοκοκαλιά των συστημάτων προτάσεων περιεχομένου σε πλατφόρμες όπως το YouTube, το Netflix, το TikTok, το Instagram και το Spotify. Τα συστήματα αυτά αναλύουν τη συμπεριφορά του χρήστη (κλικ, προβολές, likes, κοινοποιήσεις) μέσω μοντέλων βαθιάς μάθησης για να προσωποποιήσουν τη ροή περιεχομένου. Για παράδειγμα, ο αλγόριθμος του TikTok χρησιμοποιεί μοντελοποίηση ακολουθιών και ενσωματώσεις χρήστη-αντικειμένου για να προβλέψει ποια βίντεο θα αρέσουν στον χρήστη [20].

Επιπλέον, πλατφόρμες όπως το Facebook και το Twitter χρησιμοποιούν Νευρωνικά Δίκτυα για προσθήκη ετικετών σε εικόνες, μετάφραση γλώσσας και φιλτράρισμα περιεχομένου. Οι εφαρμογές αυτές συνδυάζουν CNNs (για εικόνες) και μοντέλα επεξεργασίας φυσικής γλώσσας (για κείμενο) για να βελτιώσουν την προσβασιμότητα και την ασφάλεια [30].

Στον χρηματοοικονομικό τομέα, τα Νευρωνικά Δίκτυα είναι κρίσιμα για την ανίχνευση απάτης, αξιοπιστία ως προς την αποπληρωμή δανείων και το αλγορίθμικό εμπόριο. Τράπεζες και fintech εταιρείες χρησιμοποιούν αναδρομικά νευρωνικά δίκτυα και autoencoders για να εντοπίζουν ανωμαλίες στα δεδομένα συναλλαγών. Τα μοντέλα αυτά μπορούν να ανιχνεύσουν ύποπτες δραστηριότητες, όπως κλοπή ταυτότητας ή ξέπλυμα

χρήματος, σε πραγματικό χρόνο μέσω ανάλυσης μοτίβων δαπανών, τοποθεσιών και συμπεριφορικών ενδείξεων [18]. Ταυτόχρονα, chatbots με τεχνητή νοημοσύνη βελτιώνουν την εξυπηρέτηση πελατών αναλαμβάνοντας την επίλυση ερωτημάτων μέσω live chat.

1.4.2 Μεταφορές και Συσκευές IoT

Ίσως η πιο εμφανής εφαρμογή των νευρωνικών δικτύων είναι στα αυτόνομα οχήματα. Εταιρείες όπως η Tesla, η Waymo και η NVIDIA χρησιμοποιούν βαθιά συνελικτικά νευρωνικά δίκτυα και ενισχυτική μάθηση (reinforcement learning) για να επιτρέψουν στα οχήματα να αναγνωρίζουν αντικείμενα, να παίρνουν αποφάσεις και να πλοηγούνται σε περιβάλλοντα.

Για παράδειγμα, το σύστημα Autopilot της Tesla επεξεργάζεται δεδομένα από κάμερες, ραντάρ και αισθητήρες υπερήχων μέσω ενός εκτενούς νευρωνικού δικτύου εκπαίδευμένου σε εκατομμύρια χιλιόμετρα οδήγησης [19]. Τα μοντέλα αυτά επιτρέπουν την ανίχνευση λωρίδων, πεζών και εμποδίων, καθώς και τον σχεδιασμό βέλτιστων διαδρομών. Επιπλέον, εφαρμογές ride-sharing όπως η Uber και η Lyft χρησιμοποιούν νευρωνικά δίκτυα για τη βελτιστοποίηση διαδρομών, την προσαρμογή τιμών και την πρόβλεψη χρόνου άφιξης [23].

Συστήματα ασφαλείας με κάμερες που ενσωματώνουν τεχνητή νοημοσύνη μπορούν να αναγνωρίσουν μέλη της οικογένειας, επισκέπτες ή αγνώστους μέσω αναγνώρισης προσώπου και να ειδοποιούν σε περίπτωση ύποπτης κίνησης [21].

1.5 Προκλήσεις των Τεχνητών Νευρωνικών Δικτύων

Αν και τα Τεχνητά Νευρωνικά Δίκτυα έχουν επιδείξει εντυπωσιακές δυνατότητες σε πολλούς τομείς, η ευρεία υιοθέτησή τους δεν είναι απαλλαγμένη από σημαντικούς περιορισμούς. Η κατανόηση αυτών των προκλήσεων είναι κρίσιμη για ερευνητές και επαγγελματίες που επιθυμούν να αξιοποιήσουν τα TNΔ με υπεύθυνο τρόπο σε πραγματικές εφαρμογές [31].

1.5.1 Υψηλό Υπολογιστικό Κόστος και Απαιτήσεις Πόρων

Η εκπαίδευση Deep learning νευρωνικών δικτύων απαιτεί αυξημένους υπολογιστικούς πόρους: ισχυρές GPUs ή TPUs, μεγάλο εύρος μνήμης και πολύωρη επεξεργασία. Μοντέλα όπως το GPT-4 ή το AlphaFold περιλαμβάνουν δισεκατομμύρια παραμέτρους, οδηγώντας σε υψηλή κατανάλωση ενέργειας και κόστος υποδομών. Αυτή η κατάσταση εγείρει ανησυχίες σχετικά με τη βιωσιμότητα και την προσβασιμότητα, καθώς μόνο οι οργανισμοί με σημαντικούς οικονομικούς και τεχνικούς πόρους μπορούν να αντέξουν οικονομικά την εκπαίδευση και τη συντήρηση τέτοιων μοντέλων[31].

Τα TNΔ απαιτούν τεράστιες ποσότητες δεδομένων για αποδοτική εκπαίδευση. Η συγκέντρωση επαρκών, ποιοτικών και κατηγοριοποιημένων δεδομένων είναι δαπανηρή, χρονοβόρα ή περιορισμένη από ζητήματα ιδιωτικότητας, ειδικά σε ευαίσθητους τομείς όπως η υγεία. Η απόδοση ενός μοντέλου συχνά εξαρτάται από τη διαφορετικότητα των δεδομένων. Ένα μοντέλο που εκπαιδεύεται σε προκατειλημμένα ή περιορισμένα δεδομένα ενδέχεται να μην γενικεύει καλά, με αρνητικές συνέπειες σε κρίσιμους τομείς όπως η ιατρική διάγνωση ή η απονομή δικαιοσύνης [32].

Η απόδοση των ΤΝΔ εξαρτάται από την ορθή ρύθμιση υπερπαραμέτρων, όπως ο ρυθμός μάθησης, ο αριθμός επιπέδων, το μέγεθος παρτίδας και οι συναρτήσεις ενεργοποίησης, η οποία πραγματοποιείται στο στάδιο της εκπαίδευσης. Η εύρεση της κατάλληλης διαμόρφωσης απαιτεί εκτεταμένο πειραματισμό και εμπειρία. Αυτόματες τεχνικές, όπως grid search, random search και Bayesian optimization, μπορούν να χρησιμοποιηθούν για τον σκοπό αυτό, αλλά απαιτούν σημαντικούς υπολογιστικούς πόρους [36].

1.5.2 Έλλειψη Ερμηνευσιμότητας και Διαφάνειας

Τα ΤΝΔ χαρακτηρίζονται συχνά ως "μαύρα κουτιά" λόγω της πολυπλοκότητάς τους. Είναι δύσκολο να εντοπιστεί πώς ακριβώς το μοντέλο καταλήγει σε μια απόφαση — κάτι που αποτελεί εμπόδιο σε τομείς που απαιτούν λογιδοσία, όπως η ιατρική, η χρηματοοικονομική και η νομική. Παρότι μέθοδοι όπως το LIME [33] και το SHAP [34] έχουν προσφέρει εργαλεία για την εξήγηση της συμπεριφοράς των μοντέλων, απέχουν από τη πλήρη διαφάνεια. Η εμπιστοσύνη στις αποφάσεις των ΤΝΔ είναι συνεπώς περιορισμένη [33].

Αν και τα ΤΝΔ εντοπίζουν πρότυπα, δεν κατανοούν αιτιακές σχέσεις. Δηλαδή, μπορούν να συσχετίσουν δεδομένα αλλά όχι να τα ερμηνεύσουν. Αυτό περιορίζει τη χρησιμότητά τους σε συγκεκριμένους τομείς. Γίνονται προσπάθειες για ενσωμάτωση causal inference frameworks με τα βαθιά μάθηση, αλλά βρίσκονται ακόμη σε αρχικά στάδια [37].

1.5.3 Αξιοπιστία και Γενίκευση

Η υπερεκπαίδευση εμφανίζεται όταν το δίκτυο αποδίδει καλά στα δεδομένα εκπαίδευσης αλλά αποτυγχάνει σε νέα παραδείγματα. Αυτό συμβαίνει συχνά όταν η πολυπλοκότητα του μοντέλου είναι δυσανάλογη προς τον όγκο δεδομένων. Αν και τεχνικές κατά την εκπαίδευση όπως το dropout, η πρόωρη διακοπή και η ενίσχυση δεδομένων βοηθούν, δεν προσφέρουν πανάκεια. Η επίτευξη ισορροπίας μεταξύ πολυπλοκότητας και γενίκευσης παραμένει ένα ανοιχτό ζήτημα [28].

Η δημιουργία ενός ΤΝΔ δεν αποτελεί μία μεμονωμένη ενέργεια. Καθώς τα δεδομένα αλλάζουν, τα μοντέλα καθίστανται ξεπερασμένα — φαινόμενο γνωστό ως "model drift". Η αξιόπιστη χρήση απαιτεί συνεχή παρακολούθηση, επανακατάρτιση και επικύρωση. Περιλαμβάνει προκλήσεις όπως έκδοση δεδομένων, CI/CD για ML και διακυβέρνηση μοντέλων — τομείς που εξελίσσονται ακόμα στη βιομηχανία [27].

1.5.4 Ευπάθεια σε Εχθρικές Επιθέσεις

Ειδικά στα συστήματα όρασης, τα ΤΝΔ είναι ευάλωτα σε εχθρικές επιθέσεις — εσκεμμένες παραποιήσεις εισόδων που οδηγούν σε λανθασμένες προβλέψεις. Ένα τροποποιημένο σήμα stop, για παράδειγμα, μπορεί να αναγνωριστεί ως σήμα ταχύτητας από ένα αυτόνομο όχημα [35]. Αυτή η ευπάθεια ενέχει κινδύνους ασφαλείας για κρίσιμες εφαρμογές.

Τα ΤΝΔ μπορούν να κληρονομήσουν μεροληψίες από τα δεδομένα εκπαίδευσης ή τη σχεδίαση του συστήματος. Αυτό μπορεί να οδηγήσει σε άδικη μεταχείριση ομάδων και να ενισχύσει κοινωνικές ανισότητες. Για παράδειγμα, συστήματα αναγνώρισης προσώπου έχουν παρουσιάσει υψηλότερα ποσοστά σφάλματος σε άτομα με πιο σκούρο δέρμα [32].

1.6 Η Λειτουργία ενός Νευρωνικού Δικτύου

Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) είναι υπολογιστικά μοντέλα εμπνευσμένα από τη δομή και τη λειτουργία των βιολογικών νευρωνικών δικτύων. Λειτουργούν επεξεργαζόμενα εισόδους μέσω διασυνδεδεμένων επιπέδων απλών μονάδων που ονομάζονται νευρώνες, χρησιμοποιώντας μαθηματικές συναρτήσεις για να μετασχηματίσουν και να μεταδώσουν πληροφορίες. Η κατανόηση του τρόπου λειτουργίας ενός νευρωνικού δικτύου είναι ουσιώδης για την κατανόηση των δυνατοτήτων και των περιορισμών του.

1.6.1 Συναρτήσεις

Στον πυρήνα της λειτουργίας ενός νευρωνικού δικτύου βρίσκονται οι μαθηματικές συναρτήσεις που διέπουν τον μετασχηματισμό των δεδομένων μέσω του δικτύου. Οι συναρτήσεις αυτές μπορούν να κατηγοριοποιηθούν σε τρεις κύριες κατηγορίες [28]:

α) Συνάρτηση Εισόδου Η συνάρτηση εισόδου υπολογίζει το σταθμισμένο άθροισμα των σημάτων από το προηγούμενο επίπεδο ή από τα αρχικά δεδομένα. Για έναν νευρώνα j η είσοδος υπολογίζεται ως:

$$z_j = \sum_i w_{ij}x_i + b_j \quad (01)$$

όπου:

- x_i : είσοδοι
- w_{ij} : βάρος σύνδεσης από την είσοδο i στον νευρώνα j
- b_j : όρος μετατόπισης (bias).

Αυτός ο γραμμικός συνδυασμός προετοιμάζει το σήμα για μεταγενέστερο μετασχηματισμό.

β) Συνάρτηση Ενεργοποίησης Η συνάρτηση ενεργοποίησης εισάγει μη γραμμικότητα στο δίκτυο, καθιστώντας δυνατή την εκμάθηση πολύπλοκων σχέσεων μεταξύ εισόδων και εξόδων. Εφαρμόζεται στην είσοδο Z_j και παράγει την έξοδο a_j :

$$a_j = \phi(z_j) \quad (02)$$

Συχνές συναρτήσεις ενεργοποίησης:

- **Sigmoid:** $\phi(z) = \frac{1}{1+e^{-z}}$ (03) Χρησιμοποιείται κυρίως για δυαδική ταξινόμηση.
- **Tanh:** $(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ (04) Πλεονεκτεί της sigmoid λόγω μηδενικού κέντρου.

- **ReLU**: $\phi(z) = \max(0, z)$ (05) Εισάγει αραιότητα και περιορίζει προβλήματα με μηδενικές παραγώγους [38].
- **Softmax**: Χρησιμοποιείται στην έξοδο για πολυταξική ταξινόμηση:

$$\phi(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (06)$$

γ) Συνάρτηση Απώλειας Η συνάρτηση απώλειας μετρά τη διαφορά μεταξύ προβλεπόμενης και πραγματικής εξόδου και καθοδηγεί τη μάθηση του δικτύου μέσω βελτιστοποίησης.

Συνήθεις συναρτήσεις:

- Μέσο Τετραγωνικό Σφάλμα (MSE): για παλινδρόμηση.
- Διασταυρούμενη Εντροπία (Cross-Entropy): για ταξινόμηση.

Η ελαχιστοποίηση της απώλειας επιτυγχάνεται μέσω αλγορίθμων όπως η καθοδική κλίση (gradient descent) [28]

1.6.2 Νευρώνες

Ο νευρώνας είναι η βασική υπολογιστική μονάδα ενός τεχνητού νευρωνικού δικτύου, εμπνευσμένη από τον βιολογικό νευρώνα. Κάθε νευρώνας λαμβάνει πολλαπλές εισόδους, τις επεξεργάζεται και παράγει μία έξοδο.

- **Δομή Νευρώνα:**

- **Είσοδοι (X_i):** Σήματα που λαμβάνονται είτε από τα αρχικά δεδομένα εισόδου είτε από άλλους νευρώνες.
- **Βάρη (W_i):** Κάθε είσοδος πολλαπλασιάζεται με ένα βάρος, που αντιπροσωπεύει τη σημασία της.
- **Μεροληψία (b):** Σταθερός όρος που μετατοπίζει τη συνάρτηση ενεργοποίησης, προσφέροντας μεγαλύτερη ευελιξία.
- **Άθροιση:** Τα σταθμισμένα σήματα εισόδου και η μεροληψία αθροίζονται, σχηματίζοντας την είσοδο του νευρώνα Z .
- **Ενεργοποίηση:** Το άθροισμα Z διέρχεται από μια συνάρτηση ενεργοποίησης ϕ , η οποία καθορίζει το σήμα εξόδου.

Μαθηματικά, η έξοδος ενός νευρώνα a δίνεται από:

$$a = \phi(\sum_i w_i x_i + b) \quad (07)$$

- **Τύποι Νευρώνων:**

- **Εισόδου:** δέχονται δεδομένα.
- **Κρυφών Επίπεδων:** πραγματοποιούν ενδιάμεση επεξεργασία.
- **Εξόδου:** παράγουν την τελική πρόβλεψη ή ταξινόμηση.

Κάθε νευρώνας συνεισφέρει στην τελική προσέγγιση της συνάρτησης που επιδιώκει να μάθει το μοντέλο.

1.7 Εκπαίδευση Νευρωνικών Δικτύων

Η εκπαίδευση είναι η διαδικασία μέσω της οποίας ένα τεχνητό νευρωνικό δίκτυο μαθαίνει να εκτελεί ένα καθορισμένο έργο, προσαρμόζοντας τις εσωτερικές του παραμέτρους — κυρίως τα βάρη και τις μετατοπίσεις (biases) — με βάση τα δεδομένα εισόδου και την ανατροφοδότηση (feedback). Η επαναληπτική αυτή βελτιστοποίηση επιτρέπει στο μοντέλο να γενικεύει από τα δεδομένα εκπαίδευσης και να παράγει ακριβείς προβλέψεις σε νέα, άγνωστα δείγματα.

Οι μέθοδοι εκπαίδευσης διακρίνονται κυρίως σε εποπτευόμενη (supervised), μη εποπτευόμενη (unsupervised) και ενισχυτική μάθηση (reinforcement learning), ανάλογα με τη φύση του προβλήματος και τη διαθεσιμότητα δεδομένων [23].

1.7.1 Εποπτευόμενη Μάθηση (Supervised Learning)

Η εποπτευόμενη μάθηση αποτελεί την πιο διαδεδομένη και ώριμη προσέγγιση στην εκπαίδευση νευρωνικών δικτύων. Στην περίπτωση αυτή, το μοντέλο εκπαιδεύεται σε σύνολα δεδομένων που περιέχουν ετικέτες (labels), δηλαδή κάθε δείγμα εισόδου συνοδεύεται από την αντίστοιχη επιθυμητή έξοδο [28]. Η διαδικασία λειτουργεί ως εξής: το δίκτυο λαμβάνει μια είσοδο και παράγει μια πρόβλεψη \hat{y} . Η πρόβλεψη αυτή συγκρίνεται με την πραγματική τιμή y μέσω μιας συνάρτησης απώλειας, όπως για παράδειγμα το Mean Squared Error (MSE) ή η Cross-Entropy. Το σφάλμα που προκύπτει μεταδίδεται προς τα πίσω μέσω του αλγορίθμου της οπισθοδιάδοσης (backpropagation), με τα βάρη του δικτύου να προσαρμόζονται χρησιμοποιώντας μεθόδους βελτιστοποίησης όπως η καθοδική κλίση (gradient descent).

Η εποπτευόμενη μάθηση βρίσκει εφαρμογή σε μια πληθώρα πεδίων. Ενδεικτικά παραδείγματα περιλαμβάνουν την ταξινόμηση εικόνας, όπως η κατηγοριοποίηση φωτογραφιών σε γάτες ή σκύλους, την ιατρική διάγνωση, όπου τα νευρωνικά δίκτυα μπορούν να κατατάξουν όγκους σε καλοήθεις ή κακοήθεις [10], καθώς και την αναγνώριση φωνής, δηλαδή τη μετατροπή της φωνής σε κείμενο.

Μεταξύ των βασικών πλεονεκτημάτων της εποπτευόμενης μάθησης συγκαταλέγονται η υψηλή ακρίβεια που μπορεί να επιτευχθεί όταν υπάρχουν επαρκή και ποιοτικά δεδομένα, καθώς και η ύπαρξη ώριμων εργαλείων και αλγορίθμων που διευκολύνουν την ανάπτυξη και εφαρμογή της. Ωστόσο, η μέθοδος παρουσιάζει και μειονεκτήματα. Πρωταρχική πρόκληση αποτελεί η απαίτηση για μεγάλα και σωστά ετικετοποιημένα σύνολα δεδομένων, ενώ ένα ακόμη σημαντικό ζήτημα είναι ο κίνδυνος υπερπροσαρμογής (overfitting), όταν το μοντέλο αποδίδει εξαιρετικά καλά στα δεδομένα εκπαίδευσης αλλά αποτυγχάνει σε νέα παραδείγματα.

1.7.2 Μη Εποπτευόμενη Μάθηση (Unsupervised Learning)

Η μη εποπτευόμενη μάθηση αφορά στην εκπαίδευση μοντέλων χωρίς τη χρήση ετικετών, με βασικό στόχο την ανακάλυψη κρυφών μοτίβων ή δομών στα δεδομένα. Σε αυτήν την περίπτωση, το δίκτυο δέχεται μόνο εισόδους χωρίς γνωστές εξόδους και προσπαθεί να

ομαδοποιήσει ή να αναπαραστήσει τα δεδομένα με τρόπο που να αποκαλύπτει τις υποκείμενες σχέσεις και δομές.

Κύριοι στόχοι της μεθόδου αυτής είναι η ελαχιστοποίηση του σφάλματος ανακατασκευής, όπως συμβαίνει στους αυτόματους κωδικοποιητές (autoencoders), καθώς και η μεγιστοποίηση της διαχωριστικότητας μεταξύ ομάδων, όπως παρατηρείται στα μοντέλα clustering. Οι εφαρμογές της μη εποπτευόμενης μάθησης είναι ποικίλες, με χαρακτηριστικά παραδείγματα την τμηματοποίηση πελατών με βάση την καταναλωτική συμπεριφορά, την ανίχνευση ανωμαλιών, όπως ύποπτες συναλλαγές, αλλά και τη μείωση των διαστάσεων μέσω τεχνικών όπως η PCA και η t-SNE, που διευκολύνουν την οπτικοποίηση των δεδομένων.

Όσον αφορά τα μοντέλα που χρησιμοποιούνται, ιδιαίτερα σημαντική κατηγορία αποτελούν οι αυτόματοι κωδικοποιητές, οι οποίοι χρησιμοποιούνται για συμπίεση και ανακατασκευή δεδομένων, καθώς και οι χάρτες αυτό-οργάνωσης (Self-Organizing Maps – SOMs), οι οποίοι επιτρέπουν την προβολή δεδομένων υψηλών διαστάσεων σε δισδιάστατα πλέγματα.

Η μέθοδος παρουσιάζει σημαντικά πλεονεκτήματα, καθώς είναι ιδανική σε περιπτώσεις όπου δεν υπάρχουν διαθέσιμες ετικέτες και συμβάλλει στην ανάδειξη της δομής των δεδομένων. Ωστόσο, εμφανίζει και μειονεκτήματα, με κυριότερα την περιορισμένη ερμηνευσιμότητα των αποτελεσμάτων, καθώς και τη δυσκολία αξιολόγησης της απόδοσης των μοντέλων.

1.7.3 Ενισχυτική Μάθηση (Reinforcement Learning)

Η ενισχυτική μάθηση αποτελεί μια προσέγγιση κατά την οποία ένας πράκτορας (agent) μαθαίνει μέσω της συνεχούς αλληλεπίδρασής του με το περιβάλλον, λαμβάνοντας ανταμοιβές ή ποινές ανάλογα με τις ενέργειες που εκτελεί. Ο πράκτορας παρατηρεί την κατάσταση του περιβάλλοντος, επιλέγει μια ενέργεια βάσει μιας πολιτικής (policy) και στη συνέχεια το περιβάλλον ανταποκρίνεται παρέχοντας νέα κατάσταση και σήμα ανταμοιβής. Μέσα από αυτήν τη διαδικασία, η πολιτική του πράκτορα ενημερώνεται διαρκώς, με σκοπό τη μεγιστοποίηση της μελλοντικής συνολικής ανταμοιβής.

Βασικές έννοιες που χαρακτηρίζουν την ενισχυτική μάθηση είναι ο πράκτορας, δηλαδή ο φορέας λήψης αποφάσεων, το περιβάλλον μέσα στο οποίο δρα, η πολιτική που καθορίζει τη στρατηγική επιλογής ενεργειών, καθώς και το σήμα ανταμοιβής, το οποίο λειτουργεί ως ανατροφοδότηση για την ενίσχυση της επιθυμητής συμπεριφοράς.

Η μεθοδολογία αυτή έχει βρει εφαρμογές σε πολλούς τομείς, όπως στα παιχνίδια με χαρακτηριστικό παράδειγμα το AlphaGo [22], στη ρομποτική για εκμάθηση κίνησης και πλοήγησης, αλλά και στη διαχείριση πόρων, όπως η βελτιστοποίηση του χρονοπρογραμματισμού σε data centers. Ιδιαίτερη σημασία έχει η βαθιά ενισχυτική μάθηση (Deep Reinforcement Learning – Deep RL), η οποία συνδυάζει τα νευρωνικά δίκτυα με την ενισχυτική μάθηση, επιτρέποντας την αποτελεσματική διαχείριση σύνθετων και δυναμικών περιβαλλόντων.

Ανάμεσα στα πλεονεκτήματα της ενισχυτικής μάθησης περιλαμβάνονται η ικανότητά της να λαμβάνει αποφάσεις σε δυναμικά περιβάλλοντα και το γεγονός ότι δεν απαιτεί ετικετοποιημένα δεδομένα, αλλά μόνο το σήμα ανταμοιβής. Ωστόσο, η μέθοδος συνοδεύεται και από σημαντικά μειονεκτήματα, όπως το υψηλό υπολογιστικό κόστος και η

ευαισθησία στις παραμέτρους, γεγονός που καθιστά δύσκολη τη σταθερότητα της εκπαίδευσης.

Συγκριτικός Πίνακας Παραδειγμάτων Μάθησης

Χαρακτηριστικό	Εποπτευόμενη	Μη Εποπτευόμενη	Ενισχυτική
Ετικέτες Δεδομένων	Απαιτούνται	Όχι	Όχι
Στόχος Μάθησης	Αντιστοίχιση εισόδων-εξόδων	Ανακάλυψη δομών	Μέγιστη ανταμοιβή
Τυπικές Εφαρμογές	Ταξινόμηση, Παλινδρόμηση	Clustering, Ανίχνευση ανωμαλιών	Παιχνίδια, Ρομποτική
Αποδοτικότητα Δεδομένων	Υψηλή (με ετικέτες)	Εξαρτάται από τη δομή	Χαμηλή
Κοινές Μέθοδοι	Backpropagation	Autoencoders, SOMs	Q-learning, Policy Gradients

Πίνακας 1: Συνοπτική παρουσίαση των κύριων διάφορων μεταξύ εποπτευόμενης, μη εποπτευόμενης και ενισχυτικής μάθησης.

1.8 Αλγόριθμοι Εκπαίδευσης και Βελτιστοποίηση

Η αποτελεσματική εκπαίδευση τεχνητών νευρωνικών δικτύων εξαρτάται από ισχυρούς αλγορίθμους για την ελαχιστοποίηση της συνάρτησης απώλειας. Αυτή η διαδικασία βασίζεται κυρίως στον υπολογισμό των βαθμίδων (gradients) και στην κατάλληλη προσαρμογή των βαρών. Οι τρεις πιο θεμελιώδεις αλγόριθμοι που χρησιμοποιούνται είναι η Οπισθοδιάδοση Σφάλματος (Backpropagation), η Καθοδική Κλίση (Gradient Descent) και ο Adam — καθένας με ξεχωριστό ρόλο στη διαδικασία μάθησης [41].

1.8.1 Οπισθοδιάδοση Σφάλματος (Backpropagation)

Η οπισθοδιάδοση είναι ο αλγόριθμος που υπολογίζει το βαθμωτό της συνάρτησης απώλειας ως προς κάθε παράμετρο (βάρος) του δικτύου. Αποτελεί τη ραχοκοκαλιά του υπολογισμού βαθμίδων στα πολυεπίπεδα νευρωνικά δίκτυα.

Βήματα:

- Μπροστινή Διέλευση (Forward Pass):** Υπολογισμός της εξόδου \hat{y} για δεδομένη είσοδο X
- Υπολογισμός Απώλειας:** Χρήση συνάρτησης απώλειας $L(y, \hat{y})$.

- **Οπισθοδιάδοση (Backward Pass):** Υπολογισμός βαθμίδων με τον κανόνα της αλυσίδας.
- **Ενημέρωση Παραμέτρων:** Χρήση των βαθμίδων σε αλγόριθμο όπως η καθοδική κλίση για την ενημέρωση των βαρών.

Μαθηματική Περιγραφή:

$$z^l = w^l a^{l-1} + b^l \quad (08)$$

$$a^l = f(z^l) \quad (09)$$

$$\delta^l = \left(\frac{\partial L}{\partial z^l} \right) \quad (10)$$

$$w^l := w^l - \eta \cdot \left(\frac{\partial L}{\partial w^l} \right) \quad (11)$$

Όπου η είναι ο ρυθμός μάθησης (learning rate).

[41]

1.8.2 Καθοδική Κλίση (Gradient Descent)

Η καθοδική κλίση είναι ένας αλγόριθμος πρώτης τάξης που χρησιμοποιείται για την ελαχιστοποίηση της συνάρτησης απώλειας κινούμενος προς την αντίθετη κατεύθυνση της βαθμίδας.

Κανόνας Ενημέρωσης:

$$\theta := \theta - \eta \cdot \nabla_{\theta} L(\theta) \quad (12)$$

Όπου:

- θ οι παράμετροι του μοντέλου
- η ο ρυθμός μάθησης
- $\nabla_{\theta} L(\theta)$ η βαθμίδα της απώλειας ως προς τις παραμέτρους.

Τύποι Καθοδικής Κλίσης:

- **Κανονική (Batch):** Ολόκληρο το σύνολο δεδομένων.
- **Στοχαστική (SGD):** Ένα μόνο δείγμα.
- **Μικρής Παρτίδας (Mini-Batch):** Υποσύνολο δείγματος.

Πλεονεκτήματα:

- Απλή υλοποίηση.
- Σταθερή θεωρητική βάση.

Μειονεκτήματα:

- Ευαισθησία στον ρυθμό μάθησης.
- Αργή σύγκλιση ή παγίδευση σε τοπικά ελάχιστα.

1.8.3 Adam (Adaptive Moment Estimation)

Ο Adam είναι ένας εξελιγμένος αλγόριθμος βελτιστοποίησης που συνδυάζει πλεονεκτήματα των AdaGrad και RMSProp. Έχει εδραιωθεί ως προεπιλογή για την εκπαίδευση βαθιών νευρωνικών δικτύων.

Κύρια Ιδέα:

Διατηρεί δύο εκθετικά κινητούς μέσους όρους:

- m_t : Πρώτη ροπή (μέση τιμή των βαθμίδων).
- v_t : Δεύτερη ροπή (μέση τιμή τετραγώνων βαθμίδων).

$$m_t = \beta_1 \cdot m_{\{t-1\}} + (1 - \beta_1) \cdot g_t \quad (13)$$

$$v_t = \beta_2 \cdot v_{\{t-1\}} + (1 - \beta_2) \cdot g_t^2 \quad (14)$$

Με διόρθωση μεροληψίας(Bias Correction):

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \hat{v}_t = \frac{v_t}{1-\beta_2^t} \quad (15)$$

Τελική ενημέρωση:

$$\theta_t := \theta_{t-1} - n \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (16)$$

Όπου ϵ είναι μικρή σταθερά για αποφυγή διαίρεσης με το μηδέν.

Πλεονεκτήματα:

- Λειτουργεί καλά με αραιές βαθμίδες.
- Ταχύτερη σύγκλιση.
- Ελάχιστη απαίτηση ρυθμίσεων υπερπαραμέτρων.

Μειονεκτήματα:

- Μπορεί να γενικεύσει χειρότερα από SGD.
- Πιθανή ταλάντωση χωρίς αποδυνάμωση του ρυθμού μάθησης.

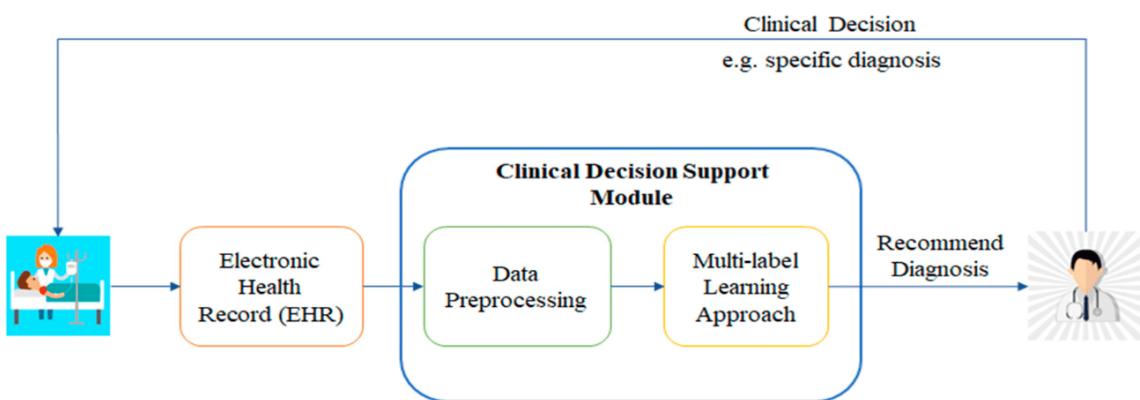
[39][40]

2 Συστήματα Υποστήριξης αποφάσεων

2.1 Εισαγωγή

Τα πληροφοριακά συστήματα αποτελούν σήμερα αναπόσπαστο τμήμα της διοικητικής και λειτουργικής υποδομής των οργανισμών, συμπεριλαμβανομένων των νοσοκομείων και των παρόχων υγειονομικής περίθαλψης. Ένα πληροφοριακό σύστημα μπορεί να ορισθεί ως ένα σύνολο αλληλοσυνδεόμενων στοιχείων τα οποία συλλέγουν, αποθηκεύουν, επεξεργάζονται και διανέμουν πληροφορία με σκοπό την υποστήριξη των επιχειρησιακών διεργασιών και της λήψης αποφάσεων. Στην πιο απλή μορφή του, το σύστημα δέχεται δεδομένα ως είσοδο, εφαρμόζει κάποιες λειτουργίες επεξεργασίας και παράγει πληροφορίες ως έξοδο — πληροφορίες που στη συνέχεια χρησιμοποιούνται από τους υπευθύνους για τη λήψη αποφάσεων. Η χρησιμότητα ενός πληροφοριακού συστήματος μετριέται από την ικανότητά του να μετατρέψει ωμά δεδομένα σε γνώση και πρακτικά αποτελέσματα για τον οργανισμό [42].

Παρακάτω παρουσιάζεται σαν παράδειγμα μια σχηματική απεικόνιση που δείχνει τη ροή πληροφορίας σε ένα σύστημα υποστήριξης κλινικής απόφασης, ξεκινώντας από τα ηλεκτρονικά αρχεία υγείας, περνώντας από στάδια προεπεξεργασίας και multi-label (πολύ-επίπεδης) ταξινόμησης, και καταλήγοντας σε προτεινόμενη διάγνωση για τον ιατρό.



Εικόνα 2: Σχηματική απεικόνιση της διαδικασίας υποστήριξης κλινικής απόφασης. Τα ηλεκτρονικά αρχεία υγείας (EHR) συλλέγονται και περνούν από στάδιο προεπεξεργασίας δεδομένων, ενώ στη συνέχεια εφαρμόζεται προσέγγιση multi-label ταξινόμησης για την πρόταση διάγνωσης στον κλινικό ιατρό, ο οποίος λαμβάνει την τελική απόφαση. (<https://botpenquin.com/glossary/decision-support-system>)

Μέσα στο ευρύ φάσμα των πληροφοριακών συστημάτων υπάρχουν ορισμένες κατηγορίες που αξίζει να διακριθούν: τα Συστήματα Επεξεργασίας Συναλλαγών (Transaction Processing Systems — TPS), τα Συστήματα Πληροφόρησης Διοίκησης (Management Information Systems — MIS) και, το αντικείμενο του κεφαλαίου αυτού, τα Συστήματα Υποστήριξης Αποφάσεων (Decision Support Systems — DSS). Τα TPS είναι επιφορτισμένα με την επεξεργασία καθημερινών συναλλαγών (π.χ. παραγγελίες, παραλαβές, πωλήσεις), ενώ τα MIS παράγουν συγκεντρωμένες αναφορές που στηρίζουν τις διοικητικές λειτουργίες [43], [44]. Τα DSS, αντίθετα, σχεδιάζονται ειδικά για τη στήριξη πολύπλοκων και μητυποποιημένων αποφάσεων, παρέχοντας ανάλυση, μοντελοποίηση και οπτικοποίηση δεδομένων που βοηθούν στην επιλογή μεταξύ εναλλακτικών λύσεων.

Υπάρχουν ποικίλες κατηγορίες πληροφοριακών συστημάτων. Παρακάτω παρατίθενται μερικές από τις σημαντικότερες κατηγορίες που μια επιχείρηση, μπορεί να αξιοποιήσει στην καθημερινή της λειτουργία. Μία από τις πλέον διαδεδομένες κατηγορίες είναι τα Συστήματα Επεξεργασίας Συναλλαγών (Transaction Processing Systems — TPS), τα οποία αναλαμβάνουν την αυτοματοποιημένη διαχείριση των καθημερινών συναλλαγών, όπως την καταχώρηση παραγγελιών, την επεξεργασία παραλαβών και τη διεκπεραίωση πωλήσεων. [43] Παράλληλα, τα Συστήματα Πληροφόρησης Διοίκησης (Management Information Systems — MIS) παρέχουν συνοπτικές αναφορές, πίνακες και γραφήματα που διευκολύνουν τις διοικητικές διαδικασίες και τη λήψη αποφάσεων σε επίπεδο οργάνωσης. [44] Το θέμα που θα απασχολήσει το παρόν κεφάλαιο, δεδομένης της ευρείας χρήσης τους στην ιατρική, είναι τα Συστήματα Υποστήριξης Αποφάσεων (Decision Support Systems — DSS), τα οποία υποβοηθούν τη λήψη σύνθετων και μη-τυποποιημένων αποφάσεων μέσω της ανάλυσης δεδομένων και της παροχής εργαλείων για την αξιολόγηση και σύγκριση εναλλακτικών λύσεων. [45]

2.2 Ιστορική Εξέλιξη των Συστημάτων Υποστήριξης Αποφάσεων

Οι πρώτες θεωρητικές προσεγγίσεις που έθεσαν τις βάσεις για τα Συστήματα Υποστήριξης Αποφάσεων (Decision Support Systems — DSS) εμφανίστηκαν στις δεκαετίες του 1950 και 1960, όταν τα Ινστιτούτα Τεχνολογίας Carnegie και Μασαχουσέτης εισήγαγαν την ιδέα της ανάπτυξης υπολογιστικών εργαλείων για τη λήψη αποφάσεων [46]. Στη δεκαετία του 1970, τα DSS άρχισαν να απασχολούν έντονα την επιχειρηματική βιβλιογραφία και περιοδικά, τα οποία δημοσίευαν άρθρα σχετικά με τα συστήματα διοίκησης αποφάσεων και τα εργαλεία στρατηγικού σχεδιασμού [47]. Ένα χαρακτηριστικό παράδειγμα αποτελεί η μελέτη των Sprague και Watson (1976), όπου παρουσιάστηκαν εφαρμογές DSS στον τραπεζικό τομέα [48].

Το 1971 σηματοδοτήθηκε από την κυκλοφορία του έργου του Michael S. Scott Morton με τίτλο *Management Decision Systems: Computer-Based Support for Decision Making*, το οποίο θεωρείται από τα πρώτα βιβλία που ανέλυσαν διεξοδικά την έννοια της υποστήριξης αποφάσεων μέσω ηλεκτρονικών υπολογιστών. Λίγα χρόνια αργότερα, οι Capon και Hulbert, μέσα από το άρθρο τους (1975), περιέγραψαν την εφαρμογή της μεθόδου Decision System Analysis (DSA) σε βασικές αποφάσεις μάρκετινγκ, όπως η τιμολόγηση, οι προβλέψεις, η διαφήμιση και η ανάπτυξη νέων προϊόντων [47]. Τα συμπεράσματά τους ανέδειξαν ασυνέπειες στη διαδικασία λήψης αποφάσεων και παρείχαν ουσιαστική κατανόηση των προβλημάτων που αντιμετώπιζαν οι επιχειρήσεις στο συγκεκριμένο πεδίο.

Σταθμός για την περαιτέρω ανάπτυξη αποτέλεσε το βιβλίο των Ralph Sprague και Eric Carlson, *Building Effective Decision Support Systems* (1982), το οποίο περιέγραφε τις αρχές σχεδίασης και υλοποίησης αποδοτικών DSS. Παρά τις αρχικές υπεραισιόδοξες προβλέψεις, τα DSS εκείνης της περιόδου περιορίζονταν σημαντικά από τις τεχνολογικές δυνατότητες της εποχής. Με την πάροδο των χρόνων, αυτοί οι περιορισμοί ξεπεράστηκαν, ιδίως μετά την εισαγωγή τεχνολογιών εξόρυξης δεδομένων (data mining) και αποθηκών δεδομένων (data warehousing) στις αρχές της δεκαετίας του 1990, γεγονός που προσέφερε νέα δυναμική στην ανάλυση δεδομένων και στη λήψη αποφάσεων.

