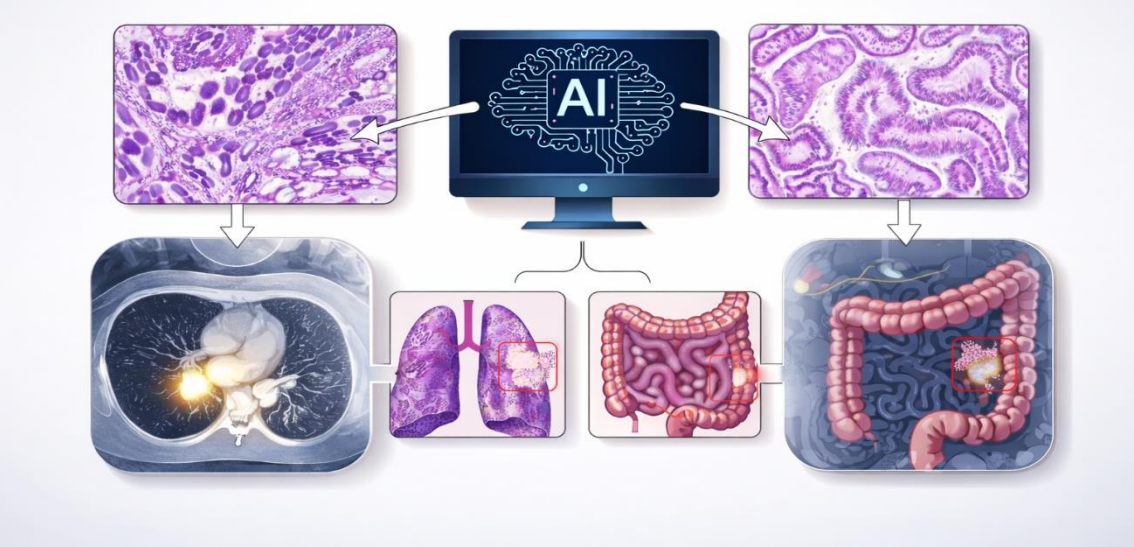
ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΗΣ ΕΛΛΑΔΟΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ανίχνευση καρκίνου του Πνεύμονα και του Παχέος Εντέρου μέσω ιστοπαθολογικών εικόνων με τη χρήση τεχνητής νοημοσύνης



Του φοιτητή
Καραπατσιά Θεοδόσιου
Αρ. Μητρώου: 2020063

Επιβλέπων
Παναγιώτης Αδαμίδης
Καθηγητής

Θεσσαλονίκη, Ιούνιος 2026

Τίτλος Δ.Ε.: Ανίχνευση καρκίνου του Πνεύμονα και του Παχέος Εντέρου μέσω ιστοπαθολογικών εικόνων με την χρήση τεχνητής νοημοσύνης

Κωδικός Δ.Ε.: 25347

Ονοματεπώνυμο φοιτητή/τών: Καραπατσιάς Θεοδόσιος

Ονοματεπώνυμο εισηγητή: Αδαμίδης Παναγιώτης

Ημερομηνία ανάληψης Δ.Ε. :02-11-2025

Ημερομηνία περάτωσης Δ.Ε. : 29-05-2026

Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή Καραπατσιά Θεοδόσιου που την εκπόνησε/αν. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης, ο συγγραφέας/δημιουργός εκχωρεί στο Διεθνές Πανεπιστήμιο της Ελλάδος άδεια χρήσης του δικαιώματος αναπαραγωγής, δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσης της εργασίας διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος. Η ανοικτή πρόσβαση στο πλήρες κείμενο της εργασίας, δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού, ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, πώληση, εμπορική χρήση, διανομή, έκδοση, μεταφόρτωση (downloading), ανάρτηση (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος, δεν υποδηλώνει απαραιτήτως και αποδοχή των απόψεων του συγγραφέα, εκ μέρους του Τμήματος.

«Σε όσους με στηρίζουν αδιάκοπα»

Πρόλογος

Ο λόγος ο οποίος επιλέχθηκε το παρών θέμα της εργασίας προέκυψε από το ενδιαφέρον μου για τον κλάδο της τεχνητής νοημοσύνης καθώς και τις εφαρμογές του σε πολλές πτυχές της ιατρικής επιστήμης. Η εργασία αυτή εκπονήθηκε με σκοπό την μελέτη και την σύγκριση διαφόρων αρχιτεκτονικών τεχνητής νοημοσύνης με επίκεντρο τους δύο πιο συνηθισμένους θανατηφόρους τύπους καρκίνου. Η επιλογή του συγκεκριμένου θέματος συνδέεται με το ενδιαφέρον μου για τις ιατρικές εφαρμογές της τεχνητής νοημοσύνης καθώς και ειδικότερα στην δυνατότητα υποστήριξης της διαγνωστικής διαδικασίας μέσω υπολογιστικών μεθόδων

Μέσα από αυτήν την διαδικασία είχα την ευκαιρία να εφαρμόσω τις θεωρητικές μου γνώσεις που απέκτησα από το τμήμα πάνω σε ένα πραγματικό πρόβλημα, αλλά και να τις επεκτείνω περαιτέρω εφαρμόζοντας νέες καινοτομίες.

Περίληψη

Στην παρούσα εργασία διερευνάται η χρήση τεχνικών βαθιάς μάθησης για την ταξινόμηση ιστοπαθολογικών εικόνων, με στόχο την συγκριτική αξιολόγηση διαφορετικών αρχιτεκτονικών και μοντέλων. Για τις ανάγκες της μελέτης χρησιμοποιήθηκε το σύνολο δεδομένων LC25000, το οποίο αποτελείται από ισορροπημένο σύνολο εικόνων διαφόρων κατηγοριών ιστών. Στο πλαίσιο της υλοποίησης, εφαρμόστηκε ενιαία μεθοδολογία εκπαίδευσης και αξιολόγησης, ώστε να διασφαλιστεί η αντικειμενικότητα της σύγκρισης. Συγκεκριμένα, εξετάστηκαν τόσο παραδοσιακά Convolutional Neural Networks (CNN), όσο και πιο εξελιγμένες αρχιτεκτονικές όπως τα VGG16, VGG19, ResNet50 και InceptionResNetV2, αξιοποιώντας τεχνικές transfer learning και fine-tuning. Παράλληλα, υλοποιήθηκε και ένα Vision Transformer (ViT), προκειμένου να μελετηθεί η απόδοσή του σε σχέση με τα convolutional μοντέλα. Η αξιολόγηση βασίστηκε σε μετρικές όπως accuracy, precision, recall και F1-score. Τα αποτελέσματα έδειξαν ότι τα CNN-based μοντέλα επιτυγχάνουν υψηλή απόδοση, με το InceptionResNetV2 να εμφανίζει την καλύτερη συνολική επίδοση ίση με 99.2%. Αντίθετα, το Vision Transformer παρουσίασε χαμηλότερη απόδοση, γεγονός που υποδηλώνει την ανάγκη για μεγαλύτερα datasets καθώς και περαιτέρω εκπαίδευση. Συνολικά, η εργασία αναδεικνύει τη σημασία της επιλογής κατάλληλης αρχιτεκτονικής για την αποτελεσματική ανάλυση ιατρικών εικόνων.

Detection of Lung and Colon Cancer from Histopathological Images using Artificial Intelligence

Karapatsias Theodosios

Abstract

This thesis explores the application of deep learning techniques for the classification of histopathological images, aiming to provide a comprehensive comparative evaluation of different neural network architectures. The study utilizes the LC25000 dataset, which comprises a balanced collection of images representing various tissue categories, enabling a consistent experimental setting for comparative evaluation. A unified training and evaluation framework was adopted to guarantee the consistency and fairness of the comparative analysis. The experimental investigation includes both conventional Convolutional Neural Networks (CNNs) and more advanced architectures, namely VGG16, VGG19, ResNet50, and InceptionResNetV2, employing transfer learning and fine-tuning strategies. Furthermore, a Vision Transformer (ViT) model was implemented to assess its effectiveness relative to convolution-based approaches. Model performance was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score, providing a comprehensive assessment of predictive capability. The results indicate that CNN-based architectures consistently achieve high performance, with InceptionResNetV2 outperforming all other models, attaining an accuracy of 99.2%. In contrast, the Vision Transformer demonstrated comparatively lower performance, suggesting that such architectures require larger-scale datasets and more extensive training to reach their full potential. Overall, the findings emphasize the critical role of architectural selection in medical image analysis and contribute to a deeper understanding of the practical strengths and limitations of contemporary deep learning models.

Ευχαριστίες

Θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες σε όλους όσοι συνέβαλαν, άμεσα ή έμμεσα, στην ολοκλήρωση της παρούσας διπλωματικής εργασίας.

Αρχικά ιδιαίτερη και βαθύτατη ευγνωμοσύνη οφείλω στους γονείς και τον αδερφό μου, οι οποίοι αποτέλεσαν το σημαντικότερο στήριγμα καθ' όλη την διάρκεια της ακαδημαϊκής και προσωπικής μου πορείας. Με συνεχή υποστήριξη, αγάπη και αληθινή πίστη στις δυνατότητές μου, μου παρείχαν τα απαραίτητα εφόδια για να συνεχίσω να ανταπεξέρχομαι στις προκλήσεις που εμφανιζόταν στον δρόμο μου. Η υπομονή και η ενθάρρυνσή τους στις δύσκολες στιγμές καθώς και σε περιόδους αμφιβολίας ήταν και θα είναι οι καθοριστικοί παράγοντες για τους οποίους μπορώ να συνεχίσω να ξεπερνώ κάθε δυσκολία που εμφανίζεται. Γι' αυτούς τους λόγους τους οφείλω ένα βαθύ και ειλικρινές «ευχαριστώ».

Ιδιαίτερη ευγνωμοσύνη οφείλω και στους φίλους μου οι οποίοι μου συμπαραστάθηκαν και βοήθησαν κατά την διάρκεια αυτής της απαιτητικής πορείας. Η παρουσία τους ήταν σημαντική τόσο σε πρακτικό αλλά και σε ψυχολογικό επίπεδο προσφέροντάς μου ενθάρρυνση και κατανόηση σε στιγμές πίεσης και αμφιβολίας. Οι στιγμές χαλάρωσης που μοιραστήκαμε αποτέλεσαν σημαντικό παράγοντα ισορροπίας ανάμεσα στις περιόδους πίεσης της εργασίας, που μου επέτρεψαν να αντιμετωπίσω με μεγάλη ψυχραιμία τις δυσκολίες που προέκυψαν κατά την εκπόνησή της. Η συνεχή παρουσία τους και η στήριξή τους ήταν ιδιαίτερα σημαντική και τους ευχαριστώ ειλικρινά μέσα από την καρδία μου.

Τέλος, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή της εργασίας κ. Αδαμίδη, ο οποίος από την πρώτη στιγμή που ανέλαβα την διπλωματική μου εργασία με υποστήριζε και με συμβούλευε καθ' όλη την διάρκειά της . Η διαθεσιμότητά του για συζήτηση και η προθυμία του να επιλύει απορίες σε κάθε στάδιο της εργασίας αποτέλεσαν σημαντικό παράγοντα προόδου. Οι εύστοχες παρατηρήσεις και οι πολύτιμες συμβουλές του συνέβαλαν ουσιαστικά στη διαμόρφωση της μεθοδολογίας και στη βελτίωση της ποιότητας της έρευνας. Για όλα τα παραπάνω, τον ευχαριστώ θερμά.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Συντομογραφίες

Δ.Ε.	Διπλωματική Εργασία
ΔΙΠΑΕ	Διεθνές Πανεπιστήμιο Ελλάδος
AI	Artificial Intelligence
ML	Machine Learning
CNN	Convolutional Neural Networks
ViT	Vision Transformers
DL	Deep Learning
XAI	Explainable AI
DNN	Deep Neural Network
MLP	Multi-Layer Perceptron
TL	Transfer Learning
CCT	Compact Convolutional Transformer
WSI	Whole Slide Images
ΠΟΥ	Παγκόσμιος Οργανισμός Υγείας
Grad-CAM	Gradient-weighted Class Activation Mapping
RGB	Red Green Blue
H&E	Hematoxylin and Eosin
IARC	International Agency for Research on Cancer
GLOBOCAN	Global Cancer Observatory
CT	Computed Tomography
NSCLC	Non-Small Cell Lung Cancer
NLP	Natural Language Processing
OOM	Out of Memory
SGD	Stochastic Gradient Descent
ReLU	Rectified Linear Unit
CLS	Classification Token
GAP	Global Average Pooling
TP	True Positive
TN	True Negative
GPU	Graphics Processing Unit

BGR	Blue Green Red
CPU	Central Processing Unit

Κεφάλαιο 1ο: Εισαγωγή

1.1 Καρκίνος του Πνεύμονα και του Παχέος Εντέρου

Σύμφωνα με τα στοιχεία GLOBOCAN 2022/IARC [25], ο καρκίνος του πνεύμονα αποτελεί την πρώτη αιτία θανάτου από καρκίνο παγκοσμίως, με περίπου 1,8 εκατομμύρια θανάτους, ενώ ο καρκίνος του παχέος εντέρου ακολουθεί ως δεύτερη αιτία θανάτου από καρκίνο, με περίπου 9,3% των συνολικών θανάτων από καρκίνο.

Ο καρκίνος του πνεύμονα, κυρίως λόγω της αργοπορημένης του διάγνωσης αλλά και της επιθετικής του φύσης, χαρακτηρίζεται από υψηλά ποσοστά θνησιμότητας. Βασικό παράγοντα κινδύνου αποτελεί η επανειλημμένη έκθεση σε καρκινογόνες ουσίες σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας [30]. Ωστόσο σημαντικό ρόλο διαδραματίζουν και περιβαλλοντικοί παράγοντες όπως η έκθεση σε αμίαντο, ινώδες υλικό που χρησιμοποιείται κυρίως στην δόμηση, σε ραδόνιο, ραδιενεργό αέριο που παράγεται από τη φυσική διάσπαση του ουρανίου και η ατμοσφαιρική ρύπανση. Επιπλέον γενετικοί παράγοντες μπορούν επίσης να αυξήσουν την προδιάθεση εμφάνισης της νόσου [30].

Από την άλλη πλευρά ο καρκίνος του παχέος εντέρου σχετίζεται κυρίως με διατροφικούς, γενετικούς και περιβαλλοντικούς παράγοντες, με μεγάλα ποσοστά εμφάνισης σε άτομα που κατοικούν σε ανεπτυγμένα κράτη που έχουν υιοθετήσει τον δυτικό τρόπο ζωής. Σύμφωνα με τον ΠΟΥ [31], οι διατροφικές συνήθειες που περιέχουν πληθώρα λιπαρών και κατανάλωση επεξεργασμένων τροφίμων, η έλλειψη φυσικής δραστηριότητας που οδηγεί στην παχυσαρκία αποτελούν τους βασικούς παράγοντες κινδύνου. Τις τελευταίες δεκαετίες παρατηρείται και ανησυχητική αύξηση κρουσμάτων σε άτομα κάτω των πενήντα ετών λόγω γενετικών συνδρόμων όπως το σύνδρομο Lynch [31].

Η έγκαιρη διάγνωση και στις δύο περιπτώσεις αποτελεί καίριο παράγοντα για την βελτίωση της πρόγνωσης και στις 2 περιπτώσεις. Παρόλα αυτά η εξέλιξη των νόσων αυτών στα αρχικά στάδια χωρίς την ένδειξη συμπτωμάτων δυσχεραίνει την διαδικασία της ανίχνευσής τους. Ο ΠΟΥ αναφέρει ότι ο καρκίνος του παχέος εντέρου συχνά δεν παρουσιάζει συμπτώματα στα αρχικά στάδια, ενώ τα προγράμματα έγκαιρης διάγνωσης και προ συμπτωματικού ελέγχου μπορούν να μειώσουν τη θνησιμότητα [30]. Αντίστοιχα, για τον καρκίνο του πνεύμονα, ο ΠΟΥ τονίζει ότι τα πρώιμα συμπτώματα μπορεί να είναι ήπια ή να υποτιμηθούν, οδηγώντας σε καθυστερημένη διάγνωση [30]. Γι' αυτόν τον λόγο καθίσταται αναγκαία η ανάπτυξη και η ενσωμάτωση αξιόπιστων μεθόδων καθώς και νέων τεχνολογιών στην ιατρική πρακτική.

1.2 Η σημασία της Ιστοπαθολογικής Ανάλυσης

Η ιστοπαθολογική ανάλυση αποτελεί τον θεμέλιο λίθο της διάγνωσης των κακοηθειών, καθώς επιτρέπει την εξονυχιστική εξέταση των ιστών σε μικροσκοπικό επίπεδο [26]. Μέσω αυτής της ανάλυσης γίνεται εφικτή η αναγνώριση αλλοιώσεων στη μορφολογία των κυττάρων, όπως οι αλλαγές στο μέγεθος, το σχήμα και την πυρηνική δομή των κυττάρων, η διαφοροποίησή τους από τα φυσιολογικά κύτταρα του ιστού από τον οποίο προήλθαν, καθώς και η διήθηση των γύρω ιστών [26].

Και στις δύο περιπτώσεις, η ιστοπαθολογική εξέταση παρέχει κρίσιμες πληροφορίες σχετικά με το είδος, για παράδειγμα αδενοκαρκίνωμα και πλακώδες καρκίνωμα, τον βαθμό της κακοήθειας αλλά και το στάδιο της νόσου [27]. Οι πληροφορίες αυτές είναι απαραίτητες για την επιλογή της κατάλληλης θεραπευτικής προσέγγισης.

Επιπλέον, η χρήση ανοσοϊστοχημικών τεχνικών, δηλαδή διαγνωστικών μεθόδων που συνδυάζουν ιστολογικές και ανοσολογικές αρχές, επιτρέπει την ανίχνευση ειδικών βιοδεικτών που σχετίζονται με τη συμπεριφορά του όγκου και την ανταπόκριση στη θεραπεία [27]. Για παράδειγμα, η αυξημένη έκφραση πρωτεϊνών (κυρίως της βιμεντίνης) μπορεί να υποδεικνύει την πιθανότητα μετάστασης του καρκίνου ή να αναδείξει την αποτελεσματικότητα της θεραπείας.

Η εξέλιξη της ψηφιακής ιστοπαθολογίας ενισχύει σημαντικά αυτήν τη διαδικασία, επιτρέποντας την αποθήκευση, την ανάλυση και τον διαμοιρασμό ψηφιακών εικόνων σε υψηλή ανάλυση. Επιπλέον, η εξέλιξή της, σε συνδυασμό με την ενσωμάτωση της τεχνητής νοημοσύνης, δημιούργησε νέες ευκαιρίες για την αυτοματοποίηση της ανάλυσης καθώς και τη βελτίωση της ακρίβειας της διάγνωσης [28].

1.3 Περιορισμοί της Παραδοσιακής Διάγνωσης

Παρόλο που η σημασία της παραδοσιακής διάγνωσης είναι καταλυτική, η προσέγγιση αυτή παρουσιάζει σημαντικούς περιορισμούς, οι οποίοι επηρεάζουν τόσο την ακρίβεια όσο και την ταχύτητα λήψης κλινικών αποφάσεων. Ένας από τους πιο βασικούς είναι η εξάρτηση από την εμπειρία και η υποκειμενική κρίση του παθολογοανατόμου. Η ιστοπαθολογική αξιολόγηση βασίζεται στην οπτική παρατήρηση και την εμπειρία του παθολογοανατόμου, γεγονός που μπορεί να οδηγήσει σε διαφοροποιήσεις μεταξύ αξιολογητών αλλά και στον ίδιο αξιολογητή σε βάθος χρόνου[29].

Η μεταβλητότητα των διαγνώσεων αποτελεί σημαντικό ζήτημα, ειδικά όταν πρόκειται για περιπτώσεις οριακών αλλοιώσεων ή αμφίβολων περιστατικών, κυρίως σε πρώιμα στάδια, όπου τα μορφολογικά χαρακτηριστικά δεν είναι ακόμα σαφώς διακριτά [29]. Αποτέλεσμα είναι είτε να προκύψουν σφάλματα κατά τη διάγνωση, είτε ψευδώς θετικά είτε ψευδώς αρνητικά αποτελέσματα, τα οποία υπάρχει πιθανότητα να έχουν σοβαρές επιπτώσεις στη διαχείριση και τη θεραπεία του ασθενούς.

Στη συνέχεια, η διαδικασία αυτή αποδεικνύεται χρονοβόρα και απαιτεί υψηλό βαθμό εξειδίκευσης. Η προετοιμασία και η λήψη δειγμάτων, η χρώση καθώς και η μικροσκοπική εξέταση είναι πρακτικές χρονοβόρες και ακριβές, με αποτέλεσμα την καθυστέρηση στη λήψη αποφάσεων. Σε περιβάλλοντα στα οποία υπάρχει μεγάλος φόρτος εργασίας, όπως μεγάλες κλινικές και νοσοκομεία, η καθυστέρηση αυτή μπορεί να επηρεάσει αρνητικά την πορεία της θεραπείας.

Επιπλέον, η παραδοσιακή διάγνωση είναι δύσκολο να χειριστεί μεγάλο όγκο δεδομένων. Αυτό συμβαίνει διότι με τη χρήση σύγχρονων ιατρικών πρακτικών, ένας ασθενής μπορεί να έχει πληθώρα αποτελεσμάτων από διαφορετικές εξετάσεις, όπως απεικονιστικά δεδομένα, ιστολογικά δείγματα και μοριακές αναλύσεις. Η ενοποίηση των αποτελεσμάτων αυτών σε συνδυασμό με την ποιοτική αξιολόγηση των δεδομένων πολλές φορές ξεπερνά τις ανθρώπινες δυνατότητες [28].

Οι περιορισμοί αυτοί στο σύνολό τους υπογραμμίζουν την ανάγκη για την ενσωμάτωση νέων τεχνολογιών που θα ενισχύσουν την ακρίβεια, θα μειώσουν τον χρόνο διάγνωσης και θα υποστηρίξουν τη λήψη κλινικών αποφάσεων [28].

1.4 Η Συμβολή της Τεχνητής Νοημοσύνης στην Ιατρική Απεικόνιση

Η τεχνητή νοημοσύνη αποτελεί μία από τις ταχύτερα αναπτυσσόμενες τεχνολογίες της σύγχρονης εποχής, με αυξανόμενη εφαρμογή σε πολλούς επιστημονικούς και επαγγελματικούς τομείς, μεταξύ των οποίων και η ιατρική. Ιδιαίτερα σημαντική επίδραση στην απεικόνιση ιατρικών δεδομένων καθώς και στη συμβολή της στη διαγνωστική διαδικασία. Η δυνατότητα των αλγορίθμων της μηχανικής μάθησης (Machine Learning – ML) να επεξεργάζονται μεγάλους όγκους δεδομένων και να εντοπίζουν

σύνθετα χαρακτηριστικά καθιστά το AI ένα εργαλείο μεγάλης σημασίας για την υποστήριξη του ιατρικού προσωπικού [28].

Σημαντικό ρόλο στη συμβολή του AI διαδραματίζουν κυρίως τα μοντέλα βαθιάς μάθησης (Deep Learning). Πιο συγκεκριμένα, τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks – CNNs), τα οποία έχουν σχεδιαστεί για την ανάλυση εικόνων, είναι ικανά να εξάγουν αυτόματα χαρακτηριστικά χωρίς να χρειάζεται χειροκίνητη επιλογή χαρακτηριστικών, ενώ παράλληλα επιτυγχάνουν υψηλή ακρίβεια σε ενέργειες όπως η ταξινόμηση, η ανίχνευση και η τμηματοποίηση εικόνων [28].

Στον κλάδο της ιστοπαθολογίας, η τεχνητή νοημοσύνη χρησιμοποιείται για την ανάλυση ψηφιακών ιστολογικών εικόνων υψηλής ανάλυσης. Μέσω της εκπαίδευσης σε κατάλληλα επισημασμένα δεδομένα, τα μοντέλα μπορούν να αναγνωρίζουν καρκινικά κύτταρα, να εντοπίζουν περιοχές ενδιαφέροντος και να διαχωρίζουν διαφορετικούς τύπους ιστών. Αυτό επιτρέπει την υποστήριξη του παθολογοανατόμου, μειώνοντας την πιθανότητα σφαλμάτων και επιταχύνοντας τη διαδικασία διάγνωσης [26].