Σήμερα, τα Συστήματα Υποστήριξης Αποφάσεων εξελίσσονται διαρκώς χάρη στη συνδυαστική χρήση τεχνητής νοημοσύνης, μηχανικής μάθησης και προηγμένων αναλυτικών

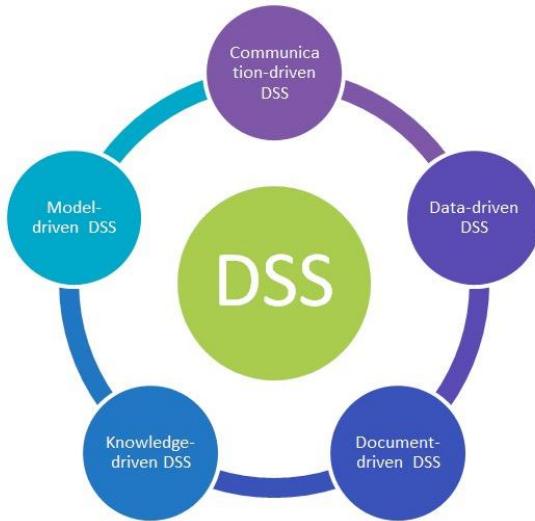
μεθόδων. Η συμβολή τους δεν περιορίζεται πλέον σε υποστηρικτικό ρόλο αλλά έχει καταστεί καθοριστική για κρίσιμους τομείς, όπως η υγειονομική περίθαλψη, η οικονομική ανάλυση και οι στρατηγικές επιχειρησιακές λειτουργίες, καθιστώντας τα DSS αναπόσπαστο κομμάτι της σύγχρονης διοίκησης και λήψης αποφάσεων.

2.3 Κατηγορίες Συστημάτων Υποστήριξης Αποφάσεων

Υπάρχουν διαφορετικές προσεγγίσεις για την ταξινόμηση των Συστημάτων Υποστήριξης Αποφάσεων (DSS), ωστόσο μία από τις πιο διαδεδομένες προσεγγίσεις διακρίνει πέντε κύριες κατηγορίες, καθεμία από τις οποίες εξυπηρετεί διαφορετικούς σκοπούς και βασίζεται σε διαφορετικές τεχνολογικές αρχές [45].

- **Model-Driven DSS** (Συστήματα βασισμένα σε μοντέλα): Αυτά τα συστήματα επικεντρώνονται στη χρήση μαθηματικών, στατιστικών και βελτιστοποιητικών μοντέλων για την ανάλυση προβλημάτων και την αξιολόγηση εναλλακτικών λύσεων. Μέσω της προσομοίωσης και της μοντελοποίησης, επιτρέπουν στον χρήστη να προβλέψει αποτελέσματα, να εξετάσει υποθετικά σενάρια και να λάβει αποφάσεις βάσει ποσοτικών αναλύσεων. Εφαρμόζονται συχνά σε τομείς όπως η διαχείριση εφοδιαστικής αλυσίδας και η επιχειρησιακή στρατηγική.
- **Data-Driven DSS** (Συστήματα βασισμένα σε δεδομένα): Η συγκεκριμένη κατηγορία δίνει έμφαση στην πρόσβαση και ανάλυση μεγάλων όγκων δεδομένων, τα οποία μπορεί να προέρχονται από εσωτερικές βάσεις δεδομένων, αποθήκες δεδομένων (data warehouses) ή εξωτερικές πηγές. Χρησιμοποιεί τεχνικές όπως OLAP και data mining για την εξαγωγή προτύπων και την υποστήριξη αποφάσεων βάσει πραγματικών στοιχείων. Είναι ιδιαίτερα χρήσιμα σε περιβάλλοντα που απαιτούν ανάλυση ιστορικών δεδομένων και τρέχουσες μετρήσεις.
- **Communication-Driven DSS** (Συστήματα βασισμένα στην επικοινωνία): Αυτή η κατηγορία εστιάζει στην ενίσχυση της συνεργασίας μεταξύ ομάδων και στη διευκόλυνση συλλογικής λήψης αποφάσεων. Ενσωματώνει τεχνολογίες επικοινωνίας, όπως συστήματα τηλεδιασκέψεων, εργαλεία ανταλλαγής αρχείων και πλατφόρμες συνεργασίας, ώστε να διαμοιράζεται αποτελεσματικά η πληροφορία και να λαμβάνονται κοινές αποφάσεις σε πραγματικό χρόνο.
- **Document-Driven DSS** (Συστήματα βασισμένα σε έγγραφα): Τα συστήματα αυτά επικεντρώνονται στην αναζήτηση και διαχείριση μεγάλου όγκου εγγράφων και μη δομημένων πληροφοριών. Χρησιμοποιούν τεχνικές ανάκτησης πληροφορίας και ανάλυσης κειμένου για να υποστηρίξουν τη λήψη αποφάσεων σε περιβάλλοντα όπου η πληροφορία αποθηκεύεται κυρίως σε μορφή εγγράφων, αναφορών ή πολυμεσικού περιεχομένου.
- **Knowledge-Driven DSS** (Συστήματα βασισμένα στη γνώση): Αυτά τα DSS αξιοποιούν βάσεις γνώσης, συστήματα κανόνων και μηχανισμούς εξαγωγής συμπερασμάτων για την παροχή συστάσεων, διαγνώσεων ή εξειδικευμένων συμβουλών. Συνήθως περιλαμβάνουν μηχανές εξαγωγής συμπερασμάτων (inference engines) και είναι ιδιαίτερα διαδεδομένα σε τομείς που απαιτούν υψηλό βαθμό εξειδίκευσης, όπως η ιατρική διάγνωση και οι εφαρμογές μηχανικής.

Αξίζει να σημειωθεί ότι στην πράξη πολλά σύγχρονα DSS υιοθετούν υβριδικές προσεγγίσεις, συνδυάζοντας χαρακτηριστικά από διαφορετικές κατηγορίες, προκειμένου να καλύψουν πολύπλοκες επιχειρησιακές ανάγκες.



Εικόνα 3: Γραφική απεικόνιση των κατηγοριών DSS που περιγράφονται παραπάνω.
(<https://mscisbolouereifidi.wordpress.com/tag/decision-support-system/>)

2.4 Εφαρμογές των Συστημάτων Υποστήριξης Αποφάσεων στην Ιατρική

Η χρήση Συστημάτων Υποστήριξης Αποφάσεων (DSS) στον χώρο της υγείας έχει γνωρίσει ραγδαία ανάπτυξη τις τελευταίες δεκαετίες, καθώς η πολυπλοκότητα των κλινικών δεδομένων και η ανάγκη για ακρίβεια στη λήψη αποφάσεων καθιστούν τα εργαλεία αυτά αναπόσπαστο τμήμα της ιατρικής πρακτικής. Στον κλινικό τομέα, τα DSS αξιοποιούνται για τη βελτίωση της ποιότητας της περιθαλψης, τη μείωση λαθών και την υποστήριξη των επαγγελματιών υγείας στη διαδικασία λήψης αποφάσεων που συχνά είναι περίπλοκες και απαιτούν την ανάλυση μεγάλου όγκου δεδομένων [48].

Τα ιατρικά DSS είναι σχεδιασμένα για να επεξεργάζονται δεδομένα που αφορούν δημογραφικά στοιχεία ασθενών, ιστορικό νόσων, αποτελέσματα εργαστηριακών εξετάσεων και απεικονιστικά ευρήματα. Μέσω της επεξεργασίας αυτών των πληροφοριών, μπορούν να προτείνουν πιθανές διαγνώσεις, να υποδείξουν κατάλληλες θεραπευτικές παρεμβάσεις ή να εκτιμήσουν τον κίνδυνο εμφάνισης συγκεκριμένων επιπλοκών. Σε περιβάλλοντα όπου η λήψη αποφάσεων πρέπει να γίνεται άμεσα και με υψηλό βαθμό ακρίβειας, τα DSS λειτουργούν ως υποστηρικτικά εργαλεία για τον ιατρό, χωρίς να αντικαθιστούν την τελική του κρίση.

Τα συστήματα αυτά ενσωματώνουν τεχνολογίες όπως βάσεις γνώσης, μηχανισμούς εξαγωγής συμπερασμάτων, καθώς και αλγορίθμους τεχνητής νοημοσύνης και μηχανικής μάθησης. Οι τεχνολογίες αυτές επιτρέπουν στα DSS να ενημερώνονται συνεχώς με νέα δεδομένα και να προσαρμόζουν τις προτάσεις τους ανάλογα με τα νεότερα επιστημονικά ευρήματα. Ενδεικτικές εφαρμογές περιλαμβάνουν:

- Υποστήριξη διαγνωστικών διαδικασιών μέσω συστημάτων που αξιολογούν συμπτώματα και κλινικές ενδείξεις.
- Εκτίμηση κινδύνου εμφάνισης ασθενειών με βάση προγνωστικά μοντέλα και ιστορικά δεδομένα.
- Προτάσεις για φαρμακευτικές αγωγές, λαμβάνοντας υπόψη πιθανές αλληλεπιδράσεις φαρμάκων και αλλεργίες.
- Διαχείριση θεραπευτικών πρωτοκόλλων και παρακολούθηση της πορείας του ασθενούς.

Η συμβολή των DSS στην ιατρική είναι καθοριστική, καθώς ενισχύουν την τεκμηριωμένη λήψη αποφάσεων και μειώνουν την πιθανότητα σφαλμάτων. Ωστόσο, απαιτείται σωστή ενσωμάτωση τους στις υγειονομικές δομές και εκπαίδευση του προσωπικού για την αποδοτική αξιοποίησή τους.

2.4.1 Δομή ενός Κλινικού Συστήματος Υποστήριξης Αποφάσεων

Ένα τυπικό Κλινικό Σύστημα Υποστήριξης Αποφάσεων (Clinical Decision Support System – CDSS) αποτελείται από μια σειρά υποσυστημάτων που συνεργάζονται με στόχο την υποβοήθηση του ιατρικού προσωπικού στη διαδικασία λήψης αποφάσεων [48]. Στον πυρήνα του συστήματος βρίσκεται η βάση γνώσης (Knowledge Base), η οποία περιλαμβάνει κλινικούς κανόνες, πρωτόκολλα θεραπείας, διαγνωστικά κριτήρια και βιβλιογραφικά δεδομένα. Η ποιότητα και η επικαιρότητα αυτών των πληροφοριών είναι καθοριστικής σημασίας για την ακρίβεια και αξιοποίηση των προτάσεων που παράγει το σύστημα.

Στη συνέχεια, ο μηχανισμός εξαγωγής συμπερασμάτων (Inference Engine) εφαρμόζει τους κανόνες της βάσης γνώσης στα δεδομένα που εισάγονται, με σκοπό την παραγωγή κατάλληλων συστάσεων ή πιθανών διαγνώσεων. Παράλληλα, ο μηχανισμός επεξήγησης (Explanation Module) ενισχύει τη διαφάνεια του συστήματος, παρέχοντας στον χρήστη επεξηγήσεις σχετικά με τον τρόπο με τον οποίο προέκυψε μια πρόταση ή διάγνωση, γεγονός που αυξάνει την εμπιστοσύνη στη χρήση του CDSS.

Τέλος, η διεπαφή χρήστη (User Interface) αποτελεί το μέσο επικοινωνίας του κλινικού προσωπικού με το σύστημα, διευκολύνοντας τόσο την εισαγωγή δεδομένων όσο και την κατανόηση των παρεχόμενων συστάσεων. Η αρμονική συνεργασία όλων αυτών των στοιχείων είναι εκείνη που καθιστά το CDSS αποτελεσματικό και αξιόπιστο εργαλείο για την καθημερινή κλινική πράξη.

2.4.2 Περιορισμοί και Προκλήσεις Εφαρμογής

Παρά τα πλεονεκτήματά τους, τα κλινικά DSS αντιμετωπίζουν σημαντικές προκλήσεις. Πρώτον, η φύση των ιατρικών δεδομένων – που συχνά είναι μη δομημένα, περιγραφικά και ανομοιογενή – δυσχεραίνει την τυποποιημένη αποθήκευση και επεξεργασία τους. Επιπλέον, η συνεχής αύξηση της πολυπλοκότητας της ιατρικής πληροφορίας καθιστά δύσκολη τη συντήρηση και την επικαιροποίηση της βάσης γνώσης.

Ένας άλλος σημαντικός παράγοντας είναι ο συντηρητισμός που εκδηλώνεται από μερίδα επαγγελματιών υγείας, οι οποίοι εκφράζουν επιφυλάξεις για την αξιοποίησία των

συστημάτων ή φοβούνται ότι η χρήση τους μπορεί να περιορίσει την κλινική τους κρίση. Είναι κρίσιμο να υπογραμμιστεί ότι τα DSS έχουν σχεδιαστεί για να υποστηρίζουν και όχι να υποκαθιστούν τον ρόλο του ιατρού.

Τέλος, η εφαρμογή εξελιγμένων DSS προϋποθέτει επαρκείς οικονομικούς και τεχνολογικούς πόρους, κάτι που μπορεί να αποτελέσει εμπόδιο για υγειονομικούς οργανισμούς με περιορισμένο προϋπολογισμό. Παρά τις δυσκολίες αυτές, η συνεχής πρόοδος της τεχνολογίας και η σωστή εκπαίδευση των χρηστών μπορούν να συμβάλουν στην αποτελεσματική ενσωμάτωση των συστημάτων αυτών στην κλινική πράξη [42].

3. Πνευμονία

3.1 Γενικές πληροφορίες

Η πνευμονία είναι φλεγμονώδης λοίμωξη του πνευμονικού παρεγχύματος που εντοπίζεται κυρίως στις κυψελίδες και το διάμεσο ιστό. Παθολογοανατομικά χαρακτηρίζεται από συλλογή φλεγμονώδους εξιδρώματος στις κυψελίδες (πύον/υγρό), πάχυνση του κυψελιδικού-τριχοειδικού φραγμού και διαταραχή του αερισμού/αιμάτωσης (V/Q mismatch), με αποτέλεσμα υποξαιμία και δύσπνοια. Η κλινική εικόνα και η βαρύτητα ποικίλλουν ανάλογα με το αίτιο, την ηλικία και τις συννοσηρότητες του ασθενούς. Προκαλείται συνήθως από λοιμώδεις παράγοντες όπως βακτήρια, ιοί, μύκητες ή παράσιτα. Αποτελεί μια από τις πιο συχνές αιτίες νοσηρότητας και θνησιμότητας παγκοσμίως, επηρεάζοντας άτομα όλων των ηλικιών, αλλά κυρίως παιδιά μικρής ηλικίας, ηλικιωμένους και άτομα με εξασθενημένο ανοσοποιητικό σύστημα. Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας (WHO), η πνευμονία συγκαταλέγεται στις κυριότερες αιτίες θανάτου στα παιδιά κάτω των 5 ετών, ενώ σε ενήλικες μπορεί να επιπλακεί με σοβαρές καταστάσεις όπως σηψαιμία ή οξεία αναπνευστική ανεπάρκεια. [49]

Παθογένεση – τι συμβαίνει στους πνεύμονες

Η παθογένεση της πνευμονίας ξεκινά με τον αποικισμό ή την είσοδο του παθογόνου μέσω εισπνοής σταγονιδίων, μικρο-εισρόφησης από τον στοματοφάρυγγα ή σπανιότερα μέσω αιματογενούς διασποράς. Όταν οι αμυντικοί μηχανισμοί υπερνικηθούν (όπως συμβαίνει με τη διαταραχή του βλεννοκροσωτού επιθηλίου, τη μειωμένη λειτουργία των κυψελιδικών μακροφάγων ή την εξασθένιση της ανοσίας) προκαλείται φλεγμονώδης απάντηση. Αυτή χαρακτηρίζεται από ενεργοποίηση κυτταροκινών, διαρροή τριχοειδών και συσσώρευση ουδετεροφίλων, οδηγώντας σε κυψελιδική συμπύκνωση (consolidation). Οι λειτουργικές συνέπειες περιλαμβάνουν μείωση της ενδοτικότητας, ελάττωση της επιφανειοδραστικής ουσίας, υποαερισμό και ανακατανομή της αιμάτωσης. Σε σοβαρές περιπτώσεις μπορεί να προκύψουν επιπλοκές όπως πλευριτική συλλογή, εμπύημα, σήψη ή σύνδρομο οξείας αναπνευστικής δυσχέρειας (ARDS).

Κύριοι τρόποι μετάδοσης

Η πνευμονία μεταδίδεται κυρίως μέσω σταγονιδίων και αερολυμάτων από συμπτωματικούς ή ασυμπτωματικούς φορείς. Άλλοι τρόποι περιλαμβάνουν την εισρόφηση, ιδιαίτερα σε ασθενείς με νευρολογικές διαταραχές, κατάχρηση αλκοόλ ή μετεγχειρητικές καταστάσεις, την άμεση επαφή και αυτοενοθαλμισμό από μολυσμένες επιφάνειες προς τον ρινοφαρυγγικό βλεννογόνο, καθώς και την αιματογενή διασπορά από απομακρυσμένες εστίες, αν και αυτό συμβαίνει σπανιότερα.

Ταξινόμηση

Η ταξινόμηση της πνευμονίας μπορεί να γίνει με βάση διάφορα κριτήρια. Κατά την προέλευση, διακρίνεται σε πνευμονία κοινότητας (CAP), νοσοκομειακή (HAP) και σχετιζόμενη με μηχανικό αερισμό (VAP). Με βάση την αιτιολογία μπορεί να είναι βακτηριακή, ιογενής, μυκητιασική ή παρασιτική (βλ. 3.2). Η κατανομή στις απεικονιστικές εξετάσεις διαφοροποιεί την πνευμονία σε λοβώδη, βρογχοπνευμονία ή διάμεση, ενώ σε σχέση με την κατάσταση του ξενιστή διακρίνεται σε ανοσοεπαρκή και ανοσοκατεσταλμένα άτομα. Τέλος, η βαρύτητα εκτιμάται κλινικά με δείκτες όπως το PSI και το CURB-65, τα οποία καθοδηγούν την απόφαση για νοσηλεία ή εισαγωγή σε ΜΕΘ.

Επιδημιολογική σημασία

Η πνευμονία κατέχει εξέχουσα θέση ως αιτία νοσηρότητας και θνησιμότητας παγκοσμίως. Ιδιαίτερα ευάλωτες ομάδες είναι τα παιδιά κάτω των πέντε ετών, οι ηλικιωμένοι άνω των 65 ετών και τα άτομα με πολλαπλές συννοσηρότητες ή ανοσοκαταστολή (π.χ. HIV, μεταμοσχευμένοι ή ασθενείς υπό χημειοθεραπεία).

Παράγοντες κινδύνου

Στους σημαντικότερους παράγοντες κινδύνου συγκαταλέγονται η ηλικία (νήπια, μικρά παιδιά και ηλικιωμένοι), οι χρόνιες παθήσεις όπως ΧΑΠ, άσθμα, καρδιακή ανεπάρκεια, σακχαρώδης διαβήτης, χρόνια νεφρική και ηπατική νόσος, καθώς και η ανοσοκαταστολή λόγω φαρμάκων, HIV ή κακοηθειών. Επιπλέον, το κάπνισμα, η κατανάλωση αλκοόλ, ο υποσιτισμός, η ατμοσφαιρική ρύπανση ή η καύση βιομάζας σε οικιακούς χώρους αποτελούν καθοριστικούς επιβαρυντικούς παράγοντες. Οι νευρολογικές διαταραχές, η αυξημένη πιθανότητα εισρόφησης, η νοσηλεία, η παρατεταμένη κατάκλιση, η χρήση αναπνευστήρα και μια πρόσφατη ιογενής λοίμωξη (π.χ. γρίπη ή RSV) που προδιαθέτει σε βακτηριακή επιμόλυνση αυξάνουν επίσης τον κίνδυνο εμφάνισης της νόσου.

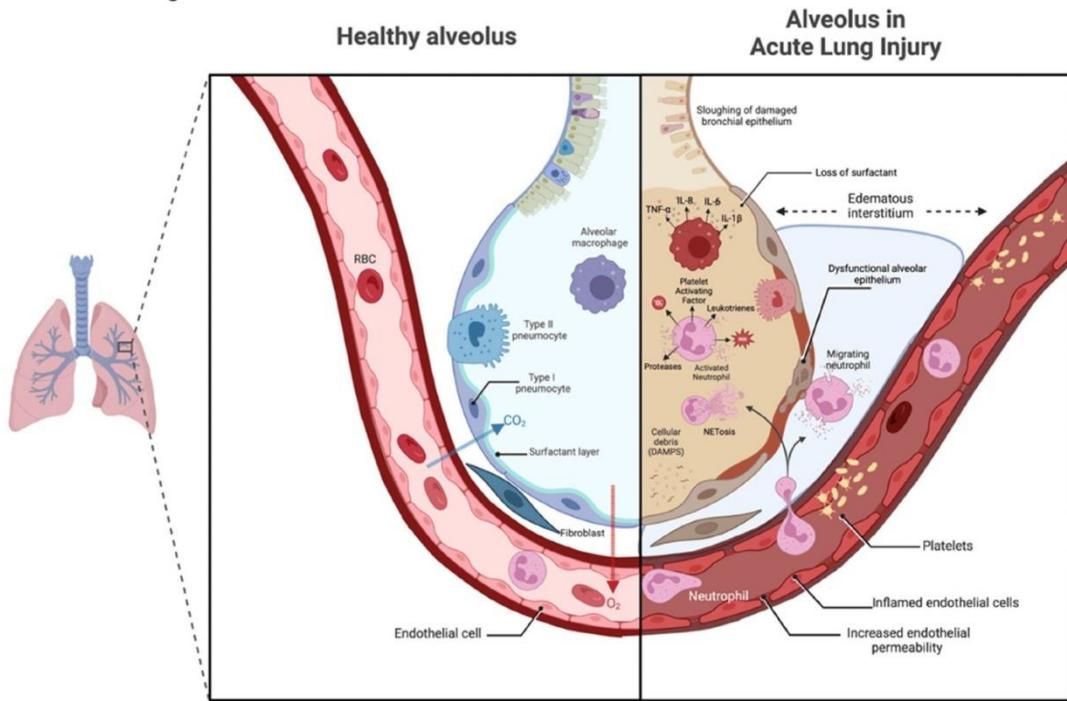
Κλινικό φάσμα

Η κλινική εικόνα της πνευμονίας ποικίλει από ήπιες μορφές έως βαριά σηψαμική νόσο με αναπνευστική ανεπάρκεια. Η «τυπική» εξωνοσοκομειακή πνευμονία (συχνότερα από πνευμονιόκοκκο) εκδηλώνεται με οξεία έναρξη, υψηλό πυρετό, παραγωγικό βήχα και πλευριτικό άλγος. Αντίθετα, η «άτυπη» μορφή, που οφείλεται σε ιούς ή μικροοργανισμούς όπως Mycoplasma, Chlamydophila και Legionella, χαρακτηρίζεται από ξηρό βήχα, μυαλγίες, κεφαλαλγία, ήπιο πυρετό και συχνά εξωπνευμονικές εκδηλώσεις.

Πρόληψη

Η πρόληψη της πνευμονίας βασίζεται σε συνδυασμό μέτρων. Ο εμβολιασμός, τόσο με τον αντιγρυπικό όσο και με τον αντιπνευμονιοκοκκικό εμβολιασμό, σύμφωνα με τις εθνικές οδηγίες, αποτελεί το σημαντικότερο προληπτικό μέτρο. Παράλληλα, η διακοπή του καπνίσματος, η μείωση της έκθεσης σε ρύπους, η αντιμετώπιση υποκείμενων νοσημάτων και η τήρηση μέτρων υγιεινής, όπως η συστηματική υγιεινή των χεριών και η αναπνευστική εθιμοτυπία, συμβάλλουν ουσιαστικά. Επιπλέον, η ορθολογική χρήση των αντιβιοτικών, η διατροφική υποστήριξη και η προαγωγή του θηλασμού στα βρέφη, ιδιαίτερα σε αναπτυσσόμενα περιβάλλοντα, έχουν σημαντικό ρόλο στη μείωση της νοσηρότητας και της θνησιμότητας από πνευμονία.

Acute Respiratory Distress Syndrome (ARDS) Alveolar Changes



Εικόνα 4: Σχηματική απεικόνιση κυψελίδας σε πνευμονία – εξίδρωμα, ουδετερόφιλα, πάχυνση μεμβράνης ανταλλαγής αερίων και V/Q mismatch ([https://www.jthjournal.org/article/S1538-7836\(22\)01667-1/fulltext](https://www.jthjournal.org/article/S1538-7836(22)01667-1/fulltext)).

3.2 Κατηγορίες Πνευμονίας

Η πνευμονία όπως προαναφέρθηκε μπορεί να προκληθεί από διάφορους παθογόνους μικροοργανισμούς και να εμφανιστεί σε διαφορετικά κλινικά πλαίσια. Η διαφοροποίηση της νόσου ανάλογα με το αίτιο είναι καθοριστική τόσο για τη διάγνωση όσο και για την επιλογή της θεραπευτικής αγωγής. Στην ενότητα αυτή παρουσιάζονται οι κύριες κατηγορίες [51][54][55]:

Βακτηριδιακή Πνευμονία

Η βακτηριδιακή πνευμονία είναι η πιο συχνή μορφή πνευμονίας και οφείλεται σε μόλυνση από βακτήρια που εγκαθίστανται στις κατώτερες αναπνευστικές οδούς. Ο πιο συχνός αιτιολογικός παράγοντας είναι ο *Streptococcus pneumoniae* (πνευμονιόκοκκος), ενώ άλλοι συχνοί μικροοργανισμοί είναι οι *Haemophilus influenzae*, *Staphylococcus aureus* και *Klebsiella pneumoniae*. Κλινικά χαρακτηρίζεται από απότομη έναρξη υψηλού πυρετού (άνω των 38°C), παραγωγικό βήχα με πυώδη πτύελα και πλευριτικό θωρακικό πόνο. Στην ακτινογραφία θώρακος παρατηρείται συχνά λοιβώδης διήθηση. Η αντιμετώπιση βασίζεται στη χορήγηση κατάλληλης αντιβιοτικής αγωγής σύμφωνα με κατευθυντήριες οδηγίες, οι οποίες περιλαμβάνουν συνήθως β-λακτάμες, μακρολίδες ή φθοριοκινολόνες, ανάλογα με το προφίλ αντοχής και τη βαρύτητα της λοιμωξης.

Ιογενής Πνευμονία

Η ιογενής πνευμονία προκαλείται από ιούς του αναπνευστικού, όπως οι ιοί γρίπης (Influenza A και B), ο αναπνευστικός συγκυτιακός ιός (RSV), ο αδενοϊός, ο ιός παραγρίπης, αλλά και ο SARS-CoV-2 που ευθύνεται για την COVID-19. Είναι συχνότερη σε παιδιά, ηλικιωμένους και ανοσοκατασταλμένα άτομα. Στις περισσότερες περιπτώσεις παρουσιάζει ήπια κλινική εικόνα, ωστόσο σε ευπαθείς ομάδες μπορεί να εξελιχθεί σε σοβαρή αναπνευστική ανεπάρκεια. Τα τυπικά συμπτώματα περιλαμβάνουν ξηρό βήχα, μέτριο πυρετό, κακουχία και δύσπνοια. Η διάγνωση υποστηρίζεται από μοριακές τεχνικές, όπως η PCR, καθώς και από απεικονιστικά ευρήματα, όπως διάχυτα διηθήματα στην αξονική τομογραφία. Η θεραπεία είναι κατά κύριο λόγο υποστηρικτική, ενώ σε περιπτώσεις σοβαρής λοίμωξης μπορεί να απαιτηθεί η χορήγηση αντιικών φαρμάκων, όπως η οσελταμιβίρη για τη γρίπη ή η ρεμδεσιβίρη για την COVID-19.

Μυκητιασική Πνευμονία

Η μυκητιασική πνευμονία εμφανίζεται πιο σπάνια και αφορά κυρίως ανοσοκατασταλμένους ασθενείς, όπως μεταμοσχευμένους, άτομα με HIV/AIDS ή ασθενείς που λαμβάνουν χημειοθεραπεία ή κορτικοστεροειδή. Οι σημαντικότεροι παθογόνοι μύκητες περιλαμβάνουν τα είδη *Aspergillus* (ασπεργίλλωση), *Histoplasma capsulatum* (ιστοπλάσμωση) και *Pneumocystis jirovecii*, που συχνά ανιχνεύεται σε HIV λοίμωξη. Η διάγνωση είναι απαιτητική και βασίζεται σε ειδικές εργαστηριακές εξετάσεις, όπως καλλιέργειες, αντιγονικές δοκιμές και μοριακές τεχνικές. Η θεραπεία στηρίζεται στη χορήγηση αντιμυκητιασικών φαρμάκων, όπως η αμφοτερικίνη B, η τριμεθοπρίμησουλφαμεθοξαζόλη (ειδικά για *Pneumocystis*) και η βορικοναζόλη.

Παρασιτική Πνευμονία

Η παρασιτική πνευμονία αποτελεί μια σπάνια μορφή, η οποία εμφανίζεται κυρίως σε τροπικές και υποτροπικές περιοχές. Οφείλεται σε παράσιτα όπως το *Toxoplasma gondii*, ο *Paragonimus westermani* (πνευμονικός τρηματώδης σκώληκας), καθώς και σε ελμινθικές λοιμώξεις όπως η ασκαρίδωση, που μπορεί να προκαλέσουν πνευμονική διήθηση. Τα κλινικά συμπτώματα είναι συχνά χρόνια και μη ειδικά, περιλαμβάνοντας βήχα, αιμόπτυση και απώλεια βάρους. Η διάγνωση βασίζεται σε ορολογικές δοκιμές και απεικονιστικά ευρήματα, τα οποία συχνά αναδεικνύουν οζίδια ή κύστεις στον πνεύμονα. Η θεραπεία περιλαμβάνει χορήγηση ανθελμινθικών ή αντιπαρασιτικών φαρμάκων, ανάλογα με τον υπεύθυνο παθογόνο.

3.3 Συμπτώματα και Διάγνωση

Η πνευμονία παρουσιάζει ένα ευρύ φάσμα κλινικών εκδηλώσεων, ανάλογα με τον αιτιολογικό παράγοντα, την ηλικία, τη γενική κατάσταση του ασθενούς και την ανοσολογική του κατάσταση [50][53][54].

Κύρια συμπτώματα

Τα κύρια συμπτώματα περιλαμβάνουν πυρετό και ρίγη, που είναι συχνά υψηλά στις βακτηριδιακές λοιμώξεις και μέτριας έντασης στις ιογενείς, καθώς και βήχα, ο οποίος μπορεί να είναι παραγωγικός με πυώδη πτύελα στις βακτηριδιακές μορφές ή ξηρός στις ιογενείς και μυκητιασικές. Η δύσπνοια και η ταχύπνοια εμφανίζονται ανάλογα με την

έκταση της πνευμονικής βλάβης, ενώ ο θωρακικός πόνος, συνήθως πλευριτικού τύπου, εντοπίζεται στην περιοχή της φλεγμονής. Συχνά παρατηρούνται επίσης γενικά συμπτώματα, όπως κόπωση, κακουχία, μυαλγίες και κεφαλαλγία, ενώ σε ευπαθείς ομάδες μπορεί να εκδηλωθούν άτυπα σημεία, όπως γαστρεντερικές διαταραχές ή σύγχυση σε ηλικιωμένους.

Διάγνωση

Η διάγνωση της πνευμονίας βασίζεται σε συνδυασμό κλινικών, εργαστηριακών και απεικονιστικών ευρημάτων. Κατά την κλινική εξέταση αξιολογούνται τα ζωτικά σημεία (θερμοκρασία, καρδιακή και αναπνευστική συχνότητα, κορεσμός O_2), ενώ η ακρόαση πνευμόνων μπορεί να αποκαλύψει σφύξεις, ρωγμές ή μειωμένους αναπνευστικούς ήχους. Από τις εργαστηριακές εξετάσεις, η γενική αίματος μπορεί να δείξει λευκοκυττάρωση ή λευκοπενία, ενώ οι φλεγμονώδεις δείκτες, όπως CRP, ESR και προκαλσιτονίνη (PCT), συμβάλλουν στην εκτίμηση βακτηριακής λοίμωξης. Η ανάλυση αερίων αίματος και το οξειοβασικό ισοζύγιο είναι χρήσιμα για την εκτίμηση της υποξαιμίας και της σοβαρότητας της νόσου.

Ο μικροβιολογικός έλεγχος περιλαμβάνει καλλιέργειες πτυέλων ή άλλων αναπνευστικών εκκριμάτων για ταυτοποίηση βακτηρίων, μοριακές τεχνικές (PCR) για ιούς και μύκητες, καθώς και ορολογικές εξετάσεις για την ανίχνευση αντισωμάτων ή αντιγόνων. Στις απεικονιστικές μεθόδους, η ακτινογραφία θώρακος αποτελεί την πρώτη γραμμή διάγνωσης και μπορεί να δείξει λοβώδη ή διάχυτα διηθήματα, ενώ η αξονική τομογραφία θώρακος χρησιμοποιείται σε αμφίβολες περιπτώσεις ή σε βαριές λοιμώξεις για τον αποκλεισμό επιπλοκών. Το υπερηχογράφημα πνεύμονα συμβάλλει στην παρακολούθηση πλευριτικών συλλογών ή σε ασθενείς που δεν είναι σε θέση να μετακινηθούν. Σε ειδικές περιπτώσεις, μπορεί να αξιοποιηθεί το σπινθηρογράφημα πνευμόνων, ενώ η εκτίμηση δεικτών ανοσοκαταστολής βοηθά στη διαστρωμάτωση κινδύνου για τις ευπαθείς ομάδες.

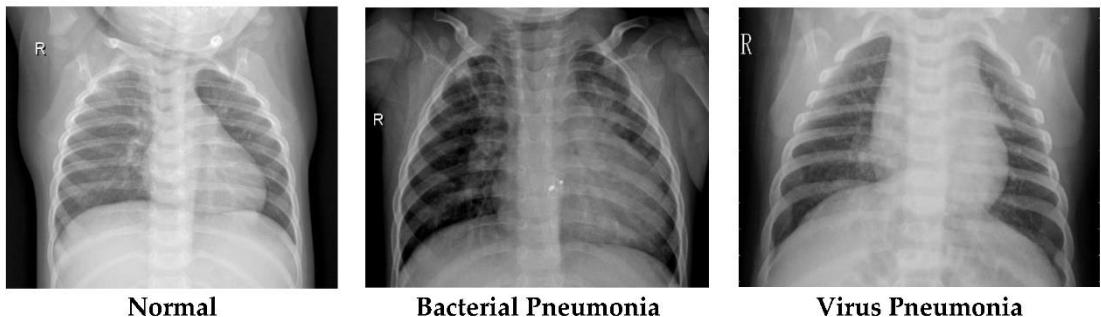
3.4 Μέθοδοι Ανίχνευσης και Ιατρική Απεικόνιση

Η διάγνωση της πνευμονίας στηρίζεται σε συνδυασμό κλινικής εκτίμησης, εργαστηριακών εξετάσεων και απεικονιστικών μεθόδων. Η ιατρική απεικόνιση αποτελεί θεμελιώδες εργαλείο, καθώς αποκαλύπτει την έκταση και τη μορφή της πνευμονικής βλάβης, εντοπίζει επιπλοκές και υποστηρίζει την παρακολούθηση της θεραπείας. Επιτρόσθετα είναι ένας σχετικά φτηνός και κυρίως άμεσος τρόπος για το σχηματισμό μιας αρχικής διάγνωσης [54][55].

Ακτινογραφία θώρακος

Η ακτινογραφία θώρακος (Chest X-Ray) χρησιμοποιείται ως πρώτη γραμμή απεικόνισης, παρέχοντας τη δυνατότητα ανίχνευσης λοβωδών ή διάχυτων διηθημάτων, consolidation και πλευριτικής συλλογής, ενώ συμβάλλει επίσης στην παρακολούθηση της εξέλιξης της νόσου και της ανταπόκρισης στη θεραπεία. Ωστόσο, παρουσιάζει περιορισμούς, καθώς δεν εντοπίζει πάντοτε πρώιμες βλάβες ή μικρές εστίες, ιδιαίτερα σε ανοσοκατασταλμένους ασθενείς.

Ενδεικτικές ακτινογραφίες θώρακος



Normal

Bacterial Pneumonia

Virus Pneumonia

Εικόνα 5: Ενδεικτικές ακτινογραφίες θώρακος. Αριστερά, μια καθαρή ακτινογραφία χωρίς ενδείξεις πνευμονίας. Στη μεση εχουμε μια περίπτωση βακτηριακής πνευμονίας, όπου παρατηρείται συνήθως εντοπισμένο, λοβώδες διήθημα με σαφή όρια. Δεξιά, περίπτωση ιογενούς πνευμονίας, με πιο διάχυτα και αμφοτερόπλευρα διηθήματα, τα οποία εμφανίζουν μεγαλύτερη ενδοκλασσιακή ποικιλία και δυσκολεύουν τη διάκριση από άλλες παθήσεις [56].

Αξονική Τομογραφία

Η αξονική τομογραφία (CT θώρακος) εφαρμόζεται κυρίως σε περιπτώσεις σοβαρής ή σύνθετης πνευμονίας, όταν η ακτινογραφία δεν παρέχει επαρκείς πληροφορίες, καθώς αναδεικνύει μικρές εστίες φλεγμονής, διάχυτα διηθήματα, αποστήματα και πλευριτικές συλλογές, ενώ ταυτόχρονα βοηθά στη διαφοροδιάγνωση μεταξύ βακτηριακής, ιογενούς και μυκητιασικής αιτιολογίας.

Υπερηχογράφημα πνεύμονα:

Το υπερηχογράφημα πνεύμονα προσφέρει σημαντικά πλεονεκτήματα σε ασθενείς που δεν μπορούν να υποβληθούν σε ακτινολογική εξέταση, επιτρέποντας την ανίχνευση πλευριτικών υγρών, συλλογών ή συγκεντρωμένων εστιών consolidation, ενώ μπορεί να χρησιμοποιηθεί και για την καθοδήγηση παρεμβατικών διαδικασιών, όπως η παροχέτευση πλευριτικών συλλογών.

Πυρηνική Ιατρική (Σπινθηρογράφημα πνευμόνων):

Σπανιότερα, χρησιμοποιούνται μέθοδοι πυρηνικής ιατρικής, όπως το σπινθηρογράφημα πνευμόνων, το οποίο συμβάλλει στην εκτίμηση της λειτουργικής συμμετοχής των πνευμονικών περιοχών, ιδίως σε υποπλεγμονώδεις ή χρόνιες καταστάσεις.

Συμπληρωματικά εργαλεία:

Παράλληλα, συμπληρωματικά εργαλεία, όπως οι μοριακές τεχνικές και η PCR, αξιοποιούνται για την ταυτοποίηση των αιτιολογικών παραγόντων (ιούς, μύκητες), ενώ οι καλλιέργειες πτυέλων και οι αιματολογικές εξετάσεις χρησιμοποιούνται κυρίως για την επιβεβαίωση της βακτηριακής αιτιολογίας. Συνολικά, η συνδυαστική αξιοποίηση κλινικών, εργαστηριακών και απεικονιστικών στοιχείων αυξάνει σημαντικά την ακρίβεια της διάγνωσης και επιτρέπει την εφαρμογή στοχευμένων θεραπευτικών προσεγγίσεων.

3.5 Προκλήσεις στη Διάγνωση Πνευμονίας

Η διάγνωση της πνευμονίας μπορεί να είναι δύσκολη, κυρίως λόγω της ποικιλομορφίας των συμπτωμάτων και της επικαλυπτόμενης κλινικής εικόνας με άλλες αναπνευστικές παθήσεις [54]. Μία από τις κυριότερες προκλήσεις αφορά τα μη ειδικά συμπτώματα, καθώς σε ηλικιωμένους ή ανοσοκατεσταλμένους ασθενείς η πνευμονία μπορεί να εμφανιστεί χωρίς τον κλασικό πυρετό ή βήχα, με κυρίαρχα συμπτώματα την κόπωση, τη σύγχυση ή γαστρεντερικές διαταραχές. Επιπλέον, η διαφοροδιάγνωση από άλλες παθήσεις, όπως η χρόνια αποφρακτική πνευμονοπάθεια (ΧΑΠ), το άσθμα, η καρδιακή ανεπάρκεια με πλευριτική συλλογή ή φλεγμονώδεις και αυτοάνοσες παθήσεις του πνεύμονα, καθιστά τη διάγνωση ακόμη πιο σύνθετη. Σημαντικές δυσκολίες προκαλούν και οι ακτινολογικές εξετάσεις, καθώς οι αρχόμενες βλάβες μπορεί να μην εντοπίζονται σε απλή ακτινογραφία, ενώ διάχυτες ή μικρές εστίες σε ιογενείς ή μυκητιασικές λοιμώξεις συχνά απαιτούν CT. Η πολυπαραγοντική αιτιολογία, όπως σπάνιες μικτές λοιμώξεις (π.χ. βακτηριακή και ιογενής ταυτόχρονα), δυσχεραίνει την επιλογή κατάλληλης θεραπείας, ενώ η παρουσία ανθεκτικών στελεχών βακτηρίων μπορεί να καθυστερήσει την ανταπόκριση στη θεραπεία και να περιορίσει την επιβεβαίωση διάγνωσης μέσω καλλιεργειών. Για την αντιμετώπιση αυτών των προκλήσεων, εφαρμόζονται στρατηγικές όπως η συνδυαστική χρήση κλινικών, εργαστηριακών και απεικονιστικών στοιχείων, η επαναλαμβανόμενη παρακολούθηση του ασθενούς, η χρήση εξειδικευμένων μοριακών δοκιμασιών για την ταυτοποίηση παθογόνων και η ομαδική εκτίμηση από πνευμονολόγο, ακτινολόγο και λοιμωξιολόγο. Η αναγνώριση και η κατανόηση αυτών των προκλήσεων είναι κρίσιμη για την έγκαιρη διάγνωση, την αποτελεσματική θεραπεία και τη μείωση των επιπλοκών της πνευμονίας.