Επιπλέον, η AI μπορεί να συμβάλει στην ποσοτικοποίηση των ευρημάτων, παρέχοντας αντικειμενικές μετρήσεις, όπως το ποσοστό καρκινικών κυττάρων ή η πυκνότητα του όγκου. Η δυνατότητα αυτή ενισχύει τη δυνατότητα παρακολούθησης της εξέλιξης της νόσου και της αξιολόγησης της ανταπόκρισης στη θεραπεία, αναδεικνύοντας τον ρόλο της AI στην υπολογιστική παθολογοανατομία (computational pathology) [28].

Ένα ακόμη χαρακτηριστικό της τεχνητής νοημοσύνης είναι η ικανότητα συνδυασμού δεδομένων από διαφορετικές πηγές, όπως εικόνες, κλινικά χαρακτηριστικά και γενετικές πληροφορίες. Μέσω αυτής της ολοκληρωμένης οπτικής καθίσταται δυνατή η δημιουργία μοντέλων μεγάλης ακρίβειας που υποστηρίζουν την εξατομικευμένη ιατρική [27].

Συνοψίζοντας, η τεχνητή νοημοσύνη έχει τη δυνατότητα να μετασχηματίσει την ιατρική απεικόνιση και τη διάγνωση του καρκίνου, προσφέροντας λύσεις στα προβλήματα της παραδοσιακής προσέγγισης και ανοίγοντας νέους δρόμους για την ανάπτυξη πιο αποδοτικών και αξιόπιστων συστημάτων υγείας [28].

1.5 Σκοπός και Στόχοι της Διπλωματικής Εργασίας

Η παρούσα διπλωματική εργασία έχει ως κύριο σκοπό την ανάπτυξη και αξιολόγηση συστημάτων τεχνητής νοημοσύνης για την αυτόματη ανίχνευση καρκίνου του πνεύμονα και του παχέος εντέρου μέσω ιστοπαθολογικών εικόνων. Συγκεκριμένα, επιδιώκεται η διερεύνηση της αποτελεσματικότητας σύγχρονων τεχνικών βαθιάς μάθησης στην ανάλυση ιατρικών εικόνων υψηλής ανάλυσης, με στόχο την υποστήριξη της διαγνωστικής διαδικασίας.

Η εργασία εστιάζει στη σύγκριση διαφορετικών αρχιτεκτονικών μοντέλων, όπως τα Συνελικτικά Νευρωνικά Δίκτυα (CNN), τα μοντέλα Transfer Learning (π.χ. ResNet50) και τα Vision Transformers (ViT), αξιολογώντας την απόδοσή τους σε ένα κοινό σύνολο δεδομένων. Η επιλογή αυτών των προσεγγίσεων βασίζεται στη σύγχρονη βιβλιογραφία, όπου έχουν παρουσιαστεί ως ιδιαίτερα αποδοτικές στην ανάλυση εικόνων και στην ιατρική διάγνωση.

Οι επιμέρους στόχοι της εργασίας διαμορφώνονται ως εξής:

- Η κατανόηση των βασικών αρχών της μηχανικής μάθησης και της βαθιάς μάθησης, με έμφαση στην ανάλυση εικόνων.

- Η μελέτη και υλοποίηση ενός βασικού μοντέλου CNN ως baseline για σύγκριση.
- Η αξιοποίηση τεχνικών Transfer Learning με προ εκπαιδευμένα μοντέλα για τη βελτίωση της απόδοσης.
- Η διερεύνηση της χρήσης Vision Transformers και η σύγκρισή τους με τα παραδοσιακά CNN.
- Η προ επεξεργασία και κατάλληλη διαχείριση του dataset LC25000, καθώς και η εφαρμογή τεχνικών data augmentation.
- Η αξιολόγηση των μοντέλων με τη χρήση κατάλληλων μετρικών, όπως Accuracy, Precision, Recall και F1-score.
- Η ανάλυση των σφαλμάτων μέσω confusion matrices και η ερμηνεία των αποτελεσμάτων.
- Η εφαρμογή τεχνικών Εξηγήσιμης Τεχνητής Νοημοσύνης (Explainable AI), όπως το Grad-CAM, για την κατανόηση των αποφάσεων των μοντέλων.

Επιπλέον, η εργασία στοχεύει στην ανάδειξη των πλεονεκτημάτων και των περιορισμών κάθε αρχιτεκτονικής, καθώς και στην εξαγωγή συμπερασμάτων σχετικά με την καταλληλόλητά τους για εφαρμογή σε πραγματικά κλινικά περιβάλλοντα. Μέσω της συγκριτικής αξιολόγησης, επιδιώκεται να προσδιοριστεί ποια προσέγγιση προσφέρει την καλύτερη ισορροπία μεταξύ ακρίβειας, υπολογιστικού κόστους και ερμηνευσιμότητας.

Τέλος, ένας σημαντικός στόχος της εργασίας είναι η γεφύρωση του χάσματος μεταξύ θεωρητικής έρευνας και πρακτικής εφαρμογής, αναδεικνύοντας τον ρόλο της τεχνητής νοημοσύνης ως εργαλείου υποστήριξης και όχι αντικατάστασης του ιατρού.

1.6 Δομή της Εργασίας

Η δομή της παρούσας διπλωματικής εργασίας έχει σχεδιαστεί με τρόπο που να εξασφαλίζει τη συστηματική παρουσίαση του θέματος, από το θεωρητικό υπόβαθρο έως την πειραματική αξιολόγηση και τα τελικά συμπεράσματα.

Στο Κεφάλαιο 1 (Εισαγωγή) παρουσιάζεται το γενικό πλαίσιο της εργασίας, αναλύεται η σημασία της έγκαιρης διάγνωσης του καρκίνου του πνεύμονα και του παχέος εντέρου, καθώς και ο ρόλος της ιστοπαθολογικής ανάλυσης. Επιπλέον, αναδεικνύονται οι περιορισμοί της παραδοσιακής διάγνωσης και εισάγεται η συμβολή της τεχνητής νοημοσύνης στην ιατρική απεικόνιση, ενώ διατυπώνονται ο σκοπός και οι στόχοι της εργασίας.

Στο Κεφάλαιο 2 (Θεωρητικό Υπόβαθρο**)** παρουσιάζονται οι βασικές έννοιες της μηχανικής μάθησης και της βαθιάς μάθησης, με ιδιαίτερη έμφαση στα συνελικτικά νευρωνικά δίκτυα, στις τεχνικές Transfer Learning και στα Vision Transformers. Επιπλέον, γίνεται αναφορά στην εξηγήσιμή τεχνητή νοημοσύνη και στις τεχνικές Grad-CAM για την οπτικοποίηση των αποφάσεων των μοντέλων.

Στο Κεφάλαιο 3 (Ανασκόπηση Βιβλιογραφίας**)** παρουσιάζονται σχετικές ερευνητικές εργασίες που αφορούν την εφαρμογή βαθιάς μάθησης στην ιστοπαθολογία, καθώς και η χρήση διαφορετικών μοντέλων για την ανίχνευση καρκίνου. Πραγματοποιείται σύγκριση των προσεγγίσεων που έχουν προταθεί στη διεθνή βιβλιογραφία.

Στο Κεφάλαιο 4 (Δεδομένα και Προεπεξεργασία**)** περιγράφεται το dataset LC25000, οι κατηγορίες των εικόνων και η διαδικασία διαχωρισμού των δεδομένων σε σύνολα εκπαίδευσης,

επικύρωσης και ελέγχου. Επιπλέον, αναλύονται οι τεχνικές προεπεξεργασίας και data augmentation, καθώς και το πειραματικό πρωτόκολλο που ακολουθήθηκε.

Στο Κεφάλαιο 5 **(**Σχεδιασμός και Υλοποίηση Μοντέλων) παρουσιάζεται η αρχιτεκτονική του baseline CNN, καθώς και η υλοποίηση των μοντέλων Transfer Learning και Vision Transformer. Περιγράφονται επίσης οι διαδικασίες εκπαίδευσης, οι υπερπαράμετροι και τα προβλήματα που προέκυψαν κατά την υλοποίηση.

Στο Κεφάλαιο 6 (Πειραματική Αξιολόγηση) παρουσιάζονται τα αποτελέσματα των μοντέλων, με βάση συγκεκριμένες μετρικές αξιολόγησης. Περιλαμβάνονται confusion matrices, ανάλυση σφαλμάτων και συγκριτική αξιολόγηση μεταξύ των διαφορετικών προσεγγίσεων.

Στο Κεφάλαιο 7 (Ερμηνεία Αποτελεσμάτων) εφαρμόζονται τεχνικές Grad-CAM για την οπτικοποίηση των περιοχών ενδιαφέροντος και την κατανόηση των αποφάσεων των μοντέλων, συνδέοντας τα αποτελέσματα με τα μορφολογικά χαρακτηριστικά του καρκίνου.

Στο Κεφάλαιο 8 (Συζήτηση) αναλύονται τα αποτελέσματα σε βάθος, επισημαίνονται τα πλεονεκτήματα και οι περιορισμοί των μοντέλων και εξετάζονται πιθανές βελτιώσεις της προτεινόμενης μεθοδολογίας.

Τέλος, στο Κεφάλαιο 9 (Συμπεράσματα και Μελλοντική Έρευνα) συνοψίζονται τα βασικά ευρήματα της εργασίας, αναδεικνύεται η συμβολή της στο επιστημονικό πεδίο και προτείνονται κατευθύνσεις για μελλοντική έρευνα.

Κεφάλαιο 2ο: Θεωρητικό Υπόβαθρο

2.1 Εισαγωγή στη Μηχανική Μάθηση και τη Βαθιά Μάθηση

Η Μηχανική Μάθηση (Machine Learning) αποτελεί κλάδο της Τεχνητής Νοημοσύνης, ο οποίος εστιάζει στην ανάπτυξη αλγορίθμων και μοντέλων που έχουν την ικανότητα να μαθαίνουν από δεδομένα και να βελτιώνουν την απόδοσή τους χωρίς να απαιτείται ρητός προγραμματισμός για κάθε συγκεκριμένη εργασία. Αντί το σύστημα να βασίζεται σε προκαθορισμένους κανόνες, η Μηχανική Μάθηση επιτρέπει την εξαγωγή και την κατανόηση μοτίβων και σχέσεων από δεδομένα, καθιστώντας εφικτή την αυτοματοποίηση της λήψης αποφάσεων που αφορούν σύνθετες διαδικασίες [6][16].

Ο θεμελιώδης κανόνας της Μηχανικής Μάθησης συνίσταται στη δημιουργία ενός μαθηματικού μοντέλου το οποίο εκπαιδεύεται πάνω σε ένα σύνολο δεδομένων (train set) με στόχο να γενικεύει σε άγνωστα δεδομένα (test set). Η διαδικασία αυτή έχει ως στόχο τη βελτιστοποίηση μιας συνάρτησης απώλειας, η οποία ποσοτικοποιεί το σφάλμα μεταξύ των προβλέψεων του μοντέλου και των πραγματικών τιμών [2].

Οι κύριοι τομείς της Μηχανικής Μάθησης αποτελούν η Επιβλεπόμενη Μάθηση (Supervised Learning), η Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning), η Ημι-επιβλεπόμενη Μάθηση (Semi-supervised Learning) και η Ενισχυτική Μάθηση (Reinforcement Learning). Στην Επιβλεπόμενη Μάθηση το μοντέλο εκπαιδεύεται με δεδομένα που συνοδεύονται από ετικέτες. Τέτοια προβλήματα αποτελούν κυρίως τα προβλήματα ταξινόμησης, καθώς και παλινδρόμησης. Στη Μη Επιβλεπόμενη Μάθηση το μοντέλο ανακαλύπτει μοτίβα χωρίς τη χρήση ετικετών, όπως στην ομαδοποίηση [3]. Η Ημι-επιβλεπόμενη Μάθηση αποτελεί έναν συνδυασμό των δύο προαναφερθέντων κατηγοριών. Τέλος, με την Ενισχυτική Μάθηση το μοντέλο μαθαίνει την αναγνώριση μοτίβων μέσα από την αλληλεπίδραση με το περιβάλλον και μέσω ενός συστήματος επιβραβεύσεων [14].

Παρά τη διαδεδομένη χρήση της, η Μηχανική Μάθηση παρουσιάζει περιορισμούς όταν χρειάζεται να αξιοποιηθεί σε προβλήματα που περιλαμβάνουν υψηλή πολυπλοκότητα, όπως η ανάλυση εικόνας, φωνής και φυσικής γλώσσας. Αυτός είναι ο λόγος για τον οποίο οι ερευνητές οδηγήθηκαν στην ανάπτυξη της Βαθιάς Μάθησης (Deep Learning – DL), η οποία βασίζεται στη δημιουργία τεχνητών νευρωνικών δικτύων με πολλαπλά επίπεδα [7][8].

2.1.1 Βαθιά Μάθηση

Η Βαθιά Μάθηση αποτελεί εξέλιξη των τεχνητών νευρωνικών δικτύων και χαρακτηρίζεται από την ύπαρξη πολλών κρυφών επιπέδων (hidden layers), τα οποία επιτρέπουν την ιεραρχική εξαγωγή χαρακτηριστικών από τα δεδομένα. Τυπικά, ένα νευρωνικό δίκτυο αποτελείται από ένα επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα και ένα επίπεδο εξόδου. Κάθε νευρώνας εκτελεί μια γραμμική πράξη και ακολουθείται από μια μη γραμμική συνάρτηση ενεργοποίησης (activation function), όπως η ReLU. Η μη γραμμικότητα είναι κρίσιμη, καθώς επιτρέπει στο δίκτυο να μοντελοποιεί πολύπλοκες σχέσεις [6][15].

Η εκπαίδευση ενός νευρωνικού δικτύου και κατ' επέκταση ενός μοντέλου εξαρτάται από τον αλγόριθμο της οπισθοδιάδοσης (Backpropagation), σε συνεργασία με μεθόδους βελτιστοποίησης όπως το Stochastic Gradient Descent (SGD) και η μέθοδος Adam optimizer [6]. Η διαδικασία της εκπαίδευσης πραγματοποιείται σε τέσσερα βήματα: την Προώθηση (Forward Pass), όπου γίνεται ο υπολογισμός της εξόδου του δικτύου, τον Υπολογισμό Σφάλματος (Loss), την Οπισθοδιάδοση (Backward Pass), όπου γίνεται ο υπολογισμός παραγώγων, και την ενημέρωση των βαρών [16].

Τα πλεονεκτήματα που καθιστούν τη Βαθιά Μάθηση πιο αποτελεσματική από την απλή Μηχανική Μάθηση είναι η αυτόματη εξαγωγή χαρακτηριστικών, όπου, σε αντίθεση με την παραδοσιακή Μηχανική Μάθηση, στην οποία τα χαρακτηριστικά επιλέγονται χειροκίνητα, στη Βαθιά Μάθηση τα νευρωνικά δίκτυα ανακαλύπτουν μόνα τους τα σημαντικά χαρακτηριστικά απευθείας από τα ακατέργαστα δεδομένα, εξοικονομώντας χρόνο και μειώνοντας τα ανθρώπινα λάθη [14][15]. Παράλληλα, σε τομείς όπως η αυτόματη οδήγηση ή η ιατρική διάγνωση από ακτινογραφίες και ιστοπαθολογικές εικόνες, οι παραδοσιακές μέθοδοι συνήθως είναι ανεπαρκείς. Η Βαθιά Μάθηση, λόγω της δομής της, μπορεί να λύσει προβλήματα με υψηλή πολυπλοκότητα, που απαιτούν χιλιάδες παραμέτρους για να μοντελοποιηθούν σωστά [12][17]. Τέλος, το ML χαρακτηρίζεται από προβλήματα διαχείρισης δεδομένων και, μετά από κάποιο σημείο, σταματά να βελτιώνεται. Αντιθέτως, με το DL η απόδοση των μοντέλων συνεχίζει να βελτιώνεται όσο αυξάνεται ο όγκος των δεδομένων [24].

2.2 Συνελικτικά Νευρωνικά Δίκτυα

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks – CNNs) είναι μια κατηγορία Βαθιών Νευρωνικών Δικτύων (Deep Neural Networks - DNNs), τα οποία έχουν σχεδιαστεί για την προσαρμοστική και αυτόματη εκμάθηση ιεραρχικών χαρακτηριστικών σε δεδομένα με χωρική δομή, όπως είναι οι εικόνες. Η ανακάλυψή τους αποτέλεσε μια σημαντική εξέλιξη στον τομέα της υπολογιστικής όρασης, διότι επιτρέπουν την επίτευξη εξαιρετικά υψηλής ακρίβειας σε προβλήματα όπως η ταξινόμηση εικόνων, η ανίχνευση αντικειμένων και η τμηματοποίηση [6][15].

Τα CNNs εντοπίζουν και αξιοποιούν την τοπική συσχέτιση των δεδομένων. Αντιθέτως, τα πλήρως συνδεδεμένα νευρωνικά δίκτυα επεξεργάζονται κάθε είσοδο ανεξάρτητα, χωρίς να συγκρατούν μνήμη [8][10], ενώ τα CNNs εκμεταλλεύονται και συγκρατούν χωρικά χαρακτηριστικά χάρη στα φίλτρα συνέλιξης, τα οποία εντοπίζουν ακμές, σχήματα και υφές [9][11].

2.2.1 Αρχιτεκτονική των CNNs

Ένα CNN αποτελείται από διαδοχικά επίπεδα, τα οποία ανήκουν σε τρεις βασικές κατηγορίες: τα Συνελικτικά Επίπεδα (Convolutional Layers), τα Επίπεδα Υποδειγματοληψίας (Pooling Layers) και τα Πλήρως Συνδεδεμένα Επίπεδα (Fully Connected Layers). Τα συνελικτικά επίπεδα αποτελούν τον πυρήνα του δικτύου και είναι υπεύθυνα για την εξαγωγή των χαρακτηριστικών από την είσοδο. Τα επίπεδα υποδειγματοληψίας χρησιμοποιούνται για τη μείωση των διαστάσεων των δεδομένων, επιταχύνοντας τους υπολογισμούς και μειώνοντας την πιθανότητα υπερπροσαρμογής. Ενώ τα πλήρως συνδεδεμένα επίπεδα βρίσκονται συνήθως στο τέλος του δικτύου και είναι υπεύθυνα για την τελική ταξινόμηση [6][10].

Ο στόχος των CNNs είναι η σταδιακή μετατροπή της εισόδου (π.χ. εικόνας) σε ένα σύνολο χαρακτηριστικών. Μέσα από αυτήν την επεξεργασία, το CNN μαθαίνει να εξάγει χαρακτηριστικά και να αναγνωρίζει από απλές γραμμές και σχήματα μέχρι και πολύπλοκα αντικείμενα, τα οποία το οδηγούν στην παραγωγή μιας τελικής πρόβλεψης [2][11].

2.2.2 Λειτουργία της Συνέλιξης

Η συνέλιξη είναι ο πυρήνας των CNNs. Είναι μια μαθηματική πράξη στην οποία ένα μικρό φίλτρο εφαρμόζεται πάνω στην εικόνα εισόδου και υπολογίζει τοπικά γινόμενα και αθροίσματα. Έστω μια εικόνα εισόδου I και ένα φίλτρο K. Η συνέλιξη υπολογίζεται με τον παρακάτω τρόπο:

$$S(i,j) = (\boldsymbol{I} * \boldsymbol{K})(\boldsymbol{i},\boldsymbol{j}) = \boldsymbol{m}\sum \boldsymbol{n}\sum \boldsymbol{I}(\boldsymbol{i}+\boldsymbol{m},\boldsymbol{j}+\boldsymbol{n}) \cdot \boldsymbol{K}(\boldsymbol{m},\boldsymbol{n}) \quad (2.1)$$

Η διαδικασία αυτή επιτρέπει στο CNN να εξάγει τοπικά χαρακτηριστικά, όπως οι ακμές, οι γωνίες και οι υφές. Κάθε φίλτρο μαθαίνει έναν διαφορετικό τύπο χαρακτηριστικών. Το γεγονός αυτό καθιστά τα CNNs αρκετά ισχυρά σε προβλήματα που αφορούν εικόνες [6][8].

Η χρήση της συνέλιξης αποδεικνύεται αποτελεσματική γι' αυτόν τον σκοπό, επειδή κάθε νευρώνας συνδέεται με ένα μικρό τμήμα της εικόνας, γεγονός που μειώνει σημαντικά τον αριθμό των παραμέτρων, δηλαδή υπάρχει τοπική συνδεσιμότητα. Παράλληλα, το ίδιο φίλτρο εφαρμόζεται σε όλη την εικόνα, που σημαίνει ότι υπάρχει κοινή χρήση βαρών, με αποτέλεσμα τη μείωση της πολυπλοκότητας και τη βελτίωση κατά τη γενίκευση. Τέλος, τα CNNs μπορούν να αναγνωρίζουν χαρακτηριστικά ανεξάρτητα από τη θέση τους στην εικόνα, δηλαδή υπάρχει μεταθετική αμεταβλητότητα [10][11].

2.2.3 Επίπεδα Υποδειγματοληψίας

Τα Επίπεδα Υποδειγματοληψίας χρησιμοποιούνται για να μειώσουν την διάσταση των δεδομένων διατηρώντας τα πιο σημαντικά χαρακτηριστικά. Υπάρχουν δύο βασικοί τύποι pooling, το max pooling που υπολογίζει τη μέγιστη τιμή μιας περιοχής και το average pooling, το οποίο υπολογίζει τον μέσο όρο. Για παράδειγμα:

$$f(x) = \mathbf{max}(\boldsymbol{x1}, \boldsymbol{x2}, \ldots., \boldsymbol{xn}) \qquad (\mathbf{2.2})$$

Με την χρήση του pooling επιτυγχάνεται η μείωση του υπολογιστικού κόστους, η αποφυγή υπερπροσαρμογής, καθώς και η διατήρηση σημαντικών χαρακτηριστικών της εικόνας [6][10].

2.2.4 Συναρτήσεις Ενεργοποίησης

Μετά τη συνέλιξη, εφαρμόζεται στο μοντέλο μια μη γραμμική συνάρτηση ενεργοποίησης. Η πιο διαδεδομένη είναι η ReLU (Rectified Linear Unit):

$$f(x) = \mathbf{max}(\mathbf{0}, \boldsymbol{x}) \qquad (\mathbf{2.3})$$

2.2.5 Πλήρως Συνδεδεμένα Επίπεδα

Στο τελευταίο στάδιο του μοντέλου, τα χαρακτηριστικά που έχουν εξαχθεί μετατρέπονται σε διάνυσμα και μεταφέρονται σε πλήρως συνδεδεμένα επίπεδα τα οποία αναλαμβάνουν να τα ταξινομήσουν. Η τελική έξοδος δίνεται μέσω της συνάρτησης Softmax:

$$P(y = i) = \frac{\boldsymbol{e}^{\boldsymbol{z_i}}}{\sum_{\boldsymbol{j}} \boldsymbol{e}^{\boldsymbol{z_j}}} \qquad (\mathbf{2.4})$$

όπου z_i είναι το score της κάθε κλάσης [10][11].

2.3 Μεταφορά Μάθησης σε προβλήματα ανάλυσης εικόνας

Το Transfer Learning είναι μία από τις πιο σημαντικές τεχνικές στη σύγχρονη Μηχανική Μάθηση και την υπολογιστική όραση, διότι επιτρέπει την αξιοποίηση της γνώσης που έχει αποκτήσει ένα μοντέλο στην επίλυση ενός προβλήματος. Η ανάγκη για τη χρήση του Transfer Learning προκύπτει

κυρίως από το γεγονός ότι η εκπαίδευση των DNN απαιτεί τεράστιο όγκο δεδομένων και μεγάλους υπολογιστικούς πόρους, που συχνά δεν είναι διαθέσιμοι σε πρακτικές εφαρμογές [14][15].

Στη Μηχανική Μάθηση η εκπαίδευση κάθε μοντέλου γίνεται από την αρχή (training from scratch). Το γεγονός αυτό αυξάνει τον χρόνο εκπαίδευσης και δημιουργεί πιθανό πρόβλημα υπερπροσαρμογής, ειδικά σε μικρά datasets. Η χρήση του Transfer Learning αντιμετωπίζει αυτά τα προβλήματα, παρέχοντας προεκπαιδευμένα μοντέλα τα οποία έχουν μάθει να γενικεύουν γενικά χαρακτηριστικά από μεγάλα datasets [12][14].