4. Επεξεργασία Δεδομένων

4.1 Συλλογή Δεδομένων

Η βάση δεδομένων που αξιοποιήθηκε στην παρούσα εργασία προέρχεται από τη δημοσίευση στο περιοδικό Cell του άρθρου: Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning ([https://www.cell.com/cell/fulltext/S0092-8674\(18\)30154-5](https://www.cell.com/cell/fulltext/S0092-8674(18)30154-5)). Το dataset περιλαμβάνει παιδιατρικές ακτινογραφίες θώρακα (Chest X-Rays) με στόχο την ανίχνευση και ταξινόμηση πνευμονίας [62]. Οι εικόνες είναι αποθηκευμένες σε μορφή JPEG και κατηγοριοποιούνται σε τρεις βασικές κατηγορίες:

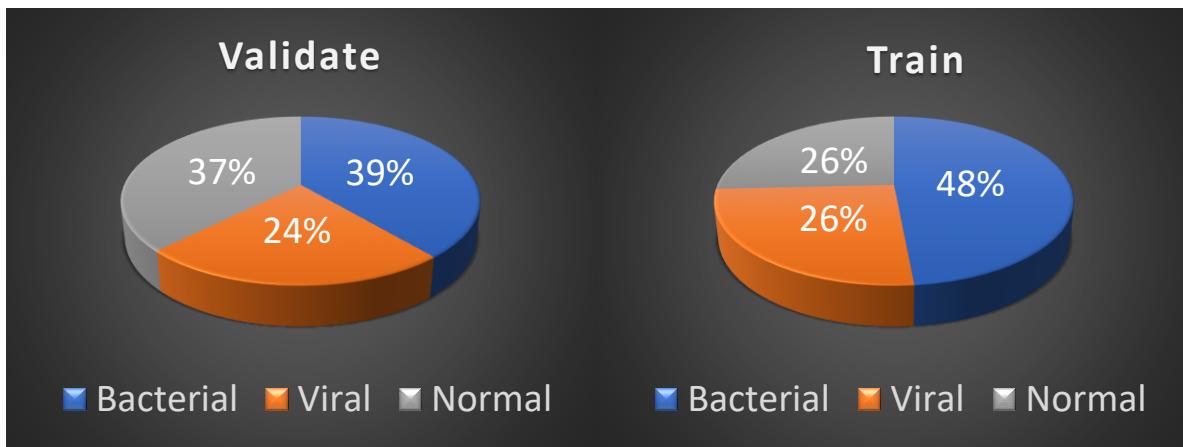
- Bacterial Pneumonia
- Viral Pneumonia
- Normal

Συνολικά περιλαμβάνονται 5.856 εικόνες, οι οποίες κατηγοριοποιούνται με βάση την κλινική διάγνωση. Οι εικόνες προέρχονται από διαφορετικές πηγές και παρουσιάζουν ανομοιογένεια ως προς την ποιότητα, τη φωτεινότητα και τις διαστάσεις, γεγονός που καθιστά απαραίτητη την επεξεργασία τους πριν από την είσοδο στο μοντέλο. Τα δεδομένα χωρίστηκαν σε εκπαίδευση (train) και επαλήθευση (validation) με αναλογία περίπου 90% - 10%. Όπως φαίνεται και παρακάτω η κατανομή των δεδομένων ανά κατηγορία παρουσιάζει ανισορροπία (imbalanced dataset), καθώς η κατηγορία bacterial pneumonia είναι πιο συχνή. Η ανισορροπία αυτή αντιμετωπίστηκε σε επόμενο στάδιο με την εφαρμογή τεχνικών όπως data augmentation και class weighting.

Αναλυτική κατανομή δεδομένων:

- Training set:
 - Bacterial: 2538
 - Normal: 1349
 - Viral: 1345
- Validation set:
 - Bacterial: 242
 - Normal: 234
 - Viral: 148

Παρακάτω ακολουθεί και μια σχηματική αναπαράσταση των ποσοστών συνόλου δεδομενων για την καλύτερη κατανόηση του.



Εικόνα 6: Ποσοστά διαχωρισμού του συνόλου δεδομένων. Αριστερά το συνολο των ποσοστών των κλάσεων του Validate που χρησιμοποιείται για την αξιολόγηση του εκπαίδευμένου μοντέλου και αντίστοιχα δεξιά το συνολο του Train που χρησιμοποιείται για την εκπαίδευση του μοντέλου.

4.2 Καθαρισμός Δεδομένων και Αντιμετώπιση Ελλιπών Τιμών

Η προεπεξεργασία των δεδομένων αποτελεί ένα από τα πιο κρίσιμα στάδια στην ανάπτυξη ενός συστήματος μηχανικής μάθησης, καθώς η ποιότητα των δεδομένων επηρεάζει άμεσα την απόδοση του μοντέλου. Στην παρούσα εργασία, ο καθαρισμός των δεδομένων και η αντιμετώπιση ελλιπών ή ατελών τιμών πραγματοποιήθηκε σε δύο επίπεδα: (α) στο επίπεδο των αρχείων εικόνας και (β) στο επίπεδο της τροφοδότησης δεδομένων σε ενα μοντέλο. Η ανίχνευση πραγματοποιήθηκε με χρήση scripts σε Python, τα οποία προσπάθησαν να ανοίξουν και να φορτώσουν κάθε εικόνα μέσω της βιβλιοθήκης Pillow (PIL) [61]. Όσες εικόνες δεν μπορούσαν να φορτωθούν απορρίφθηκαν και οι μη έγκυρες εικόνες διαγράφηκαν από το dataset. Η διαδικασία αυτή εξασφάλισε ότι το τελικό σύνολο δεδομένων περιείχε έγκυρες και μοναδικές εικόνες ανά κλάση.

Έλεγχος και Καθαρισμός Δεδομένων Εικόνων

Αρχικά, τα δεδομένα που χρησιμοποιήθηκαν προέρχονταν από έναν φάκελο εικόνων ταξινομημένων σε υποφακέλους ανάλογα με την κατηγορία (π.χ. πνευμονία, φυσιολογικές ακτινογραφίες). Προτού ξεκινήσει η διαδικασία εκπαίδευσης, πραγματοποιήθηκε έλεγχος για:

- Ελλιπή ή κατεστραμμένα αρχεία:** Οποιοδήποτε αρχείο εικόνας που δεν μπορούσε να φορτωθεί ή δεν πληρούσε τις ελάχιστες απαιτήσεις διαστάσεων απορρίφθηκε.
- Ασυνέπειες στην ονοματοδοσία:** Επιβεβαιώθηκε ότι οι φάκελοι αντιστοιχούσαν σε μοναδικές κλάσεις και ότι δεν υπήρχαν εικόνες σε λάθος κατηγορία.
- Διπλότυπα αρχεία:** Έγινε έλεγχος για διπλές εικόνες, ώστε να αποφεύγεται η μεροληψία του μοντέλου.

Αντιμετώπιση Διπλότυπων Εικόνων

Για να αποφευχθεί η υπερεκπαίδευση, έγινε έλεγχος για τυχόν διπλότυπα αρχεία. Χρησιμοποιήθηκαν τεχνικές hashing (π.χ., MD5 checksum) για τον εντοπισμό αρχείων με ταυτόσημο περιεχόμενο.

Έλεγχος και Διαχείριση Ελλιπών Τιμών

Αν και οι εικόνες δεν περιλαμβάνουν πίνακες χαρακτηριστικών με κενές τιμές, υπήρχε πιθανότητα ελλιπών μεταδεδομένων (π.χ., ετικέτες κατηγορίας). Για το σκοπό αυτό:

- Ελέγχθηκε η ύπαρξη αντιστοίχισης κάθε εικόνας με την κατάλληλη ετικέτα (label).
- Όπου δεν υπήρχε διαθέσιμη ετικέτα, η εικόνα αφαιρέθηκε από το dataset για να διασφαλιστεί η ορθότητα της εκπαίδευσης.

Αντιμετώπιση Ανομοιόμορφης Κατανομής Κλάσεων

Σε περιπτώσεις ανισορροπίας μεταξύ των κλάσεων, υπολογίστηκαν βάρη κλάσεων (class weights) ώστε να μειωθεί η μεροληφία προς τις πολυπληθέστερες κλάσεις κατά την εκπαίδευση του μοντέλου [8].

Η τεχνική των class weights εφαρμόζεται σε προβλήματα ταξινόμησης όταν το dataset παρουσιάζει ανισορροπία μεταξύ των κλάσεων. Σε τέτοιες περιπτώσεις, αν δεν γίνει καμία ενέργεια, το μοντέλο τείνει να ευνοεί τις πλειοψηφικές κλάσεις, αγνοώντας σε μεγάλο βαθμό τις μειοψηφικές. Αυτό οδηγεί σε μειωμένη απόδοση για τις κατηγορίες που αντιπροσωπεύονται λιγότερο.

Όταν ο αριθμός δειγμάτων ανά κλάση διαφέρει σημαντικά, ο υπολογισμός των βαρών (weights) γίνεται έτσι ώστε να αυξάνεται η σημασία των μειοψηφικών κλάσεων κατά την εκπαίδευση. Αυτό εξισορροπεί την επίδραση κάθε κλάσης στη συνάρτηση κόστους, μειώνοντας το bias προς την πλειοψηφική κλάση

Η πιο συνηθισμένη προσέγγιση είναι η αντιστρόφως ανάλογη σχέση με τη συχνότητα των δειγμάτων κάθε κλάσης. Δηλαδή, όσο λιγότερα δείγματα έχει μια κλάση, τόσο μεγαλύτερο βάρος αποκτά. Στο παράδειγμά μας, υπολογίζουμε τον αριθμό δειγμάτων για κάθε κλάση και εφαρμόζουμε:

$$class_weight_i = \frac{1}{count(i)} \quad (17)$$

Τα weights κανονικοποιούνται ώστε το άθροισμά τους να ισούται με 1, εξασφαλίζοντας σταθερότητα κατά την εκπαίδευση.

Εναρμόνιση Δομής Φακέλων

Όλες οι εικόνες οργανώθηκαν σε φάκελο ανά κατηγορία (Normal, Bacterial Pneumonia, Viral Pneumonia), ώστε να είναι συμβατές με τον μηχανισμό φόρτωσης δεδομένων που χρησιμοποιεί η συνάρτηση **ImageDataGenerator** της βιβλιοθήκης TensorFlow Keras.

Η διαδικασία καθαρισμού διασφάλισε ότι το τελικό dataset ήταν απαλλαγμένο από σφάλματα και κατάλληλα οργανωμένο για την επόμενη φάση της προεπεξεργασίας.

4.3 Κανονικοποίηση και Μετασχηματισμός Εικόνων

Η κανονικοποίηση (Regularization) και ο κατάλληλος μετασχηματισμός των δεδομένων εισόδου αποτελούν κύρια βήματα πριν την εκπαίδευση ενός μοντέλου βαθιάς μάθησης. Η ποιότητα και η ομοιομορφία των εικόνων επηρεάζουν άμεσα τόσο την ταχύτητα σύγκλισης του αλγορίθμου όσο και την τελική ακρίβεια του συστήματος. Πρέπει επίσης να εξασφαλιστεί ότι τα δεδομένα εισόδου θα είναι συμβατά με τα προεκπαιδευμένα μοντέλα (Xception και VGG16) [7][60] που χρησιμοποιήθηκαν στην παρούσα εργασία.

Κανονικοποίηση Εικονοστοιχείων (Pixel Normalization)

Οι αρχικές εικόνες του συνόλου δεδομένων είχαν τιμές εικονοστοιχείων (pixels) στην κλίμακα [0, 255]. Για να επιτευχθεί σταθερότητα και βελτίωση της απόδοσης κατά την εκπαίδευση, εφαρμόστηκε κανονικοποίηση των τιμών (μέσω διαίρεσης με το 255) ώστε οι τιμές να βρίσκονται στο διάστημα [0, 1]. Το βήμα αυτό είναι απαραίτητο για να αποτραπεί η υπερχείλιση των τιμών στα επίπεδα ενεργοποίησης και να διατηρηθεί η αριθμητική σταθερότητα κατά τον υπολογισμό των συναρτήσεων κόστους.

$$x' = \frac{x}{255} \quad (18)$$

όπου x είναι η αρχική τιμή pixel και x' η κανονικοποιημένη τιμή.

Η κανονικοποίηση πραγματοποιήθηκε μέσω της παραμέτρου rescale=1./255 στη συνάρτηση ImageDataGenerator της βιβλιοθήκης Keras.

Προσαρμογή Διαστάσεων (Resizing) και Bilinear Interpolation

Το προεκπαιδευμένα μοντέλα Xception και VGG16 απαιτούν συγκεκριμένες διαστάσεις εισόδου [7][60]:

- **Xception:** 299 × 299 × 3
- **VGG16:** 224 × 224 × 3

Επομένως, όλες οι εικόνες επαναδειγματοληπτήκαν σε αυτό το μέγεθος χρησιμοποιώντας τη μέθοδο bilinear interpolation, διατηρώντας την αναλογία όποτε ήταν δυνατό. Η επιλογή του συγκεκριμένου μεγέθους προέρχεται από τις προδιαγραφές του μοντέλου και συμβάλλει στη μέγιστη αξιοποίηση των προεκπαιδευμένων βαρών.

Η bilinear interpolation υπολογίζει την τιμή κάθε νέου pixel ως σταθμισμένο μέσο όρο των 4 πλησιέστερων pixels στο αρχικό πλέγμα. Σε αντίθεση με την nearest-neighbor μέθοδο, η οποία απλώς αντιγράφει την τιμή του κοντινότερου pixel, η bilinear προσφέρει πιο ομαλό αποτέλεσμα και μειώνει τις παραμορφώσεις.

Μαθηματικά, για ένα νέο σημείο (x, y) η τιμή του υπολογίζεται ως:

$$f(x, y) = \sum_{i=0}^1 \sum_{j=0}^1 w_{ij} \cdot f(x_i, y_j) \quad (19)$$

Όπου w_{ij} είναι τα βάρη που εξαρτώνται από την απόσταση του νέου σημείου από τα γειτονικά.

Μετασχηματισμός Χρωματικού Χώρου

Όλες οι εικόνες μετατράπηκαν σε τριών καναλιών RGB, ανεξαρτήτως αν οι αρχικές ακτινογραφίες ήταν ασπρόμαυρες. Αυτό επιτεύχθηκε μέσω της λειτουργίας `ImageDataGenerator.flow_from_directory`, ώστε οι διαστάσεις να είναι συμβατές με με τα προεκπαιδευμένα μοντέλα Xception και VGG16 (3 κανάλια εισόδου).

4.4 Τεχνικές Αύξησης Δεδομένων (Data Augmentation)

Η αύξηση δεδομένων (Data Augmentation) αποτελεί κρίσιμη τεχνική για τη βελτίωση της γενίκευσης των μοντέλων βαθιάς μάθησης, ιδίως όταν το διαθέσιμο σύνολο εκπαίδευσης είναι περιορισμένο ή παρουσιάζει ανισορροπία μεταξύ των κλάσεων. Η μέθοδος αυτή συνίσταται στη δημιουργία νέων δειγμάτων από τα υπάρχοντα, μέσω εφαρμογής τυχαίων αλλά ελεγχόμενων μετασχηματισμών, χωρίς να απαιτείται συλλογή πρόσθετων δεδομένων [57][58][59].

Στόχοι του Data Augmentation

- Μείωση Overfitting:** Εμπλουτίζοντας τα δεδομένα εκπαίδευσης με διαφοροποιημένες εκδοχές των ίδιων εικόνων, μειώνεται η πιθανότητα υπερεκπαίδευσης.
- Αύξηση Γενίκευσης:** Το μοντέλο μαθαίνει να αναγνωρίζει χαρακτηριστικά ανεξάρτητα από αλλαγές θέσης, κλίμακας ή φωτισμού.
- Αντιμετώπιση Ανισορροπίας Κλάσεων:** Εφαρμογή εντονότερου augmentation στις κλάσεις με μικρότερο πλήθος δειγμάτων.

Τεχνικές Data Augmentation που χρησιμοποιήθηκαν

Η ανισορροπία μεταξύ κλάσεων (class imbalance) αποτελεί συχνό πρόβλημα σε ιατρικά δεδομένα, όπου ο αριθμός των θετικών δειγμάτων (π.χ. ασθενών) είναι συνήθως μικρότερος σε σχέση με τα αρνητικά. Το Data Augmentation συμβάλλει σημαντικά στην αντιμετώπιση ενός Αντιμετώπιση Imbalanced Dataset όπως αυτου που εχουμε. Οι τεχνικές που εφαρμόστηκαν μέσω της κλάσης `ImageDataGenerator` της Keras περιλαμβάνουν:

- Περιστροφή εικόνας (rotation):** Έως $\pm 15^\circ$, ώστε να προσομοιωθούν διαφορετικές θέσεις λήψης.
- Οριζόντια αναστροφή (horizontal flip):** Για την ενίσχυση συμμετρίας δεδομένων.
- Μετατόπιση (width & height shift):** Έως 5% της διάστασης.

- **Διατμητική παραμόρφωση (shear):** Μέχρι 5%.
- **Εστίαση (zoom):** Έως 15%.
- **Μεταβολές φωτεινότητας (brightness range):** Από 85% έως 115% της αρχικής τιμής.
- **Συμπλήρωση (fill mode):** nearest για αποφυγή δημιουργίας μαύρων περιοχών μετά τους μετασχηματισμούς.

Παρακάτω παραθέτω και έναν συγκριτικό πίνακα Τεχνικών Data Augmentation και της επίδρασης τους.

Τεχνική	Περιγραφή	Επίδραση στην Γενίκευση	Επίδραση στο Imbalanced Dataset
Rotation	Περιστροφή εικόνας έως $\pm 15^\circ$	Αυξάνει ανθεκτικότητα σε αλλαγές θέσης	Μέτρια, προσθέτει ποικιλία στη μειοψηφική κλάση
Horizontal Flip	Αντιστροφή εικόνας οριζόντια	Καλή για συμμετρικά αντικείμενα	Χαμηλή, δεν αλλάζει τη βασική γεωμετρία
Width/Height Shift	Μετατόπιση έως 5%	Βελτιώνει robustness σε κακή ευθυγράμμιση	Μέτρια
Shear	Διατμητική παραμόρφωση	Αυξάνει διαφοροποίηση χαρακτηριστικών	Υψηλή, χρήσιμη για μειοψηφικές
Zoom	Εστίαση $\pm 15\%$	Προσομοιώνει διαφορετικές αποστάσεις	Υψηλή, ειδικά σε μειοψηφικές κλάσεις
Brightness Variation	Μεταβολή φωτεινότητας $\pm 15\%$	Αυξάνει ανοχή σε διαφορετικό φωτισμό	Μέτρια

Πίνακας 2: Σύνοψη βασικών τεχνικών αύξησης δεδομένων (data augmentation) και της επίδρασή τους στη γενίκευση και σε ανισόρροπα σύνολα δεδομένων.

Σημαντικές Παρατηρήσεις

- Το Data Augmentation εφαρμόζεται μόνο στο σύνολο εκπαίδευσης, ενώ το σύνολο επικύρωσης παραμένει αμετάβλητο για να παρέχει αντικειμενική μέτρηση απόδοσης.

- Οι τιμές των παραμέτρων επιλέχθηκαν πειραματικά ώστε να μην αλλοιώνουν σημαντικά τα διαγνωστικά χαρακτηριστικά των ακτινογραφιών.

4.5 Ετεροσυσχέτιση Δεδομένων

Αξίζει να σημειωθεί ότι οι είσοδοι του μοντέλου, δηλαδή τα εικονοστοιχεία των ακτινογραφιών θώρακος, δεν είναι ανεξάρτητες μεταβλητές αλλά παρουσιάζουν υψηλό βαθμό ετεροσυσχέτισης. Η συσχέτιση αυτή οφείλεται στη χωρική συνέχεια και στη δομή των ιατρικών εικόνων, όπου γειτονικά pixels εμφανίζουν παρόμοια τιμή φωτεινότητας ή υφής. Η ιδιαιτερότητα αυτή αξιοποιείται από τα Συνελικτικά Νευρωνικά Δίκτυα, τα οποία μέσω των συνελικτικών φίλτρων αναγνωρίζουν τοπικά μοτίβα και τα συνδυάζουν σε ανώτερα επύπεδα αφαίρεσης. Έτσι, η ετεροσυσχέτιση των δεδομένων εισόδου δεν αποτελεί πρόβλημα αλλά, αντιθέτως, βασικό χαρακτηριστικό που επιτρέπει στο δίκτυο να επιτυγχάνει υψηλή ακρίβεια στην ταξινόμηση. Επιπλέον, οι διαφορετικές αρχιτεκτονικές παρουσιάζουν διαφοροποιημένη ικανότητα εκμετάλλευσης αυτής της σχέσης. Το μοντέλο Xception, που χρησιμοποιεί depthwise separable convolutions (βλ. 5.6) καταφέρνει να συλλάβει πιο αποδοτικά τις τοπικές συσχετίσεις σε σχέση με το VGG16. Η παρατήρηση αυτή εξηγεί εν μέρει τις διαφοροποιήσεις που καταγράφηκαν στα πειραματικά αποτελέσματα (βλ. 7.6).

4.6 Εργαλεία που χρησιμοποιήθηκαν

Κατά την εκπόνηση της παρούσας διπλωματικής εργασίας χρησιμοποιήθηκε κυρίως η γλώσσα προγραμματισμού Python. Η συγκεκριμένη γλώσσα κρίθηκε κατάλληλη για τον σκοπό της εργασίας, καθώς παρέχει μια πληθώρα βιβλιοθηκών που διευκολύνουν την ανάπτυξη εφαρμογών για την υλοποίηση Τεχνητών Νευρωνικών Δικτύων, την επεξεργασία εικόνων και την αξιολόγηση μοντέλων. Παρακάτω παρουσιάζονται οι σημαντικότερες βιβλιοθήκες που αξιοποιήθηκαν [61]:

TensorFlow & Keras

Οι βιβλιοθήκες TensorFlow και Keras αποτελούν τον βασικό κορμό της υλοποίησης του νευρωνικού δικτύου. Η Keras παρέχει υψηλού επιπέδου API για την κατασκευή και εκπαίδευση μοντέλων, ενώ η TensorFlow προσφέρει υποστήριξη για GPU, επιταχύνοντας τη διαδικασία εκπαίδευσης. Χρησιμοποιήθηκαν για τη δημιουργία του CNN, την εφαρμογή τεχνικών κανονικοποίησης, την προεπεξεργασία εικόνων και τον ορισμό callbacks όπως το Early Stopping.

Scikit-learn (Sklearn)

Η βιβλιοθήκη Scikit-learn χρησιμοποιήθηκε για τον διαχωρισμό των δεδομένων σε σύνολα εκπαίδευσης, επικύρωσης και δοκιμής, καθώς και για τον υπολογισμό μετρικών απόδοσης όπως το classification report, η ROC-AUC και οι καμπύλες ROC. Επιπλέον, χρησιμοποιήθηκε η δυνατότητα stratified splitting για τη διατήρηση της ισορροπίας μεταξύ κλάσεων.

OpenCV (cv2)

Η βιβλιοθήκη OpenCV χρησιμοποιήθηκε για την επεξεργασία εικόνων πριν από την είσοδό τους στο νευρωνικό δίκτυο. Εφαρμόστηκαν λειτουργίες όπως αλλαγή μεγέθους (resizing), μετατροπή χρωματικού χώρου και χειρισμός εικόνων για augmentation.

Pandas

Η βιβλιοθήκη Pandas χρησιμοποιήθηκε για τη διαχείριση και ανάλυση δομημένων δεδομένων, όπως πίνακες και ετικέτες εικόνων. Παρείχε ευκολία στη φόρτωση και οργάνωση των δεδομένων, καθώς και στη δημιουργία DataFrames για τον έλεγχο στατιστικών και μετρικών απόδοσης.

Pillow (PIL)

Η βιβλιοθήκη Pillow αξιοποιήθηκε για την ανάγνωση, αποθήκευση και βασική επεξεργασία εικόνων σε διάφορες μορφές (formats). Παρείχε λειτουργίες όπως αλλαγή μεγέθους, περιστροφή και προσαρμογή χρωματικών χαρακτηριστικών, διευκολύνοντας την προεπεξεργασία των δεδομένων πριν από την εκπαίδευση.

Matplotlib

Η Matplotlib αξιοποιήθηκε για την οπτικοποίηση των αποτελεσμάτων, όπως οι καμπύλες ακρίβειας και απώλειας κατά την εκπαίδευση, οι καμπύλες ROC, καθώς και για την παρουσίαση των heatmaps από τεχνικές ερμηνευσιμότητας (π.χ. Grad-CAM).

NumPy

Η NumPy χρησιμοποιήθηκε για αριθμητικούς υπολογισμούς και την αποτελεσματική διαχείριση πολυδιάστατων πινάκων δεδομένων.

Σύνοψη Εργαλείων

Βιβλιοθήκη Χρήση

TensorFlow / Keras Κατασκευή και εκπαίδευση CNN, GPU επιτάχυνση

Scikit-learn Διαχωρισμός δεδομένων, μετρικές απόδοσης

OpenCV Επεξεργασία εικόνων

Matplotlib Οπτικοποίηση αποτελεσμάτων

NumPy Αριθμητικοί υπολογισμοί

Πίνακας 3: Σύνοψη εργαλείων και της χρήσης τους στην παρούσα εργασία

5. Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN)

5.1 Εισαγωγή

Η παρούσα ενότητα εστιάζει στην αξιολόγηση της απόδοσης των συνελικτικών νευρωνικών δικτύων (CNN). Τα δίκτυα αυτά αποτελούν μία από τις πλέον καθιερωμένες αρχιτεκτονικές στον χώρο της βαθιάς μάθησης, λόγω της ικανότητάς τους να επιλύουν σύνθετα προβλήματα επεξεργασίας δεδομένων. Ουσιαστικά, συνιστούν μια εξειδικευμένη εκδοχή τεχνητών νευρωνικών δικτύων εμπρόσθιας διάδοσης και εφαρμόζονται εκτενώς σε καθήκοντα όπως η πρόβλεψη και η ανίχνευση προτύπων σε διαφορετικά είδη δεδομένων, μεταξύ των οποίων συγκαταλέγονται το κείμενο, οι εικόνες και το ηχητικό σήμα [63]. Η βασική αρχή λειτουργίας τους βασίζεται στη χρήση συνελικτικών φίλτρων (kernels), τα οποία εξάγουν κρίσιμες πληροφορίες από τα δεδομένα. Αν και η αρχική τους ανάπτυξη στόχευε κυρίως στην ανάλυση εικόνων, η εφαρμογή τους έχει πλέον επεκταθεί σε πεδία όπως η αναγνώριση φωνής, η ανάλυση βίντεο και η επεξεργασία φυσικής γλώσσας. Παρά το γεγονός ότι η κύρια διαφοροποιητική τους ιδιότητα σε σχέση με άλλες αρχιτεκτονικές έγκειται στην υψηλή αποτελεσματικότητα σε δεδομένα εικόνας και ήχου, είναι εξίσου ικανά να εφαρμοστούν σε δεδομένα χωρίς εμφανή χωρική δομή [64].

5.2 Αρχιτεκτονικές CNNs – Εξέλιξη και μοντέλα αναφοράς

Η εξέλιξη των συνελικτικών νευρωνικών δικτύων (CNNs) έχει σημειώσει σημαντική πρόοδο από την αρχική τους πρόταση έως τις σύγχρονες αρχιτεκτονικές. Τα πρώτα CNNs, όπως το LeNet-5, σχεδιάστηκαν για την αναγνώριση χειρόγραφων ψηφίων και αποτέλεσαν το θεμέλιο για την ανάπτυξη πιο σύνθετων δικτύων. Το LeNet-5 χρησιμοποιούσε δύο συνελικτικά στρώματα ακολουθούμενα από πλήρως συνδεδεμένα στρώματα, εισάγοντας την έννοια της ιεραρχικής εξαγωγής χαρακτηριστικών [67].

Η επόμενη σημαντική εξέλιξη παρουσιάστηκε με την αρχιτεκτονική AlexNet, η οποία εισήγαγε την εκτεταμένη χρήση της συνάρτησης ReLU για την επιτάχυνση της εκπαίδευσης και την τεχνική dropout για την πρόληψη υπερπροσαρμογής. Η AlexNet έδειξε ότι τα CNNs μπορούν να επιτύχουν εξαιρετικά αποτελέσματα σε μεγάλα σύνολα δεδομένων, όπως το ImageNet [7].

Στη συνέχεια, τα VGG δίκτυα εισήγαγαν τη χρήση μικρών φίλτρων 3x3 και την αύξηση του βάθους του δικτύου, προσφέροντας καλύτερη αναπαράσταση των χαρακτηριστικών χωρίς υπερβολική αύξηση των παραμέτρων [68]. Ακολούθησαν τα δίκτυα GoogLeNet (Inception), τα οποία εισήγαγαν τα inception modules, επιτρέποντας την επεξεργασία χαρακτηριστικών σε πολλαπλές κλίμακες μέσα στο ίδιο επίπεδο [69].

Πιο πρόσφατα, τα ResNet δίκτυα εισήγαγαν τις residual συνδέσεις, οι οποίες αντιμετώπισαν το πρόβλημα της εξαφάνισης του gradient σε πολύ βαθιά δίκτυα, επιτρέποντας τη δημιουργία CNNs με εκατοντάδες στρώματα [70]. Επιπλέον, αρχιτεκτονικές όπως το DenseNet και το Xception έχουν βελτιστοποιήσει την αποδοτικότητα των συνδέσεων μεταξύ των στρωμάτων, ενισχύοντας την εκμάθηση χαρακτηριστικών και μειώνοντας την πολυπλοκότητα του μοντέλου.

Σύγκριση Μοντέλων Αναφοράς

Μοντέλο	Έτος	Βάθος	Παράμετροι	Καινοτομίες
LeNet-5	1998	7	~60 χιλιάδες	Πρώτο CNN για αναγνώριση ψηφίων
AlexNet	2012	8	~60 εκατομμύρια	ReLU, GPU training, Dropout
VGG16	2014	16	~138 εκατομμύρια 3x3 φίλτρα, απλή αρχιτεκτονική	
GoogLeNet	2014	22	~6,8 εκατομμύρια	Inception modules
ResNet-50	2015	50	~25 εκατομμύρια	Residual connections
Xception	2017	71	~22 εκατομμύρια	Depthwise separable convolutions
EfficientNet-B0	2019	-	~5,3 εκατομμύρια	Compound scaling

Η συνεχής εξέλιξη των αρχιτεκτονικών CNN στοχεύει στη βελτίωση της ακρίβειας, στη μείωση της πολυπλοκότητας και στην καλύτερη γενίκευση. Τα νεότερα μοντέλα επικεντρώνονται στην αποδοτικότητα και στην προσαρμοστικότητα, καλύπτοντας τις αυξανόμενες ανάγκες εφαρμογών σε τομείς όπως η ιατρική διάγνωση, η αυτόνομη οδήγηση και η ανάλυση εικόνας υψηλής ανάλυσης.

5.3 Δομή και λειτουργία Συνελικτικών Τεχνητών Νευρωνικών Δικτυών

Τα συνελικτικά ΤΝΔ αποτελούνται από επίπεδα νευρώνων τα οποία στοχεύουν στη μείωση του σφάλματος μέσω της εφαρμογής του αλγορίθμου εκπαίδευσης backpropagation. Στο τελικό στάδιο, τα δίκτυα περιλαμβάνουν ένα πλήρως συνδεδεμένο επίπεδο (dense layer), το οποίο είναι υπεύθυνο για την παραγωγή της τελικής πρόβλεψης [64]. Όσον αφορά στη συνάρτηση ενεργοποίησης, η πλέον συνηθισμένη επιλογή είναι η ReLU (βλ. εισαγωγή 1.6.1), η οποία μηδενίζει όλες τις αρνητικές τιμές, συμβάλλοντας έτσι στην ταχύτερη σύγκλιση του δικτύου [65].

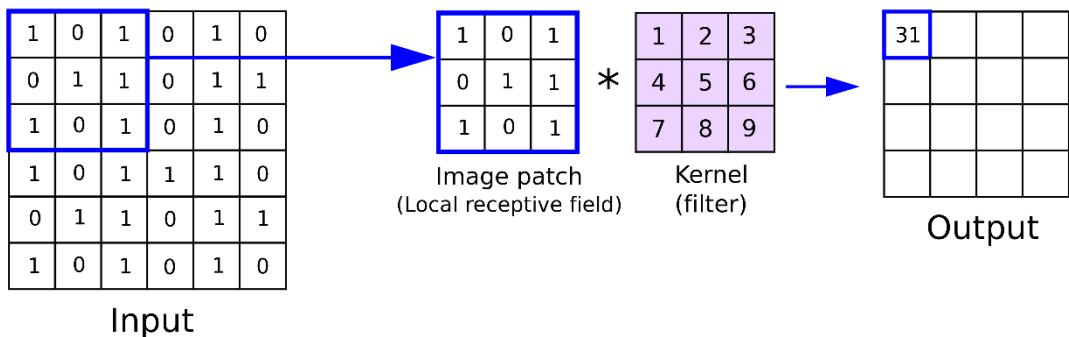
Η βασική διαφορά μεταξύ των Fully Connected Networks (MLPs) δηλαδη των παραδοσιακών feedforward δικτύων χωρίς convolution και των CNN, εντοπίζεται στη διάταξη των επιπέδων που προηγούνται του πλήρως συνδεδεμένου τμήματος του τελικού στρώματος. Τα επίπεδα αυτά λειτουργούν ως φίλτρα σε συγκεκριμένα υποσύνολα των δεδομένων και κατηγοριοποιούνται σε τρείς βασικές κατηγορίες, οι οποίες θα αναλυθούν στη συνέχεια.

5.3.1 Συνελικτικό επίπεδο (Convolutional Layer)

Το συνελικτικό επίπεδο, όπως υποδηλώνει και η ονομασία του, αποτελεί θεμελιώδες στοιχείο στη λειτουργία των Συνελικτικών Νευρωνικών Δικτύων (CNN). Ο κύριος ρόλος του είναι η εφαρμογή πυρήνων (kernels), οι οποίοι, παρά το μικρό τους μέγεθος σε χωρική διάσταση, καλύπτουν ολόκληρο το βάθος των δεδομένων εισόδου. Όταν η είσοδος περνά μέσα από το συνελικτικό επίπεδο, δημιουργείται ένας δισδιάστατος χάρτης ενεργοποίησης (activation map).

Κατά τη διαδικασία αυτή, ο υπολογισμός του γινομένου κατά βαθμίδες πραγματοποιείται για κάθε στοιχείο του πυρήνα καθώς αυτός διασχίζει την είσοδο. Μέσα από αυτή τη λειτουργία, το δίκτυο μαθαίνει φίλτρα τα οποία ενεργοποιούνται όταν ανιχνεύουν συγκεκριμένα χαρακτηριστικά σε ορισμένες περιοχές της εισόδου. Η διαδικασία αυτή είναι γνωστή ως ενεργοποίηση (activation).

Στο παρακάτω σχήμα παρουσιάζεται η απεικόνιση της λειτουργίας του συνελικτικού επιπέδου: το κεντρικό στοιχείο του πυρήνα τοποθετείται πάνω από το τμήμα της εισόδου και αντικαθίσταται με ένα σταθμισμένο άθροισμα των τιμών του ίδιου και των γειτονικών εικονοστοιχείων. [63]



Εικόνα 7: Οπτική αναπαράσταση Convolutional Layer (<https://anhreynolds.com/blogs/cnn.html>)

Η εκπαίδευση Τεχνητών Νευρωνικών Δικτύων σε δεδομένα όπως οι εικόνες οδηγεί στη δημιουργία μοντέλων με πολύ μεγάλο αριθμό παραμέτρων, γεγονός που καθιστά την αποτελεσματική εκπαίδευσή τους ιδιαίτερα απαιτητική. Ο κύριος λόγος είναι η πλήρης συνδεσιμότητα που χαρακτηρίζει τα κλασικά Νευρωνικά Δίκτυα, όπου κάθε νευρώνας συνδέεται με όλους τους νευρώνες του προηγούμενου επιπέδου. Για την αντιμετώπιση αυτού του ζητήματος, στα συνελικτικά στρώματα κάθε νευρώνας συνδέεται μόνο με μια περιορισμένη τοπική περιοχή του όγκου εισόδου. Το μέγεθος αυτής της περιοχής ορίζεται ως μέγεθος του πεδίου υποδοχής (receptive field) του εκάστοτε νευρώνα. Η συνδεσιμότητα κατά μήκος της διάστασης του βάθους συνήθως ισούται με το πλήρες βάθος της εισόδου, ώστε να λαμβάνονται υπόψη όλα τα χαρακτηριστικά καναλιών.

Για παράδειγμα, αν η είσοδος στο δίκτυο είναι μία εικόνα διαστάσεων $128 \times 128 \times 3$ (μια έγχρωμη εικόνα RGB με πλάτος και ύψος 128 pixels) και ορίσουμε το μέγεθος του πεδίου υποδοχέα ως 8×8 , τότε κάθε νευρώνας στο συνελικτικό στρώμα θα διαθέτει συνολικά 192 βάρη ($8 \times 8 \times 3$, όπου το 3 αντιστοιχεί στη συνδεσιμότητα κατά μήκος του βάθους). Αν συγκρίνουμε με ένα τυπικό νευρώνα σε πλήρως συνδεδεμένο TNΔ, εκείνο θα είχε 49.152 βάρη.

Τα συνελικτικά στρώματα μπορούν επίσης να μειώσουν αισθητά την πολυπλοκότητα του μοντέλου μέσω της βελτιστοποίησης της εξόδου τους. Αυτό επιτυγχάνεται με τρεις υπερπαραμέτρους: το βάθος (depth), το βήμα (stride) και την μηδενική συμπλήρωση (zero padding).

Το βάθος του όγκου εξόδου που παράγεται από τα συνελικτικά στρώματα μπορεί να οριστεί χειροκίνητα μέσω του αριθμού των φίλτρων (νευρώνων) στο στρώμα για την ίδια περιοχή εισόδου. Αυτή η ρύθμιση μπορεί να μειώσει τον συνολικό αριθμό νευρώνων του δικτύου, αλλά ταυτόχρονα ενδέχεται να περιορίσει τις δυνατότητες του μοντέλου στην αναγνώριση χαρακτηριστικών.

Επιπλέον, μπορούμε να καθορίσουμε το βήμα, δηλαδή το πόσο θα μετακινηθεί το φίλτρο κατά μήκος της χωρικής διάστασης της εισόδου ώστε να τοποθετήσει το πεδίο υποδοχέα. Για παράδειγμα, αν το βήμα οριστεί ως 1, τα πεδία υποδοχέα θα επικαλύπτονται έντονα, με αποτέλεσμα να παράγονται πολύ μεγάλες ενεργοποιήσεις. Αντιθέτως, αν το βήμα έχει μεγαλύτερη τιμή, η επικάλυψη μειώνεται και η έξοδος αποκτά μικρότερες χωρικές διαστάσεις.

Τέλος, η μηδενική συμπλήρωση (zero padding) είναι η διαδικασία επικάλυψης των άκρων της εισόδου με μηδενικά, επιτρέποντας τον έλεγχο των διαστάσεων της εξόδου και την αποφυγή απώλειας πληροφορίας στις άκρες της εικόνας.

Είναι σημαντικό να γίνει σαφές ότι η χρήση αυτών των υπερπαραμέτρων μεταβάλλει τις χωρικές διαστάσεις της εξόδου ενός συνελικτικού στρώματος. Οι διαστάσεις μπορούν να υπολογιστούν από τον τύπο:

$$\frac{(V-R)+2Z}{S} + 1 \quad (20)$$

όπου:

- V: μέγεθος χωρικής διάστασης της εισόδου (ύψος ή πλάτος)
- R: μέγεθος πεδίου υποδοχέα (kernel size)
- Z: ποσότητα μηδενικής συμπλήρωσης (zero padding)
- S: βήμα μετατόπισης του φίλτρου (stride).

Σημείωση: Ο τύπος εφαρμόζεται ανεξάρτητα σε ύψος και πλάτος, ενώ το βάθος της εξόδου αντιστοιχεί στον αριθμό των φίλτρων του στρώματος. Αν το αποτέλεσμα δεν είναι ακέραιος αριθμός, τότε ο επιλεγμένος συνδυασμός υπερπαραμέτρων είναι ασύμβατος: το φίλτρο δεν «ταιριάζει» στο επίπεδο εισόδου και απαιτείται προσαρμογή του S, του Z ή του R.

5.3.2 Στρώμα Pooling

Τα στρώματα pooling έχουν ως στόχο τη σταδιακή μείωση των διαστάσεων της αναπαράστασης, μειώνοντας έτσι τον αριθμό παραμέτρων και την υπολογιστική πολυπλοκότητα του μοντέλου [67].