2.3.1 Θεμελιώδης Ιδέα της Μεταφοράς Μάθησης

Η βασική αρχή του Transfer Learning στηρίζεται στην υπόθεση ότι τα πρώτα επίπεδα ενός νευρωνικού δικτύου μαθαίνουν γενικά χαρακτηριστικά, όπως οι ακμές, οι υφές και τα βασικά σχήματα. Αυτά τα χαρακτηριστικά είναι ανεξάρτητα από κάθε πρόβλημα και μπορούν να μεταφερθούν και σε άλλα προβλήματα. Αντίθετα, τα βαθύτερα επίπεδα του δικτύου μαθαίνουν πιο εξειδικευμένα χαρακτηριστικά, τα οποία σχετίζονται περισσότερο με το dataset του προβλήματος που έχουν κληθεί να επιλύσουν [14][17].

2.3.2 Μαθηματική Προσέγγιση Μεταφοράς Μάθησης

Το Transfer Learning μπορεί να θεωρηθεί ως προσαρμογή μιας συνάρτησης πρόβλεψης:

$$f(x;\theta_s) \rightarrow f(x;\theta_t) \tag{2.5}$$

Όπου θ_s είναι οι παράμετροι του προεκπαιδευμένου μοντέλου και θ_t οι παράμετροι του νέου προβλήματος. Η διαδικασία αυτή βοηθά στην μεταφορά γνώσης από το ήδη επιλυμένο πρόβλημα στο νέο πρόβλημα δηλαδή από το source domain στο target domain [14].

2.3.3 Τρόποι εφαρμογής Μεταφοράς Μάθησης

Η μέθοδος του Transfer Learning χρησιμοποιείται κυρίως για δύο βασικούς λόγους: το Feature Extraction και το Fine-Tuning. Στην προσέγγιση του Feature Extraction, τα convolutional layers παραμένουν σταθερά (frozen) και εκπαιδεύεται μόνο ο τελικός classifier. Η χρήση του Feature Extraction βοηθά στην ταχεία εκπαίδευση του μοντέλου, μειώνει τον κίνδυνο overfitting και είναι ιδανική για εκπαίδευση σε μικρά datasets [12].

Στην περίπτωση του fine-tuning, ξεπαγώνουν κάποια ή όλα τα layers, γεγονός το οποίο οδηγεί το μοντέλο στο να προσαρμόζεται πλήρως στο νέο dataset. Για τον λόγο αυτό, η μέθοδος του fine-tuning επιτυγχάνει μεγάλα ποσοστά ακρίβειας καθώς και πολύ καλύτερη προσαρμογή σε εξειδικευμένα δεδομένα [17].

2.3.4 Ρόλος των Προεκπαιδευμένων Μοντέλων

Τα πιο συχνά προεκπαιδευμένα μοντέλα που χρησιμοποιούνται είναι το VGG16/VGG19, το ResNet50, το InceptionResNetV2 και το MobileNet. Τα μοντέλα αυτά έχουν εκπαιδευτεί σε μεγάλα datasets με εκατομμύρια εικόνες και έχουν μάθει να αναπαριστούν και να εξάγουν χαρακτηριστικά από μια πληθώρα εικόνων [18][24].

2.4 Vision Transformers και μηχανισμοί προσοχής

Τα Vision Transformers αποτελούν μια πιο πρόσφατη καινοτομία στην ανάλυση εικόνας, βασισμένη στην αρχιτεκτονική των transformers, η οποία αναπτύχθηκε αρχικά για εφαρμογές

επεξεργασίας φυσικής γλώσσας (Natural Language Processing – NLP). Σε αντίθεση με τα CNNs, που βασίζονται σε τοπικά φίλτρα, τα ViTs χρησιμοποιούν μηχανισμούς προσοχής (attention mechanisms) για την επεξεργασία μιας εικόνας [7][23].

2.4.1 Βασική Αρχιτεκτονική

Η βασική λειτουργία των ViTs περιλαμβάνει τέσσερα στάδια. Αρχικά γίνεται διαχωρισμός της εικόνας σε patches. Δηλαδή, η εικόνα χωρίζεται σε μικρότερα κομμάτια (π.χ. 16x16 pixels), μετατρέποντας την εικόνα σε μια ακολουθία από patches. Εφόσον τα patches αποτελούνται από pixels, μέσω του embedding το μοντέλο τα μετατρέπει σε αριθμητικά διανύσματα, ώστε να μπορέσει να τα επεξεργαστεί. Έτσι, κάθε patch μετατρέπεται σε ένα διάνυσμα, το οποίο περιέχει την οπτική πληροφορία του τμήματος που αντιστοιχεί.

Επειδή οι transformers δεν γνωρίζουν τη σειρά των δεδομένων, μέσω μηχανισμών θεσιακής κωδικοποίησης προστίθεται ένα «σήμα» σε κάθε embedding, που υποδηλώνει τη θέση του στην εικόνα. Τέλος, τα embeddings περνούν μέσα από πολλαπλά encoder layers, τα οποία περιέχουν Multi-Head Self-Attention layers και MLPs (Multi-Layer Perceptrons). Τα Multi-Head Self-Attention layers επιτρέπουν σε κάθε patch να παρατηρεί όλα τα υπόλοιπα patches και να κατανοεί τις σχέσεις μεταξύ τους, ενώ τα MLPs χρησιμοποιούνται για την περαιτέρω επεξεργασία χαρακτηριστικών [26].

2.4.2 Ο Μηχανισμός Αυτοπροσοχής

Ο μηχανισμός αυτός αποτελεί το πιο σημαντικό κομμάτι των transformers. Με τον μηχανισμό αυτό το κάθε patch της εικόνας πηγαίνει στο σημείο το οποίο του αντιστοιχεί

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2.6}$$

Στην πράξη με τον συσχετισμό QK^T πολλαπλασιάζουμε το Query με το Key. Αν ταιριάζουν τα patches το αποτέλεσμα είναι ένας μεγάλος αριθμός. Η διαίρεση γίνεται για να κρατούνται οι αριθμοί σε λογικά πλαίσια και να σταθεροποιηθεί η εκπαίδευση. Η χρήση του softmax μετατρέπει τους αριθμούς σε ποσοστά. Τέλος γίνεται πολλαπλασιασμός με το Value παίρνουμε την πληροφορία από τα patches στα οποία αποφασίσε να δώσει προσοχή ο μηχανισμός [7].

2.4.3 Πολυκεφαλική Προσοχή

Με το Self-Attention το μοντέλο να κατανοεί τη συνολική δομή του αντικειμένου πολύ πιο γρήγορα.

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_n)W^O \tag{2.7}$$

Όπου το Concat μαζεύει τις απαντήσεις από τα κεφάλια και τις ενώνει ενώ το W^O είναι ένας τελικός πίνακας που συνδυάζει όλες αυτές τις διαφορετικές οπτικές σε μια ενιαία κατανόηση [7][23].

2.5 Εξηγήσιμη Τεχνητή Νοημοσύνη

Η ραγδαία ανάπτυξη της Μηχανικής Μάθησης και ιδιαίτερα της Βαθιάς Μάθησης έχει οδηγήσει στη δημιουργία μοντέλων με εξαιρετικά υψηλή απόδοση σε πλήθος εφαρμογών. Ωστόσο, τα μοντέλα αυτά συχνά λαμβάνουν αποφάσεις που δεν είναι άμεσα κατανοητές από τον άνθρωπο. Το γεγονός αυτό

δημιουργεί σημαντικά προβλήματα, ιδιαίτερα σε κρίσιμους τομείς όπως η υγεία και η κυβερνοασφάλεια, όπου η διαφάνεια και η ερμηνευσιμότητα των αποφάσεων είναι απαραίτητες [15].

Η Εξηγήσιμη Τεχνητή Νοημοσύνη (Explainable Artificial Intelligence) αποτελεί ένα πεδίο που στοχεύει στην ανάπτυξη μεθόδων και τεχνικών που καθιστούν τα μοντέλα Μηχανικής Μάθησης πιο διαφανή, κατανοητά και ερμηνεύσιμα [21].

Η ανάγκη για ερμηνευσιμότητα προκύπτει αρχικά από την ανάγκη για αξιοπιστία, διότι οι χρήστες πρέπει να είναι ικανοί να εμπιστεύονται τις αποφάσεις του μοντέλου. Παράλληλα, είναι σημαντικό να υπάρχει διαφάνεια, έτσι ώστε οι ερευνητές να μπορούν να κατανοούν τον τρόπο λήψης αποφάσεων του μοντέλου και να επεμβαίνουν, κάνοντας αλλαγές αν χρειαστεί. Τέλος, ένας ακόμα σημαντικός λόγος είναι η ανίχνευση σφαλμάτων, όπου μπορεί να γίνει εντοπισμός bias ή ακόμα και λανθασμένων προβλέψεων [8].

Στην ιατρική απεικόνιση, η ανάγκη για επεξηγήσιμα μοντέλα είναι ιδιαίτερα σημαντική, καθώς οι αποφάσεις ενός συστήματος τεχνητής νοημοσύνης μπορούν να επηρεάσουν άμεσα τη διαγνωστική διαδικασία και την κλινική πρακτική. Για τον λόγο αυτό, οι τεχνικές XAI χρησιμοποιούνται ώστε να αναδεικνύονται οι περιοχές της εικόνας που επηρεάζουν περισσότερο την τελική πρόβλεψη του μοντέλου.

2.5.1 Κατηγορίες Ερμηνευσιμότητας

Τα XAI χωρίζονται σε δύο βασικές κατηγορίες: τα Intrinsic Interpretability (Ενδογενής Ερμηνευσιμότητα) και τα Post-hoc Explainability (Εκ των Υστέρων Ερμηνεία). Τα ενδογενή είναι μοντέλα τα οποία είναι από τη «φύση» τους ερμηνεύσιμα, όπως η Linear Regression και τα Decision Trees. Αυτά, αν και έχουν περιορισμένη απόδοση σε πολύπλοκα προβλήματα, είναι πολύ απλά στην κατανόηση και βοηθούν στη διαφάνεια. Τα Post-hoc μοντέλα αφορούν τεχνικές που εφαρμόζονται μετά την εκπαίδευση ενός μοντέλου, όπως το LIME (Local Interpretable Model-agnostic Explanations) και το Grad-CAM (Gradient-weighted Class Activation Mapping) [8].

Τα μοντέλα Deep Learning παρουσιάζουν δυσκολία στην ερμηνεία τους, διότι αποτελούνται από μεγάλο αριθμό παραμέτρων, είναι μη γραμμικά και περιλαμβάνουν πολλά layers, γεγονός το οποίο συμβάλλει στην πολυπλοκότητά τους. Με αφορμή τις δυσκολίες αυτές, έχουν δημιουργηθεί μηχανισμοί που βοηθούν στην οπτικοποίηση του τρόπου με τον οποίο ένα μοντέλο λαμβάνει τις αποφάσεις του. Τέτοιοι μηχανισμοί είναι τα Integrated Gradients, το Attention Visualization καθώς και το Grad-CAM, το οποίο θα αναφερθεί πιο διεξοδικά στην Ενότητα 2.6 [21].

2.5.2 Πλεονεκτήματα και περιορισμοί Εξηγήσιμης Τεχνητής Νοημοσύνης

Όπως γίνεται κατανοητό, η χρήση των XAI μπορεί να βελτιστοποιήσει την ερμηνευσιμότητα ενός μοντέλου. Ως αποτέλεσμα, υπάρχει αυξημένη εμπιστοσύνη των χρηστών στο μοντέλο, διότι πλέον όλες οι αποφάσεις του είναι εμφανείς. Για τον ίδιο λόγο, ο χρήστης μπορεί να εντοπίσει αν υπάρχει πιθανή μεροληψία προς κάποια κατηγορία απαντήσεων, αλλά και να υποστηρίξει τη λήψη αποφάσεων του μοντέλου, γεγονός το οποίο οδηγεί στη βελτίωση ολόκληρου του μοντέλου [8][21].