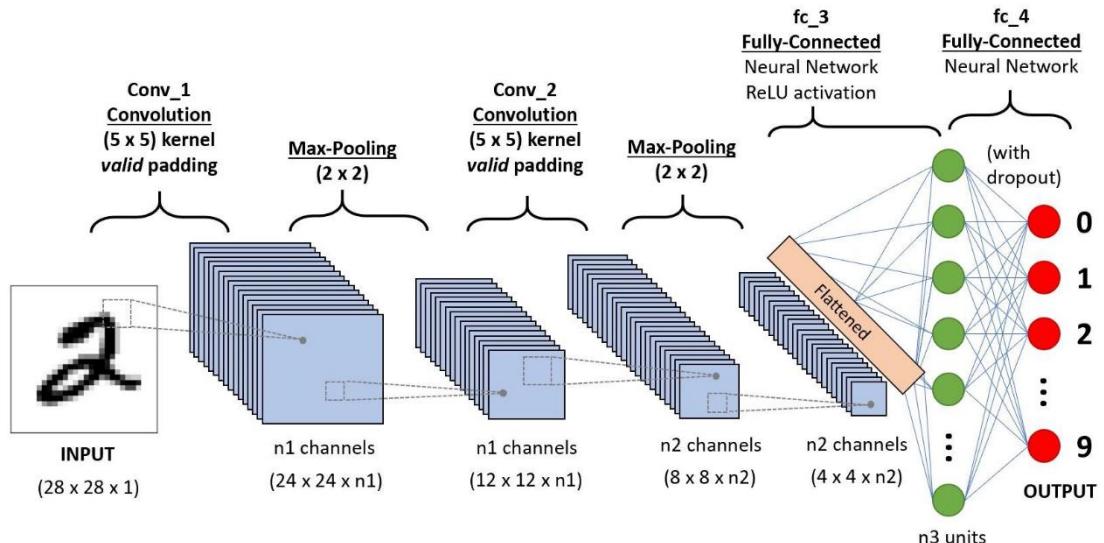
Κάθε στρώμα pooling εφαρμόζεται ξεχωριστά σε κάθε χάρτη ενεργοποίησης της εισόδου και περιορίζει τις διαστάσεις του χρησιμοποιώντας συνήθως τη λειτουργία Max-Pooling. Στα περισσότερα CNNs, το max-pooling πραγματοποιείται με πυρήνες μεγέθους 2×2 και βήμα 2 κατά μήκος των χωρικών διαστάσεων της εισόδου. Αυτό έχει ως αποτέλεσμα τη μείωση του μεγέθους της ενεργοποίησης στο 25% του αρχικού, διατηρώντας ωστόσο τον όγκο βάθους αναλλοίωτο.

Λόγω της φύσης του pooling, οι δύο πιο συνήθεις παραμετροποιήσεις που παρατηρούνται είναι:

- Πυρήνες: 2×2
- Βήμα: 2

Με αυτόν τον τρόπο, κάθε στρώμα pooling καλύπτει πλήρως τις χωρικές διαστάσεις της εισόδου, μειώνοντας σημαντικά τον αριθμό χαρακτηριστικών χωρίς να εξαλείφει τις κρίσιμες πληροφορίες.

Στο παρακάτω σχήμα παρουσιάζεται ένα διάγραμμα αρχιτεκτονικής τυπικού συνελικτικού νευρωνικού δικτύου, στο οποίο φαίνεται η διάταξη των συνελικτικών και pooling επιπέδων.



Εικόνα 8: Απεικόνιση της αρχιτεκτονικής ενός CNN (<https://www.analyticsvidhya.com/blog/2020/10/what-is-the-convolutional-neural-network-architecture/>)

5.3.3 Πλήρως Συνδεδεμένα Στρώματα (Fully Connected Layers)

Τα Πλήρως Συνδεδεμένα Στρώματα (Fully Connected Layers - FC) αποτελούν ένα από τα βασικά δομικά στοιχεία στα Τεχνητά Νευρωνικά Δίκτυα και ιδιαίτερα στα Συνελικτικά Νευρωνικά Δίκτυα (CNNs). Σε αντίθεση με τα συνελικτικά στρώματα, τα οποία επεξεργάζονται χωρικές συσχετίσεις, τα FC στρώματα λειτουργούν ως κλασικοί νευρωνικοί συνδετικοί μηχανισμοί, όπου κάθε νευρώνας συνδέεται με όλους τους νευρώνες του προηγούμενου στρώματος [67].

Ο κύριος ρόλος τους είναι η ενσωμάτωση των χαρακτηριστικών που έχουν εξαχθεί από τα προηγούμενα συνελικτικά και pooling στρώματα και η μετατροπή τους σε μια υψηλού επιπέδου αναπαράσταση κατάλληλη για την τελική διαδικασία ταξινόμησης ή πρόβλεψης. Πρακτικά, το FC στρώμα λειτουργεί ως ταξινομητής, λαμβάνοντας ως είσοδο ένα μονοδιάστατο διάνυσμα, το οποίο προκύπτει από την επιπέδωση (flattening) των τρισδιάστατων χαρτών χαρακτηριστικών.

Σε ένα πλήρως συνδεδεμένο στρώμα, κάθε νευρώνας έχει ξεχωριστό βάρος για κάθε είσοδο, καθώς και έναν όρο μεροληψίας (*bias*). Ο υπολογισμός της εξόδου κάθε νευρώνα γίνεται με τον τύπο:

$$y = f(\sum_{i=1}^n w_i x_i + b) \quad (21)$$

όπου είναι τα βάρη, οι είσοδοι, ο όρος μεροληψίας και η συνάρτηση ενεργοποίησης.

Παρά την αποτελεσματικότητά τους, τα FC στρώματα εισάγουν μεγάλο αριθμό παραμέτρων, ειδικά όταν το προηγούμενο επίπεδο έχει μεγάλο αριθμό χαρακτηριστικών. Για παράδειγμα, αν μετά το τελευταίο στρώμα pooling έχουμε ένα διάνυσμα 4096 χαρακτηριστικών και το πλήρως συνδεδεμένο στρώμα περιέχει 1024 νευρώνες, τότε ο συνολικός αριθμός παραμέτρων είναι:

Αυτό αυξάνει σημαντικά τον κίνδυνο υπερεκπαίδευσης (overfitting) και τον χρόνο υπολογισμού.

Για την αντιμετώπιση των παραπάνω προβλημάτων, συνήθως εφαρμόζονται τεχνικές βελτιστοποίησης όπως Dropout, Κανονικοποίηση για τις οποίες θα μιλήσουμε πιο αναλυτικά παρακάτω.

Τα FC στρώματα αποτελούν κρίσιμο τμήμα των CNNs, καθώς συνδέουν τα εξαγόμενα χαρακτηριστικά με την τελική έξοδο, αλλά η χρήση τους πρέπει να είναι προσεκτικά σχεδιασμένη ώστε να διατηρείται η ισορροπία μεταξύ πολυπλοκότητας και απόδοσης.

5.4 Επιλογή Αρχιτεκτονικής CNN

Για την υλοποίηση της παρούσας εργασίας δοκιμάστηκαν δύο διαφορετικές αρχιτεκτονικές συνελικτικών νευρωνικών δικτύων (CNNs), το VGG16 και το Xception. Η επιλογή των συγκεκριμένων μοντέλων δεν ήταν τυχαία, καθώς πρόκειται για δύο από τις πιο ευρέως χρησιμοποιούμενες αρχιτεκτονικές σε προβλήματα ταξινόμησης ιατρικών εικόνων, που συνδυάζουν υψηλή ακρίβεια και υποστήριξη μέσω προκατασκευασμένων βαρών (pre-trained weights) από το σύνολο δεδομένων ImageNet. Αυτό επιτρέπει την εφαρμογή της τεχνικής μεταφοράς μάθησης (transfer learning, βλέπε 5.8), ώστε το μοντέλο να επωφελείται από ήδη ανεπτυγμένα χαρακτηριστικά και να προσαρμόζεται γρηγορότερα στα δεδομένα των ακτινογραφιών θώρακος.

Επιπλέον, η επιλογή του VGG16 δικαιολογείται από την απλότητα και την καθιερωμένη χρήση του στη βιβλιογραφία, γεγονός που το καθιστά ιδανικό ως baseline μοντέλο για συγκρίσεις. Από την άλλη πλευρά, το Xception αντιπροσωπεύει μια πιο σύγχρονη και πολύπλοκη αρχιτεκτονική, η οποία εισάγει τη χρήση των depthwise separable convolutions, μειώνοντας τον αριθμό παραμέτρων χωρίς να θυσιάζει την ακρίβεια. Με αυτόν τον τρόπο, καλύπτεται ένα ευρύ φάσμα: από μια κλασική και αξιόπιστη λύση (VGG16) έως μια βελτιστοποιημένη και υπολογιστικά αποδοτική προσέγγιση (Xception).

Η σύγκριση των δύο μοντέλων επιτρέπει την αξιολόγηση των πλεονεκτημάτων και των περιορισμών τους σε σχέση με την ταξινόμηση πνευμονίας, προσφέροντας σαφή εικόνα για το ποια αρχιτεκτονική ανταποκρίνεται καλύτερα στις απαιτήσεις της παρούσας εργασίας.

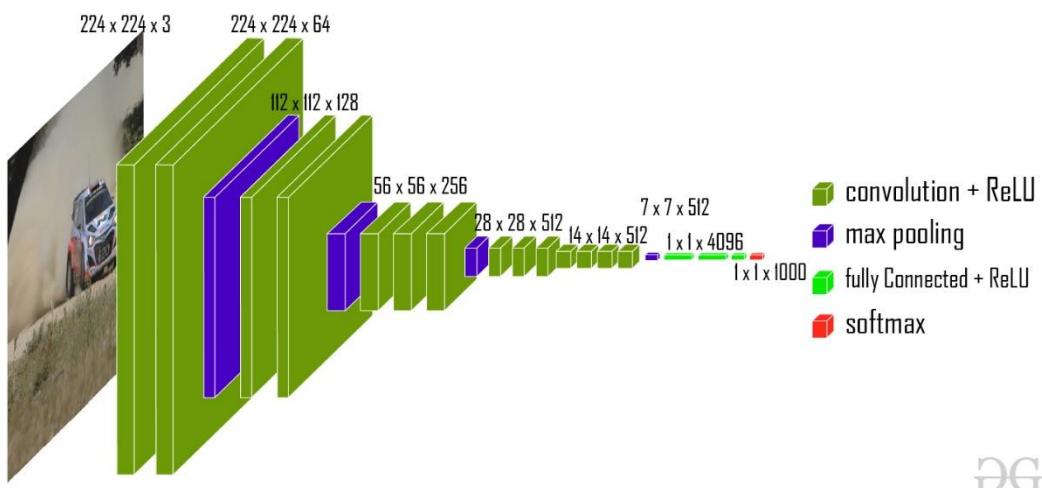
5.5 VGG16: Δομή και χαρακτηριστικά

Η αρχιτεκτονική VGG16 αποτελεί ένα από τα πιο διαδεδομένα συνελικτικά νευρωνικά δίκτυα λόγω της απλότητας και της αποδοτικότητάς της στην εξαγωγή χαρακτηριστικών από εικόνες. Το δίκτυο αυτό προτάθηκε από τους K. Simonyan και A. Zisserman το 2014 και βασίστηκε στην ιδέα χρήσης πολλών διαδοχικών συνελικτικών στρωμάτων με μικρά φίλτρα 3x3 για τη σταδιακή εξαγωγή ιεραρχικών χαρακτηριστικών [68].

Η VGG16 περιλαμβάνει συνολικά 16 στρώματα με παραμέτρους, εκ των οποίων τα 13 είναι συνελικτικά στρώματα και τα 3 πλήρως συνδεδεμένα στρώματα στο τέλος. Η δομή της είναι οργανωμένη σε πέντε συνελικτικά μπλοκ, καθένα από τα οποία περιέχει δύο ή τρία συνελικτικά στρώματα ακολουθούμενα από ένα στρώμα max pooling 2x2. Το βήμα του pooling είναι 2, μειώνοντας κατά το ήμισυ τις διαστάσεις των χαρακτηριστικών μετά από κάθε μπλοκ.

Η αρχιτεκτονική των συνελικτικών μπλοκ της VGG16 αναλύεται παρακάτω:

- Μπλοκ 1: 2 συνελικτικά στρώματα (3x3) με 64 φίλτρα το καθένα, ακολουθούμενα από max pooling.
- Μπλοκ 2: 2 συνελικτικά στρώματα (3x3) με 128 φίλτρα το καθένα, ακολουθούμενα από max pooling.
- Μπλοκ 3: 3 συνελικτικά στρώματα (3x3) με 256 φίλτρα, ακολουθούμενα από max pooling.
- Μπλοκ 4: 3 συνελικτικά στρώματα (3x3) με 512 φίλτρα, ακολουθούμενα από max pooling.
- Μπλοκ 5: 3 συνελικτικά στρώματα (3x3) με 512 φίλτρα, ακολουθούμενα από max pooling.



Εικόνα 9: Γραφική απεικόνιση της αρχιτεκτονικής του αλγορίθμου VGG-16 ([VGG-16 | CNN model | GeeksforGeeks](#))

Με αυτή τη διάταξη, η VGG16 επιτυγχάνει σταδιακή αύξηση του αριθμού των χαρακτηριστικών (channels) καθώς προχωράμε σε βαθύτερα στρώματα, ενώ ταυτόχρονα μειώνονται οι χωρικές διαστάσεις των feature maps μέσω του pooling. Αυτή η δομή επιτρέπει στο δίκτυο να εξάγει πρώτα βασικά μοτίβα όπως ακμές και γωνίες στα πρώτα μπλοκ και πιο σύνθετα χαρακτηριστικά στα βαθύτερα μπλοκ [71].

Η ομοιομορφία της VGG16 διευκολύνει την επέκταση και την τροποποίηση του δικτύου για διάφορες εργασίες, ενώ η χρήση αποκλειστικά μικρών φίλτρων 3x3 μειώνει σημαντικά τον αριθμό των παραμέτρων συγκριτικά με μεγάλα φίλτρα, χωρίς να θυσιάζεται η ικανότητα αναπαράστασης των χαρακτηριστικών [72].

Το VGG16 σχεδιάστηκε για να επιδείξει υψηλή ακρίβεια σε μεγάλα σύνολα δεδομένων εικόνων, όπως το ImageNet. Το ImageNet (14+ εκατομμύρια εικόνες, >20.000 κατηγορίες) είναι μια τεράστια βάση δεδομένων εικόνων, η οποία περιλαμβάνει εκατομμύρια εικόνες κατηγοριοποιημένες σε χλιάδες κατηγορίες αντικειμένων. Κάθε χρόνο διεξάγεται ο διαγωνισμός ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (εκεί χρησιμοποιείται ένα υποσύνολο με ~1,2 εκατ. εικόνες, 1000 κατηγορίες), όπου τα δίκτυα CNN αξιολογούνται στην ικανότητά τους να ταξινομούν και να εντοπίζουν αντικείμενα [73].

Τα προεκπαίδευμένα βάρη (pre-trained weights) του VGG16 που διατίθενται από βιβλιοθήκες όπως το Keras και το TensorFlow βασίζονται ακριβώς σε αυτή την εκπαίδευση στο ImageNet, επιτρέποντας τη χρήση τους ως σημείο εκκίνησης για νέα προβλήματα μέσω της τεχνικής μεταφοράς μάθησης (transfer learning). Η συγκεκριμένη τεχνική αξιοποιείται ευρέως σε εφαρμογές όπου τα δεδομένα εκπαίδευσης είναι περιορισμένα, καθώς μειώνει τον χρόνο εκπαίδευσης και βελτιώνει την απόδοση των μοντέλων. Στο πλαίσιο της παρούσας εργασίας, θα ακολουθήσουμε την ίδια προσέγγιση, χρησιμοποιώντας τα προεκπαίδευμένα βάρη ως αφετηρία για την ανάπτυξη του μοντέλου μας.

Το VGG16 ήταν ένα από τα δίκτυα που απέδειξαν την αποτελεσματικότητα της βαθιάς μάθησης στην αναγνώριση εικόνας, σημειώνοντας εξαιρετικές επιδόσεις στον διαγωνισμό.

Η ερευνητική σημασία της VGG16 είναι πολλαπλή. Αρχικά, λειτουργεί ως βασικό μοντέλο αναφοράς, καθώς χρησιμοποιείται συχνά ως σημείο εκκίνησης (baseline) για την αξιολόγηση νέων αρχιτεκτονικών αλλά και για την εφαρμογή τεχνικών μεταφοράς μάθησης (transfer learning). Επιπλέον, η δομή της επιτρέπει την κατανόηση της ιεραρχικής εξέλιξης των χαρακτηριστικών που ανιχνεύει το δίκτυο, ξεκινώντας από απλά μοτίβα, όπως γραμμές και ακμές, και καταλήγοντας σε πιο σύνθετες αναπαραστάσεις. Τέλος, η απλή και ομοιόμορφη αρχιτεκτονική της διευκολύνει την προσαρμογή και την επέκταση σε διαφορετικά προβλήματα, τόσο στην αναγνώριση εικόνας γενικής χρήσης όσο και σε εφαρμογές της ιατρικής διάγνωσης [71].

Η VGG16 συνέβαλε στην καθιέρωση της βαθιάς μάθησης ως κυρίαρχης προσέγγισης στην επεξεργασία εικόνας και αποτέλεσε πρότυπο για τη δημιουργία ακόμη πιο σύνθετων και βαθιών δικτύων, όπως τα ResNet και DenseNet.

5.6 Xception: Δομή και χαρακτηριστικά

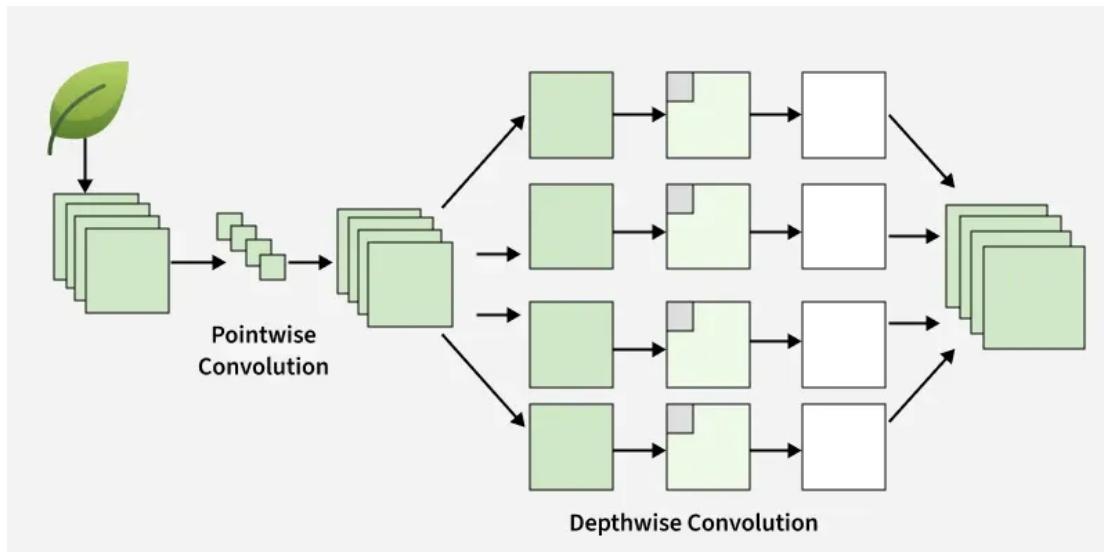
Η αρχιτεκτονική Xception αποτελεί μία από τις πιο σύγχρονες και αποδοτικές προσεγγίσεις συνελικτικών νευρωνικών δικτύων (CNNs), βασισμένη στην ιδέα της απόλυτης αποδοτικότητας των χωρικών και κανάλι-ειδικών χαρακτηριστικών μέσω της τεχνικής depthwise separable convolution. Το δίκτυο αυτό προτάθηκε από τον François Chollet το 2017 και αποτελεί ουσιαστικά μία επέκταση της αρχιτεκτονικής Inception, αντικαθιστώντας τα παραδοσιακά inception modules με χωριστές συνελικτικές λειτουργίες βάθους (depthwise) και σημείου (pointwise) [60].

Το Inception εισήχθη από την ομάδα του Szegedy το 2014 (γνωστή και ως GoogLeNet), στηρίχθηκε στην ιδέα ότι διαφορετικά μεγέθη συνελικτικών φίλτρων (π.χ. 1×1 , 3×3 , 5×5) μπορούν να εφαρμοστούν παράλληλα στο ίδιο επίπεδο, επιτρέποντας στο δίκτυο να “αποφασίζει” ποια κλίμακα χαρακτηριστικών είναι πιο χρήσιμη. Τα λεγόμενα *Inception modules* συνδύαζαν αυτές τις διαφορετικές συνελίξεις μαζί με max pooling, παρέχοντας μεγαλύτερη εκφραστικότητα στο μοντέλο χωρίς εκρηκτική αύξηση παραμέτρων. Μια κρίσιμη βελτίωση ήταν η χρήση των 1×1 συνελίξεων ως “bottlenecks”, ώστε να μειώνεται η διαστατικότητα των χαρακτηριστικών πριν από πιο κοστοβόρες συνελίξεις.

Το μοντέλο Xception επεκτείνει αυτή την ιδέα, αθώντας τη λογική του Inception στο «απόλυτο όριο»: αντί να ομαδοποιεί φίλτρα διαφορετικών μεγεθών, διαχωρίζει πλήρως την επεξεργασία χωρικών συσχετίσεων από τις συσχετίσεις μεταξύ καναλιών. Το Xception περιλαμβάνει περισσότερα από 36 συνελικτικά στρώματα οργανωμένα σε 14 κύρια modules, τα οποία ακολουθούνται από πλήρως συνδεδεμένα στρώματα στο τέλος για την τελική ταξινόμηση. Η αρχιτεκτονική της βασίζεται στην αρχή ότι η χωρική εξαγωγή χαρακτηριστικών (spatial correlations) και η ανακατανομή χαρακτηριστικών μεταξύ καναλιών (cross-channel correlations) μπορούν να εκπαιδευτούν ξεχωριστά, γεγονός που μειώνει σημαντικά τον αριθμό παραμέτρων και αυξάνει την αποδοτικότητα του μοντέλου.

Στο πρώτο στάδιο, τη depthwise convolution, κάθε κανάλι εισόδου υποβάλλεται σε ξεχωριστή συνελικτική επεξεργασία· δηλαδή, εφαρμόζεται ένα φίλτρο ανά κανάλι χωρίς άμεσο συνδυασμό μεταξύ των καναλιών. Αυτή η προσέγγιση επιτρέπει στο δίκτυο να συλλαμβάνει τοπικά, χωρικά μοτίβα μέσα σε κάθε κανάλι με μειωμένο υπολογιστικό κόστος σε σχέση με την κλασική, πλήρη συνέλιξη. Η ανεξαρτησία της επεξεργασίας ανά κανάλι σημαίνει επίσης λιγότερες παραμέτρους και ταχύτερη εκτέλεση, παράλληλα με ικανότητα εντοπισμού βασικών οπτικών χαρακτηριστικών.

Στη συνέχεια, η pointwise convolution (συνέλιξη 1×1) αναλαμβάνει να συνδυάσει και να αναδιανείμει την πληροφορία που εξήχθη από τα επιμέρους κανάλια, δημιουργώντας πλουσιότερες και πολυδιάστατες αναπαραστάσεις. Σε αυτό το στάδιο επιτυγχάνεται η ενσωμάτωση των χαρακτηριστικών σε επίπεδο καναλιών, κάτι που επιτρέπει την ανίχνευση πιο σύνθετων μοτίβων. Μετά από κάθε συνελικτική λειτουργία εφαρμόζονται τεχνικές όπως *batch normalization* για τη σταθεροποίηση της κατανομής των ενεργοποιήσεων και *ReLU* ως μη γραμμική συνάρτηση ενεργοποίησης, οι οποίες συμβάλλουν στη σταθερότερη και ταχύτερη σύγκλιση κατά την εκπαίδευση. Ο συνδυασμός depthwise + pointwise καθιστά την Xception μια αποδοτική αρχιτεκτονική που συνδυάζει υψηλή απόδοση με μικρότερο υπολογιστικό και χωρικό αποτύπωμα σε σχέση με παραδοσιακά CNNs.

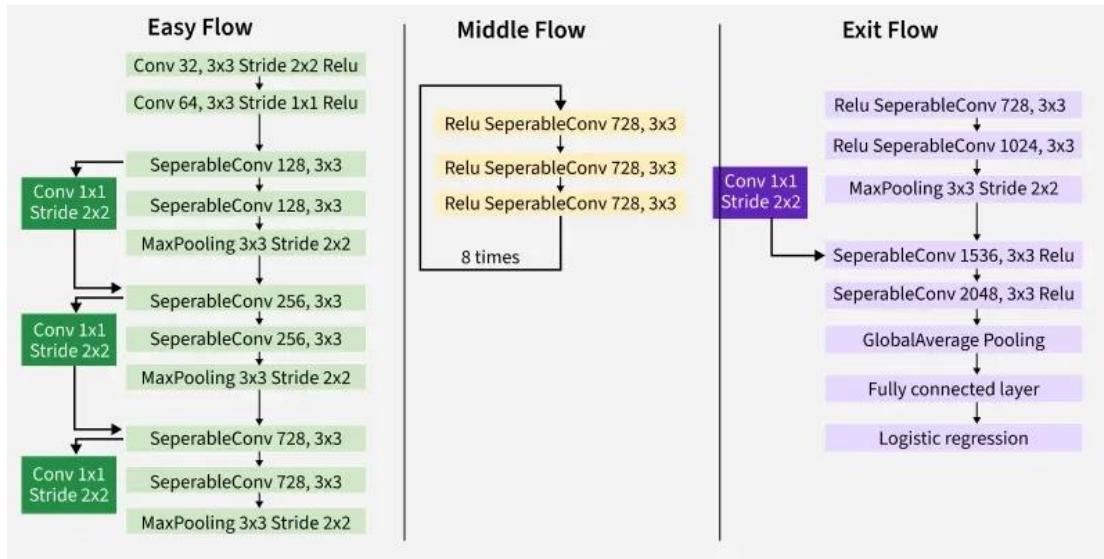


Εικόνα 10: Η φιλοσοφία πίσω από την αρχιτεκτονική του Xception ([Xception - GeeksforGeeks](#))

Η αρχιτεκτονική των συνελικτικών μπλοκ της Xception αναλύεται ως εξής:

- **Είσοδος & Αρχικά Στρώματα:**
 - 1 κανονικό συνελικτικό στρώμα (32 φίλτρα, 3x3)
 - 1 κανονικό συνελικτικό στρώμα (64 φίλτρα, 3x3)
 - Ακολουθεί max pooling.
- **Entry Flow (Μπλοκ 1–3):**
 - Μπλοκ 1: Depthwise separable conv (128 φίλτρα), batch norm, ReLU → max pooling.
 - Μπλοκ 2: Depthwise separable conv (256 φίλτρα), batch norm, ReLU → max pooling.
 - Μπλοκ 3: Depthwise separable conv (728 φίλτρα), batch norm, ReLU → max pooling.

Κάθε μπλοκ περιλαμβάνει επίσης residual connections.
- **Middle Flow (Μπλοκ 4–11):**
 - 8 επαναλήψεις του ίδιου μοτίβου: depthwise separable conv (728 φίλτρα, τριπλό stack), batch norm, ReLU.
 - Κάθε επανάληψη έχει skip connection.
- **Exit Flow (Μπλοκ 12–14):**
 - Μπλοκ 12: Depthwise separable conv (728 → 1024 φίλτρα), batch norm, ReLU → max pooling.
 - Μπλοκ 13: Depthwise separable conv (1536 φίλτρα), batch norm, ReLU.
 - Μπλοκ 14: Depthwise separable conv (2048 φίλτρα), batch norm, ReLU.
- **Τελικά Στρώματα:**
 - Global average pooling.
 - Fully connected layer με softmax για ταξινόμηση.



Εικόνα 11: Γραφική απεικόνιση της αρχιτεκτονικής του αλγορίθμου Xception ([Xception - GeeksforGeeks](#))

Η Xception επιτυγχάνει υψηλή απόδοση σε μεγάλα σύνολα δεδομένων, όπως το ImageNet (ILSVRC), όπου αξιολογούνται δίκτυα CNN στην ταξινόμηση και ανίχνευση αντικειμένων (βλέπε 5.4). Τα προεκπαιδευμένα βάροι του Xception που διατίθενται σε βιβλιοθήκες όπως το Keras και το TensorFlow έχουν εκπαιδευτεί ακριβώς σε αυτό το σύνολο δεδομένων, επιτρέποντας τη χρήση τους ως σημείο εκκίνησης για νέα προβλήματα μέσω μεταφοράς μάθησης (transfer learning) όμοια με το VGG16. Η χρήση των προεκπαιδευμένων βαρών μειώνει σημαντικά τον χρόνο εκπαίδευσης και βελτιώνει την απόδοση των μοντέλων σε εφαρμογές με περιορισμένα δεδομένα [68].

Η ερευνητική σημασία της Xception είναι πολυδιάστατη. Η χωριστή επεξεργασία χαρακτηριστικών ανά κανάλι έχει δείξει ότι μπορεί να ξεπεράσει παραδοσιακές αρχιτεκτονικές Inception, παρέχοντας μεγαλύτερη αποδοτικότητα και μικρότερο αριθμό παραμέτρων. Η αρχιτεκτονική της διευκολύνει την κατανόηση της ιεραρχικής εξέλιξης των χαρακτηριστικών, από απλά χωρικά μοτίβα έως πολύπλοκες αναπαραστάσεις, ενώ αποτελεί ένα ισχυρό εργαλείο για εφαρμογές γενικής χρήσης στην αναγνώριση εικόνας και στην ιατρική διάγνωση [71].

Η Xception έχει καθιερωθεί ως πρότυπο μοντέλο αναφοράς και χρησιμοποιείται ευρέως ως βάση για πιο εξειδικευμένα ή βαθύτερα δίκτυα, προσφέροντας παράλληλα βελτιώσεις στην απόδοση και την αποδοτικότητα συγκριτικά με παραδοσιακά CNNs, όπως το VGG16 και τα Inception δίκτυα.

5.7 Τεχνικές βελτίωσης απόδοσης (Dropout, Regularization, Batch Normalization)

Η απόδοση των συνελικτικών νευρωνικών δικτύων μπορεί να επηρεαστεί αρνητικά από φαινόμενα όπως η υπερεκπαίδευση (overfitting), όπου το δίκτυο μαθαίνει υπερβολικά καλά τα δεδομένα εκπαίδευσης χάνοντας τη δυνατότητα γενίκευσης, ή από την αστάθεια κατά την εκπαίδευση, η οποία συχνά οφείλεται σε μεγάλες διακυμάνσεις των βαρών και των βαθμίδων. Για την αντιμετώπιση αυτών των προκλήσεων έχουν αναπτυχθεί διάφορες τεχνικές βελτιστοποίησης της μάθησης τις οποίες και θα αναλύσουμε στην συνέχεια.

Dropout:

Το Dropout είναι μια τεχνική που εφαρμόζεται κατά τη διάρκεια της εκπαίδευσης, όπου τυχαία "απενεργοποιείται" ένα ποσοστό των νευρώνων (π.χ 50%) σε κάθε στρώμα σε κάθε βήμα εκπαίδευσης. Μειώνοντας έτσι την εξάρτηση μεταξύ των μονάδων (το δίκτυο δεν βασίζεται σε συγκεκριμένους νευρώνες) και αναγκάζοντας το δίκτυο να μάθει πιο ανθεκτικές και γενικεύσιμες αναπαραστάσεις. Με αυτόν τον τρόπο αποτρέπεται η υπερεξάρτηση των νευρώνων μεταξύ τους και μειώνεται η πιθανότητα υπερεκπαίδευσης. Η τεχνική αυτή ενισχύει τη γενίκευση του μοντέλου, επιτρέποντας ταυτόχρονα τη διατήρηση της απόδοσης κατά την αξιολόγηση σε νέα δεδομένα [25].

Regularization:

Η κανονικοποίηση εισάγει ένα πρόσθετο όρο στο κόστος εκπαίδευσης, που τιμωρεί μεγάλες τιμές βαρών. Οι πιο συνηθισμένες μορφές είναι η L1 (sum of absolute weights) και η L2 (sum of squared weights). Η χρήση της L2 regularization περιορίζει την αύξηση των βαρών, με αποτέλεσμα ένα πιο σταθερό και γενικεύσιμο μοντέλο, μειώνοντας την πιθανότητα υπερεκπαίδευσης [74].

Batch Normalization:

Το Batch Normalization εφαρμόζεται για να σταθεροποιεί τις εισόδους κάθε στρώματος κατά την εκπαίδευση, κανονικοποιώντας τα δεδομένα με βάση τη μέση τιμή και την τυπική απόκλιση του batch. Η τεχνική αυτή επιτρέπει μεγαλύτερα learning rates, επιταχύνει τη σύγκλιση και μειώνει την ευαισθησία σε αρχικές τυχαίες τιμές των παραμέτρων. Επιπλέον, έχει παρατηρηθεί ότι λειτουργεί ως ένα είδος regularization, βελτιώνοντας τη γενίκευση του μοντέλου.

Συνδυάζοντας αυτές τις τεχνικές, τα συνελικτικά δίκτυα μπορούν να επιτύχουν υψηλότερη απόδοση και μεγαλύτερη σταθερότητα κατά την εκπαίδευση, ενώ μειώνεται η πιθανότητα υπερεκπαίδευσης και βελτιώνεται η γενίκευση σε νέα δεδομένα [75].

5.8 Transfer Learning και Fine-tuning

Η τεχνική της μεταφοράς μάθησης (Transfer Learning) αποτελεί μία από τις πλέον αποτελεσματικές προσεγγίσεις στην εκπαίδευση συνελικτικών νευρωνικών δικτύων, ιδιαίτερα όταν τα δεδομένα εκπαίδευσης είναι περιορισμένα. Η βασική ιδέα είναι η αξιοποίηση γνώσεων που έχουν ήδη αποκτηθεί από ένα δίκτυο εκπαίδευμένο σε μεγάλο και γενικό σύνολο δεδομένων, όπως το ImageNet, για να επιταχυνθεί η εκπαίδευση και να βελτιωθεί η απόδοση σε μια νέα εργασία [76].

Στην πράξη, αυτό επιτυγχάνεται χρησιμοποιώντας προεκπαίδευμένα βάρη (pre-trained weights) και προσαρμόζοντάς τα στο νέο πρόβλημα. Η διαδικασία μπορεί να περιλαμβάνει δύο βασικές στρατηγικές:

Feature Extraction:

Στην συγκεκριμένη προσέγγιση, τα προεκπαίδευμένα βάρη χρησιμοποιούνται ως σταθερά χαρακτηριστικά (feature extractor). Τα πρώτα στρώματα του δικτύου διατηρούνται ανέπαφα, ενώ τα τελευταία πλήρως συνδεδεμένα στρώματα αντικαθίστανται ώστε να ταιριάζουν στον αριθμό κατηγοριών της νέας εργασίας. Η εκπαίδευση επικεντρώνεται στα νέα στρώματα, μειώνοντας τον χρόνο εκπαίδευσης και την πιθανότητα υπερεκπαίδευσης [77].

Fine-tuning:

Σε αυτήν την περίπτωση, ένα μέρος ή ολόκληρο το προεκπαίδευμένο δίκτυο επιτρέπει την εκπαίδευση (backpropagation) με νέο dataset, προσαρμόζοντας τα βάρη ώστε να βελτιστοποιηθεί η απόδοση για την ειδική εργασία. Η μέθοδος αυτή είναι πιο ευέλικτη και μπορεί να οδηγήσει σε υψηλότερη ακρίβεια, ειδικά όταν το νέο dataset είναι μεγάλο ή οι απαιτήσεις ταξινόμησης διαφέρουν σημαντικά από το αρχικό σύνολο δεδομένων [78].

Η χρήση μεταφοράς μάθησης και fine-tuning έχει εφαρμογές σε πολλούς τομείς, όπως αναγνώριση εικόνας, ιατρική διάγνωση, επεξεργασία φυσικής γλώσσας και ανάλυση βίντεο. Παράλληλα, μειώνει σημαντικά τον απαιτούμενο χρόνο εκπαίδευσης, περιορίζει τις ανάγκες σε μεγάλο όγκο δεδομένων και αυξάνει τη γενικευσιμότητα του μοντέλου [68].

Στο πλαίσιο της παρούσας υλοποίησης, η μεταφορά μάθησης χρησιμοποιείται για να αξιοποιηθούν τα προεκπαίδευμένα βάρη των VGG16 και Xception μοντέλων, ενώ η προσαρμογή τους μέσω fine-tuning επιτρέπει τη βελτιστοποίηση της απόδοσης στις συγκεκριμένες εικόνες του dataset.

6. Υλοποίηση και εκπαίδευση των δικτυών

6.1 Περιβάλλον Ανάπτυξης (Hardware & Software)

Η ανάπτυξη και εκπαίδευση του μοντέλου πραγματοποιήθηκε σε απομακρυσμένο υπολογιστικό περιβάλλον υψηλών επιδόσεων, το οποίο παρέχεται από το υπολογιστικό κέντρο της σχολής των Μηχανικών Υπολογιστών Πάτρας. Η επιλογή ενός τέτοιου περιβάλλοντος κρίθηκε απαραίτητη λόγω των απαιτήσεων σε υπολογιστική ισχύ και μνήμη που συνοδεύουν την εκπαίδευση βαθιών νευρωνικών δικτύων στις οποίες ένα μηχάνημα προσωπικής χρήσης δεν θα μπορούσε εύκολα να ανταπεξέλθει.

Αναγκαίοτητα Χρήσης Υπολογιστικού Συστήματος Υψηλών Επιδόσεων

Η εκπαίδευση ενός μοντέλου βαθιάς μάθησης δεν είναι απλώς χρονοβόρα αλλά και εξαιρετικά απαιτητική σε πόρους. Ενώ ένας προσωπικός υπολογιστής μπορεί να διαθέτει πολυπύρηνο επεξεργαστή και μια σύγχρονη κάρτα γραφικών, συνήθως δεν επαρκεί για την εκπαίδευση σύνθετων αρχιτεκτονικών σε ρεαλιστικούς χρόνους. Η διαδικασία εκπαίδευσης περιλαμβάνει εκατομμύρια παραμέτρους πολλαπλές εποχές (epochs) που πρέπει να υπολογιστούν και να ενημερωθούν, γεγονός που απαιτεί υψηλή υπολογιστική ισχύ και επαρκή μνήμη.

Η επιλογή ενός υπολογιστικού συστήματος υψηλών επιδόσεων, όπως ο πανεπιστημιακός εξυπηρετητής(server) με ισχυρή CPU και GPU, εξασφαλίζει:

- Μείωση χρόνου εκπαίδευσης μέσω παράλληλης επεξεργασίας.
- Διαχείριση μεγάλων όγκων δεδομένων χωρίς προβλήματα μνήμης.
- Εκτέλεση σύνθετων πειραμάτων που δεν θα ήταν εφικτά σε προσωπικό υπολογιστή.

Χαρακτηριστικά Υλικού και Λογισμικού

Λειτουργικό Σύστημα Ubuntu 20.04.6 LTS

Πυρήνας (Kernel) Linux 5.15.0-84-generic

Αρχιτεκτονική CPU x86_64

Επεξεργαστής Intel Xeon Silver 4116 (48 πυρήνες)

Συνολική Μνήμη RAM 125 GB

GPU NVIDIA Tesla V100 (32 GB HBM2)

CUDA Version 12.2

Δίσκος Υψηλής ταχύτητας SSD

Πίνακας 4: Χαρακτηριστικά Υλικού και Λογισμικού

Επιτάχυνση μέσω κάρτας γραφικών (GPU)

Όπως προαναφέραμε εκπαίδευση βαθιών νευρωνικών δικτύων είναι μια διαδικασία με τεράστιες υπολογιστικές απαιτήσεις, καθώς περιλαμβάνει την εκτίμηση και ενημέρωση εκατομμυρίων παραμέτρων, πιο συγκεκριμένα στην δική μας εφαρμογή το δίκτυο VGG16 περιλαμβάνει ~138 εκατομμύρια παραμέτρους, ενώ το Xception ~23 εκατομμύρια (βλέπε 5.2). Η χρήση γραφικών επεξεργαστών (GPUs) είναι καθοριστική για την επιτάχυνση αυτών των διεργασιών, καθώς οι GPUs διαθέτουν αρχιτεκτονική μαζικής παραλληλοποίησης που επιτρέπει τον ταυτόχρονο υπολογισμό μεγάλου αριθμού πράξεων. Η κάρτα γραφικών NVIDIA Tesla V100 που είναι εγκατεστημένη στο παρόν υπολογιστικό περιβάλλον, είναι ειδικά σχεδιασμένη για εφαρμογές τεχνητής νοημοσύνης και παρέχει επιτάχυνση μέσω CUDA, μειώνοντας δραστικά τον χρόνο εκπαίδευσης [79].

CUDA

Η τεχνολογία CUDA (Compute Unified Device Architecture) της NVIDIA παρέχει ένα περιβάλλον προγραμματισμού που αξιοποιεί στο έπακρο τις δυνατότητες της GPU, επιτρέποντας την εκτέλεση λειτουργιών σε πολλαπλούς πυρήνες ταυτόχρονα. Αυτό έχει ως αποτέλεσμα σημαντική μείωση του χρόνου εκπαίδευσης συγκριτικά με την αποκλειστική χρήση CPU [80].

Προβλήματα μνήμης

Ωστόσο, η μνήμη της GPU αποτελεί έναν κρίσιμο περιοριστικό παράγοντα. Για παράδειγμα, σε δίκτυα μεγάλης κλίμακας ή όταν χρησιμοποιούνται μεγάλα batch sizes, η μνήμη μπορεί να γεμίσει πλήρως, οδηγώντας σε σφάλματα Out of Memory (OOM). Σε έναν προσωπικό υπολογιστή με περιορισμένη GPU (π.χ., 4–8 GB VRAM), η εκπαίδευση τέτοιων μοντέλων γίνεται πρακτικά αδύνατη χωρίς σημαντικούς συμβιβασμούς στο μέγεθος των batch ή στο ίδιο το μοντέλο. Αντίθετα, η ύπαρξη 32 GB μνήμης GPU στον πανεπιστημιακό εξυπηρετητή καθιστά δυνατή την εκπαίδευση πολύπλοκων δικτύων χωρίς συνεχείς διακοπές ή ανάγκη τροποποίησης του αρχιτεκτονικού σχεδιασμού.

Λογισμικό server (Software)

Το λειτουργικό σύστημα του server είναι το Ubuntu 20.04 LTS, μία σταθερή και ευρέως χρησιμοποιούμενη διανομή Linux για επιστημονικές εφαρμογές. Για την υλοποίηση αξιοποιήθηκε ένα απομονωμένο περιβάλλον conda, στο οποίο εγκαταστάθηκαν τα απαραίτητα εργαλεία και βιβλιοθήκες που αναφέρθηκαν σε προηγούμενα κεφάλαια (π.χ., TensorFlow, Keras, NumPy). Το περιβάλλον αυτό εξασφαλίζει τη σωστή διαχείριση των εξαρτήσεων και τη συμβατότητα των εκδόσεων. Παράλληλα, σε συνδυασμό με το CUDA toolkit και τους αντίστοιχους NVIDIA drivers, γίνεται δυνατό να αξιοποιηθεί πλήρως η υπολογιστική ισχύς της GPU.

6.2 Εκπαίδευση μοντέλων

Στο παρόν εδάφιο θα μιλήσουμε για την εκπαίδευση των μοντέλων VGG16, Xception. Και στις δύο περιπτώσεις χρησιμοποιήθηκε η προεκπαίδευμένη εκδοχή του μοντέλου με βάρη από το ImageNet, χωρίς το τελικό πλήρως συνδεδεμένο στρώμα (include_top=False). Προστέθηκαν επιπλέον στρώματα ταξινόμησης, αποτελούμενα από Global Average Pooling, ένα πλήρως συνδεδεμένο στρώμα με 512 νευρώνες (ReLU ενεργοποίηση) και Dropout 0.5 για μείωση του overfitting, καθώς και ένα τελικό dense στρώμα με softmax ενεργοποίηση. Για βελτίωση της προσαρμογής στο εκάστοτε σύνολο δεδομένων, κάνουμε unfreeze τα τελευταία οκτώ συνελικτικά στρώματα (βλέπε 6.3) του βασικού μοντέλου ώστε να εκπαίδευτούν εκ νέου (fine-tuning βλέπε 5.8). Η επιλογή αυτή εγινε διοτι τα ανώτερα επίπεδα των CNN είναι εκείνα που εξειδικεύονται περισσότερο στα χαρακτηριστικά του αρχικού συνόλου ImageNet. Με τον τρόπο αυτό, το δίκτυο διατηρεί τα γενικά χαρακτηριστικά χαμηλού επιπέδου (όπως ακμές και υφές), ενώ ταυτόχρονα προσαρμόζει τις πιο σύνθετες αναπαραστάσεις του στις ιδιαιτερότητες του νέου ιατρικού συνόλου δεδομένων. Το VGG16 χρησιμοποιήθηκε κυρίως ως baseline μοντέλο, ώστε να συγκριθεί η απόδοσή του με το πιο σύνθετο Xception.

6.2.1 Υπερπαράμετροι Εκπαίδευσης: Learning Rate και Adam Optimizer

Ρυθμός μάθησης (Learning Rate)

Η επιλογή του ρυθμού μάθησης (learning rate) αποτελεί μία από τις σημαντικότερες υπερπαραμέτρους στη διαδικασία εκπαίδευσης των βαθιών νευρωνικών δικτύων, καθώς επηρεάζει άμεσα τη σύγκλιση του μοντέλου και την ποιότητα των παραγόμενων αποτελεσμάτων [28]. Ένας πολύ υψηλός ρυθμός μάθησης μπορεί να οδηγήσει σε αστάθεια και αποτυχία σύγκλισης, ενώ ένας πολύ χαμηλός ρυθμός μάθησης ενδέχεται να επιβραδύνει υπερβολικά τη διαδικασία εκπαίδευσης [78]. Για τον λόγο αυτό επιλέγεται συνήθως μία ενδιάμεση τιμή, η οποία εξισορροπεί την ταχύτητα και τη σταθερότητα εκπαίδευσης. Στην παρούσα υλοποίηση επιλέχθηκε ρυθμός μάθησης ίσος με 0.0001. Στη σταδιακή προσαρμογή των βαρών κατά το fine-tuning προεκπαίδευμένων μοντέλων τα αρχικά βάρη έχουν ήδη μάθει γενικά χαρακτηριστικά. Ένα υψηλό learning rate θα μπορούσε να οδηγήσει σε απότομες μεταβολές, καταστρέφοντας τις προεκπαίδευμένες παραμέτρους. Η χαμηλή τιμή επιτρέπει ομαλή προσαρμογή στα νέα δεδομένα. Επιπρόσθετα βοηθάει στην αποφυγή τοπικών ελαχίστων και υπερβολικών ταλαντώσεων. Ένα μικρό learning rate μειώνει την πιθανότητα απότομων διακυμάνσεων της συνάρτησης κόστους, εξασφαλίζοντας σταθερή και ελεγχόμενη σύγκλιση [61].

Βελτιστοποιητής Adam

Όσον αφορά τον βελτιστοποιητή, επιλέχθηκε ο Adam (Adaptive Moment Estimation), ο οποίος συνδυάζει τα πλεονεκτήματα των μεθόδων Momentum και RMSProp, επιτυγχάνοντας γρήγορη και αποδοτική σύγκλιση [39]. Ο Adam χρησιμοποιεί εκθετικά κινούμενους μέσους όρους των πρώτων και δεύτερων ροπών των παραγώγων, γεγονός που προσφέρει προσαρμοστικό ρυθμό μάθησης για κάθε παράμετρο ξεχωριστά [81]. Το χαρακτηριστικό αυτό τον καθιστά ιδιαίτερα αποτελεσματικό σε προβλήματα βαθιάς μάθησης με μεγάλα και ετερογενή σύνολα δεδομένων, όπως οι ιατρικές εικόνες ακτινογραφίας θώρακα που χρησιμοποιούμε εδώ. Επιπλέον, απαιτεί λιγότερη ρύθμιση

υπερπαραμέτρων σε σχέση με άλλους αλγορίθμους, γεγονός που τον καθιστά δημοφιλή επιλογή για πειράματα με περιορισμένο χρόνο εκπαίδευσης [82].

Η συγκεκριμένη επιλογή (Adam με learning rate 0.0001) αποσκοπεί στην εξασφάλιση σταθερής εκπαίδευσης κατά τη διαδικασία fine-tuning, όπου είναι απαραίτητο να προσαρμόζονται οι τελευταίοι στρώματα του προεκπαίδευμένου μοντέλου χωρίς να αλλοιωθούν σημαντικά τα ήδη εκπαίδευμένα χαρακτηριστικά.

6.2.2 Επιλογή Batch Size και Επίδρασή του στην Εκπαίδευση

Το μέγεθος του batch (batch size) αποτελεί μία από τις πιο κρίσιμες υπερπαραμέτρους κατά την εκπαίδευση ενός νευρωνικού δικτύου, καθώς επηρεάζει άμεσα τόσο την ταχύτητα εκπαίδευσης όσο και τη γενίκευση του μοντέλου. Το batch size καθορίζει πόσα δείγματα δεδομένων επεξεργάζεται το δίκτυο πριν ενημερώσει τα βάρη του μέσω backpropagation [28]. Συνολικά, η επιλογή batch size εξαρτάται από παράγοντες όπως η μνήμη της GPU, η πολυπλοκότητα και το μέγεθος του μοντέλου, και η ισορροπία μεταξύ ταχύτητας εκπαίδευσης και ικανότητας γενίκευσης.

Επίδραση του Batch Size

Ένα μικρό batch size (π.χ., 8 ή 16) οδηγεί σε πιο συχνές ενημερώσεις των βαρών, εισάγοντας στοχαστικότητα στη διαδικασία εκπαίδευσης. Αυτό μπορεί να βοηθήσει το μοντέλο να αποφύγει τοπικά ελάχιστα και να επιτύχει καλύτερη γενίκευση, αν και αυξάνει τον χρόνο εκπαίδευσης λόγω περισσότερων ενημερώσεων [83]. Αντίθετα, ένα μεγάλο batch size (π.χ., 32 ή 64) επιτρέπει πιο γρήγορη εκπαίδευση ανά εποχή, καθώς εκτελείται λιγότερος αριθμός ενημερώσεων. Ωστόσο, η χρήση μεγάλων batches μπορεί να οδηγήσει σε χειρότερη γενίκευση, επειδή η διαδικασία εκπαίδευσης γίνεται πιο ντετερμινιστική [84].

Διαφορές Batch Size για Xception και VGG16

Στην παρούσα εργασία, για το δίκτυο Xception επιλέχθηκε batch size ίσο με 8. Η απόφαση αυτή βασίζεται στο γεγονός ότι η αρχιτεκτονική Xception διαθέτει μεγάλο αριθμό παραμέτρων (περίπου 22 εκατομμύρια) και υψηλή υπολογιστική απαίτηση καθώς χρησιμοποιεί depthwise separable convolutions και μεγαλύτερο μέγεθος εισόδου ($299 \times 299 \times 3$ έναντι $224 \times 224 \times 3$ στο VGG16), κάτι που αυξάνει σημαντικά τον αριθμό των υπολογισμών και τη χρήση μνήμης GPU κατά την εκπαίδευση. Έτσι, η μείωση του batch size σε 8 εξασφαλίζει ότι η μνήμη της GPU (Tesla V100 32GB) δεν υπερφορτώνεται, επιτρέποντας ταυτόχρονα την επιτυχή ολοκλήρωση της διαδικασίας εκπαίδευσης χωρίς σφάλματα μνήμης (out of memory errors) [79] ιδιαίτερα όταν χρησιμοποιούνται τεχνικές όπως fine-tuning, που απαιτούν την εκπαίδευση επιπλέον παραμέτρων στα ξεπαγωμένα στρώματα.

Αντίθετα, στο δίκτυο VGG16 επιλέχθηκε batch size ίσο με 32, καθώς η αρχιτεκτονική αυτή, αν και διαθέτει επίσης μεγάλο αριθμό παραμέτρων (περίπου 138 εκατομμύρια), απαιτεί λιγότερη μνήμη GPU ανά δείγμα λόγω της διαφορετικής διάρθρωσης των layers. Έτσι, η χρήση μεγαλύτερου batch size είναι εφικτή, επιταχύνοντας την εκπαίδευση και αξιοποιώντας καλύτερα τους διαθέσιμους πόρους [68].

6.3 Προσαρμογή Τελικής Κεφαλής μέσω Μεταφοράς Μάθησης

Custom Head Ανάλυση

Η ενσωμάτωση ενός custom head σε ένα προεκπαιδευμένο δίκτυο, όπως το Xception ή το VGG16, βοηθάει στη προσαρμογή του μοντέλου σε ένα νέο πρόβλημα ταξινόμησης [68]. Στην πράξη, το προεκπαιδευμένο δίκτυο λειτουργεί ως εξαγωγέας χαρακτηριστικών (feature extractor), δηλαδή αναλαμβάνει να μάθει και να απομονώσει τα γενικά οπτικά μοτίβα που έχουν ήδη ενσωματωθεί μέσω της εκπαίδευσής του στο ImageNet. Αντί να εκπαιδεύεται εκ νέου όλο το δίκτυο από το μηδέν –μια διαδικασία ιδιαίτερα χρονοβόρα και απαιτητική σε δεδομένα– διατηρούμε το βασικό μοντέλο «παγωμένο» ή μερικώς προσαρμόσιμο και αντικαθιστούμε τα τελευταία επίπεδα με ένα custom head που προσαρμόζεται ειδικά στις ανάγκες του εκάστοτε dataset.

Με αυτό τον τρόπο, το προεκπαιδευμένο δίκτυο προσφέρει ένα ισχυρό υπόβαθρο γνώσης, ενώ η νέα κεφαλή (head) αναλαμβάνει την τελική ταξινόμηση, λαμβάνοντας υπόψη τις ιδιαιτερότητες του νέου προβλήματος. Η διαδικασία αυτή ανήκει στη γνωστή στρατηγική της μεταφοράς μάθησης (transfer learning), η οποία επιτρέπει την αξιοποίηση έτοιμων αναπαραστάσεων και μειώνει σημαντικά τον κίνδυνο overfitting, ειδικά όταν το διαθέσιμο dataset δεν είναι ιδιαίτερα μεγάλο.

Τα βασικά στοιχεία αυτής της κεφαλής είναι:

Global Average Pooling (GAP)

Η τεχνική του Global Average Pooling (GAP) χρησιμοποιείται ευρέως σε σύγχρονα συνελικτικά νευρωνικά δίκτυα για να αντικαταστήσει τα πλήρως συνδεδεμένα (fully connected) layers που συναντώνται σε παλαιότερες αρχιτεκτονικές [85]. Ο κύριος λόγος για αυτήν την επιλογή είναι η σημαντική μείωση του αριθμού των παραμέτρων του δικτύου, γεγονός που οδηγεί σε μικρότερο υπολογιστικό κόστος και σε μειωμένο κίνδυνο υπερεκπαίδευσης (overfitting).

Συγκεκριμένα, ενώ ένα πλήρως συνδεδεμένο layer θα «ισιώσει» (flatten) το tensor σε ένα μεγάλο διάνυσμα, εισάγοντας έτσι εκατοντάδες χιλιάδες παραμέτρους, το GAP εφαρμόζει έναν απλό μέσο όρο σε κάθε feature map ξεχωριστά. Αυτό σημαίνει ότι κάθε feature map, ανεξαρτήτως μεγέθους, συμπυκνώνεται σε μία και μοναδική τιμή, η οποία αντιπροσωπεύει την ένταση ενεργοποίησης του συγκεκριμένου χαρακτηριστικού στο σύνολο της εικόνας.

Το αποτέλεσμα είναι ένα πολύ πιο συμπαγές και αποδοτικό διάνυσμα χαρακτηριστικών, το οποίο τροφοδοτείται στα επόμενα επίπεδα του δικτύου. Με τον τρόπο αυτό επιτυγχάνεται όχι μόνο εξοικονόμηση πόρων, αλλά και βελτίωση της γενίκευσης του μοντέλου, αφού το GAP ενθαρρύνει το δίκτυο να εστιάζει σε ολόκληρες περιοχές της εικόνας αντί σε μεμονωμένα τοπικά χαρακτηριστικά, καθιστώντας το πιο ανθεκτικό σε παραλλαγές και θόρυβο.

Dense Layer με ReLU και L2 Regularization

Το πρώτο πλήρως συνδεδεμένο επίπεδο του custom head αποτελείται από 512 νευρώνες και αξιοποιεί τη συνάρτηση ενεργοποίησης ReLU (Rectified Linear Unit) [38]. Ο αριθμός των νευρώνων δεν είναι τυχαιος αλλα είναι μια επιλογη που προσφέρει ικανή υπολογιστική ισχύ για την εκμάθηση σύνθετων αναπαραστάσεων, ενώ ταυτόχρονα διατηρεί τον συνολικό αριθμό παραμέτρων σε επίπεδο που περιορίζει το overfitting και εξασφαλίζει αποδοτικούς χρόνους εκπαίδευσης [28]. Η ReLU θεωρείται ιδιαίτερα αποτελεσματική στα συνελικτικά νευρωνικά δίκτυα, καθώς εισάγει μη γραμμικότητα με έναν απλό και υπολογιστικά αποδοτικό τρόπο. Σε αντίθεση με συναρτήσεις όπως η sigmoid ή η tanh, η ReLU δεν υποφέρει έντονα από το πρόβλημα του vanishing gradient, γεγονός που επιτρέπει ταχύτερη και πιο σταθερή εκπαίδευση σε βαθιά δίκτυα. Επιπλέον, η χρήση της ReLU προάγει τη σπανιότητα (sparsity) στην ενεργοποίηση των νευρώνων, βελτιώνοντας έτσι τη δυνατότητα γενίκευσης του μοντέλου.

Παράλληλα, εφαρμόζεται L2 regularization (γνωστό και ως weight decay) [86] στα βάρη του επιπέδου. Το L2 regularization επιβάλλει ποινή ανάλογη με το τετράγωνο του μεγέθους των βαρών, αποθαρρύνοντας την ανάπτυξη υπερβολικά μεγάλων τιμών που μπορεί να οδηγήσουν σε overfitting. Με αυτόν τον τρόπο, ενισχύεται η σταθερότητα του μοντέλου και εξασφαλίζεται ότι τα χαρακτηριστικά που μαθαίνει δεν είναι υπερβολικά προσαρμοσμένα στο εκπαιδευτικό σύνολο. Ο συνδυασμός ReLU και L2 regularization μας δίνει μια πιο αποδοτική εκμάθηση αναπαραστάσεων, ισορροπώντας την εκφραστικότητα του δικτύου με την ανάγκη για καλή γενίκευση.

Τελικό Dense Layer με Softmax

Το τελικό πλήρως συνδεδεμένο επίπεδο του μοντέλου είναι υπεύθυνο για την τελική διαδικασία ταξινόμησης. Ο αριθμός των νευρώνων σε αυτό το επίπεδο αντιστοιχεί άμεσα στον αριθμό των κλάσεων του προβλήματος, που στην περίπτωσή μας είναι τρεις. Κάθε νευρώνας σχετίζεται με μία από τις πιθανές κατηγορίες, παρέχοντας έτσι την έξοδο του δικτύου σε μορφή κατανομής πιθανοτήτων.

Για την παραγωγή αυτών των πιθανοτήτων χρησιμοποιείται η συνάρτηση ενεργοποίησης Softmax. Η Softmax μετατρέπει τις ακατέργαστες τιμές εξόδου (logits) σε πιθανότητες, κανονικοποιώντας τις ώστε να αθροίζουν στο 1 [23]. Αυτό επιτρέπει την ερμηνεία της εξόδου του δικτύου ως εκτιμώμενη πιθανότητα για κάθε κατηγορία, κατι ουσιώδες σε προβλήματα πολυκατηγορικής ταξινόμησης. Έτσι, η κατηγορία με τη μεγαλύτερη πιθανότητα θεωρείται ως η προβλεπόμενη από το μοντέλο. Αυτό θα το δουμε και αργότερα όταν θα αξιολογούμε τα αποτελέσματα.

Η χρήση της Softmax καθιστά δυνατή τη λήψη ξεκάθαρων και ερμηνεύσιμων προβλέψεων, προσφέροντας ένα μέτρο βεβαιότητας για κάθε κατηγορία και κάνοντας ευκολότερη την αξιολόγηση. Στο πλαίσιο της κατηγοριοποίησης των ειδών πνευμονίας, η λειτουργία αυτή είναι καθοριστική, καθώς παρέχει στους ιατρούς όχι μόνο μια πρόβλεψη αλλά και την πιθανότητα που αποδίδει το μοντέλο σε κάθε διάγνωση.

Dropout

Το Dropout είναι μια τεχνική βελτίωσης απόδοσης (βλέπε 5.8) που απενεργοποιεί τυχαία νευρώνες κατά την εκπαίδευση. Εδώ έχουμε δώσει την τιμή 0.5, που σημαίνει ότι το 50% των νευρώνων θα μηδενιστούν σε κάθε forward pass, μειώνοντας έτσι την εξάρτηση του μοντέλου από συγκεκριμένους νευρώνες και βελτιώνοντας τη γενίκευση. Η επιλογή αυτής της τιμής θεωρείται αρκετά ισχυρή μορφή regularization, συμβάλλοντας στην αποφυγή εξάρτησης του μοντέλου από συγκεκριμένα μοτίβα και στην καλύτερη γενίκευση σε μη-γνωστά δεδομένα. Επιπλέον, η χρήση του Dropout μπορεί να έχει ελαφρώς επιβαρυντική επίδραση στον χρόνο εκπαίδευσης, καθώς το δίκτυο χρειάζεται περισσότερα epochs για να συγκλίνει. Ωστόσο, το όφελος που προσφέρει σε επίπεδο απόδοσης και αξιοπιστίας καθιστά τη χρήση του αναπόσπαστο κομμάτι σε σύγχρονα CNNs, ειδικά σε συνδυασμό με άλλες τεχνικές όπως το L2 regularization και το Early Stopping που επισης χρησιμοποιούνται στην υλοποίηση.

Unfreezing Τελευταίων Layers

Η τεχνική του ξεπαγώματος (unfreezing) ορισμένων επιπέδων σε ένα προεκπαιδευμένο δίκτυο επιτρέπει στο μοντέλο να προσαρμόσει τα χαρακτηριστικά του σε ένα νέο dataset. Στην περίπτωσή μας, επιλέξαμε να αποδεσμεύσουμε τα τελευταία 8 layers του μοντέλου, ώστε να επιτρέψουμε στο δίκτυο να μάθει πιο εξειδικευμένα χαρακτηριστικά σχετιζόμενα με τα δεδομένα ακτινογραφιών. Με αυτόν τον τρόπο αποφεύγεται η ανάγκη να επανεκπαιδευτεί ολόκληρο το δίκτυο, κάτι που θα απαιτούσε τεράστια υπολογιστική ισχύ και θα αύξανε τον κίνδυνο υπερεκπαίδευσης [60].

Στο Xception, τα τελευταία αυτά layers αντιστοιχούν σε πιο αφηρημένα και σύνθετα χαρακτηριστικά, τα οποία είναι ιδιαίτερα χρήσιμα για την προσαρμογή σε εξειδικευμένες κλινικές εικόνες. Το fine-tuning τους ενισχύει την ικανότητα του μοντέλου να αναγνωρίζει τις λεπτές διαφορές ανάμεσα σε τύπους πνευμονίας.

Σε σύγκριση, στο VGG16 τα τελικά layers έχουν πιο παραδοσιακή δομή, βασισμένη σε επαναλαμβανόμενα blocks 3x3 convolutions. Αυτό σημαίνει ότι η προσαρμογή μέσω unfreezing είναι πιο περιορισμένη σε όρους αφαιρετικότητας, καθώς τα υψηλού επιπέδου χαρακτηριστικά του VGG16 δεν έχουν την ίδια εκφραστική ισχύ με εκείνα του Xception. Ως αποτέλεσμα, το fine-tuning στο VGG16 μπορεί να προσφέρει βελτίωση, αλλά όχι στον ίδιο βαθμό αποδοτικότητας όπως στο Xception, το οποίο εκμεταλλεύεται τις depthwise separable convolutions για πιο πλούσια αναπαράσταση.

6.4 Callbacks στην Εκπαίδευση Νευρωνικών Δικτύων

Κατά την εκπαίδευση βαθιών νευρωνικών δικτύων, η διαδικασία της εκπαίδευσης μπορεί να είναι ιδιαίτερα απαιτητική τόσο σε χρόνο όσο και σε υπολογιστικούς πόρους. Για την αποτελεσματικότερη διαχείριση αυτής της διαδικασίας, χρησιμοποιούνται τα callbacks, τα οποία αποτελούν μηχανισμούς παρακολούθησης και παρέμβασης κατά τη διάρκεια της εκπαίδευσης. Οι callbacks επιτρέπουν την αυτοματοποίηση ενεργειών όπως η διακοπή της εκπαίδευσης όταν δεν υπάρχει βελτίωση, η δυναμική προσαρμογή της ταχύτητας εκμάθησης (learning rate) και η αποθήκευση του καλύτερου μοντέλου [61].

Early Stopping

To Early Stopping είναι ένας από τους πιο διαδεδομένους μηχανισμούς, ο οποίος τερματίζει την εκπαίδευση όταν η απόδοση του μοντέλου στο σύνολο επικύρωσης (validation set) πάψει να βελτιώνεται. Η λογική πίσω από αυτή την προσέγγιση είναι η αποφυγή του overfitting, δηλαδή της υπερπροσαρμογής του μοντέλου στα δεδομένα εκπαίδευσης, που οδηγεί σε χαμηλότερη γενίκευση σε άγνωστα δεδομένα [78].

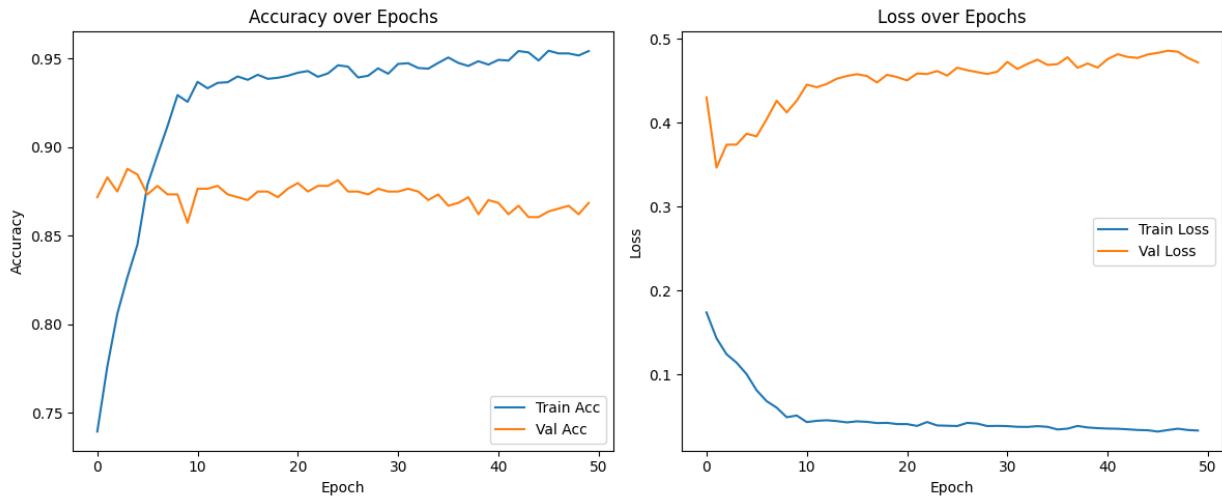
Στην πράξη, το Early Stopping παρακολουθεί μια μετρική, όπως η συνάρτηση απώλειας (loss) στο validation set, και αν για έναν προκαθορισμένο αριθμό εποχών (patience) δεν υπάρχει βελτίωση, η εκπαίδευση διακόπτεται. Επιπλέον, παρέχεται η δυνατότητα επαναφοράς των βαρών στην καλύτερη κατάσταση που επιτεύχθηκε κατά τη διάρκεια της εκπαίδευσης, γεγονός που εξασφαλίζει ότι το τελικό μοντέλο δεν θα είναι υποδεέστερο [28].

ReduceLROnPlateau

Ένας άλλος σημαντικός μηχανισμός είναι το ReduceLROnPlateau, το οποίο προσαρμόζει δυναμικά την τιμή του learning rate όταν παρατηρείται στασιμότητα στη βελτίωση της απόδοσης του μοντέλου. Η μείωση του learning rate καθιστά τη διαδικασία εκμάθησης πιο σταθερή και επιτρέπει στο μοντέλο να συγκλίνει σε καλύτερα τοπικά ελάχιστα [87].

Συγκεκριμένα, το ReduceLROnPlateau παρακολουθεί μια καθορισμένη μετρική (συνήθως το validation loss) και, αν δεν παρατηρηθεί βελτίωση για ορισμένο αριθμό εποχών, μειώνει το learning rate κατά έναν προκαθορισμένο παράγοντα. Αυτή η στρατηγική είναι ιδιαίτερα χρήσιμη σε περιπτώσεις όπου το μοντέλο έχει ήδη μάθει τις κύριες δομές των δεδομένων και χρειάζεται πιο λεπτομερή προσαρμογή για να επιτύχει υψηλότερη ακρίβεια [39].

Παρακάτω φαίνεται ένα παράδειγμα στην περίπτωση που δεν χρησιμοποιείται η μείωση του learning rate.



Εικόνα 12: Εξέλιξη της ακρίβειας και της απώλειας κατά την εκπαίδευση και την επικύρωση. Παρατηρείται ότι μετά από έναν αριθμό εποχών η απόδοση στο *Training set* αυξανεται συνεχόμενα όσο μαθαίνει καλύτερα τα δεδομένα. Αντίθετα στο *validation set* σταθεροποιείται αρχικά και αργότερα μειώνεται ελαφρώς. Το γεγονός αυτο δικαιολογεί τη χρήση του μηχανισμού *ReduceLROnPlateau* για τη δυναμική προσαρμογή του ρυθμού μάθησης και την επίτευξη καλύτερης σύγκλισης αποφεύγοντας το *overfitting*.

7. Αποτελέσματα και Μετρικές αξιολόγησης

7.1 Εισαγωγή στα αποτελέσματα

Στο κεφάλαιο αυτό παρουσιάζονται αναλυτικά τα αποτελέσματα που προέκυψαν από την εκπαίδευση και αξιολόγηση των συνελικτικών νευρωνικών δικτύων (CNN) που εφαρμόστηκαν για την ταξινόμηση ακτινογραφιών θώρακος στις κατηγορίες Bacterial Pneumonia, Viral Pneumonia και Normal. Στόχος της ανάλυσης είναι να αποτυπωθεί η ικανότητα των μοντέλων να διακρίνουν με αξιοπιστία μεταξύ των διαφορετικών κλινικών καταστάσεων, λαμβάνοντας υπόψη τόσο την ακρίβεια όσο και τη γενικότερη συμπεριφορά τους σε συνθήκες εκπαίδευσης και επικύρωσης. Με τον τρόπο αυτό επιδιώκεται μια ολοκληρωμένη αποτίμηση της απόδοσης των μοντέλων, όχι μόνο σε επίπεδο αριθμητικών μεγεθών αλλά και σε ό,τι αφορά τη σταθερότητα, τη γενίκευση και τη χρησιμότητά τους σε πραγματικές συνθήκες κλινικής πράξης.

Παρακάτω παρατίθεται ενδεικτικό στιγμιότυπο από το περιβάλλον του terminal, όπου αποτυπώνεται ο τρόπος παρουσίασης των αποτελεσμάτων μετά την εκπαίδευση.

```
Training history saved as 'training_history.png'.
Validation started...
78/78 [=====] - 7s 83ms/step - loss: 0.3518 - accuracy: 0.9071
Validation Loss: 0.3518, Validation Accuracy: 0.9071
78/78 [=====] - 8s 85ms/step

Classification Report:
precision    recall    f1-score   support
bacterial    0.91     0.97      0.94      242
normal        0.93     0.92      0.92      234
viral         0.85     0.78      0.82      148
accuracy          0.91      0.91      0.91      624
macro avg       0.90     0.89      0.89      624
weighted avg    0.91     0.91      0.91      624

Sensitivity and Specificity per Class:
Class      Sensitivity      Specificity
-----
bacterial    0.97          0.94
normal        0.92          0.96
viral         0.78          0.96

AUC-ROC Score: 0.9782
Confusion matrix saved as 'confusion_matrix.png'.
ROC curves saved as 'roc_curves.png'.

Total number of validation inputs: 624
Number of output classes: 3
```

Εικόνα 13: Στιγμιότυπο αποτελεσμάτων στο terminal.

7.2 Μετρικές Αξιολόγησης

Η αξιολόγηση ενός μοντέλου μηχανικής μάθησης δεν μπορεί να βασιστεί αποκλειστικά στην τιμή της ακρίβειας (accuracy), καθώς αυτή δεν αποτυπώνει πάντα με ακρίβεια την πραγματική του απόδοση, ειδικά σε προβλήματα με μη ισορροπημένες κλάσεις όπως η ιατρική διάγνωση. Για τον λόγο αυτό χρησιμοποιούνται πολλαπλές μετρικές, οι οποίες παρέχουν μια πιο ολοκληρωμένη εικόνα της συμπεριφοράς του συστήματος. Οι μετρικές αυτές εξετάζουν διαφορετικές πτυχές της απόδοσης, όπως η ικανότητα σωστής ανίχνευσης θετικών περιστατικών (recall/sensitivity), η ακρίβεια των προβλέψεων (precision), η συνολική ισορροπία μεταξύ των δύο (F1-score), αλλά και η διάκριση ανάμεσα στις κατηγορίες μέσω της ROC καμπύλης και της AUC τιμής. Η χρήση πολλαπλών μετρικών είναι βασική για να διασφαλιστεί ότι το μοντέλο είναι όχι μόνο ακριβές, αλλά και κλινικά αξιόπιστο. Παρακάτω θα αναλύσουμε αυτές που χρησιμοποιήθηκαν στην παρούσα εργασία.

7.2.1 Accuracy και Loss

Η ακρίβεια (Accuracy) και η συνάρτηση κόστους (Loss) αποτελούν τις πιο βασικές μετρικές αξιολόγησης ενός νευρωνικού δικτύου. Η ακρίβεια μετρά το ποσοστό των σωστών προβλέψεων του μοντέλου επί του συνολικού αριθμού δειγμάτων. Στην περίπτωση της διάκρισης τύπων πνευμονίας, εκφράζει το ποσοστό των ακτινογραφιών που ταξινομήθηκαν σωστά στις κατηγορίες «Bacterial», «Viral» και «Normal» [28].

Ο μαθηματικός ορισμός της ακρίβειας δίνεται από τον τύπο:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (22)$$

όπου:

- TP (True Positives): σωστές θετικές προβλέψεις
- TN (True Negatives): σωστές αρνητικές προβλέψεις
- FP (False Positives): λανθασμένες θετικές προβλέψεις
- FN (False Negatives): λανθασμένες αρνητικές προβλέψεις.

Ωστόσο, η χρήση της ακρίβειας ως μοναδικό κριτήριο δεν είναι πάντα αξιόπιστη, ειδικά σε μη ισορροπημένα datasets, όπου μια κλάση μπορεί να υπερεκπροσωπεύται. Σε αυτές τις περιπτώσεις, το μοντέλο μπορεί να εμφανίζει υψηλή ακρίβεια απλώς επειδή προβλέπει συχνότερα την πλειοψηφική κλάση.

Η συνάρτηση κόστους (Loss), και πιο συγκεκριμένα η Categorical Cross-Entropy, χρησιμοποιείται για την ποσοτικοποίηση της απόστασης μεταξύ των προβλεπόμενων πιθανοτήτων και των πραγματικών ετικετών [68]. Ο μαθηματικός ορισμός είναι:

$$L = -\sum_{i=1}^N \sum_{c=1}^C y_{i,c} \cdot \log(\hat{y}_{i,c}) \quad (23)$$

όπου:

- N είναι ο αριθμός των δειγμάτων
- C είναι ο αριθμός των κλάσεων
- $\gamma_{i,c}$ είναι η πραγματική ετικέτα (1 αν το δείγμα i ανήκει στην κλάση c, αλλιώς 0)
- $\hat{\gamma}_{i,c}$ είναι η προβλεπόμενη πιθανότητα του μοντέλου για το δείγμα i να ανήκει στην κλάση c.

Χαμηλότερες τιμές loss υποδεικνύουν καλύτερη εκμάθηση του μοντέλου, ενώ υψηλές τιμές δηλώνουν ότι οι προβλέψεις του αποκλίνουν σημαντικά από την πραγματικότητα. Η παρακολούθηση της loss τόσο στο training όσο και στο validation set είναι κρίσιμη, καθώς επιτρέπει τον εντοπισμό φαινομένων όπως το overfitting (χαμηλή loss στο training αλλά υψηλή στο validation) [91].

7.2.2 Precision, Recall, F1-score και Support

Για την αξιολόγηση της απόδοσης του μοντέλου σε κάθε κατηγορία («Bacterial», «Viral», «Normal»), χρησιμοποιήθηκαν οι μετρικές Precision, Recall, F1-score και Support. Αυτές οι μετρικές παρέχουν πιο λεπτομερή εικόνα της συμπεριφοράς του ταξινομητή, ειδικά σε περιπτώσεις μη ισορροπημένων δεδομένων [88][89].

Precision

Η ακρίβεια κατηγορίας (Precision) ορίζεται ως το ποσοστό των σωστών θετικών προβλέψεων σε σχέση με το σύνολο των προβλέψεων που έγιναν θετικές. Μαθηματικά:

$$Precision = \frac{TP}{TP+FP} \quad (24)$$

Η Precision δείχνει πόσο αξιόπιστες είναι οι θετικές προβλέψεις για κάθε κατηγορία.

Recall

Η ανάκληση (Recall) μετρά το ποσοστό των σωστών θετικών προβλέψεων σε σχέση με όλα τα πραγματικά θετικά δείγματα. Ορίζεται ως:

$$Recall = \frac{TP}{TP+FN} \quad (25)$$

Η Recall δείχνει πόσο καλά το μοντέλο αναγνωρίζει όλα τα δείγματα μιας συγκεκριμένης κλάσης.

F1-score

Το F1-score είναι ο αρμονικός μέσος του Precision και του Recall, και παρέχει μία ισορροπημένη μέτρηση μεταξύ των δύο:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (26)$$

Η μετρική αυτή είναι ιδιαίτερα σημαντική όταν υπάρχει ανισορροπία στις κλάσεις.

Sensitivity (ή True Positive Rate)

Η ευαισθησία (Sensitivity) ισοδυναμεί με το Recall και δείχνει το ποσοστό των πραγματικών θετικών που εντοπίζονται σωστά:

$$Sensitivity = \frac{TP}{TP+FN} \quad (27)$$

Specificity (ή True Negative Rate)

Η ειδικότητα (Specificity) μετρά το ποσοστό των πραγματικών αρνητικών που ταξινομούνται σωστά:

$$Specificity = \frac{TN}{TN+FP} \quad (28)$$

Η Specificity είναι ιδιαίτερα κρίσιμη σε κλινικές εφαρμογές, όπου η εσφαλμένη ταξινόμηση υγιών ως ασθενών (false positives) μπορεί να οδηγήσει σε περιπτές ιατρικές παρεμβάσεις.

Support

Το Support αναφέρεται στον συνολικό αριθμό δειγμάτων που ανήκουν σε κάθε κατηγορία στο validation set. Δεν αποτελεί μετρική απόδοσης, αλλά παρέχει πληροφορία για την κατανομή των δεδομένων.

7.2.3 ROC Curves και AUC-ROC Score

Εκτός από τις βασικές μετρικές ταξινόμησης, ιδιαίτερα σημαντική είναι και η ανάλυση μέσω των ROC curves (Receiver Operating Characteristic curves) και της συνοδευτικής μετρικής AUC-ROC Score. Αυτά τα εργαλεία επιτρέπουν την οπτικοποίηση και την ποσοτική αξιολόγηση της ικανότητας ενός μοντέλου να διακρίνει μεταξύ κλάσεων [88].

ROC Curve

Η ROC curve (καμπήλη) αναπαριστά τη σχέση μεταξύ του True Positive Rate (TPR) και του False Positive Rate (FPR) για διαφορετικά κατώφλια απόφασης. Ορίζονται ως:

$$TPR = \frac{TP}{TP+FN} \quad (30) \qquad FPR = \frac{FP}{TN+FP} \quad (31)$$

Μια ιδανική ROC curve θα βρισκόταν όσο πιο κοντά γίνεται στην πάνω αριστερή γωνία του γραφήματος, υποδεικνύοντας υψηλή ευαισθησία (Sensitivity) και ταυτόχρονα χαμηλά ψευδώς θετικά [91].

AUC-ROC Score

Η περιοχή κάτω από την καμπύλη ROC (Area Under the Curve, AUC) συνοψίζει την απόδοση του ταξινομητή σε μία μόνο τιμή. Η τιμή αυτή κυμαίνεται μεταξύ 0 και 1 [91]:

- AUC = 1: τέλεια διάκριση μεταξύ των κλάσεων.
- AUC = 0.5: τυχαία απόδοση, χωρίς ικανότητα διάκρισης.

Μια υψηλή τιμή AUC (>0.9) θεωρείται ένδειξη πολύ καλού μοντέλου, ειδικά σε ιατρικά δεδομένα όπου η διάκριση μεταξύ ασθενών και υγιών είναι κρίσιμη.

Πολυκατηγορική Επέκταση

Στην περίπτωση πολυκατηγορικής ταξινόμησης, η ROC curve υπολογίζεται με την τεχνική one-vs-rest (OvR), όπου κάθε κλάση συγκρίνεται έναντι όλων των υπολοίπων. Στη συνέχεια, το AUC μπορεί να υπολογιστεί ως macro-average (μέσος όρος μεταξύ όλων των κλάσεων) ή weighted-average (μέσος όρος με βάρη ανάλογα με το support κάθε κλάσης).

7.2.4 Confusion Matrix

Το confusion matrix αποτελεί ένα από τα πιο χρήσιμα εργαλεία για την ανάλυση της απόδοσης ενός ταξινομητή, καθώς παρέχει μια αναλυτική εικόνα των προβλέψεων σε επίπεδο κλάσης. Ουσιαστικά, παρουσιάζει σε μορφή πίνακα τις πραγματικές κλάσεις σε σχέση με τις προβλεπόμενες [88].