Η XAI, παρά τη συμβολή της στη διαφάνεια, αντιμετωπίζει σημαντικές προκλήσεις. Αρχικά, το υπολογιστικό κόστος είναι υψηλό, διότι οι μέθοδοι ερμηνείας συχνά απαιτούν πρόσθετους πόρους πέρα από την ίδια την εκπαίδευση του μοντέλου. Η προσεγγιστική φύση των απαντήσεων σημαίνει ότι ο ερευνητής βλέπει μια απλουστευμένη μορφή των απαντήσεων που προήλθαν από μια σύνθετη διαδικασία, γεγονός που μπορεί να οδηγήσει σε πιθανή παραπλανητική ερμηνεία, η οποία αρχικά

φαίνεται λογική αλλά στην πραγματικότητα δεν ανταποκρίνεται με ακρίβεια στον τρόπο με τον οποίο το μοντέλο έλαβε την τελική απόφαση [21].

2.6 Τεχνικές Grad-CAM για οπτικοποίηση αποφάσεων μοντέλων

Όπως αναφέρθηκε στην προηγούμενη ενότητα, η ανάγκη για ερμηνευσιμότητα οδήγησε στην ανάπτυξη τεχνικών οπτικοποίησης των αποφάσεων των μοντέλων, που επιτρέπουν την κατανόηση της διαδικασίας με την οποία ένα μοντέλο καταλήγει σε αυτές. Μία από τις πιο διαδεδομένες και αποτελεσματικές μεθόδους είναι η Grad-CAM (Gradient-weighted Class Activation Mapping), η οποία παρέχει οπτικές εξηγήσεις μέσω θερμικών χαρτών (heatmaps) [8].

Η Grad-CAM αποτελεί τεχνική Explainable AI που χρησιμοποιείται κυρίως σε προβλήματα ταξινόμησης εικόνων και επιτρέπει τον εντοπισμό των περιοχών της εικόνας που επηρεάζουν περισσότερο την πρόβλεψη ενός μοντέλου [21].

2.6.1 Βασική Ιδέα Grad-CAM

Το Grad-CAM βασίζεται στην υπόθεση ότι οι σημαντικές περιοχές μιας εικόνας μπορούν να εντοπιστούν μέσα από την ανάλυση των gradients της εξόδου ως προς τα feature maps ενός συνελικτικού επιπέδου. Σε αντίθεση με πιο απλές τεχνικές, το Grad-CAM δεν απαιτεί τροποποίηση της αρχιτεκτονικής του μοντέλου, μπορεί να εφαρμοστεί σε προεκπαιδευμένα μοντέλα και παρέχει class-specific explanations [8][21].

2.6.2 Μαθηματική Ανάλυση Grad-CAM

Έστω ένα CNN και μία κλάση στόχος c. Η Grad-CAM βασίζεται στον υπολογισμό των gradients της εξόδου y^c ως προς τα feature maps A^k ενός convolutional layer.

Τα βάρη a_k^c υπολογίζονται ως:

$$a_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\theta y^c}{\theta A_{ij}^k} \tag{2.9}$$

Όπου Z ο αριθμός pixel στο feature map. Στην συνέχεια το heatmap υπολογίζεται με τον παρακάτω τρόπο[8]:

$$L_{Grad-CAM}^c = ReLU\left(\sum_k a_k^c\, A^k\right) \tag{2.10}$$

Οι βαθμίδες (gradients) χρησιμοποιούνται για τον υπολογισμό της σημασίας κάθε feature map ως προς την προβλεπόμενη κλάση, επιτρέποντας έτσι την παραγωγή ενός χάρτη θερμότητας που αναδεικνύει τις σημαντικότερες περιοχές της εικόνας.

2.6.3 Διαδικασία Υλοποίησης Grad-CAM

Η χρήση του Grad-CAM περιλαμβάνει επτά βήματα. Αρχικά γίνεται προώθηση της εικόνας στο μοντέλο. Στη συνέχεια γίνεται επιλογή της κλάσης ενδιαφέροντος. Υπολογίζονται τα gradients και

τα weights a_k^c. Η διαδικασία συνεχίζεται με συνδυασμό των feature maps και τη δημιουργία των heatmaps. Τέλος, γίνεται επικάλυψη του αποτελέσματος πάνω στην αρχική εικόνα [8].

Το αποτέλεσμα αυτής της διαδικασίας είναι ένας θερμικός χάρτης, όπου οι θερμές περιοχές (κίτρινο/κόκκινο χρώμα) δείχνουν μεγάλη συμβολή, ενώ οι ψυχρές περιοχές (μπλε χρώμα) δείχνουν μικρή ή μηδενική συμβολή. Η οπτικοποίηση αυτή συμβάλλει σημαντικά στην κατανόηση της απόφασης, στον εντοπισμό σημαντικών χαρακτηριστικών πάνω στην εικόνα, καθώς και στη διευκόλυνση της αξιολόγησης της αξιοπιστίας [21].

2.6.4 Πλεονεκτήματα και Περιορισμοί

Η εφαρμογή του Grad-CAM και η χρήση των θερμικών χαρτών έχουν σημαντικά πλεονεκτήματα στη Βαθιά Μάθηση. Αρχικά, μπορεί να εφαρμοστεί σε διαφορετικές αρχιτεκτονικές και δεν απαιτεί επανεκπαίδευση του μοντέλου. Είναι ικανό να παρέχει στους ερευνητές class-specific εξηγήσεις και οι πληροφορίες του είναι εύκολα κατανοητές οπτικά [8].

Παρά τα πλεονεκτήματά του, το Grad-CAM παρουσιάζει και κάποιους περιορισμούς. Η ανάλυση των heatmaps είναι χαμηλή και εξαρτώμενη από το επίπεδο των feature maps που χρησιμοποιούνται. Οι απαντήσεις που παρέχει εξαρτώνται από ένα επιλεγμένο layer, οπότε δεν παρέχεται πλήρης κατανόηση του μοντέλου. Επιπλέον, το Grad-CAM παρέχει προσεγγιστική και όχι πλήρως αιτιοκρατική ερμηνεία. Παράλληλα, υπάρχει πιθανότητα εμφάνισης θορύβου, που σημαίνει ότι το μοντέλο εστιάζει σε περισσότερες περιοχές από ό,τι θα έπρεπε το οποίο σημαίνει ότι, σε ιστοπαθολογικές εικόνες υψηλής πολυπλοκότητας μπορεί να εμφανίζονται διάχυτες ενεργοποιήσεις [8][21]. Τέλος το Grad-CAM δεν υποκαθιστά την κλινική ερμηνεία ειδικού ιατρού και απλώς αποτελεί ένα εργαλείο.

Κεφάλαιο 3ο: Ανασκόπηση Σχετικής Βιβλιογραφίας

3.1 Εφαρμογές βαθιάς μάθησης στην ιστοπαθολογία

Η εξέλιξη και η χρήση της βαθιάς μάθησης στον τομέα της ιστοπαθολογίας έχει αλλάξει τον τρόπο που πλέον προσεγγίζεται η διάγνωση κακοηθειών. Τα υπολογιστικά συστήματα είναι δυνατό να μάθουν ιεραρχικές αναπαραστάσεις από ψηφιακές εικόνες μικροσκοπίου, χωρίς να εξαρτώνται από μορφολογικά χαρακτηριστικά τα οποία έχουν σχεδιαστεί χειροκίνητα. Στα πεδία του καρκίνου του πνεύμονα και του παχέος εντέρου, αυτό το γεγονός είναι σύνηθες, διότι οι ιστοπαθολογικές εικόνες εμφανίζουν σημαντικές διαφοροποιήσεις μεταξύ υποτύπων του ίδιου οργάνου αλλά και μικρότερες μεταβολές, που είναι κρίσιμες για τη διάκριση μεταξύ ενός καλοήθους και ενός κακοήθους ιστού. Η γενική αυτή ιδέα αποτυπώνεται τόσο στη μελέτη ανασκόπησης των Rabia Javed, Tahir Abbas, Ali Haider Khan, Ali Daud, Amal Bukhari και Riad Alharbey, με τίτλο Deep Learning for Lungs Cancer Detection: a Review, όσο και στις νεότερες εφαρμοσμένες εργασίες πάνω σε ιστοπαθολογικά δεδομένα πνεύμονα και παχέος εντέρου, όπου τα CNNs αποτελούν τη βασική αρχιτεκτονική μάθησης [5][15].

Ένα μεγάλο μέρος των μελετών της συγκεκριμένης βιβλιογραφίας βασίζεται στο σύνολο δεδομένων LC25000, το οποίο περιλαμβάνει 25.000 ιστοπαθολογικές εικόνες πέντε κατηγοριών: lung adenocarcinoma, lung squamous cell carcinoma, lung benign tissue, colon adenocarcinoma και colon benign tissue. Η ευρεία χρήση του συγκεκριμένου dataset επιτρέπει τη σύγκριση διαφορετικών μεθόδων μέσα σε ένα σχετικά κοινό πειραματικό πλαίσιο. Για παράδειγμα, στις έρευνες των Shahid Mehmood, Mohamed M. Neamah, καθώς και των Md. Bipul Hossain και Mohamed Shaban, το LC25000 λειτουργεί ως βασικό σημείο αναφοράς για την ανάπτυξη και αξιολόγηση μοντέλων ταξινόμησης ιστοπαθολογικών εικόνων πνεύμονα και παχέος εντέρου [16][22].

Παρά την υψηλή ακρίβεια που αναφέρεται στη βιβλιογραφία, η βαθιά μάθηση έχει να προσφέρει περισσότερα στον τομέα της ιστοπαθολογίας απ' όσα είναι εμφανή. Μέσα από τη βιβλιογραφία αναδεικνύονται άλλοι δύο σημαντικοί άξονες, πέρα από τα υψηλά ποσοστά διαγνωστικής απόδοσης. Ο δεύτερος άξονας αφορά τη μείωση της υπολογιστικής πολυπλοκότητας, έτσι ώστε τα μοντέλα να είναι πρακτικά αξιοποιήσιμα. Ο τρίτος αφορά τον τρόπο με τον οποίο λαμβάνονται οι αποφάσεις, δηλαδή την ερμηνευσιμότητά τους και τη διαφάνεια των αποφάσεων, κρίσιμα στοιχεία ειδικά όταν πρόκειται για ιατρικά περιβάλλοντα [7][19]. Για παράδειγμα, η έρευνα Deep Learning and Texture Analysis for Lung and Colon Cancer Predicting των Mohamed M. Neamah, Laith A. Al-Ani και Loay E. George συνδυάζει χαρακτηριστικά βαθιάς μάθησης με texture analysis και τη χρήση Grad-CAM για οπτική εξήγηση των αποτελεσμάτων, δείχνοντας ότι η σύγχρονη βιβλιογραφία μετακινείται σταδιακά από την απλή ταξινόμηση προς περισσότερο ερμηνεύσιμα και κλινικά αποδεκτά συστήματα [17][19].

Από τη συνολική μελέτη παράλληλα προκύπτει επιπλέον ότι η βαθιά μάθηση δεν αξιοποιείται αποκλειστικά για μια απλή διάκριση ανάμεσα σε καλοήθεια και κακοήθεια αλλά και για πολυκατηγορική ταξινόμηση ανάμεσα σε ιστολογικούς υποτύπους. Ένα τέτοιο παράδειγμα αποτελεί η έρευνα Classification and Mutation Prediction from Non-Small Cell Lung Cancer Histopathology Images using Deep Learning των Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyö, Andre L. Moreira, Narges Razavian και Aristotelis Tsirigos, στην οποία χρησιμοποιήθηκε μια αρχιτεκτονική Transfer Learning με το μοντέλο Inception v3 σε whole-slide images για τη διάγνωση μεταξύ LUAD (Lung Adenocarcinoma), LUSC (Lung Squamous Cell Carcinoma) και φυσιολογικού πνευμονικού ιστού, επιτυγχάνοντας επίδοση συγκρίσιμη

με αυτήν των παθολογοανατόμων [14]. Η συγκεκριμένη μελέτη αποδεικνύεται ιδιαίτερα σημαντική, διότι δείχνει ότι οι βαθιές αρχιτεκτονικές δεν περιορίζονται σε απλή υποβοήθηση της διάγνωσης, αλλά μπορούν να επεκταθούν και σε πιο σύνθετα καθήκοντα, όπως η πρόβλεψη μοριακών μεταλλάξεων από μορφολογική πληροφορία.

Συνολικά, η εικόνα που σχηματίζεται είναι ότι η βαθιά μάθηση στην ιστοπαθολογία (στην προκειμένη περίπτωση του πνεύμονα και του παχέος εντέρου) εξελίσσεται γύρω από την υπόθεση ότι οι ιστολογικές δομές φέρουν επαρκή μορφολογική πληροφορία, ώστε ένα καλά εκπαιδευμένο μοντέλο να αντλήσει διακριτά πρότυπα για την αυτόματη αναγνώριση καρκινικών υποτύπων. Μια πιο ειδική και ευρέως χρησιμοποιούμενη έκφανση αυτής της υπόθεσης είναι οι συνελικτικές νευρωνικές δομές, οι οποίες αποτέλεσαν το κυρίαρχο ερευνητικό ρεύμα πριν από την ισχυρή εμφάνιση των transformer-based προσεγγίσεων [5].

3.2 Ανίχνευση καρκίνου του πνεύμονα και του παχέος εντέρου με CNN

Οι προσεγγίσεις που βασίζονται σε Convolutional Neural Networks αποτελούν τον θεμέλιο λίθο της σχετικής βιβλιογραφίας. Το γεγονός της επικράτησής τους δεν είναι τυχαίο, διότι τα CNNs έχουν σχεδιαστεί για να εκμεταλλεύονται χωρικά μοτίβα, ιεραρχική σύνθεση χαρακτηριστικών και επαναχρησιμοποίηση φίλτρων, στοιχεία που ταιριάζουν στη μορφολογία των ιστοπαθολογικών εικόνων [2][9]. Στην πράξη, τα CNNs χρησιμοποιούνται είτε ως προτεινόμενες αρχιτεκτονικές που μαθαίνουν από την αρχή είτε ως βάση για πιο εξελιγμένες ή ελαφριές παραλλαγές με attention μηχανισμούς.

Μια περίπτωση custom CNN εξετάζει η μελέτη Deep Learning-Based Classification of Lung and Colon Cancer using a Proposed CNN Architecture της Goldy Verma, στην οποία ένα CNN μοντέλο χρησιμοποιείται για την ταξινόμηση πέντε κατηγοριών ιστών. Στη μελέτη αυτή δίνεται έμφαση στην κατασκευή συμπαγών αρχιτεκτονικών, ικανών να διακρίνουν με σαφήνεια καλοήθη από κακοήθη ιστό, χωρίς να απαιτείται ιδιαίτερα σύνθετο pipeline προεπεξεργασίας ή εξωτερική εξαγωγή χαρακτηριστικών [2].

Μια ακόμα περίπτωση αποδεικνύει ότι τα CNNs είναι αποτελεσματικά όχι μόνο σε πεντακατηγορική ταξινόμηση συνδυασμένων lung-colon δεδομένων, αλλά και σε πιο εξειδικευμένα σενάρια ιστοπαθολογίας. Στην εργασία Deep Learning and Texture Analysis for Lung and Colon Cancer Predicting των Neamah, Al-Ani και George, αναφέρεται ότι προγενέστερες CNN προσεγγίσεις, όπως εκείνες των Mangal και Hatuwal & Thapa, πέτυχαν τιμές ακρίβειας της τάξης του 0.96–0.97 για lung και colon classification, στοιχείο που επιβεβαιώνει ότι ακόμη και σχετικά απλούστερες συνελικτικές δομές μπορούν να αποδώσουν ικανοποιητικά σε καλά επιμελημένα ιστοπαθολογικά datasets [17].

Η εξέλιξη, ωστόσο, δεν περιορίστηκε μόνο στα κλασικά συνελικτικά νευρωνικά δίκτυα. Μια νεότερη ερευνητική κατεύθυνση στράφηκε προς πιο αποδοτικές αρχιτεκτονικές με σκοπό να μειώσει τον αριθμό των παραμέτρων και το υπολογιστικό κόστος. Η εργασία Resource Efficient Attention Guided Deep Learning for Histopathology Based Lung and Colon Cancer Classification των Md. Bipul Hossain και Mohamed Shaban είναι ενδεικτική αυτής της μετατόπισης. Οι ερευνητές προτείνουν ένα attention-guided, ultra-lightweight CNN με depthwise και pointwise convolutions, καθώς και channel/spatial attention, επιτυγχάνοντας ακρίβεια άνω του 99% με εξαιρετικά μικρό αριθμό παραμέτρων και FLOPS (Floating Point Operations Per Second). Έτσι, μετατοπίζεται το κριτήριο αξιολόγησης από απλώς τη μέγιστη ακρίβεια σε μια ισορροπία ανάμεσα στην απόδοση και την υπολογιστική αποδοτικότητα [22].

Εξίσου σημαντική είναι και η διάσταση της κλινικής αξιοποίησης. Οι CNN-based μελέτες δείχνουν με συνέπεια ότι η αυτόματη ανάλυση ιστοπαθολογικών εικόνων μπορεί να λειτουργήσει ως

υποστηρικτικός μηχανισμός για τον παθολογοανατόμο, μειώνοντας τον χρόνο εξέτασης και ενισχύοντας την αναπαραγωγιμότητα της διάγνωσης. Όμως, μέσα από τη βιβλιογραφία μπορούν σαφώς να αναγνωριστούν και οι περιορισμοί. Οι υψηλές αποδόσεις συνδέονται συνήθως με datasets τα οποία είναι ισορροπημένα, όπως το LC25000 [10]. Αυτό σημαίνει ότι υπάρχει ένα ανοιχτό ερώτημα σε ετερογενή κλινικά περιβάλλοντα ως προς την απόδοση των μηχανισμών αυτών. Παράλληλα, πολλές μελέτες αναφέρουν προβλήματα γενίκευσης, κίνδυνο overfitting και ανάγκη για μεγαλύτερη ερμηνευσιμότητα των αποφάσεων.

Σύμφωνα με τη βιβλιογραφία, η βασική συνεισφορά των CNN-based ερευνών είναι ότι απέδειξαν πως η μορφολογική πληροφορία των ιστών αρκεί για εξαιρετικά υψηλής ακρίβειας αυτόματη ταξινόμηση. Υπάρχει όμως διαφοροποίηση ακόμα και ανάμεσα στις ίδιες τις έρευνες, διότι άλλα μοντέλα δίνουν έμφαση στην καθαρή ταξινομητική επίδοση, άλλα στη μείωση της πολυπλοκότητας και άλλα στην ενσωμάτωση texture analysis ή explainability. Το γεγονός αυτό θα αποτελέσει τη βάση για τη συγκριτική ανάλυση των CNN προσεγγίσεων σε επόμενο τμήμα.

3.3 Χρήση Μεταφοράς Μάθησης σε ιατρικές εικόνες

Η χρήση του Transfer Learning στην ιστοπαθολογία αποτελεί μία από τις πιο διαδεδομένες και αποτελεσματικές μεθόδους στη σημερινή εποχή. Η χρήση του TL επιτρέπει την αξιοποίηση προ εκπαιδευμένων μοντέλων, που έχουν εκπαιδευτεί σε μεγάλα datasets γενικού σκοπού, τα οποία στη συνέχεια προσαρμόζονται (fine-tuning) σε εξειδικευμένα ιατρικά προβλήματα [11].

Η ανάγκη για τη χρήση του προκύπτει από την έλλειψη μεγάλων, πλήρως επισημασμένων (labeled) συνόλων δεδομένων. Σε αντίθεση με εφαρμογές στην υπολογιστική όραση, όπου υπάρχουν τεράστιοι όγκοι εικόνων, στην ιστοπαθολογία η διαδικασία συλλογής και επισημείωσης δεδομένων απαιτεί εξειδικευμένη γνώση και πολύ χρόνο. Με το Transfer Learning επιτρέπεται η μεταφορά γνώσης από γενικά χαρακτηριστικά σε πιο εξειδικευμένα χαρακτηριστικά, τα οποία απαιτούνται στην ιστοπαθολογία, εξοικονομώντας χρόνο στους ερευνητές και στους ειδικούς [13].

Στο μεγαλύτερο μέρος των μελετών που εξετάστηκαν για τη συγγραφή της εργασίας αυτής χρησιμοποιούνται γνωστές αρχιτεκτονικές όπως το ResNet50, τα VGG16 / VGG19, το InceptionResNetV2, το InceptionV3 και το EfficientNet. Οι αρχιτεκτονικές αυτές λειτουργούν με έναν από τους δύο τρόπους που αναφέρθηκαν στην παρούσα εργασία στο Κεφάλαιο 2.3.3. Η επιλογή μεταξύ των δύο μεθόδων εξαρτάται από το μέγεθος του dataset, την πολυπλοκότητα του προβλήματος και την ανάγκη για γενίκευση [18].

3.3.1 Εφαρμογές σε καρκίνο πνεύμονα και παχέος εντέρου

Η αποτελεσματικότητα του TL φαίνεται μέσα από την πληθώρα μελετών που υπάρχουν για τη χρήση του σε ιστοπαθολογικά δεδομένα πνεύμονα και παχέος εντέρου. Για παράδειγμα, στην εργασία A Transfer Learning Approach for Classification of Lung Carcinoma, οι συγγραφείς εφαρμόζουν προεκπαιδευμένα μοντέλα για την ταξινόμηση τύπων καρκίνου του πνεύμονα, επιτυγχάνοντας σημαντικά υψηλή ακρίβεια μέσω fine-tuning. Μέσα από αυτήν τη μελέτη επιβεβαιώνεται ότι τα γενικά χαρακτηριστικά που έχουν μάθει τα προεκπαιδευμένα μοντέλα σε μια πληθώρα datasets εικόνων μπορούν να μεταφερθούν αρκετά αποτελεσματικά σε ιστοπαθολογικά δεδομένα, παρά τις μορφολογικές διαφορές [13].

Παράλληλα, στη μελέτη Deep Transfer Learning for Colon Cancer των Nils Gessert, Marcel Bengs και Lukas Wittig, το Transfer Learning χρησιμοποιείται για την ταξινόμηση ιστοπαθολογικών εικόνων του παχέος εντέρου. Οι ερευνητές καταλήγουν στο συμπέρασμα ότι τα προεκπαιδευμένα

μοντέλα υπερτερούν έναντι των CNNs που εκπαιδεύονται από την αρχή, ιδιαίτερα σε περιπτώσεις περιορισμένων δεδομένων. Επιπλέον, αναφέρεται ότι το TL δεν βελτιώνει μόνο την ακρίβεια αλλά και τη σταθερότητα της εκπαίδευσης [18].

Επιπλέον, στη μελέτη των Hidayah και Wisesy [12], προτείνεται μια προσέγγιση βασισμένη σε ensemble learning με τη χρήση του EfficientNet για την ταξινόμηση ιστοπαθολογικών εικόνων καρκίνου του πνεύμονα. Η χρήση πολλαπλών προ εκπαιδευμένων μοντέλων σε συνδυασμό με τεχνικές ensemble επιτρέπει τη βελτίωση της συνολικής απόδοσης και της σταθερότητας του συστήματος. Οι συγγραφείς καταδεικνύουν ότι η αξιοποίηση ελαφριών αλλά αποδοτικών αρχιτεκτονικών, όπως το EfficientNet, μπορεί να επιτύχει υψηλά επίπεδα ακρίβειας, μειώνοντας παράλληλα το υπολογιστικό κόστος. Το γεγονός αυτό αναδεικνύει ότι το Transfer Learning δεν περιορίζεται μόνο σε βαριά μοντέλα, αλλά μπορεί να συνδυαστεί και με αποδοτικές αρχιτεκτονικές για πρακτικές εφαρμογές σε ιατρικά δεδομένα.

Παράλληλα, στη μελέτη των Adekunle et al. [20], εξετάζεται η εφαρμογή προεκπαιδευμένων μοντέλων ResNet για την ανίχνευση καρκίνου του πνεύμονα σε δεδομένα αξονικής τομογραφίας (Computed Tomography – CT). Τα αποτελέσματα δείχνουν ότι η χρήση Transfer Learning σε συνδυασμό με βαθιές συνελικτικές αρχιτεκτονικές οδηγεί σε υψηλή διαγνωστική απόδοση ακόμη και σε διαφορετικό τύπο ιατρικών δεδομένων σε σχέση με την ιστοπαθολογία. Η συγκεκριμένη μελέτη είναι ιδιαίτερα σημαντική, καθώς αποδεικνύει ότι οι τεχνικές μεταφοράς μάθησης μπορούν να γενικευτούν σε διαφορετικές μορφές ιατρικής απεικόνισης, ενισχύοντας τη συνολική αξιοπιστία της προσέγγισης.