Δομή

Για ένα πρόβλημα πολυκατηγορικής ταξινόμησης, το confusion matrix είναι ένας πίνακας, όπου είναι ο αριθμός των κλάσεων. Κάθε γραμμή αντιπροσωπεύει τα πραγματικά δείγματα μιας κλάσης, ενώ κάθε στήλη αντιπροσωπεύει τις προβλέψεις του μοντέλου.

- **Διαγώνια στοιχεία:** Σωστές ταξινομήσεις (True Positives για κάθε κλάση).
- **Εκτός διαγωνίου στοιχεία:** Λάθος ταξινομήσεις (False Positives ή False Negatives ανάλογα με την οπτική).

Χρησιμότητα

Το confusion matrix επιτρέπει τον εντοπισμό συγκεκριμένων αδυναμιών του μοντέλου. Για παράδειγμα, αν μια κλάση συστηματικά συγχέεται με μια άλλη, αυτό φαίνεται καθαρά στον πίνακα. Σε ιατρικές εφαρμογές, αυτή η ανάλυση είναι κρίσιμη, καθώς μπορεί να αναδείξει περιπτώσεις όπου ο ταξινομητής μπερδεύει, π.χ., υγιείς με ασθενείς, κάτι που έχει ιδιαίτερα σοβαρές συνέπειες [88].

Για πρόβλημα δυαδικής ταξινόμησης, το confusion matrix έχει τη μορφή:

$$CM = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}$$

Από το confusion matrix μπορούν να εξαχθούν οι περισσότερες μετρικές, όπως Accuracy, Precision, Recall, F1-score, Sensitivity, και Specificity.

7.3 Ακρίβεια και απώλεια (Accuracy/Loss)

Στις υποενότητες που ακολουθούν παρουσιάζονται τα αποτελέσματα της ακριβείας και απώλειας της εκπαίδευσης για τα δίκτυα των αρχιτεκτονικών Xception και VGG16

7.3.1 Πίνακες ακρίβειας και απώλειας

Παρακάτω παρουσιάζονται οι πίνακες στους οποίους εμφανίζονται οι ακρίβειες και οι απώλειες που επιτευχθήκαν από τα μοντέλα σε 10 παραδείγματα εκπαίδευσης. Στις στήλες φαίνεται ο αριθμός της εκάστοτε δοκιμής. Η τελευταία στήλη είναι ο μέσος ορός των δέκα δοκίμων.

Ακρίβεια/Accuracy

1	2	3	4	5	6	7	8	9	10	Avg
0.9006	0.8958	0.8910	0.8862	0.9020	0.8985	0.8895	0.8932	0.8970	0.8901	0.8944

Πίνακας 5: Αποτελέσματα ακριβείας για το μοντέλο VGG16.

1	2	3	4	5	6	7	8	9	10	Avg
0.8798	0.8958	0.8974	0.8958	0.8965	0.8980	0.8945	0.8860	0.8950	0.8970	0.8936

Πίνακας 6: Αποτελέσματα ακριβείας για το μοντέλο Xception.

Απώλεια/Loss

1	2	3	4	5	6	7	8	9	10	Avg
0.3524	0.3591	0.3667	0.3548	0.3702	0.3785	0.3629	0.3654	0.3841	0.3687	0.3663

Πίνακας 7: Αποτελέσματα απώλειας για το μοντέλο VGG16.

1	2	3	4	5	6	7	8	9	10	Avg
0.3635	0.3289	0.3507	0.3471	0.3520	0.3495	0.3510	0.3465	0.3480	0.3500	0.3487

Πίνακας 8: Αποτελέσματα απώλειας για το μοντέλο Xception.

Ανάλυση πινάκων Ακρίβειας/Απώλειας

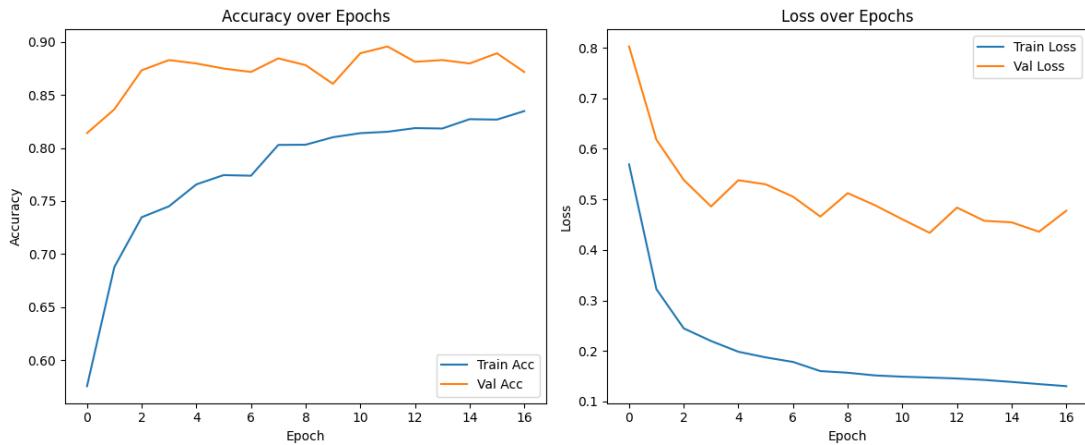
Από την ανάλυση των παραπάνω πινάκων προκύπτει ότι και τα δύο μοντέλα έχουν εκπαιδευτεί ικανοποιητικά, καθώς οι τιμές της ακρίβειας κυμαίνονται σε υψηλά επίπεδα (κοντά στο 0.90 και για τα 2 μοντέλα), ενώ οι τιμές της απώλειας παραμένουν χαμηλές (κάτω από 0.37). Επιπλέον, τόσο στην ακρίβεια όσο και στην απώλεια δεν παρατηρούνται μεγάλες διακυμάνσεις μεταξύ των διαφορετικών επαναλήψεων εκπαίδευσης. Η απόκλιση από δοκιμή σε δοκιμή είναι περιορισμένη (της τάξης ±0.01–0.02), γεγονός που καταδεικνύει τη σταθερότητα των αποτελεσμάτων και υποδηλώνει ότι τα μοντέλα δεν παρουσιάζουν σημάδια υπερπροσαρμογής(overfitting), αλλά αντίθετα εμφανίζουν ικανοποιητική ικανότητα γενίκευσης σε νέα δεδομένα.

Σε γενικές γραμμές, όταν η απόδοση ενός μοντέλου κυμαίνεται στο 85–90%, θεωρείται επαρκής για την επίλυση προβλημάτων ταξινόμησης εικόνων. Παρά ταύτα, στο πλαίσιο της παρούσας εργασίας, όπου εξετάζεται η εφαρμογή βαθιών νευρωνικών δικτύων σε ιατρικές διαγνώσεις, τα απαιτούμενα επίπεδα ακρίβειας ίσως είναι ακόμη υψηλότερα, δεδομένου ότι ένα σφάλμα ενδέχεται να έχει σοβαρές κλινικές επιπτώσεις. Συνεπώς, παρόλο που τα αποτελέσματα κρίνονται ιδιαίτερα θετικά, είναι επιθυμητή η περαιτέρω βελτίωση της απόδοσης.

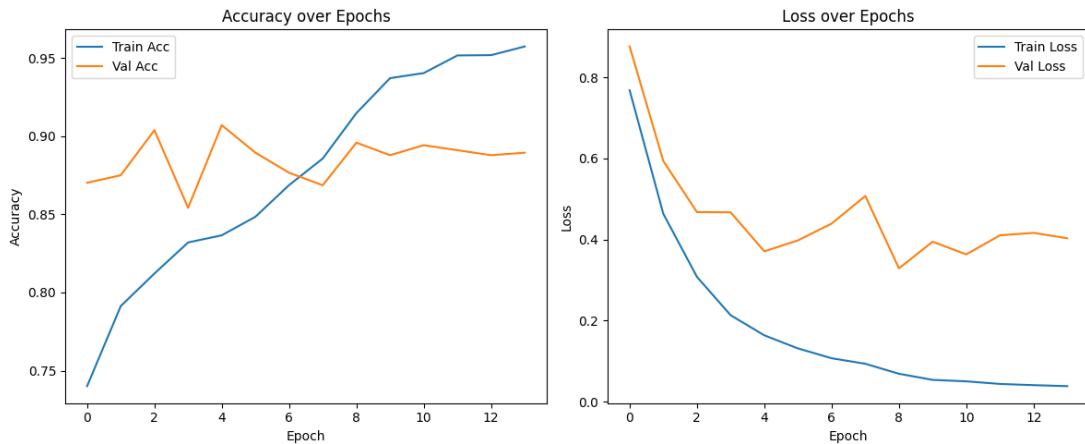
Όσον αφορά τη συγκριτική αξιολόγηση των δύο αρχιτεκτονικών, παρατηρείται ότι το VGG16 και το Xception εμφανίζουν πολύ κοντινά επίπεδα ακρίβειας (μέσος όρος 0.8944 και 0.8936 αντίστοιχα), γεγονός που υποδηλώνει πως και οι δύο προσεγγίσεις είναι κατάλληλες για το συγκεκριμένο πρόβλημα. Η ουσιαστική διαφορά εντοπίζεται στις τιμές της απώλειας, όπου το Xception υπερτερεί, επιτυγχάνοντας χαμηλότερη μέση τιμή (0.3487 έναντι 0.3663 του VGG16). Η διαφοροποίηση αυτή δείχνει ότι το Xception καταφέρνει να προσαρμόζεται καλύτερα στα δεδομένα, με μειωμένα σφάλματα πρόβλεψης και μεγαλύτερη σταθερότητα κατά την εκπαίδευση. Κατά συνέπεια ενώ τα δύο μοντέλα εμφανίζουν παρόμοια συνολική ακρίβεια, το Xception παρουσιάζει ανώτερη συμπεριφορά σε όρους γενίκευσης, γεγονός που το καθιστά καταλληλότερο για την παρούσα εφαρμογή.

7.3.2 Διαγράμματα ακριβείας και απώλειας

Στο παρόν υποκεφάλαιο παρουσιάζονται τα γραφήματα που αποτυπώνουν την πορεία της ακριβείας (accuracy) και της απώλειας (loss) κατά την εκπαίδευση του μοντέλου, προσφέροντας μια σαφή εικόνα της συνολικής απόδοσης του μοντέλου. Τα διαγράμματα περιλαμβάνουν τις καμπύλες τόσο για τα δεδομένα εκπαίδευσης όσο και για τα δεδομένα επικύρωσης, επιτρέποντας την άμεση σύγκριση της συμπεριφοράς του μοντέλου κατά τη διαδικασία μάθησης.



Εικόνα 14: Σύγκριση ακριβείας εκπαίδευσης (μπλε γραμμή)/επικύρωσης (πορτοκαλί γραμμή) στα αριστερά και στα δεξιά Σύγκριση απώλειας εκπαίδευσης (μπλε γραμμή)/επικύρωσης (πορτοκαλί γραμμή) για το μοντέλο VGG16.



Εικόνα 15: Σύγκριση ακριβείας εκπαίδευσης (μπλε γραμμή)/επικύρωσης (πορτοκαλί γραμμή) στα αριστερά και στα δεξιά Σύγκριση απώλειας εκπαίδευσης (μπλε γραμμή)/επικύρωσης (πορτοκαλί γραμμή) για το μοντέλο Xception.

Τα αποτελέσματα των διαγραμμάτων Ακρίβειας και Απώλειας ερμηνεύονται παρακάτω.

Διαγράμματα ακριβείας

Μπλε καμπύλη: Απεικονίζει τη μέγιστη ακρίβεια που επιτεύχθηκε σε κάθε επανάληψη κατά τη φάση της εκπαίδευσης.

Πορτοκαλί καμπύλη: Αντιστοιχεί στη μέγιστη ακρίβεια που καταγράφηκε σε κάθε επανάληψη κατά τη φάση της επικύρωσης. Τα δεδομένα επικύρωσης δεν συμμετέχουν στη διαδικασία εκπαίδευσης, αλλά αξιοποιούνται για να εκτιμηθεί η ικανότητα του μοντέλου να γενικεύει σε καινούρια, άγνωστα δεδομένα.

Κατά τη διάρκεια της εκπαίδευσης, το μοντέλο τροποποιεί προοδευτικά τα βάρη του με σκοπό να ελαχιστοποιήσει τη διαφορά ανάμεσα στις πραγματικές και τις προβλεπόμενες εξόδους. Ιδανικά, η ακρίβεια επικύρωσης δεν θα πρέπει να υπολείπεται σημαντικά της ακρίβειας εκπαίδευσης, καθώς μια τέτοια απόκλιση θα μπορούσε να αποτελεί ένδειξη υπερπροσαρμογής (overfitting).

Για τα παραπάνω παραδείγματα, παρατηρείται ότι οι τιμές της ακρίβειας, τόσο για το σύνολο εκπαίδευσης όσο και για το σύνολο επικύρωσης, παραμένουν σε υψηλά επίπεδα. Το γεγονός αυτό δείχνει ότι το μοντέλο έχει εκπαιδευτεί με επιτυχία και παρουσιάζει ικανοποιητική ικανότητα επίλυσης του προβλήματος. Για μια πιο ολοκληρωμένη αποτίμηση της απόδοσης του νευρωνικού δικτύου, είναι απαραίτητο να εξεταστούν και τα διαγράμματα της απώλειας που παρουσιάζονται στη συνέχεια.

Διαγράμματα απώλειας

Μπλε καμπύλη: Παρουσιάζει την ελάχιστη τιμή απώλειας που καταγράφηκε σε κάθε επανάληψη κατά τη φάση εκπαίδευσης.

Πορτοκαλί καμπύλη: Αντιστοιχεί στην ελάχιστη τιμή απώλειας που επιτεύχθηκε σε κάθε επανάληψη κατά τη φάση επικύρωσης. Τα δεδομένα επικύρωσης δεν συμμετέχουν στη διαδικασία εκπαίδευσης, αλλά χρησιμεύουν για την εκτίμηση της ικανότητας του μοντέλου να γενικεύει σε νέα, μη γνωστά δεδομένα.

Κατά την ανάλυση των διαγραμμάτων απώλειας, το αναμενόμενο αποτέλεσμα είναι τόσο η απώλεια εκπαίδευσης όσο και η απώλεια επικύρωσης να μειώνονται προοδευτικά με την πάροδο των εποχών εκπαίδευσης. Όπως και στην περίπτωση της ακρίβειας, η αξιολόγηση της απόδοσης απαιτεί τη συνεξέταση και των δύο μετρικών, ώστε να εξαχθούν ασφαλή συμπεράσματα.

Στην παρούσα περίπτωση, παρατηρείται ότι οι καμπύλες απώλειας εκπαίδευσης και επικύρωσης παραμένουν σε χαμηλά επίπεδα και παρουσιάζουν μικρή μεταξύ τους απόσταση. Το γεγονός αυτό καταδεικνύει ότι το μοντέλο έχει καταφέρει να μάθει αποτελεσματικά από τα δεδομένα εκπαίδευσης χωρίς να εμφανίζει σημάδια υπερπροσαρμογής (overfitting). Δηλαδή το νευρωνικό δίκτυο δεν περιορίζεται στην απλή απομνημόνευση των δεδομένων εκπαίδευσης, αλλά επιδεικνύει ικανοποιητική ικανότητα γενίκευσης σε νέα δεδομένα.

7.3.3 Συγκριτική Ανάλυση Διαγραμμάτων VGG16 και Xception

Στις εικόνες X1 (VGG16), X2(Xception) παρουσιάζονται οι καμπύλες ακρίβειας και απώλειας για τα δύο μοντέλα κατά τη διάρκεια της εκπαίδευσης. Παρατηρούνται ομοιότητες αλλά και ουσιαστικές διαφορές στη συμπεριφορά τους, που αντικατοπτρίζουν τις αρχιτεκτονικές τους ιδιαιτερότητες.

Ακρίβεια (Accuracy)

Για το VGG16, η καμπύλη εκπαίδευσης ξεκινάει χαμηλότερα αλλά βελτιώνεται σταδιακά, φτάνοντας περίπου το 0.83 όπως επιβιώνουμε και από τα δεδομένα εκπαίδευσης. Ενώ η καμπύλη επικύρωσης κυμαίνεται σε σχετικά υψηλά επίπεδα (0.85–0.90) σχεδόν από τις πρώτες εποχές. Αυτό δείχνει ότι το VGG16 κατορθώνει να μάθει με σταθερό ρυθμό, χωρίς μεγάλες διακυμάνσεις στην επικύρωση.

Στο Xception, η ακρίβεια εκπαίδευσης ανεβαίνει πιο γρήγορα και ξεπερνά το 0.95 μετά από λίγες εποχές. Ωστόσο, η ακρίβεια επικύρωσης εμφανίζει μεγαλύτερες διακυμάνσεις, παραμένοντας γύρω στο 0.88–0.91. Αυτό σημαίνει ότι, αν και το Xception προσαρμόζεται καλύτερα στα δεδομένα εκπαίδευσης, η γενίκευσή του σε άγνωστα δεδομένα είναι λιγότερο σταθερή συγκριτικά με το VGG16.

Απώλεια (Loss)

Στο δίκτυο VGG16 η απώλεια εκπαίδευσης μειώνεται ομαλά και φτάνει σε χαμηλά επίπεδα (~0.12), ενώ η απώλεια επικύρωσης ακολουθεί παρόμοια πορεία, παραμένοντας κοντά στις τιμές της εκπαίδευσης χωρίς έντονη απόκλιση. Αυτό δείχνει ότι το μοντέλο μαθαίνει με σταθερό τρόπο και δεν παρουσιάζει σημαντικά σημάδια υπερπροσαρμογής.

Αντίθετα στο δίκτυο Xception η απώλεια εκπαίδευσης μειώνεται πιο απότομα και φτάνει σχεδόν στο μηδέν, γεγονός που αποδεικνύει την υψηλή του ικανότητα προσαρμογής. Ωστόσο, η απώλεια επικύρωσης παρουσιάζει διακυμάνσεις και δεν ακολουθεί τόσο ομαλή πτωτική πορεία όσο στο VGG16, δείχνοντας μια σχετική αστάθεια στη γενίκευση.

Ερμηνεία διαφορών μεταξύ των διαγραμμάτων

Οι παρατηρούμενες διαφορές μπορούν να εξηγηθούν από τη φύση των δύο αρχιτεκτονικών:

Το VGG16 είναι ένα βαθύ αλλά πιο «παραδοσιακό» CNN, το οποίο μαθαίνει με πιο συντηρητικό και σταθερό τρόπο, αποφεύγοντας απότομες αυξομειώσεις. Αυτό το καθιστά πιο αξιόπιστο σε επίπεδο επικύρωσης.

Το Xception, χάρη στη χρήση depthwise separable convolutions, μπορεί να συλλάβει πιο πολύπλοκα χαρακτηριστικά και να μειώσει την απώλεια εκπαίδευσης σε πολύ χαμηλά επίπεδα. Ωστόσο, αυτή η υψηλή προσαρμοστικότητα ενδέχεται να οδηγήσει σε μικρότερη σταθερότητα κατά την επικύρωση, δηλαδή μια ελαφριά τάση προς υπερπροσαρμογή που όμως εδώ αποφεύγεται.

Συνολικα θα λεγαμε πως το VGG16 εμφανίζεται πιο σταθερό και συνεπές στη σχέση εκπαίδευσης-επικύρωσης, με ισορροπημένες καμπύλες. Το Xception υπερέχει σε όρους ελαχιστοποίησης της απώλειας εκπαίδευσης και μέγιστης ακρίβειας στο train set, αλλά παρουσιάζει διακυμάνσεις στα validation αποτελέσματα, γεγονός που υποδηλώνει υψηλότερη ευαισθησία στη δομή των δεδομένων.

7.4 Σύγκριση χρόνου Εκπαίδευσης και εποχών

Στο παρόν κεφάλαιο εξετάζεται η επίδραση του χρόνου εκπαίδευσης και των εποχών (epochs) στην απόδοση των μοντέλων. Παρουσιάζονται συγκριτικά αποτελέσματα που αναδεικνύουν πώς οι διαφορετικές ρυθμίσεις εκπαίδευσης επηρεάζουν την ακρίβεια, την απώλεια και τη σταθερότητα των μοντέλων, παρέχοντας μια ολοκληρωμένη εικόνα της συμπεριφοράς τους κατά τη διαδικασία μάθησης.

7.4.1 Αριθμοί εποχών και Πίνακες

Ο αριθμός των εποχών (epochs) αναφέρεται στο πόσες φορές το σύνολο δεδομένων εκπαίδευσης περνά εξ ολοκλήρου μέσα από το νευρωνικό δίκτυο. Αποτελεί βασική υπερπαράμετρο που επηρεάζει σημαντικά την απόδοση του μοντέλου.

Σε λίγες εποχές, το μοντέλο μπορεί να μην έχει προλάβει να μάθει επαρκώς τα χαρακτηριστικά των δεδομένων (underfitting), με αποτέλεσμα χαμηλή ακρίβεια. Αντίθετα, σε υπερβολικά πολλές εποχές υπάρχει ο κίνδυνος το μοντέλο να υπερπροσαρμοστεί (overfitting), δηλαδή να μάθει υπερβολικά καλά τα δεδομένα εκπαίδευσης εις βάρος της γενίκευσης σε άγνωστα δεδομένα.

Στους πίνακες που ακολουθούν εμφανίζεται ο αριθμός των εποχών στον οποίο το κάθε μοντέλο σταμάτησε να εκπαιδεύεται λόγο early stopping (βλ. κεφ. 6.4) για τα 10 παραδείγματα που χρησιμοποιούμε. Η τελευταία στήλη είναι ο μέσος ορός των εποχών.

Πίνακες Αριθμού εποχών

Run 1 2 3 4 5 6 7 8 9 10 Avg

Epochs 17 17 18 21 18 19 20 17 18 19 18.4

Πίνακας 9: Αριθμός εποχών οπού το μοντέλο σταμάτησε να εκπαιδεύεται για το μοντέλο VGG16.

Run 1 2 3 4 5 6 7 8 9 10 Avg

Epochs 13 14 14 15 14 14 13 14 12 13 13.9

Πίνακας 10: Αριθμός εποχών οπού το μοντέλο σταμάτησε να εκπαιδεύεται για το μοντέλο Xception.

7.4.2 Χρόνος εκπαίδευσης και Πίνακες

Ο χρόνος εκπαίδευσης σε ένα νευρωνικό δίκτυο αποτελεί σημαντικό παράγοντα αξιολόγησης, καθώς επηρεάζει άμεσα την αποδοτικότητα και τη χρηστικότητα του μοντέλου. Εξαρτάται από διάφορους παράγοντες όπως η πολυπλοκότητα της αρχιτεκτονικής (π.χ. VGG16 ή Xception), το μέγεθος του συνόλου δεδομένων, ο αριθμός των εποχών (epochs), καθώς και οι υπολογιστικοί πόροι που χρησιμοποιούνται (GPU, CPU, μνήμη).

Σε γενικές γραμμές, πιο περίπλοκα δίκτυα με περισσότερες παραμέτρους απαιτούν μεγαλύτερο χρόνο εκπαίδευσης, ενώ η αύξηση των εποχών αυξάνει γραμμικά τον συνολικό χρόνο. Ωστόσο, η χρήση ισχυρών GPUs μπορεί να μειώσει σημαντικά τη διάρκεια εκπαίδευσης. Ο χρόνος εκπαίδευσης δεν πρέπει να αξιολογείται μεμονωμένα, αλλά σε συνδυασμό με την τελική απόδοση του μοντέλου, ώστε να επιτευχθεί η βέλτιστη ισορροπία ανάμεσα σε ακρίβεια και υπολογιστικό κόστος.

Στις εικόνες που ακολουθούν μπορούμε να δούμε πως εξελίσσεται μια τυπική εκπαίδευση των μοντελών μας μαζί με το τον αριθμό των εποχών, το χρόνο της εκπαίδευσης καθώς και την επιρροή του earlystopping/ReduceLROnPlateau.

```
I0000 00:00:1757886773.728587 1697250 device_compiler.h:186] Compiled cluster using XLA! This line is logged at most once for the lifetime of the process
654/654 [=====] - 183s 250ms/step - loss: 0.7680 - accuracy: 0.7401 - val_loss: 0.8760 - val_accuracy: 0.8702 - lr: 1.0000e-04
Epoch 2/50
654/654 [=====] - 165s 252ms/step - loss: 0.4636 - accuracy: 0.7913 - val_loss: 0.5936 - val_accuracy: 0.8750 - lr: 1.0000e-04
Epoch 3/50
654/654 [=====] - 165s 252ms/step - loss: 0.3080 - accuracy: 0.8121 - val_loss: 0.4678 - val_accuracy: 0.9038 - lr: 1.0000e-04
Epoch 4/50
654/654 [=====] - 166s 253ms/step - loss: 0.2135 - accuracy: 0.8320 - val_loss: 0.4672 - val_accuracy: 0.8542 - lr: 1.0000e-04
Epoch 5/50
654/654 [=====] - 166s 254ms/step - loss: 0.1639 - accuracy: 0.8366 - val_loss: 0.3710 - val_accuracy: 0.9071 - lr: 1.0000e-04
Epoch 6/50
654/654 [=====] - 164s 250ms/step - loss: 0.1317 - accuracy: 0.8484 - val_loss: 0.3980 - val_accuracy: 0.8894 - lr: 1.0000e-04
Epoch 7/50
654/654 [=====] - 164s 251ms/step - loss: 0.1074 - accuracy: 0.8687 - val_loss: 0.4395 - val_accuracy: 0.8766 - lr: 1.0000e-04
Epoch 8/50
654/654 [=====] - ETA: 0s - loss: 0.0938 - accuracy: 0.8857
Epoch 8: ReduceLROnPlateau reducing learning rate to 1.999999494757503e-05.
654/654 [=====] - 165s 252ms/step - loss: 0.0938 - accuracy: 0.8857 - val_loss: 0.5075 - val_accuracy: 0.8686 - lr: 1.0000e-04
Epoch 9/50
654/654 [=====] - 165s 252ms/step - loss: 0.0692 - accuracy: 0.9148 - val_loss: 0.3289 - val_accuracy: 0.8958 - lr: 2.0000e-05
Epoch 10/50
654/654 [=====] - 166s 254ms/step - loss: 0.0541 - accuracy: 0.9371 - val_loss: 0.3947 - val_accuracy: 0.8878 - lr: 2.0000e-05
Epoch 11/50
654/654 [=====] - 166s 254ms/step - loss: 0.0505 - accuracy: 0.9404 - val_loss: 0.3634 - val_accuracy: 0.8942 - lr: 2.0000e-05
Epoch 12/50
654/654 [=====] - ETA: 0s - loss: 0.0441 - accuracy: 0.9516
Epoch 12: ReduceLROnPlateau reducing learning rate to 3.999999898951501e-06.
654/654 [=====] - 166s 253ms/step - loss: 0.0441 - accuracy: 0.9516 - val_loss: 0.4105 - val_accuracy: 0.8910 - lr: 2.0000e-05
Epoch 13/50
654/654 [=====] - 167s 256ms/step - loss: 0.0411 - accuracy: 0.9518 - val_loss: 0.4166 - val_accuracy: 0.8878 - lr: 4.0000e-06
Epoch 14/50
654/654 [=====] - 165s 252ms/step - loss: 0.0386 - accuracy: 0.9574 - val_loss: 0.4032 - val_accuracy: 0.8894 - lr: 4.0000e-06
Training history saved as 'training_history.png'.
Validation started...
78/78 [=====] - 6s 75ms/step - loss: 0.3289 - accuracy: 0.8958
Validation Loss: 0.3289, Validation Accuracy: 0.8958
78/78 [=====] - 7s 79ms/step
```

Εικόνα 16: Απεικόνιση μιας τυχαίας τυπικής εκπαίδευσης στο μοντέλο VGG16.

```

WARNING: All log messages before absl::InitializeLog() is called are written to STDERR
I0000 00:00:1757877420.406467 1429098 device_compiler.h:186] Compiled cluster using XLA! This line is logged at most once for the lifetime of the process
327/327 [=====] - 117s 313ms/step - loss: 0.5820 - accuracy: 0.5310 - val_loss: 0.9036 - val_accuracy: 0.7532 - lr: 1.0000e-04
Epoch 0/50
327/327 [=====] - 111s 340ms/step - loss: 0.3190 - accuracy: 0.6747 - val_loss: 0.6343 - val_accuracy: 0.7965 - lr: 1.0000e-04
Epoch 3/50
327/327 [=====] - 111s 339ms/step - loss: 0.2344 - accuracy: 0.7450 - val_loss: 0.5802 - val_accuracy: 0.8558 - lr: 1.0000e-04
Epoch 4/50
327/327 [=====] - 111s 338ms/step - loss: 0.2045 - accuracy: 0.7594 - val_loss: 0.6205 - val_accuracy: 0.8157 - lr: 1.0000e-04
Epoch 5/50
327/327 [=====] - 112s 343ms/step - loss: 0.1892 - accuracy: 0.7726 - val_loss: 0.5388 - val_accuracy: 0.8446 - lr: 1.0000e-04
Epoch 6/50
327/327 [=====] - 110s 336ms/step - loss: 0.1790 - accuracy: 0.7764 - val_loss: 0.4414 - val_accuracy: 0.8942 - lr: 1.0000e-04
Epoch 7/50
327/327 [=====] - 111s 340ms/step - loss: 0.1729 - accuracy: 0.7873 - val_loss: 0.4908 - val_accuracy: 0.8590 - lr: 1.0000e-04
Epoch 8/50
327/327 [=====] - 111s 338ms/step - loss: 0.1658 - accuracy: 0.7802 - val_loss: 0.4665 - val_accuracy: 0.8750 - lr: 1.0000e-04
Epoch 9/50
327/327 [=====] - ETA: 0s - loss: 0.1620 - accuracy: 0.7873
Epoch 9: ReduceLROnPlateau reducing learning rate to 1.9999999494757503e-05.
327/327 [=====] - 110s 336ms/step - loss: 0.1620 - accuracy: 0.7873 - val_loss: 0.4597 - val_accuracy: 0.8814 - lr: 1.0000e-04
Epoch 10/50
327/327 [=====] - 111s 339ms/step - loss: 0.1430 - accuracy: 0.8133 - val_loss: 0.4856 - val_accuracy: 0.8782 - lr: 2.0000e-05
Epoch 11/50
327/327 [=====] - 111s 340ms/step - loss: 0.1372 - accuracy: 0.8171 - val_loss: 0.4167 - val_accuracy: 0.8990 - lr: 2.0000e-05
Epoch 12/50
327/327 [=====] - 111s 339ms/step - loss: 0.1376 - accuracy: 0.8215 - val_loss: 0.3999 - val_accuracy: 0.8910 - lr: 2.0000e-05
Epoch 13/50
327/327 [=====] - 110s 338ms/step - loss: 0.1350 - accuracy: 0.8211 - val_loss: 0.4625 - val_accuracy: 0.8830 - lr: 2.0000e-05
Epoch 14/50
327/327 [=====] - 111s 339ms/step - loss: 0.1295 - accuracy: 0.8240 - val_loss: 0.5918 - val_accuracy: 0.8558 - lr: 2.0000e-05
Epoch 15/50
327/327 [=====] - ETA: 0s - loss: 0.1298 - accuracy: 0.8244
Epoch 15: ReduceLROnPlateau reducing learning rate to 3.999999898951501e-06.
327/327 [=====] - 110s 337ms/step - loss: 0.1298 - accuracy: 0.8244 - val_loss: 0.4813 - val_accuracy: 0.8878 - lr: 2.0000e-05
Epoch 16/50
327/327 [=====] - 110s 337ms/step - loss: 0.1247 - accuracy: 0.8354 - val_loss: 0.4860 - val_accuracy: 0.8862 - lr: 4.0000e-06
Epoch 17/50
327/327 [=====] - 111s 339ms/step - loss: 0.1209 - accuracy: 0.8373 - val_loss: 0.4381 - val_accuracy: 0.8878 - lr: 4.0000e-06
Training history saved as 'training_history.png'.
Validation started...

```

Εικόνα 17: Απεικόνιση μιας τυχαίας τυπικής εκπαίδευσης στο μοντέλο Xception.

Από τα παραδείγματα των παραπάνω εικόνων μπορούμε να βγάλουμε τον μεσο όρο για το χρόνο που παίρνει η εκπαίδευση μιας εποχής στο κάθε δίκτυο. Χρησιμοποιώντας το παράδειγμα της εικόνας 16 ο μέσος ορός ισούται με 110,65 και αντίστοιχα για την εικόνα 17 είναι 166,64. Αν κάνουμε το ίδιο για κάθε ένα από τα 10 παραδείγματα με τα οποία δουλεύουμε και στα δυο μοντέλα, θα καταλήξουμε στους παρακάτω πίνακες. Κάθε στήλη του πίνακα αντιστοιχεί σε μια εκτέλεση του εκάστοτε μοντέλου. Η τελευταία στήλη παρουσιάζει τον συνολικό μεσο ορό για το μοντέλο.

Πίνακες μέσου ορού χρόνου εκπαίδευσης

Run	1	2	3	4	5	6	7	8	9	Avg
Time/Epoch	110.65	110.59	110.73	111.14	110.76	110.81	110.94	110.62	110.00	110.73

Πίνακας 11: Μέσος ορός χρόνου εκπαίδευσης μιας εποχής για το μοντέλο VGG16.

Run	1	2	3	4	5	6	7	8	9	Avg
Time/Epoch	166.15	166.64	166.62	167.10	166.59	166.64	166.17	166.61	165.84	166.43

Πίνακας 12: Μέσος ορός χρόνου εκπαίδευσης μιας εποχής για το μοντέλο Xception.

7.4.3 Σχολιασμός Αποτελεσμάτων

Από την ανάλυση των προηγούμενων πινάκων προκύπτει ότι το μοντέλο VGG16 ολοκληρώνει την εκπαίδευσή του σε μεγαλύτερο αριθμό εποχών (μέσος όρος 18.4), σε σύγκριση με το Xception, το οποίο σταματά νωρίτερα (μέσος όρος 13.9). Η διαφορά αυτή αποδίδεται κυρίως στον τρόπο με τον οποίο κάθε αρχιτεκτονική «μαθαίνει» τα χαρακτηριστικά των δεδομένων. Το VGG16, με την πιο παραδοσιακή και βαθιά δομή του, χρειάζεται περισσότερες επαναλήψεις για να συγκλίνει σε ικανοποιητικά επίπεδα ακρίβειας και χαμηλής απώλειας. Αντίθετα, το Xception, αξιοποιώντας depthwise separable convolutions, διαθέτει αυξημένη εκφραστική ικανότητα και επιτυγχάνει ταχύτερη προσαρμογή, γεγονός που επιτρέπει την επίτευξη συγκρίσιμης ακρίβειας σε μικρότερο αριθμό εποχών.

Ωστόσο, η παραπάνω τάση διαφοροποιείται όταν εξεταστεί ο χρόνος εκπαίδευσης ανά εποχή. Παρόλο που το Xception εκπαιδεύεται σε λιγότερες εποχές, ο μέσος χρόνος ανά εποχή είναι σημαντικά υψηλότερος (166.43 δευτερόλεπτα) σε σχέση με το VGG16 (110.73 δευτερόλεπτα). Η διαφορά αυτή οφείλεται στην αυξημένη υπολογιστική πολυπλοκότητα του Xception, καθώς οι depthwise separable convolutions, αν και πιο αποδοτικές ως προς την εκφραστικότητα, απαιτούν περισσότερους υπολογισμούς ανά βήμα. Αντίθετα, το VGG16, αν και βαθύτερο, χρησιμοποιεί απλούστερες κλασικές συνελικτικές πράξεις, οι οποίες είναι πιο «φιλικές» σε επίπεδο εκτέλεσης.

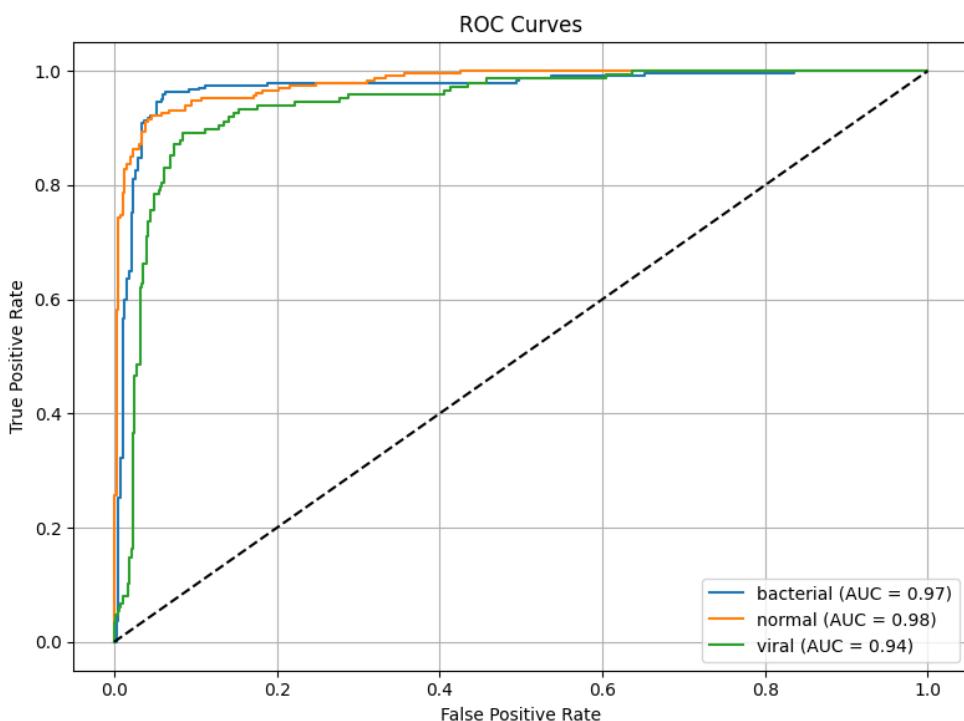
Συνδυάζοντας τους δύο παράγοντες (αριθμό εποχών και χρόνο ανά εποχή), παρατηρείται ότι το VGG16 χρειάζεται περίπου 34 λεπτά για την πλήρη εκπαίδευσή του, ενώ το Xception περίπου 38 λεπτά. Αυτό δείχνει ότι, στην πράξη, η εκπαίδευση του Xception είναι πιο αργή συνολικά, παρόλο που απαιτεί λιγότερες εποχές, γεγονός που το καθιστά πιο απαιτητικό σε υπολογιστικούς πόρους.

Συνοψίζοντας, το VGG16 χαρακτηρίζεται από μεγαλύτερη διάρκεια εκπαίδευσης σε επίπεδο εποχών αλλά ταχύτερη εκτέλεση ανά εποχή, ενώ το Xception επιτυγχάνει ταχύτερη σύγκλιση σε λιγότερες εποχές αλλά με μεγαλύτερο υπολογιστικό κόστος ανά εποχή. Η διαφοροποίηση αυτή αντικατοπτρίζει τον σχεδιασμό των δύο αρχιτεκτονικών: το VGG16 είναι πιο «συντηρητικό» και σταθερό στην εκπαίδευση, ενώ το Xception είναι πιο αποδοτικό ως προς την αξιοποίηση των δεδομένων, αλλά απαιτεί μεγαλητερη υπολογιστική ισχύ.

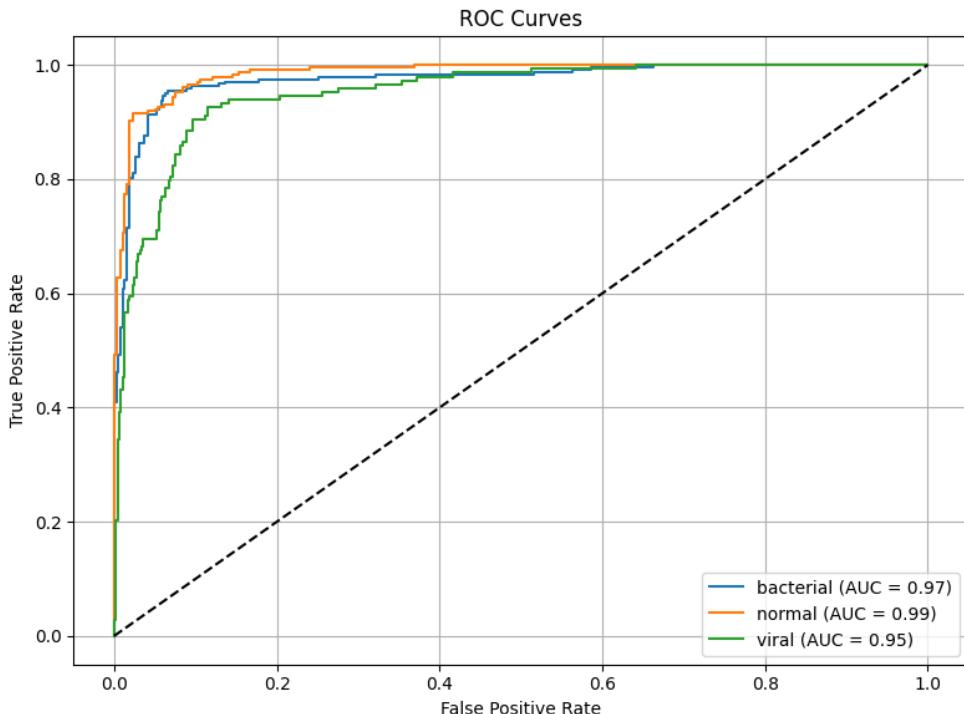
7.5 Καμπύλες ROC και AUC-ROC score

Η καμπύλη ROC (Receiver Operating Characteristic) και ο δείκτης AUC-ROC (Area Under the Curve) αποτελούν θεμελιώδη εργαλεία για την αξιολόγηση της απόδοσης ταξινομητικών μοντέλων. Η ROC αποτυπώνει τη σχέση μεταξύ του ποσοστού αληθώς θετικών προβλέψεων (True Positive Rate – TPR) και του ποσοστού ψευδώς θετικών (False Positive Rate – FPR) σε διαφορετικά κατώφλια απόφασης. Ο δείκτης AUC εκφράζει το συνολικό εμβαδόν κάτω από την καμπύλη ROC και παρέχει μια συνοπτική μέτρηση της διακριτικής ικανότητας του μοντέλου: όσο πιο κοντά βρίσκεται στο 1.0, τόσο καλύτερη είναι η απόδοση του μοντέλου στη διάκριση μεταξύ των διαφορετικών κατηγοριών.

7.5.1 Καμπύλες ROC



Εικόνα 18: Γραφική απεικόνιση καμπυλών ROC, *bacterial*(μπλε γραμμή), *normal*(πορτοκαλί γραμμή), *viral*(πράσινη γραμμή) για το μοντέλο VGG16.



Εικόνα 19: Γραφική απεικόνιση καμπυλών ROC, bacterial(μπλε γραμμή), normal(πορτοκαλί γραμμή), viral(πράσινη γραμμή) για το μοντέλο Xception.

Στα παραπάνω διαγράμματα στον άξονα Y απεικονίζεται το ποσοστό αληθώς θετικών (True Positive Rate – TPR ή Sensitivity), ενώ στον άξονα X το ποσοστό ψευδώς θετικών (False Positive Rate – FPR = 1 – Specificity). Έτσι, η ROC επιτρέπει τη μελέτη της ισορροπίας μεταξύ ευαισθησίας και ειδικότητας.

Έχουμε τρεις γραμμές με διαφορετικά χρώματα όπου κάθε γραμμή αντιστοιχεί σε μία κατηγορία διάγνωσης (bacterial-μπλέ, normal-πορτοκαλί, viral-πράσινο). Όσο πιο «ψηλά» και αριστερά βρίσκεται η γραμμή, τόσο καλύτερη η επίδοση του μοντέλου στη συγκεκριμένη κατηγορία.

Η μαύρη διακεκομμένη διαγώνιος (από το σημείο (0,0) στο (1,1)) είναι η γραμμή αναφοράς. Αν η ROC καμπύλη ενός μοντέλου έπεφτε πάνω σε αυτήν τη γραμμή, αυτό θα σήμαινε ότι το μοντέλο κάνει τυχαίες προβλέψεις (σαν να μάντευε με κέρμα). Όσο πιο πολύ η καμπύλη απομακρύνεται προς τα πάνω και αριστερά, τόσο καλύτερη είναι η απόδοση του μοντέλου (δηλαδή έχει μεγαλύτερη ευαισθησία χωρίς να αυξάνει πολύ τα ψευδώς θετικά).

Ανάλυση των ROC curves

Στην Εικόνα 18 παρουσιάζεται η καμπύλη ROC για το μοντέλο VGG16, με τις αντίστοιχες τιμές AUC για τις τρεις κατηγορίες: bacterial (0.97), normal (0.98) και viral (0.94). Παρατηρείται ότι όλες οι καμπύλες βρίσκονται σημαντικά πάνω από τη διαγώνιο αναφοράς (τυχαίος ταξινομητής), γεγονός που καταδεικνύει υψηλή απόδοση του μοντέλου. Η κατηγορία normal παρουσιάζει την υψηλότερη τιμή AUC (0.98) γεγονός που σημαίνει ότι το μοντέλο διακρίνει με μεγαλύτερη ακρίβεια τις φυσιολογικές ακτινογραφίες από τις

υπόλοιπες. Η κατηγορία bacterial ακολουθεί με AUC=0.97 ενώ η viral εμφανίζει ελαφρώς χαμηλότερη επίδοση (0.94) κάτι που υποδηλώνει ότι η διάκριση μεταξύ ιογενούς πνευμονίας και των υπόλοιπων περιπτώσεων είναι πιο απαιτητική, αυτό είναι αναμενόμενο αφού και στους «κλασσικούς» τρόπους διάγνωσης παρουσιάζει δυσκολία.

Αντίστοιχα, στην Εικόνα 19, που αφορά το μοντέλο Xception, οι τιμές AUC είναι bacterial=0.97, normal=0.99 και viral=0.95. Βλέπουμε ότι οι καμπύλες του Xception βρίσκονται ακόμα πιο κοντά στην ιδανική γωνία (0,1) του διαγράμματος, γεγονός που επιβεβαιώνει την ανώτερη διακριτική ικανότητά του. Ιδιαίτερα για την κατηγορία normal, το Xception επιτυγχάνει AUC=0.99, σχεδόν άριστη ταξινόμηση, ενώ και για την κατηγορία viral σημειώνεται μικρή βελτίωση σε σχέση με το VGG16 (0.95 έναντι 0.94).

7.5.2 Πίνακες AUC-ROC score

Στους παρακάτω πίνακες παρουσιάζονται τα αποτελέσματα για τα AUC-ROC score που συλλέχθηκαν από 10 παραδείγματα εκπαίδευσης για το κάθε μοντέλο.

AUC-ROC score

1	2	3	4	5	6	7	8	9	10	Avg
0.9638	0.9662	0.9674	0.9633	0.9651	0.9668	0.9645	0.9670	0.9655	0.9660	0.9656

Πίνακας 13: Αποτελέσματα για το AUC-ROC score του μοντέλου VGG16.

1	2	3	4	5	6	7	8	9	10	Avg
0.9603	0.9708	0.9604	0.9668	0.9740	0.9725	0.9715	0.9730	0.9720	0.9710	0.9702

Πίνακας 14: Αποτελέσματα για το AUC-ROC score του μοντέλου Xception.

Σχολιασμός AUC-ROC score

Οι μέσες τιμές AUC-ROC που υπολογίστηκαν σε 10 ανεξάρτητα παραδείγματα εκπαίδευσης επιβεβαιώνουν τα αποτελέσματα των ROC curves. Για το VGG16, η μέση τιμή είναι 0.9656, με μικρή διακύμανση, γεγονός που καταδεικνύει τη σταθερότητα του μοντέλου. Αντίστοιχα για το Xception, η μέση τιμή είναι υψηλότερη, 0.9702, γεγονός που αναδεικνύει τη βελτιωμένη συνολική ικανότητα διάκρισης των κατηγοριών.

Παρόλο που η διαφορά μεταξύ των δύο μοντέλων είναι σχετικά μικρή, το Xception υπερτερεί συστηματικά σε όλα σχεδόν τα παραδείγματα, ειδικά στις πιο απαιτητικές κατηγορίες.

7.6 Μετρικές και Confusion Matrix

Στο κεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα της αξιολόγησης των μοντέλων VGG16 και Xception για την ταξινόμηση των περιστατικών πνευμονίας, με έμφαση στις βασικές μετρικές απόδοσης. Οι μετρικές που χρησιμοποιούνται περιλαμβάνουν το Precision, το Recall, το F1-score και το Specificity, οι οποίες παρέχουν μια ολοκληρωμένη εικόνα της ικανότητας των μοντέλων να εντοπίζουν σωστά κάθε κατηγορία και να περιορίζουν τα σφάλματα ταξινόμησης. Παράλληλα, η ανάλυση των Confusion Matrices προσφέρει λεπτομερή πληροφορία σχετικά με τις περιπτώσεις όπου τα μοντέλα προβαίνουν σε ψευδώς θετικές ή ψευδώς αρνητικές προβλέψεις επιτρέποντας την αναγνώριση των κατηγοριών που αποτελούν πρόκληση για την ταξινόμηση.

Στους πίνακες των Precision, Recall, F1-score υπαρχουν και 2 γραμμες μετα τις κλασεις με το ονομα macro avg και weighted avg. Το macro average (δωσε ελληνικο ορο) υπολογίζει τον μέσο όρο της κάθε μετρικής χωρίς να λαμβάνει υπόψη τον αριθμό δειγμάτων σε κάθε κατηγορία, συνεπώς κάθε κατηγορία έχει ίση βαρύτητα ανεξαρτήτως του πόσα δείγματα περιέχει. Αντιθετα το weighted average υπολογίζει τον μέσο όρο της κάθε μετρικής λαμβάνοντας υπόψη τον αριθμό δειγμάτων κάθε κατηγορίας. Δηλαδή, οι μεγαλύτερες κατηγορίες έχουν μεγαλύτερη συνεισφορά στον μέσο όρο.

7.6.1 Precision (Ακρίβεια Θετικών Προβλέψεων)

Class / Metric	1	2	3	4	5	6	7	8	9	10	Avg
Bacterial	0.91	0.87	0.88	0.91	0.89	0.90	0.88	0.92	0.90	0.89	0.89
Normal	0.97	0.95	0.94	0.95	0.96	0.95	0.94	0.96	0.95	0.94	0.95
Viral	0.77	0.83	0.82	0.82	0.80	0.81	0.83	0.79	0.81	0.82	0.81
Macro avg	0.89	0.89	0.88	0.89	0.88	0.89	0.88	0.89	0.89	0.88	0.89
Weighted avg	0.90	0.89	0.89	0.90	0.89	0.90	0.89	0.90	0.90	0.89	0.90

Πίνακας 15: Αποτελέσματα Precision ανά κλάση για το μοντέλο VGG16.

Class / Metric	1	2	3	4	5	6	7	8	9	10	Avg
Bacterial	0.90	0.93	0.88	0.92	0.91	0.89	0.92	0.91	0.90	0.93	0.91
Normal	0.93	0.94	0.94	0.95	0.93	0.94	0.95	0.94	0.94	0.95	0.94
Viral	0.77	0.78	0.84	0.78	0.79	0.80	0.81	0.79	0.78	0.80	0.79
Macro avg	0.87	0.88	0.89	0.89	0.88	0.88	0.89	0.88	0.87	0.89	0.88
Weighted avg	0.88	0.90	0.89	0.90	0.89	0.89	0.90	0.89	0.88	0.90	0.89

Πίνακας 16: Αποτελέσματα Precision ανά κλάση για το μοντέλο Xception.

Το precision εκφράζει την ακρίβεια των θετικών προβλέψεων, δηλαδή το ποσοστό των περιπτώσεων που προβλέφθηκαν ως θετικές και ήταν πράγματι σωστές. Με άλλα λόγια, δείχνει πόσο «σύγουρο» είναι το μοντέλο όταν ταξινομεί ένα δείγμα σε μια συγκεκριμένη κλάση. Από τα αποτελέσματα φαίνεται ότι και τα δύο μοντέλα παρουσιάζουν πολύ υψηλό precision για τις κλάσεις Bacterial και Normal, με τιμές που κυμαίνονται περίπου στο 0.90–0.95, γεγονός που υποδηλώνει αξιόπιστη αναγνώριση αυτών των κατηγοριών. Αντίθετα, στην κλάση Viral οι τιμές είναι χαμηλότερες (0.79–0.81), φανερώνοντας ότι και οι δύο αρχιτεκτονικές αντιμετωπίζουν μεγαλύτερη δυσκολία στη σωστή ταξινόμηση των ιογενών περιστατικών. Σε επίπεδο σύγκρισης, το μοντέλο VGG16 παρουσιάζει μεγαλύτερη συνέπεια στις χαμηλότερες κατηγορίες, όπως το Viral ενώ το Xception υπερέχει ελαφρώς στη διάκριση της κλάσης Normal με υψηλότερο precision και σταθερότερη απόδοση. Συνεπώς, το VGG16 φαίνεται να είναι πιο ισορροπημένο ενώ το Xception δείχνει καλύτερη εξειδίκευση στην αναγνώριση των υγιών δειγμάτων.

7.6.2 Recall (Ανάκληση)

Class / Metric	1	2	3	4	5	6	7	8	9	10	Avg
Bacterial	0.95	0.97	0.97	0.97	0.96	0.95	0.96	0.97	0.95	0.96	0.96
Normal	0.89	0.83	0.89	0.87	0.85	0.88	0.86	0.89	0.87	0.88	0.87
Viral	0.84	0.88	0.77	0.78	0.81	0.80	0.82	0.79	0.83	0.80	0.81
Macro avg	0.89	0.89	0.88	0.87	0.87	0.88	0.88	0.88	0.88	0.88	0.88
Weighted avg	0.90	0.90	0.89	0.89	0.89	0.89	0.89	0.90	0.89	0.89	0.89

Πίνακας 17: Αποτελέσματα Recall ανά κλάση για το μοντέλο VGG16.

Class / Metric	1	2	3	4	5	6	7	8	9	10	Avg
Bacterial	0.86	0.90	0.95	0.93	0.91	0.92	0.94	0.91	0.90	0.93	0.92
Normal	0.95	0.92	0.93	0.90	0.91	0.92	0.94	0.91	0.92	0.93	0.92
Viral	0.80	0.84	0.75	0.84	0.79	0.80	0.78	0.82	0.81	0.80	0.80
Macro avg	0.87	0.89	0.88	0.89	0.87	0.88	0.89	0.88	0.89	0.88	0.88
Weighted avg	0.88	0.90	0.90	0.90	0.89	0.89	0.90	0.89	0.89	0.90	0.90

Πίνακας 18: Αποτελέσματα Recall ανά κλάση για το μοντέλο Xception.

Το recall μετρά την ικανότητα του μοντέλου να εντοπίζει σωστά όλες τις θετικές περιπτώσεις, δηλαδή το ποσοστό των πραγματικών θετικών που αναγνωρίστηκαν επιτυχώς. Πρόκειται για μετρική που δείχνει πόσο «ευαίσθητο» είναι το μοντέλο στην ανίχνευση περιστατικών μιας κλάσης. Στα αποτελέσματα, παρατηρείται ότι το VGG16 εμφανίζει εξαιρετικά υψηλό recall στην κατηγορία Bacterial (≈ 0.96), γεγονός που σημαίνει ότι αναγνωρίζει σχεδόν όλα τα περιστατικά βακτηριακής πνευμονίας. Αντίθετα, το recall στην κατηγορία Normal είναι χαμηλότερο (≈ 0.87), γεγονός που υποδεικνύει πως ορισμένα υγιή δείγματα ταξινομούνται εσφαλμένα ως παθολογικά. Στο μοντέλο Xception, οι τιμές recall είναι πιο ισορροπημένες ανάμεσα στις κατηγορίες, με ιδιαίτερα καλή επίδοση στην Normal (≈ 0.92), χωρίς όμως να φτάνει τα επίπεδα του VGG16 στην Bacterial. Συνολικά, το VGG16 δείχνει υπεροχή στην ανίχνευση βακτηριακής πνευμονίας, ενώ το Xception παρουσιάζει πιο ομοιόμορφη κατανομή της απόδοσης ανάμεσα στις κατηγορίες, με σημαντικό πλεονέκτημα στην αναγνώριση των υγιών δειγμάτων.

7.6.3 F1-score

Class/Metric	1	2	3	4	5	6	7	8	9	10	Avg
Bacterial	0.93	0.94	0.92	0.92	0.91	0.93	0.92	0.94	0.93	0.92	0.93
Normal	0.92	0.90	0.92	0.91	0.90	0.91	0.90	0.92	0.91	0.90	0.91
Viral	0.83	0.82	0.80	0.80	0.81	0.82	0.81	0.83	0.82	0.81	0.82
Accuracy	0.90	0.90	0.89	0.89	0.90	0.90	0.89	0.90	0.90	0.89	0.90
Macro avg	0.89	0.89	0.89	0.89	0.87	0.89	0.88	0.90	0.89	0.88	0.89
Weighted avg	0.90	0.90	0.89	0.89	0.89	0.90	0.89	0.90	0.90	0.89	0.90

Πίνακας 19: Αποτελέσματα F1-score ανά κλάση για το μοντέλο VGG16.

Class/Metric	1	2	3	4	5	6	7	8	9	10	Avg
Bacterial	0.88	0.92	0.92	0.92	0.91	0.90	0.93	0.91	0.90	0.92	0.91
Normal	0.94	0.93	0.94	0.93	0.94	0.93	0.95	0.94	0.93	0.94	0.94
Viral	0.79	0.81	0.79	0.81	0.80	0.79	0.81	0.80	0.80	0.79	0.80
Accuracy	0.88	0.90	0.90	0.90	0.89	0.89	0.91	0.90	0.89	0.90	0.90
Macro avg	0.87	0.89	0.88	0.89	0.88	0.87	0.90	0.88	0.87	0.88	0.88
Weighted avg	0.88	0.90	0.89	0.90	0.89	0.89	0.91	0.89	0.88	0.90	0.89

Πίνακας 20: Αποτελέσματα F1-score ανά κλάση για το μοντέλο Xception.

Το F1-score αποτελεί τον αρμονικό μέσο όρο του precision και του recall και προσφέρει μια ισορροπημένη αποτίμηση της απόδοσης ενός μοντέλου, ειδικά σε περιπτώσεις όπου υπάρχει ανισορροπία ανάμεσα στις κλάσεις. Για το VGG16, το F1-score είναι ιδιαίτερα υψηλό στην κατηγορία Bacterial (≈ 0.93), γεγονός που υποδηλώνει ότι το μοντέλο επιτυγχάνει καλή ισορροπία μεταξύ σωστής αναγνώρισης και περιορισμού ψευδών ταξινομήσεων. Η απόδοση στις κατηγορίες Normal (≈ 0.91) και Viral (≈ 0.82) είναι επίσης ικανοποιητική, με το χαμηλότερο σκορ να εντοπίζεται στα περιστατικά ιογενούς πνευμονίας. Αντίστοιχα, το Xception εμφανίζει πολύ κοντινά αποτελέσματα: υψηλή τιμή F1-score για την Normal (≈ 0.94), ελαφρώς χαμηλότερη για την Bacterial (≈ 0.91) και πάλι χαμηλότερη για την Viral (≈ 0.80). Η γενική εικόνα δείχνει ότι το VGG16 υπερτερεί ελαφρώς στην Bacterial, ενώ το Xception εμφανίζει ανώτερη απόδοση στην Normal. Σε όλες τις περιπτώσεις, η κατηγορία Viral αποτελεί τη μεγαλύτερη πρόκληση και για τα δύο μοντέλα, επιβεβαιώνοντας τη δυσκολία διάκρισης ιογενούς από βακτηριακή πνευμονία.

7.6.4 Specificity (Ειδικότητα)

Class	1	2	3	4	5	6	7	8	9	10	Avg
Bacterial	0.94	0.94	0.91	0.91	0.92	0.93	0.92	0.94	0.93	0.92	0.93
Normal	0.97	0.99	0.97	0.97	0.98	0.97	0.96	0.98	0.97	0.97	0.97
Viral	0.94	0.92	0.95	0.95	0.94	0.94	0.93	0.95	0.94	0.94	0.94

Πίνακας 21: Αποτελέσματα Specificity ανά κλάση για το μοντέλο VGG16.

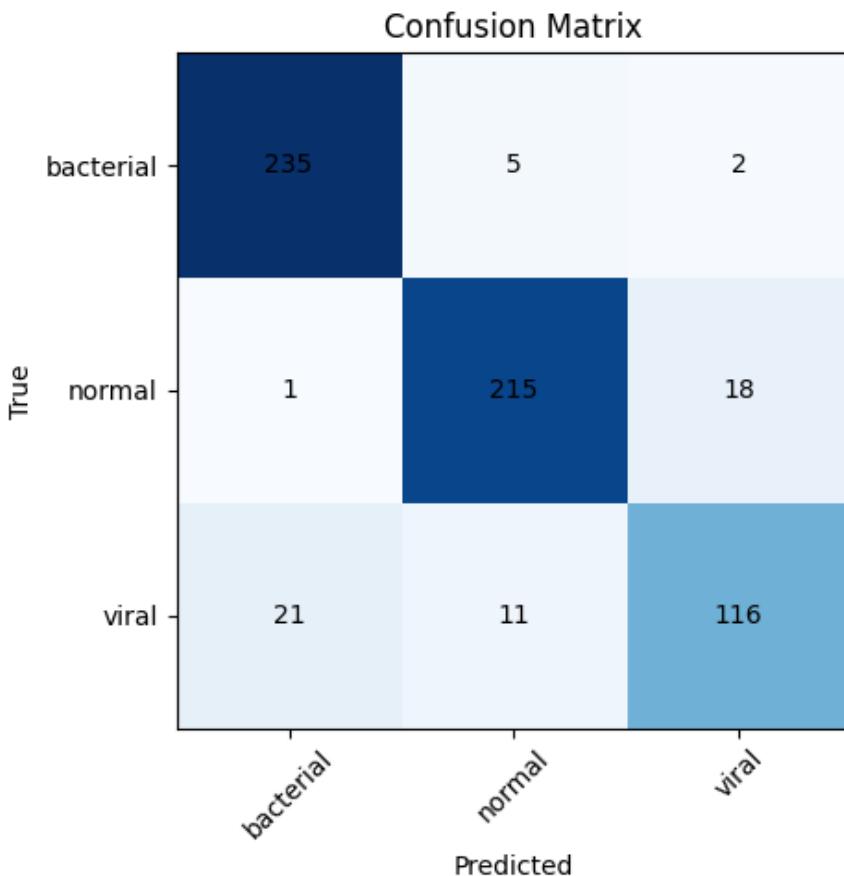
Class	1	2	3	4	5	6	7	8	9	10	Avg
Bacterial	0.94	0.96	0.92	0.95	0.94	0.95	0.96	0.95	0.94	0.95	0.95
Normal	0.96	0.96	0.96	0.97	0.96	0.96	0.97	0.96	0.96	0.97	0.96
Viral	0.92	0.93	0.96	0.93	0.94	0.94	0.95	0.93	0.94	0.94	0.94

Πίνακας 22: Αποτελέσματα Specificity ανά κλάση για το μοντέλο Xception.

Το specificity μετρά το ποσοστό των πραγματικά αρνητικών δειγμάτων που αναγνωρίζονται σωστά, δηλαδή την ικανότητα του μοντέλου να αποφεύγει τα ψευδώς θετικά. Στο πλαίσιο της πνευμονίας, υψηλό specificity σημαίνει ότι το μοντέλο ταξινομεί με ακρίβεια τα υγιή περιστατικά, μειώνοντας τον κίνδυνο λανθασμένης διάγνωσης και άσκοπης θεραπείας. Στα αποτελέσματα, τόσο το VGG16 όσο και το Xception εμφανίζουν πολύ υψηλές τιμές specificity σε όλες τις κατηγορίες ($\approx 0.93\text{--}0.97$). Το VGG16 παρουσιάζει ελαφρώς χαμηλότερη τιμή στην κατηγορία Bacterial (≈ 0.93), σε σχέση με το Xception (≈ 0.95), γεγονός που υποδηλώνει ότι το δεύτερο είναι πιο αποτελεσματικό στο να διακρίνει πότε δεν υπάρχει βακτηριακή λοίμωξη. Αντίστοιχα, στην κατηγορία Normal, και τα δύο μοντέλα επιτυγχάνουν εξαιρετικές επιδόσεις ($\approx 0.96\text{--}0.97$), με μικρά περιθώρια διαφοροποίησης. Συνολικά, το Xception φαίνεται να υπερέχει ελαφρώς στο specificity, καθώς περιορίζει καλύτερα τα ψευδώς θετικά σε όλες τις κατηγορίες, προσφέροντας μεγαλύτερη αξιοπιστία σε περιπτώσεις όπου η λανθασμένη διάγνωση θα μπορούσε να οδηγήσει σε περιττή ιατρική παρέμβαση.

7.6.5 Confusion Matrix

Το διάγραμμα Confusion Matrix αποτελεί ένα εξαιρετικά χρήσιμο εργαλείο για την οπτικοποίηση της απόδοσης ενός ταξινομητικού μοντέλου, καθώς παρουσιάζει τον αριθμό των σωστών και λανθασμένων προβλέψεων ανά κατηγορία. Κάθε γραμμή της μήτρας αντιστοιχεί στις πραγματικές κλάσεις, ενώ κάθε στήλη στις προβλεπόμενες. Με αυτόν τον τρόπο, μπορούμε να εντοπίσουμε ποιες κατηγορίες συγχέονται συχνότερα και πού εμφανίζονται τα περισσότερα ψευδώς θετικά ή ψευδώς αρνητικά αποτελέσματα. Η ανάλυση του Confusion Matrix λειτουργεί σαν επιβεβαίωση στα ευρήματα των προηγούμενων μετρικών



Εικόνα 20: Διάγραμμα Confusion Matrix για το μοντέλο Xception.

Ανάλυση Confusion Matrix

Η ανάλυση του διαγράμματος Confusion Matrix για το μοντέλο Xception δείχνει ότι η κατηγορία των βακτηριδιακών περιστατικών (Bacterial) αναγνωρίζεται εξαιρετικά καλά με το μοντέλο να ταξινομεί σωστά 235 από τα 242 περιστατικά δηλαδή περίπου 97%. Λίγα δείγματα ταξινομήθηκαν λανθασμένα, με πέντε να χαρακτηρίζονται ως Normal και δύο ως Viral γεγονός που υποδηλώνει πολύ χαμηλά ψευδώς θετικά και αρνητικά για αυτή την κατηγορία.

Η κατηγορία των υγιών περιστατικών (Normal) επιτυγχάνει σωστή αναγνώριση σε 215 από τα 234 δείγματα, περίπου 92%, με μόνο ένα δείγμα να ταξινομείται ως Bacterial και 18 ως Viral. Παρατηρείται ότι τα περισσότερα λάθη κατευθύνονται προς την κατηγορία Viral, υποδηλώνοντας ότι το μοντέλο δυσκολεύεται να διαχωρίσει ορισμένα υγιή δείγματα από τα ιογενή.

Στην κατηγορία των ιογενών περιστατικών (Viral), το μοντέλο αναγνωρίζει σωστά 116 από τα 148 δείγματα, περίπου 78%, ενώ 21 ταξινομήθηκαν ως Bacterial και 11 ως Normal. Αυτό δείχνει ότι η κατηγορία Viral παραμένει η πιο προβληματική, με σημαντικό αριθμό ψευδώς αρνητικών, κυρίως προς την κατηγορία Bacterial, επιβεβαιώνοντας τη γνωστή δυσκολία διάκρισης μεταξύ ιογενούς και βακτηριακής πνευμονίας που έχουμε δει σε όλη την εργασία.

Συνολικά το μοντέλο Xception αποδίδει εξαιρετικά στην αναγνώριση των Bacterial περιστατικών και καλά στην κατηγορία Normal, όπως επιβεβαιώνεται και από τα υψηλά Precision και Recall (βλ. 7.6.1, 7.6.2). Η κατηγορία Viral παραμένει η πιο δύσκολη, με αρκετά δείγματα να συγχέονται με Bacterial.

8. Οπτικοποίηση Αποτελεσμάτων και Ερμηνευσιμότητα Προβλέψεων

8.1 Εισαγωγή στην Οπτικοποίηση με Heatmaps

Η ανάπτυξη Συνελικτικών Νευρωνικών Δικτύων (CNNs) έχει συμβάλει καθοριστικά στην επίλυση προβλημάτων αναγνώρισης προτύπων, και ιδιαίτερα στην ιατρική απεικόνιση. Ωστόσο, ένα από τα βασικά μειονεκτήματα των CNNs είναι ότι συχνά θεωρούνται «μαύρα κουτιά», καθώς η διαδικασία λήψης αποφάσεων δεν είναι άμεσα κατανοητή από τον άνθρωπο [23]. Για την αντιμετώπιση αυτού του ζητήματος έχουν αναπτυχθεί μέθοδοι οπτικοποίησης, με κυριότερη την παραγωγή heatmaps.

Τα heatmaps αποτελούν χάρτες θερμότητας που επικαλύπτονται πάνω στην αρχική εικόνα και δείχνουν τις περιοχές στις οποίες το δίκτυο εστίασε περισσότερο για να καταλήξει στην τελική του πρόβλεψη. Με τον τρόπο αυτό επιτρέπουν την κατανόηση του «πώς» και «πού» ένα CNN κατευθύνει την προσοχή του κατά τη διαδικασία ταξινόμησης [33]. Στην περίπτωση της ιατρικής διάγνωσης, οι χάρτες αυτοί είναι ιδιαίτερα χρήσιμοι, καθώς προσφέρουν ένα επιπλέον επίπεδο διαφάνειας και ενισχύουν την εμπιστοσύνη των κλινικών ιατρών στα αποτελέσματα του μοντέλου. Η δυνατότητα ερμηνείας δεν είναι μόνο επιστημονικά επιθυμητή, αλλά αποτελεί και πρακτική προϋπόθεση για την κλινική υιοθέτηση της τεχνητής νοημοσύνης, όπου η λανθασμένη διάγνωση μπορεί να έχει σοβαρές συνέπειες [92].

Ανάμεσα στις μεθόδους που έχουν προταθεί, η Grad-CAM (Gradient-weighted Class Activation Mapping) έχει καθιερωθεί ως μία από τις πιο αποτελεσματικές και ταυτόχρονα υπολογιστικά απλές τεχνικές [93]. Για τον λόγο αυτό επιλέχθηκε και στην παρούσα εργασία, προκειμένου να διερευνηθεί η εσωτερική λειτουργία των μοντέλων VGG16 και Xception στις προβλέψεις τους για τους τύπους πνευμονίας.

8.2 Μεθοδολογία Grad-CAM

Η μέθοδος Grad-CAM βασίζεται στη χρήση των κλίσεων που προκύπτουν κατά τη διαδικασία της οπισθοδιάδοσης σφάλματος, ειδικά στο τελευταίο συνελικτικό επίπεδο του νευρωνικού δικτύου. Οι κλίσεις αυτές αποτυπώνουν πόσο σημαντική είναι κάθε χωρική θέση για την τελική πρόβλεψη της κλάσης. Με βάση αυτές τις πληροφορίες δημιουργείται ένας χάρτης ενεργοποίησης, ο οποίος προβάλλεται ως heatmap πάνω στην αρχική εικόνα, αναδεικνύοντας τις περιοχές που επηρέασαν περισσότερο την απόφαση του μοντέλου [93].

Στην παρούσα εργασία, τα heatmaps που προέκυψαν από την εφαρμογή του Grad-CAM επικαλύφθηκαν στις ακτινογραφίες θώρακος, ώστε να καταστεί ορατό σε ποιες περιοχές εστίαζαν τα μοντέλα κατά την ταξινόμηση των κατηγοριών (Bacterial, Viral, Normal). Κάθε εικόνα συνοδεύεται από το ποσοστό πρόβλεψης της επικρατέστερης κλάσης, προσφέροντας έτσι μια πιο άμεση κατανόηση της «σιγουριάς» του μοντέλου.

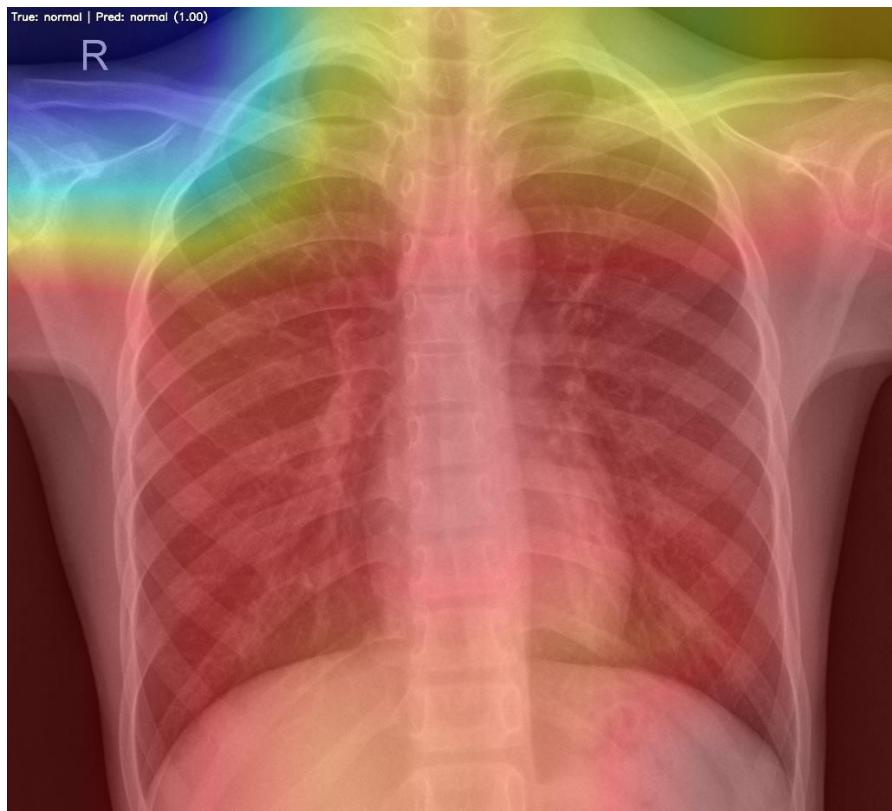
Ο κώδικας που αναπτύχθηκε για το Grad-CAM περιλαμβάνει τυχαία επιλογή δειγμάτων από το σύνολο δεδομένων για την κάθε κλάση, προκειμένου να παραχθεί ένα αντιτροσωπευτικό σύνολο παραδειγμάτων. Στην πράξη, για κάθε αρχιτεκτονική δημιουργήθηκαν heatmaps για δεκάδες εικόνες, καλύπτοντας όλες τις κλάσεις, με στόχο την εξαγωγή αξιόπιστων συμπερασμάτων σχετικά με τη συμπεριφορά των μοντέλων [94].

8.3 Παρουσίαση Αποτελεσμάτων

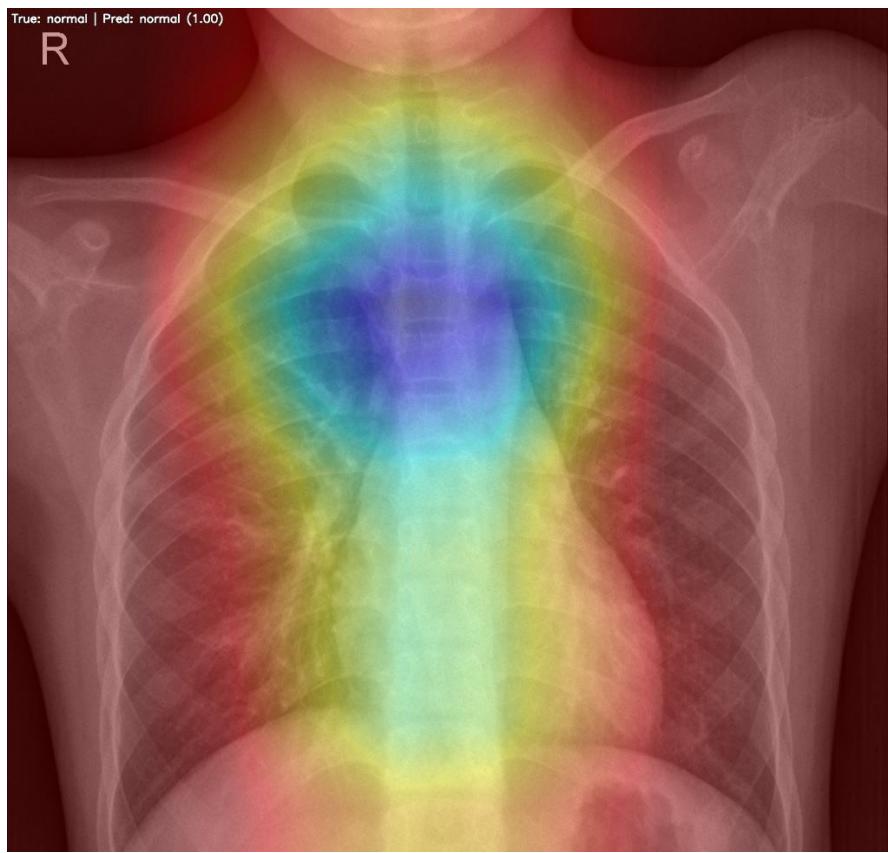
Η χρήση της μεθόδου Grad-CAM επέτρεψε την οπτικοποίηση της «προσοχής» των μοντέλων VGG16 και Xception στις ακτινογραφίες θώρακος, προσφέροντας πολύτιμη εικόνα για τον τρόπο με τον οποίο λαμβάνονται οι αποφάσεις. Στο παρόν εδάφιο παρουσιάζονται ενδεικτικά heatmaps για κάθε κλάση (Normal, Bacterial, Viral), καθώς και μια συγκριτική ανάλυση των δύο αρχιτεκτονικών.

8.3.1 Ενδεικτικά Παραδείγματα Heatmaps

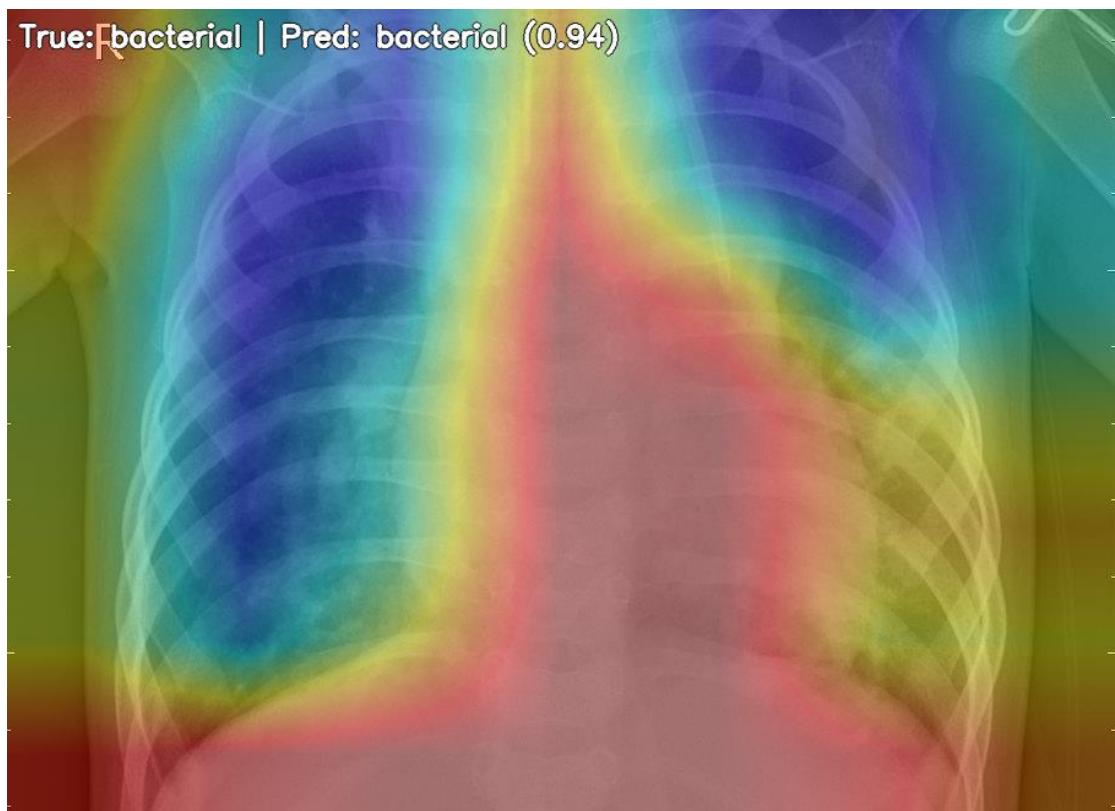
Ακολουθούν μερικά ενδεικτικά παραδείγματα των heatmaps για κάθε κλάση. Σε κάθε εικόνα στην πάνω αριστερή γωνιά αναγράφονται, η πραγματική κλάση, η προβλεπόμενη κλάση, καθώς και το ποσοστό «αυτοπεποίθησης» με το οποίο κάνει την πρόβλεψη το μοντέλο. Οι εικόνες που παρουσιάζονται επιλέχθηκαν επειδή συνοδεύονταν από υψηλά ποσοστά πρόβλεψης, ώστε να αποτυπωθεί με μεγαλύτερη σαφήνεια το πού εστιάζουν τα μοντέλα κατά τη διαδικασία ταξινόμησης. Αντίθετα, οι λανθασμένες ή οι χαμηλής βεβαιότητας προβλέψεις δεν θα προσέφεραν ουσιαστικές πληροφορίες για τη μελέτη της εσωτερικής λειτουργίας των δικτύων, καθώς η εστίαση σε αυτές τις περιπτώσεις είναι λιγότερο αντιπροσωπευτική.



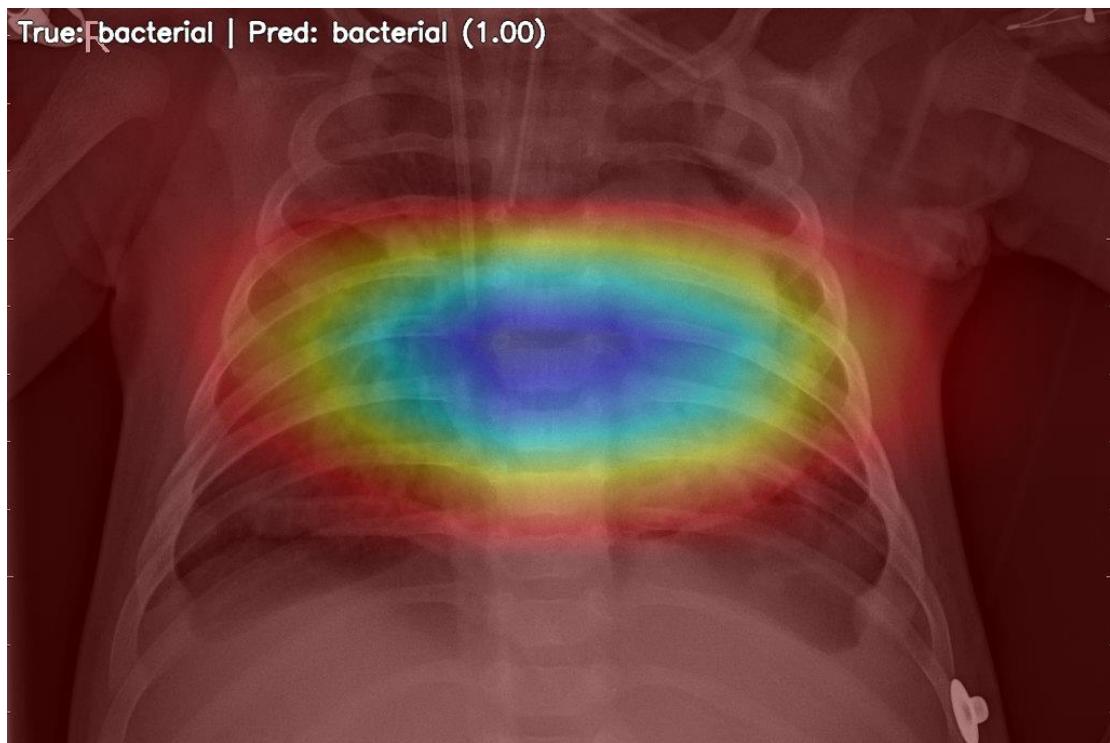
Εικόνα 21: Παράδειγμα Grad-CAM heatmap για την κλάση *normal* του μοντέλου VGG16



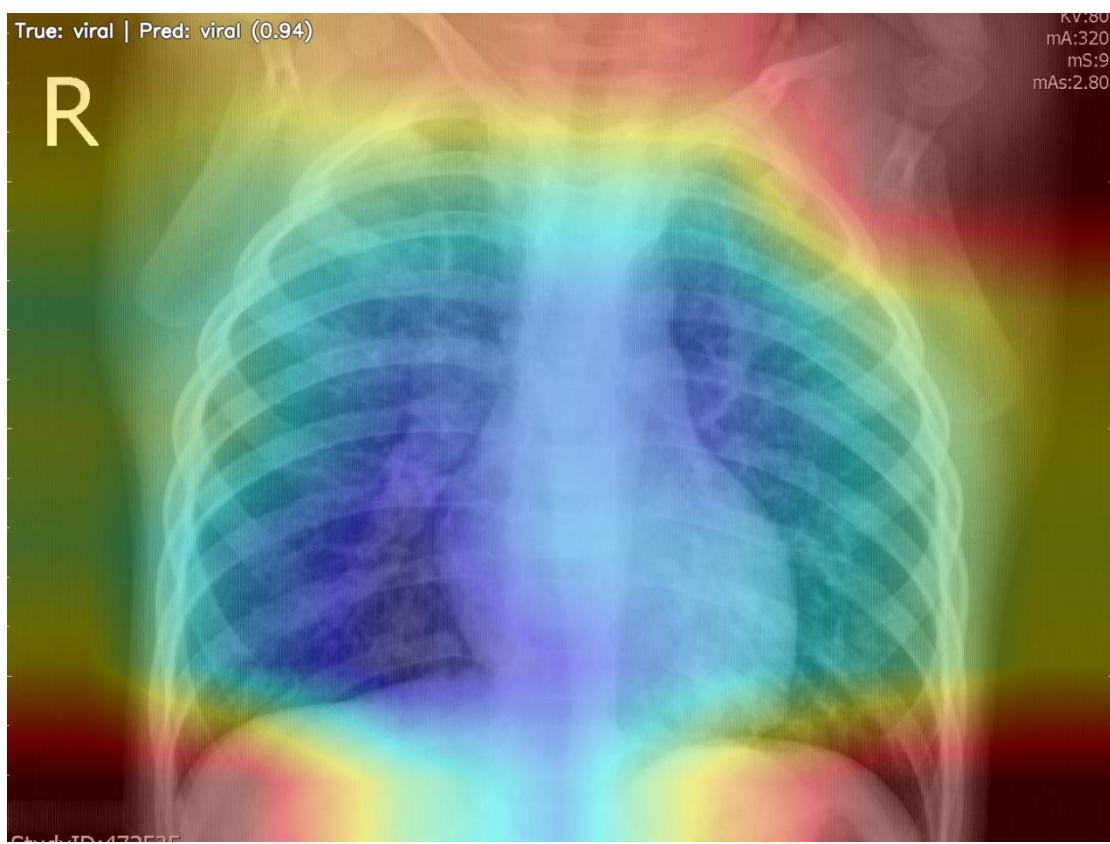
Εικόνα 22: Παράδειγμα Grad-CAM heatmap για την κλάση *normal* του μοντέλου *Xception*



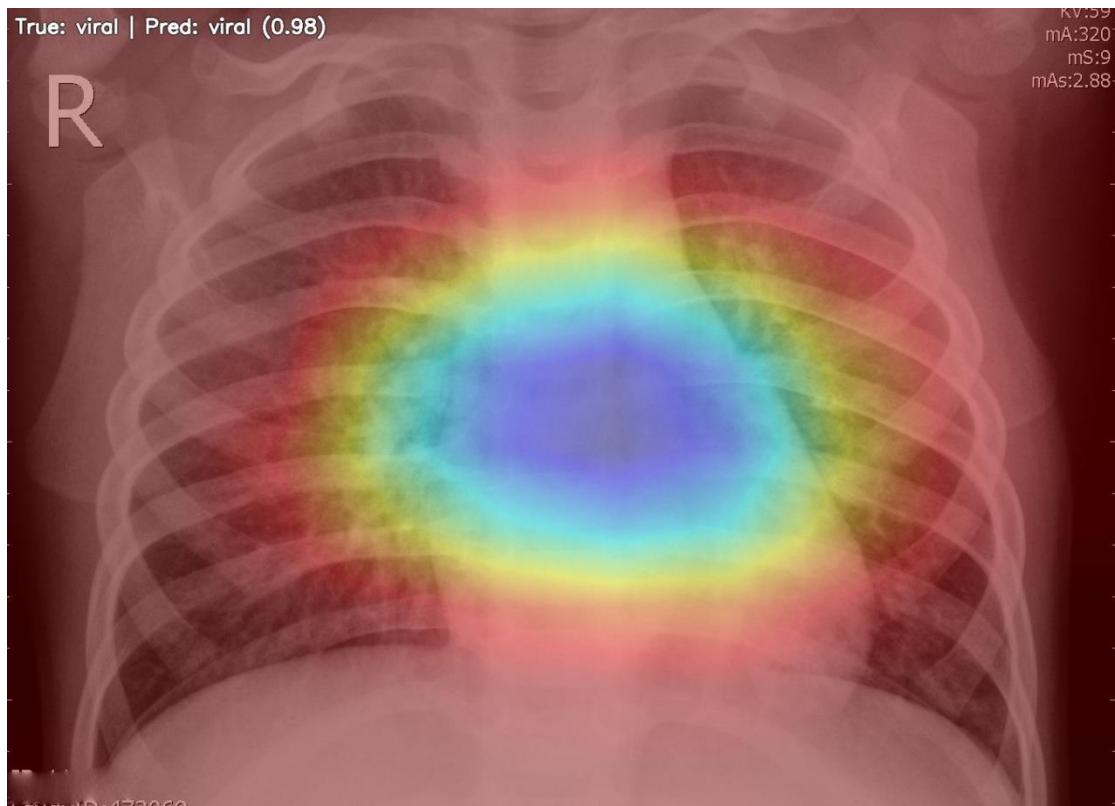
Εικόνα 23: Παράδειγμα Grad-CAM heatmap για την κλάση *bacterial* του μοντέλου *VGG16*



Εικόνα 24: Παράδειγμα Grad-CAM heatmap για την κλάση bacterial του μοντέλου Xception



Εικόνα 25: Παράδειγμα Grad-CAM heatmap για την κλάση viral του μοντέλου VGG16



Εικόνα 26: Παράδειγμα Grad-CAM heatmap για την κλάση viral του μοντέλου Xception

Ανάλυση heatmaps

Στις παραπάνω εικόνες παρουσιάζονται τα heatmaps που προέκυψαν μέσω της μεθόδου Grad-CAM, επικαλυμμένα πάνω στις αρχικές ακτινογραφίες. Τα χρώματα που βλέπουμε αντιστοιχούν στις περιοχές στις οποίες το μοντέλο εστίασε για να καταλήξει στην πρόβλεψή του. Γενικά το ποσό «σημασία» δίνει το μοντέλο πηγαίνει από τα πιο θερμά στα πιο ψυχρά χρώματα. Οι κόκκινες και κίτρινες περιοχές υποδεικνύουν υψηλή συμβολή στη λήψη απόφασης, ενώ οι μπλε περιοχές μικρότερη σημασία. Με τον τρόπο αυτό μπορούμε να κατανοήσουμε πού «κοιτάζει» το δίκτυο κατά την ταξινόμηση και να ερμηνεύσουμε τις προβλέψεις του. Για παράδειγμα, σε περιπτώσεις βακτηριδιακής πνευμονίας παρατηρείται έντονη εστίαση σε περιοχές με πύκνωση του πνευμονικού παρεγχύματος, ενώ σε φυσιολογικές ακτινογραφίες η προσοχή κατανέμεται πιο ομοιόμορφα.

Συμπεράσματα αποτελεσμάτων heatmaps

Normal κλάση: Για την κατηγορία Normal, το δίκτυο VGG16 προβάλλει εστίαση σε συμμετρικές περιοχές των πνευμόνων, επιβεβαιώνοντας την ορθότητα της πρόβλεψης με υψηλό ποσοστό εμπιστοσύνης. Αντίστοιχα, το μοντέλο Xception εμφανίζει ομοιόμορφη εστίαση, η οποία περιορίζεται σε μη παθολογικές περιοχές, γεγονός που οδηγεί σε σωστή διάγνωση με απόλυτη εμπιστοσύνη.

Bacterial κλάση: Στην περίπτωση της κατηγορίας bacterial, το δίκτυο VGG16 εστιάζει σε συγκεκριμένες περιοχές του πνευμονικού πεδίου· ωστόσο η προσοχή του παραμένει περιορισμένη και δεν ευθυγραμμίζεται πάντοτε πλήρως με τις παθολογικές περιοχές που θα αναγνώριζαν οι ακτινολόγοι. Αντίθετα, το μοντέλο Xception παρουσιάζει εκτεταμένη και

πιο ιατρικά λογική εστίαση, καλύπτοντας με συνέπεια τις περιοχές όπου εμφανίζονται τα διηθήματα.

Viral κλάση: Για την κατηγορία Viral Pneumonia, το δίκτυο VGG16 παρουσιάζει διάχυτη εστίαση, η οποία συχνά κατευθύνεται και σε μη κλινικά σημαντικά σημεία. Το γεγονός αυτό εξηγεί τα χαμηλότερα ποσοστά επιτυχίας που καταγράφονται στην κατηγορία αυτή. Αντίθετα, το μοντέλο Xception εμφανίζει επίσης διάχυτη, αλλά πιο ακριβή εστίαση, καθώς επικεντρώνεται σε περιοχές που πράγματι σχετίζονται με ιογενείς λοιμώξεις, αν και η σαφής κλινική διαφοροποίηση παραμένει μια απαιτητική διαδικασία.

8.3.2 Ανάλυση Ποσοστών Πρόβλεψης Heatmaps

Εκτός από τη χωρική εστίαση που αναδεικνύουν τα heatmaps, είναι σημαντικό να εξετάσουμε και τα ποσοστά πρόβλεψης που συνοδεύουν κάθε κατηγορία. Τα ποσοστά αυτά αποτυπώνουν το επίπεδο «βεβαιότητας» του μοντέλου στις αποφάσεις του, αναδεικνύοντας τόσο τα σημεία όπου η πρόβλεψη είναι ξεκάθαρη όσο και τις περιπτώσεις στις οποίες το μοντέλο δυσκολεύεται. Στη συνέχεια, παρουσιάζονται συγκεντρωτικοί πίνακες με τα ποσοστά πρόβλεψης για ένα σύνολο 100 δειγμάτων από κάθε κλάση, τόσο για το VGG16 όσο και για το Xception.

Avg per class	Pred. Bacterial	Pred. Normal	Pred. Viral
True Bacterial	0.87 (87)	–	0.61 (13)
True Normal	0.52 (1)	0.96 (95)	0.57 (4)
True Viral	0.74 (15)	0.84 (2)	0.86 (83)

Πίνακας 23: Μέσο Confidence προβλέψεων για το μοντέλο VGG16

Avg per class	Pred. Bacterial	Pred. Normal	Pred. Viral
True Bacterial	0.88 (94)	–	0.57 (6)
True Normal	0.48 (1)	0.97 (98)	0.64 (1)
True Viral	0.81 (15)	0.69 (2)	0.85 (83)

Πίνακας 24: Μέσο Confidence προβλέψεων για το μοντέλο Xception

Σχολιασμός Προβλέψεων

Στους παραπανω πινακες παρουσιάζονται οι μεσοι οροι των confidence scores ανα κλαση (True class) και ανα προβλεπομενη κατηγοραι (Predicted class), μαζι με τον αριθμο των δειγματων.

Ξεκινώντας με το VGG16, παρατηρούμε ότι στην κατηγορία Bacterial το μοντέλο ταξινομεί σωστά τη μεγάλη πλειονότητα των δειγμάτων (94), εμφανίζοντας υψηλό μέσο confidence (0.88). Ωστόσο, καταγράφονται και αρκετές λανθασμένες προβλέψεις προς την κατηγορία Viral (6), με μέσο confidence (0.57), γεγονός που φανερώνει μια σχετική σύγχυση μεταξύ

αυτών των δύο τύπων πνευμονίας. Στην κατηγορία Normal, η απόδοση είναι εξαιρετική, με τη συντριπτική πλειονότητα των δειγμάτων να ταξινομούνται σωστά (98) και με μεγάλη βεβαιότητα (0.97), ενώ τα λίγα λάθη που παρατηρούνται κατανέμονται προς τις άλλες δύο κατηγορίες (1 προς Bacterial με 0.52 και 1 προς Viral με 0.64). Στην κατηγορία Viral, το VGG16 επιτυγχάνει ικανοποιητικά αποτελέσματα, με τα περισσότερα δείγματα να αναγνωρίζονται σωστά (83 με 0.85), αλλά εμφανίζεται και πάλι μια συχνή σύγχυση προς την κατηγορία Bacterial (15 με 0.74) και λιγότερο προς την κατηγορία Normal (2 με 0.69), με σχετικά υψηλή μέση βεβαιότητα στις λανθασμένες ταξινομήσεις.

Το Xception, από την άλλη πλευρά, δείχνει πιο σταθερή απόδοση. Στην κατηγορία Bacterial επιτυγχάνει υψηλή ακρίβεια με 87 σωστές ταξινομήσεις (0.87) και ακόμη λιγότερες λανθασμένες προβλέψεις προς Viral (13 με 0.61) σε σχέση με το VGG16. Στην κατηγορία Normal καταγράφει την καλύτερη απόδοση από όλες, με σχεδόν όλα τα δείγματα να αναγνωρίζονται σωστά (95 με 0.96) και τα ελάχιστα λάθη να συνοδεύονται από χαμηλά confidence scores (1 προς Bacterial με 0.48 και 4 προς Viral με 0.57). Στην κατηγορία Viral, η απόδοση είναι επίσης πολύ καλή, με το μοντέλο να αναγνωρίζει τα περισσότερα δείγματα σωστά (83 με 0.86), ενώ οι λανθασμένες ταξινομήσεις προς Bacterial (15 με 0.81) και Normal (2 με 0.84) είναι περιορισμένες και με χαμηλότερη μέση βεβαιότητα σε σχέση με το VGG16.

Συγκρίνοντας τα δύο μοντέλα, το Xception υπερτερεί στη συνολική του ακρίβεια και εμφανίζει μεγαλύτερη σταθερότητα στις προβλέψεις του. Η κατηγορία Normal αναγνωρίζεται με τον πιο αξιόπιστο τρόπο και από τα δύο μοντέλα, ωστόσο η κύρια πρόκληση εντοπίζεται στη διάκριση μεταξύ Bacterial και Viral, όπου το VGG16 παρουσιάζει πιο έντονη σύγχυση. Συνεπώς, το Xception φαίνεται να αποτελεί την πιο αποτελεσματική επιλογή για το συγκεκριμένο πρόβλημα ταξινόμησης. Τα αποτελέσματα σε αυτό το εδάφιο συνάδουν με τα ευρήματα του confusion matrix (βλ. 7.6.5), δίνοντάς μας έτσι ακόμα καλύτερη κατανόηση της συμπεριφοράς των μοντέλων.

9. Συμπεράσματα και μελλοντικές προοπτικές

9.1 Συμπεράσματα

Η παρούσα εργασία δεν περιορίστηκε απλώς στην επιβεβαίωση ότι τα σύγχρονα Συνελικτικά Νευρωνικά Δίκτυα αποτελούν ισχυρό εργαλείο για την ανίχνευση και τη διάκριση τύπων πνευμονίας από ακτινογραφίες θώρακος. Προχώρησε ένα βήμα παραπέρα, διερευνώντας συστηματικά τον ρόλο της προσαρμογής προεκπαίδευσης μοντέλων σε συγκεκριμένα ιατρικά δεδομένα και αποδεικνύοντας ότι κατάλληλες επιλογές στις τεχνικές επεξεργασίας και στη ρύθμιση υπερπαραμέτρων μπορούν να βελτιώσουν ουσιαστικά την απόδοση στις πιο απαιτητικές κλάσεις, όπως η ιογενής πνευμονία. Με αυτόν τον τρόπο, η εργασία προσέθεσε πειραματικά δεδομένα και ανάλυση σε ένα πεδίο όπου η βιβλιογραφία παραμένει ακόμα περιορισμένη, ενισχύοντας τη συζήτηση για τη χρησιμότητα των CNNs στην κλινική πράξη.

Η χρήση προεκπαίδευσης μοντέλων αρχιτεκτονικών (VGG16, Xception) με τεχνικές μεταφοράς μάθησης και στοχευμένης περαιτέρω εκπαίδευσης («fine-tuning» → «λεπτομερής επαναπροσαρμογή») απέδωσε σταθερά και αξιόπιστα αποτελέσματα. Η αξιοποίηση των βαρών του ImageNet μείωσε τον χρόνο εκπαίδευσης και βελτίωσε την απόδοση σε ιατρικά σύνολα δεδομένων περιορισμένου μεγέθους, ενώ η λεπτομερής ανάλυση της συμπεριφοράς των μοντέλων ανά κλάση ανέδειξε τα δυνατά και τα αδύνατα σημεία τους. Τα αποτελέσματα έδειξαν υψηλή επίδοση στη διάκριση βακτηριακής πνευμονίας και φυσιολογικών περιπτώσεων, αλλά σαφώς χαμηλότερες επιδόσεις στην κλάση της ιογενούς πνευμονίας, όπως φάνηκε τόσο από τις χαμηλότερες τιμές F1-score (βλ. 7.6.3) όσο και από τα σφάλματα που καταγράφηκαν στις confusion matrices (βλ. 7.6.5). Το φαινόμενο αυτό μπορεί να αποδοθεί στην ενδοκλασσιακή ποικιλία και στην απεικονιστική ομοιότητα με άλλες καταστάσεις, που καθιστούν δυσκολότερη την ταξινόμηση.

Ένα από τα στοιχεία που διαφοροποιούν την παρούσα εργασία είναι η συστηματική ανάλυση της κλάσης Viral. Ενώ οι περισσότερες μελέτες αναφέρουν τις δυσκολίες ταξινόμησης μεταξύ ιογενούς και βακτηριδιακής πνευμονίας, η εργασία αυτή έδειξε με μετρικούς και γραφικούς τρόπους (F1-score, confusion matrices, Grad-CAM heatmaps) πού ακριβώς εντοπίζονται τα προβλήματα και ποιες περιοχές εικόνων οδηγούν σε λανθασμένες προβλέψεις. Με αυτόν τον τρόπο δημιουργείται μια πιο λεπτομερής βάση για μελλοντική βελτίωση της απόδοσης στη συγκεκριμένη κατηγορία.

Μεταξύ των δύο αρχιτεκτονικών, το Xception εμφάνισε καλύτερη συμπεριφορά όσον αφορά την απώλεια (loss) και την ικανότητα γενίκευσης, ενώ το VGG16 λειτούργησε ως αξιόπιστο σημείο αναφοράς (baseline). Η τελική επιλογή κατάλληλου μοντέλου δεν μπορεί να βασιστεί αποκλειστικά στην ακρίβεια (accuracy), ειδικά όταν οι διαφορές είναι μικρές. Η αξιολόγηση οφείλει να λαμβάνει υπόψη επιπλέον μετρικές, όπως precision, recall, F1-score, καμπύλες ROC/AUC και δείκτες βαθμονόμησης πιθανοτήτων, καθώς και πρακτικούς παράγοντες όπως το υπολογιστικό κόστος, ο χρόνος εκπαίδευσης και η απόδοση σε πραγματικές συνθήκες χρήσης («inference latency» → «καθυστέρηση κατά την πρόβλεψη»).

Επιπλέον, σε αντίθεση με αρκετές προγενέστερες μελέτες που αξιολογούν μόνο την ακρίβεια, εδώ δόθηκε έμφαση σε ένα πλήρες πλαίσιο μετρικών, ώστε να υπάρχει μια πολυδιάστατη εικόνα της συμπεριφοράς των μοντέλων. Η μεθοδολογία αυτή ανέδειξε ότι η «καλύτερη» αρχιτεκτονική δεν είναι μόνο εκείνη που επιτυγχάνει την υψηλότερη συνολική

ακρίβεια, αλλά εκείνη που εξισορροπεί την απόδοση μεταξύ των διαφορετικών κλάσεων και μπορεί να ενσωματωθεί πρακτικά στην κλινική ροή.

Η εφαρμογή τεχνικών οπτικοποίησης (π.χ. Grad-CAM) έδειξε ότι τα μοντέλα βασίζουν τις προβλέψεις τους σε περιοχές της εικόνας που αντιστοιχούν συχνά σε κλινικά συναφείς ανατομικές δομές, όπως τα πνευμονικά πεδία ή οι περιοχές με εμφανείς διηθήσεις. Το εύρημα αυτό ενισχύει την αξιοπιστία των δικτύων ως υποστηρικτικών εργαλείων, χωρίς ωστόσο να αναιρεί την ανάγκη για ιατρική κρίση και ανθρώπινη εποπτεία. Η παρατήρηση ότι ορισμένες φορές οι χάρτες εστίασης περιλαμβαναν και περιοχές εκτός του αναμενόμενου κλινικού πλαισίου, υπογραμμίζει την ανάγκη για περαιτέρω μελέτη στην κατεύθυνση της βελτίωσης της ερμηνευσιμότητας.

Η εργασία προχώρησε σε μια πιο ενδελεχή μελέτη της ερμηνευσιμότητας μέσω Grad-CAM σε σύγκριση με άλλες μελέτες που συχνά περιορίζονται σε μετρικούς δείκτες. Η συσχέτιση των χαρτών θερμότητας με ιατρικά λογικές περιοχές των ακτινογραφιών ανέδειξε ότι τα μοντέλα δεν «μαθαίνουν τυχαία», αλλά συχνά στηρίζουν τις αποφάσεις τους σε περιοχές που συνάδουν με κλινικά σημεία. Αυτό το εύρημα ενισχύει την εμπιστοσύνη στη χρήση τέτοιων εργαλείων από γιατρούς και ανοίγει τον δρόμο για περαιτέρω συνεργασία μεταξύ μηχανικών και κλινικών ερευνητών.

Τέλος, η εργασία ανέδειξε με σαφήνεια ότι η μετάβαση από το στάδιο της πειραματικής εφαρμογής σε πραγματική κλινική αξιοποίηση απαιτεί μια σειρά από βήματα: (α) επέκταση και επικύρωση των αποτελεσμάτων με πολυκεντρικά δεδομένα, (β) ανάπτυξη στρατηγικών αντιμετώπισης της ανισορροπίας των κλάσεων, (γ) μείωση του υπολογιστικού κόστους μέσω συμπίεσης και βελτιστοποίησης μοντέλων για χρήση σε νοσοκομειακό περιβάλλον, και (δ) ενσωμάτωση μηχανισμών διαφάνειας και ελέγχου ώστε να διασφαλίζεται η δίκαιη και ασφαλής χρήση τους.

Συνολικά, η διπλωματική αυτή εργασία απέδειξε ότι τα CNNs μπορούν να συνεισφέρουν ουσιαστικά στη διάγνωση πνευμονίας, ενώ ταυτόχρονα ανέδειξε τα σημεία στα οποία απαιτείται περαιτέρω έρευνα. Με την εστίαση σε πιο ισορροπημένα δεδομένα, πιο διαφανή και ερμηνεύσιμα μοντέλα και πρακτικές προσαρμοσμένες στις ανάγκες του κλινικού χώρου, η μελλοντική αξιοποίηση των μεθόδων αυτών μπορεί να μετατραπεί σε πραγματικό εργαλείο υποστήριξης για την ιατρική κοινότητα.

9.2 Μελλοντικές προοπτικές

Έχοντας καταλήξει στο συμπέρασμα ότι η τεχνητή νοημοσύνη και ειδικότερα τα συνελικτικά νευρωνικά δίκτυα αποτελούν ένα πολλά υποσχόμενο εργαλείο για τη διάγνωση ασθενειών, κρίνεται απαραίτητο να σχεδιαστεί ένα συνεκτικό πλάνο για τα επόμενα στάδια της έρευνας. Αρχικά, αξίζει να επεκτείνουμε και να εμπλουτίσουμε το σύνολο δεδομένων με εικόνες από διαφορετικά κλινικά κέντρα, ηλικιακές ομάδες και τεχνολογίες απεικόνισης με αυτόν τον τρόπο μειώνεται η πιθανότητα μεροληψίας και αυξάνεται η ικανότητα γενίκευσης των μοντέλων. Παράλληλα, είναι σκόπιμο να προστεθούν νέες περιπτώσεις στο υπάρχον σύνολο ώστε το δίκτυο να «δει» μεγαλύτερη ποικιλία κλινικών εικόνων και, κατά συνέπεια, να βελτιωθεί η απόδοσή του σε σπάνιες ή δυσδιάκριτες κατηγορίες.

Στο τεχνικό επόπεδο, προτείνεται η εφαρμογή πιο στοχευμένων τεχνικών αύξησης δεδομένων (augmentation) και μεθόδων εξισορρόπησης κλάσεων για τη βελτίωση της

αναγνώρισης της κλάσης Viral, που όπως φάνηκε αποτελεί τον κύριο περιορισμό της παρούσας μελέτης. Επιπλέον, η δοκιμή εναλλακτικών στρατηγικών εκπαίδευσης —όπως weighted loss, focal loss ή ensemble approaches— μπορεί να οδηγήσει σε ουσιαστική βελτίωση χωρίς απαραίτητα να απαιτεί πολύ μεγαλύτερους πόρους. Η χρήση ensemble (συνδυασμός πολλαπλών μοντέλων), ειδικότερα, μπορεί να συνδυάσει τα δυνατά σημεία του VGG16 και του Xception και να μειώσει τις αδυναμίες κάθε μεμονωμένης αρχιτεκτονικής.

Για να αξιοποιηθούν μεγάλοι όγκοι μη ετικετοποιημένων εικόνων, αξίζει να εξεταστούν μέθοδοι self-supervised ή semi-supervised pretraining, οι οποίες μπορούν να βελτιώσουν τις αναπαραστάσεις πριν το τελικό fine-tuning με ετικέτες. Παράλληλα, τεχνικές domain-adaptation θα βοηθήσουν ώστε ένα μοντέλο που εκπαιδεύτηκε σε μία πηγή να προσαρμοστεί ευκολότερα σε εικόνες από άλλα νοσοκομεία ή διαφορετικά μηχανήματα απεικόνισης.

Σε ό,τι αφορά την πρακτική εφαρμογή, χρειάζεται να σχεδιαστούν ελαφρύτερες εκδόσεις των μοντέλων (pruning, quantization, knowledge distillation) για να μπορούν να αναπτυχθούν σε ιατρικό εξοπλισμό με περιορισμένες δυνατότητες. Η μέτρηση του latency και της κατανάλωσης μνήμης στο περιβάλλον στόχο θα πρέπει να γίνει μέρος της αξιολόγησης πριν από οποιαδήποτε παραγωγική υλοποίηση. Ταυτόχρονα, η ανάπτυξη φιλικών διεπαφών χρήστη και εκπαιδευτικών προγραμμάτων για το ιατρικό προσωπικό είναι αναγκαία, ώστε τα εργαλεία να χρησιμοποιηθούν σωστά και με ασφάλεια.

Τέλος, δεν πρέπει να παραβλέπονται τα ηθικά και ρυθμιστικά ζητήματα: απαιτείται συμμόρφωση με τοπικούς και ευρωπαϊκούς κανονισμούς (π.χ. GDPR), διαδικασίες audit, μέτρηση fairness ανά υποομάδα και έλεγχος για πιθανές μεροληψίες. Η πραγματική αξία ενός τέτοιου συστήματος θα επιβεβαιωθεί μόνο μετά από πολυκεντρική, εξωτερική επικύρωση και, σε ιδανική περίπτωση, προοπτικές κλινικές δοκιμές που θα αξιολογήσουν την επίδρασή του στην κλινική διαδικασία και στα αποτελέσματα για τον ασθενή.

Βιβλιογραφία

- [1] M. Turing, *Computing machinery and intelligence.*, Netherlands: Springer, 2009.
- [2] Moor, James. "The Dartmouth College artificial intelligence conference: The next fifty years." *AI Magazine* 27.4 (2006): 87-87.
- [3] Gugerty, Leo. "Newell and Simon's logic theorist: Historical background and impact on cognitive modeling." *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 50. No. 9. Sage CA: Los Angeles, CA: SAGE Publications, 2006.
- [4] Weizenbaum, Joseph. "ELIZA—a computer program for the study of natural language communication between man and machine." *Communications of the ACM* 9.1 (1966): 36–45.
- [5] Winograd, Terry. "What does it mean to understand language?." *Cognitive science* 4.3 (1980): 209–241.
- [6] Roumeliotis, Konstantinos I., and Nikolaos D. Tselikas. "Chatgpt and open-ai models: A preliminary review." *Future Internet* 15.6 (2023): 192.
- [7] I. S. a. G. E. H. Krizhevsky, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, 2012.
- [8] Rajpurkar, P., Irvin, J., Zhu, K., et al. (2017). *CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning*. arXiv preprint arXiv:1711.05225.
- [9] De Fauw, J., Ledsam, J. R., Romera-Paredes, B., et al. (2018). *Clinically applicable deep learning for diagnosis and referral in retinal disease*. Nature Medicine, 24(9), 1342–1350.
- [10] Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). *Dermatologist-level classification of skin cancer with deep neural networks*. Nature, 542(7639), 115–118.
- [11] Rajkomar, A., Oren, E., Chen, K., et al. (2018). *Scalable and accurate deep learning with electronic health records*. NPJ Digital Medicine, 1(1), 18.
- [12] Kandoth, C., McLellan, M. D., Vandin, F., et al. (2013). *Mutational landscape and significance across 12 major cancer types*. Nature, 502(7471), 333–339.
- [13] Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., et al. (2019). *Deep learning enables rapid identification of potent DDR1 kinase inhibitors*. Nature Biotechnology, 37(9), 1038–1040.
- [14] Attanasio, A., et al. (2020). *Robotic Surgery in Urology: State of the Art and Future Perspectives*. Minerva Urology and Nephrology, 72(3), 355–366.
- [15] Shademan, A., Decker, R. S., Opfermann, J. D., et al. (2016). *Supervised autonomous robotic soft tissue surgery*. Science Translational Medicine, 8(337), 337ra64.
- [16] Wang, Y., Wang, L., Rastegar-Mojarad, M., et al. (2018). *Clinical information extraction applications: A literature review*. Journal of Biomedical Informatics, 77, 34–49.
- [17] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). *Dissecting racial bias in an algorithm used to manage the health of populations*. Science, 366(6464), 447–453.
- [18] Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). *Data mining for credit card fraud: A comparative study*. Decision Support Systems, 50(3), 602–613.
- [19] Karpathy, A. (2020). *Tesla Autopilot: Full Self-Driving and AI*. Tesla AI Day, Presentation.
- [20] Zhao, W. X., Mu, X., Zhang, Y., et al. (2019). *YouTube Watch Next Recommendations: A Deep Learning Approach*. ACM Transactions on the Web (TWEB), 13(4), 1–27.
- [21] Zeng, Y., Zhang, M., & Wang, M. (2018). *Smart home technologies based on machine learning and IoT*. Journal of Sensor and Actuator Networks, 7(3), 23.
- [22] Silver, D., Huang, A., Maddison, C. J., et al. (2016). *Mastering the game of Go with deep neural networks and tree search*. Nature, 529(7587), 484–489.
- [23] LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*. Nature, 521(7553).
- [24] Zhang, G. P., Eddy Patuwo, B., & Hu, M. Y. (1998). *Forecasting with artificial neural networks: The state of the art*. International Journal of Forecasting, 14(1), 35–62.
- [25] Srivastava N., Hinton G., Krizhevsky A., Sutskever I., & Salakhutdinov R. (2014). *Dropout: A simple way to prevent neural networks from overfitting*. Journal of Machine Learning Research, 15(1), 1929–1958.

- [26] Pan, S. J., & Yang, Q. (2010). *A survey on transfer learning*. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359.
- [27] Jouppi, N. P., Young, C., Patil, N., et al. (2021). *Ten Lessons from Three Generations Shaped Google's TPUv4i*. IEEE Micro, 41(5), 50–59.
- [28] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- [29] Hinton, G. E., & Salakhutdinov, R. R. (2006). *Reducing the dimensionality of data with neural networks*. Science, 313(5786), 504–507.
- [30] Bahdanau, D., Cho, K., & Bengio, Y. (2014). *Neural machine translation by jointly learning to align and translate*. arXiv preprint arXiv:1409.0473.
- [31] Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and policy considerations for deep learning in NLP*. arXiv:1906.02243
- [32] Buolamwini, J., & Gebru, T. (2018). *Gender shades: Intersectional accuracy disparities in commercial gender classification*. FAT* Conference
- [33] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": *Explaining the predictions of any classifier*. Proceedings of the 22nd ACM SIGKDD
- [34] Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions*. NeurIPS
- [35] Szegedy, C., Zaremba, W., Sutskever, I., et al. (2014). *Intriguing properties of neural networks*. arXiv:1312.6199
- [36] Bergstra, J., & Bengio, Y. (2012). *Random search for hyper-parameter optimization*. Journal of Machine Learning Research
- [37] Pearl, J. (2009). *Causality: Models, reasoning and inference*. Cambridge University Press
- [38] Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. ICML.
- [39] Kingma D. P., & Ba, J. (2015). *Adam: A method for stochastic optimization*. ICLR.
- [40] Wilson, A. C., Roelofs, R., Stern, M., Srebro, N., & Recht, B. (2017). *The marginal value of adaptive gradient methods in machine learning*. NeurIPS.
- [41] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536.
- [42] World Health Organization, *Health Information Systems*, NCBI Bookshelf, National Center for Biotechnology Information, Bethesda, MD, USA, 2022.
- [43] P. A. Bernstein and E. Newcomer, *Principles of Transaction Processing*, Morgan Kaufmann, 2009.
- [44] J. A. O'Brien and G. M. Marakas, *Management Information Systems*, 6th ed., McGraw-Hill Irwin, New York, NY, USA, 2006.
- [45] R. H. Sprague Jr. and H. J. Watson (eds.), *Decision Support Systems: Putting Theory into Practice*, Prentice-Hall, 1993.
- [46] M. S. Scott Morton, *Management Decision Systems: Computer-Based Support for Decision Making*, North-Holland, 1971.
- [47] P. G. W. Keen, "Decision support systems: a research perspective," in *Decision Support Systems: Issues and Challenges*, IIASA Proceedings Series 11, 1980, pp. 23–27.
- [48] E. H. Shortliffe and J. J. Cimino (eds.), *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*, 4th ed., Springer, 2014.
- [49] WHO – Pneumonia (γενικές πληροφορίες/ομάδες υψηλού κινδύνου). [Pneumonia](#)
- [50] CDC – About Pneumonia (ορισμός, τύποι, πρόληψη/παράγοντες κινδύνου). [About Pneumonia | Pneumonia | CDC](#)
- [51] European Respiratory Society – What is pneumonia? (παθολογοανατομικός ορισμός/συμπύκνωση). [What is pneumonia? | European Respiratory Society](#)
- [52] UNICEF DATA – Pneumonia in children (παγκόσμια επιβάρυνση στα παιδιά). [Pneumonia in Children Statistics - UNICEF DATA](#)
- [53] ECDC – Pneumococcal disease factsheet (πνευμονιόκοκκος και κλινικά χαρακτηριστικά).

[Factsheet for health professionals about pneumococcal disease](#)

- [54] BMJ Best Practice – Overview of pneumonia (ορισμοί/ταξινόμηση). [Overview of pneumonia - Summary of relevant conditions | BMJ Best Practice](#)
- [55] StatPearls – Pneumonia Pathology (παθογένεση/ταξινόμηση). [Pneumonia Pathology - StatPearls - NCBI Bookshelf](#)
- [56] Chouhan, V.; Singh, S.K.; Khamparia, A.; Gupta, D.; Tiwari, P.; Moreira, C.; Damaševičius, R.; de Albuquerque, V.H.C. A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images. *Appl. Sci.* 2020, *10*, 559. <https://doi.org/10.3390/app10020559>
- [57] A. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 60, 2019, doi:10.1186/s40537-019-0197-0.
- [58] S. Perez and A. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," *Convolutional Neural Networks Vis. Recognit.*, 2017.
- [59] C. Lemley, S. Bazrafkan, and P. Corcoran, "Smart Augmentation Learning an Optimal Data Augmentation Strategy," *IEEE Access*, vol. 5, pp. 5858-5869, 2017, doi:10.1109/ACCESS.2017.2696121.
- [60] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 1251-1258, doi:10.1109/CVPR.2017.195.
- [61] F. Chollet, "Deep Learning with Python," Manning Publications, 2021.
- [62] D. Kermany, K. Zhang and M. Goldbaum, "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification," *Cell*, vol. 172, no. 5, pp. 1122-1131, 2018. [Online]. Available: [https://www.cell.com/cell/fulltext/S0092-8674\(18\)30154-5](https://www.cell.com/cell/fulltext/S0092-8674(18)30154-5).
- [63] Li, Zewen, et al. "A survey of convolutional neural networks: analysis, applications, and prospects." *IEEE transactions on neural networks and learning systems* 33.12 (2021): 6999-7019.
- [64] O'Shea, K. "An introduction to convolutional neural networks." arXiv preprint arXiv: 1511.08458 (2015).
- [65] Wang, Yingying, et al. "The influence of the activation function in a convolution neural network model of facial expression recognition." *Applied Sciences* 10.5 (2020): 1897.
- [66] Cui, Yin, et al. "Kernel pooling for convolutional neural networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- [67] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [68] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*, 2014.
- [69] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
- [70] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [71] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.
- [72] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
- [73] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.
- [74] A. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *Proceedings of the Twenty-first International Conference on Machine Learning (ICML)*, 2004, p. 78.

- [75] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, 2015, pp. 448–456.
- [76] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014, pp. 3320–3328.
- [77] A. Karpathy, *CS231n Convolutional Neural Networks for Visual Recognition*, Stanford University, 2016.
- [78] S. Ruder, "An overview of gradient descent optimization algorithms," arXiv:1609.04747, 2016.
- [79] NVIDIA, "NVIDIA Tesla V100 GPU Architecture," 2023. [Online]. Available: <https://www.nvidia.com/>
- [80] NVIDIA, "CUDA Toolkit Documentation," [Online]. Available: <https://docs.nvidia.com/cuda>.
- [81] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural Networks: Tricks of the Trade*. Springer, 2012.
- [82] S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of Adam and beyond," arXiv:1904.09237, 2019.
- [83] L. Bottou, "Stochastic Gradient Descent Tricks," in *Neural Networks: Tricks of the Trade*, Springer, 2012.
- [84] S. Keskar et al., "On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima," ICLR, 2017.
- [85] M. Lin, Q. Chen, and S. Yan, "Network In Network," arXiv:1312.4400, 2013.
- [86] A. Krogh and J. Hertz, "A Simple Weight Decay Can Improve Generalization," NIPS, 1992.
- [87] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in Proc. IEEE ICCV, 2015.
- [88] S. Raschka, V. Mirjalili, *Python Machine Learning*, 3rd ed. Packt Publishing, 2019.
- [89] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer, 2009.
- [90] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [91] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [92] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv:1702.08608, 2017.
- [93] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017, pp. 618–626.
- [94] W. Samek, T. Wiegand, and K. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," *IT Professional*, vol. 21, no. 3, pp. 82–88, 2019.