Στην εργασία Classification of Colon Cancer through Analysis of Histopathology Images using Transfer Learning, οι συγγραφείς χρησιμοποιούν μοντέλα όπως τα VGG και ResNet για την εξαγωγή χαρακτηριστικών από ιστοπαθολογικές εικόνες, επιτυγχάνοντας υψηλά ποσοστά επιτυχίας στην πολυκατηγορική ταξινόμηση [3]. Παρομοίως, στη μελέτη Malignancy Detection in Lung and Colon Histopathology Images Using Transfer Learning With Class Selective Image Processing, η χρήση προεκπαιδευμένων μοντέλων σε συνδυασμό με τεχνικές προεπεξεργασίας οδηγεί σε βελτιωμένη απόδοση [16].

Τέλος, στη μελέτη Prediction of Lung and Colon Cancer using Pre-trained Visualization διαπιστώνεται πως η χρήση visualization τεχνικών σε pretrained μοντέλα βοηθά στην καλύτερη κατανόηση των περιοχών ενδιαφέροντος [8][19].

3.4 Vision Transformers σε ιατρική απεικόνιση

Η εισαγωγή των ViTs στον τομέα της υπολογιστικής όρασης ξεκίνησε τη μετατόπιση από τις παραδοσιακές αρχιτεκτονικές που βασίζονταν στη συνέλιξη προς αρχιτεκτονικές που βασίζονται αποκλειστικά σε μηχανισμούς προσοχής. Η μελέτη An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale των Alexey Dosovitskiy, Alexander Kolesnikov κ.ά. αναφέρεται στην ιδέα της αναπαράστασης της εικόνας ως μια ακολουθία από κομμάτια (patches), τα οποία επεξεργάζονται από το μοντέλο με αντίστοιχο τρόπο όπως τα tokens στη φυσική γλώσσα. Έτσι, η αρχιτεκτονική αυτή επιτρέπει τη μοντελοποίηση μακρινών εξαρτήσεων χωρίς την ανάγκη βαθιάς στοίβαξης συνελικτικών φίλτρων. Η προσέγγιση αυτή διαφοροποιεί τα ViTs από τα CNNs, τα οποία βασίζονται σε τοπικά φίλτρα και χωρική συνέλιξη [7].

Η διαφορά αυτή είναι ιδιαίτερα σημαντική στον τομέα της ιστοπαθολογίας, διότι η διάγνωση δεν εξαρτάται αποκλειστικά από τοπικά χαρακτηριστικά αλλά και από σύνθετες χωρικές σχέσεις μεταξύ κυττάρων. Στην ιστοπαθολογία, η ιδιαιτερότητα των δεδομένων καθιστά αυτή την προσέγγιση αρκετά σημαντική. Επειδή οι ιστοπαθολογικές εικόνες έχουν υψηλή χωρική ανάλυση, έντονη ετερογένεια

ιστών και πολύπλοκες μορφολογικές σχέσεις μεταξύ κυττάρων, η μελέτη Vision Transformers for Computational Histopathology αποκαλύπτει ότι πολλές φορές τα CNNs δυσκολεύονται να αποτυπώσουν τέτοιες σχέσεις χωρίς πολύ βαθιές αρχιτεκτονικές, ενώ τα ViTs επιτρέπουν την άμεση σύλληψη του global context μέσω μηχανισμών self-attention [24].

3.4.1 Εφαρμογές Vision Transformers σε καρκίνο πνεύμονα και παχέος εντέρου

Η χρήση των ViTs στην ανάλυση ιστοπαθολογικών εικόνων για την ανίχνευση και την ταξινόμηση καρκίνου αποτελεί μια νέα και ταχέως αναπτυσσόμενη ερευνητική κατεύθυνση, με τη συνεχή εμφάνιση νέων ερευνών που επιβεβαιώνουν την αποτελεσματικότητα των ViTs σε σύγκριση με τις συνελικτικές προσεγγίσεις [24].

Στη μελέτη του Arvind Kumar κ.ά., Vision Transformer Based Effective Model for Early Detection and Classification of Lung Cancer, δημιουργείται ένα μοντέλο βασισμένο σε ViT για την ταξινόμηση καρκίνου του πνεύμονα χρησιμοποιώντας ιστοπαθολογικές εικόνες από το dataset LC25000. Οι συγγραφείς εφαρμόζουν διαφορετικές ρυθμίσεις patch size και καταδεικνύουν ότι η επιλογή patch size 16×16 οδηγεί στη βέλτιστη απόδοση, επιτυγχάνοντας ακρίβεια 98.84%. Με το αποτέλεσμα αυτό επιβεβαιώνεται πως τα ViTs μπορούν να αξιοποιηθούν αποτελεσματικά σε πραγματικά δεδομένα ιστοπαθολογίας χωρίς να υπάρχει ανάγκη για ιδιαίτερα σύνθετες αρχιτεκτονικές τροποποιήσεις [21].

Επιπλέον, οι συγγραφείς συγκρίνουν την προσέγγιση αυτή με άλλες μεθόδους Deep Learning, συμπεριλαμβανομένων CNNs και υβριδικών μοντέλων, καταλήγοντας στο συμπέρασμα ότι τα transformer-based μοντέλα παρουσιάζουν ανταγωνιστική ή ανώτερη απόδοση, ιδίως όταν συνδυάζονται με κατάλληλο fine-tuning [23].

Ακόμη πιο προχωρημένη εφαρμογή παρουσιάζεται στη μελέτη των Ji κ.ά., Automated Lung and Colon Cancer Classification Using Histopathological Images, στην οποία χρησιμοποιείται ο Swin Transformer V2 για την ταυτόχρονη ταξινόμηση καρκίνου πνεύμονα και παχέος εντέρου. Η αρχιτεκτονική αυτή αποτελεί εξέλιξη των απλών ViTs και επιτυγχάνει υψηλή απόδοση με 100% accuracy, precision, recall και F1-score στο dataset LC25000 μέσω 5-fold cross-validation. Έτσι, η μελέτη αυτή αποδεικνύει ότι τα transformer-based μοντέλα μπορούν να επιτύχουν άριστη απόδοση σε multi-class προβλήματα ιστοπαθολογίας και δείχνει ότι οι εξελιγμένες αρχιτεκτονικές, όπως ο Swin Transformer, μπορούν να υπερτερούν τόσο των κλασικών CNNs όσο και των αρχικών ViTs. Ωστόσο, οι ίδιοι οι συγγραφείς επισημαίνουν ότι τα αποτελέσματα αυτά βασίζονται σε benchmark datasets και απαιτείται περαιτέρω αξιολόγηση σε πιο ετερογενή κλινικά δεδομένα για την επιβεβαίωση της γενικευσιμότητας του μοντέλου [23].

Όσον αφορά τον καρκίνο του παχέος εντέρου, στην εργασία των Zeid κ.ά., Multiclass Colorectal Cancer Histology Image Classification Using Vision Transformers, εξετάζεται η απόδοση των ViTs σε ένα πολυκατηγορικό πρόβλημα που αποτελείται από 8 κλάσεις. Οι συγγραφείς συγκρίνουν ένα κλασικό ViT με έναν Compact Convolutional Transformer (CCT) και καταλήγουν ότι το CCT επιτυγχάνει ανώτερη απόδοση (~95% accuracy), ενώ το ViT φθάνει το 93.3%. Τα αποτελέσματα αυτά υπερτερούν των προηγούμενων CNN-based προσεγγίσεων στο ίδιο dataset, αποκαλύπτοντας ότι τα transformers είναι ικανά να διαχειρίζονται πιο σύνθετα προβλήματα από τις κλασικές αρχιτεκτονικές [4].

Τέλος, η έρευνα Vision Transformers for Computational Histopathology παρέχει μια συνολική παρουσίαση των ViTs στην ιστοπαθολογία, όπου τονίζεται η χρήση τους στην ιατρική απεικόνιση και το γεγονός ότι πλέον αποτελούν ένα υποσχόμενο εργαλείο για τον τομέα της ιατρικής [24].

3.4.2 Vision Transformers σε πολυκατηγορική ιστοπαθολογική ταξινόμηση

Η εφαρμογή των Vision Transformers σε πολυκατηγορικά προβλήματα ιστοπαθολογίας αποτελεί σημαντικό πεδίο αξιολόγησης της ικανότητάς τους να διαχειρίζονται ετερογενείς μορφολογικές δομές. Σε αντίθεση με δυαδικά ή απλούστερα classification tasks, τα πολυκατηγορικά προβλήματα απαιτούν διακριτοποίηση μεταξύ διαφορετικών τύπων ιστών με συχνά λεπτές μορφολογικές διαφορές [21].

Στην εργασία των Zeid, El-Bahnasy και Abo-youssef, Multiclass Colorectal Cancer Histology Image Classification Using Vision Transformers, όπως προαναφέρθηκε εξετάζονται δύο transformer-based μοντέλα ένα κλασικό Vision Transformer και ένα Compact Convolutional Transformer (CCT), στο dataset ιστοπαθολογικών εικόνων του Kather που περιέχονται 8 διαφορετικές κατηγορίες ιστών. Τα αποτελέσματα δείχνουν ότι το κλασικό ViT επιτυγχάνει 93.3% accuracy και το CCT επιτυγχάνει περίπου 95% accuracy [4].

Η σημασία αυτών των αποτελεσμάτων βασίζεται στο γεγονός πως το dataset αυτό είναι αρκετά πολύπλοκο καθώς περιλαμβάνει τόσο κακοήθεις όσο και φυσιολογικούς ιστούς διαφορετικών τύπων. Τέλος, η σύγκριση μεταξύ ViT και CCT δείχνει ότι η ενσωμάτωση συνελικτικών στοιχείων στο transformer πλαίσιο μπορεί να βελτιώσει περαιτέρω την απόδοση, γεγονός που επιβεβαιώνει τη σημασία των hybrid αρχιτεκτονικών [21].

3.4.3 Απόδοση Vision Transformers σε σχέση με το είδος δεδομένων

Οι μελέτες στο σύνολό τους δείχνουν ότι η απόδοση των Vision Transformers εξαρτάται σε μεγάλο βαθμό από το είδος των δεδομένων και τη δομή του προβλήματος. Για παράδειγμα, στην εργασία των Arvind Kumar κ.ά., Vision Transformer Based Effective Model for Early Detection and Classification of Lung Cancer, η εφαρμογή ViT στο dataset LC25000 (5 κλάσεις) οδηγεί σε υψηλή ακρίβεια (98.84%), υποδηλώνοντας ότι τα transformer-based μοντέλα μπορούν να λειτουργήσουν ιδιαίτερα αποδοτικά σε σχετικά καλά δομημένα datasets. Αντίστοιχα, στη μελέτη των Ji κ.ά., η χρήση του Swin Transformer στο ίδιο dataset οδηγεί σε ακόμη υψηλότερη απόδοση (100%), γεγονός που αποδίδεται στην αρχιτεκτονική βελτιστοποίηση του μοντέλου [23].

Όμως, σε πιο σύνθετα datasets, όπως αυτό της μελέτης του Zeid, το οποίο αποτελείται από 8 κλάσεις, το accuracy των ViTs μειώνεται (93–95%), γεγονός που υποδηλώνει ότι η δυσκολία του προβλήματος επηρεάζει σημαντικά την τελική απόδοση. Συνεπώς, μέσα από τη βιβλιογραφία μπορούμε να συμπεράνουμε ότι τα ViTs αποδίδουν εξαιρετικά σε καλά δομημένα datasets, ενώ η απόδοσή τους μειώνεται σημαντικά όταν τα προβλήματα είναι πολυκατηγορικά και πιο ετερογενή. Το μειονέκτημα αυτό, ωστόσο, μπορεί να αντιμετωπιστεί μέσω της ενσωμάτωσης hybrid μοντέλων [4][23].

3.4.4 Περιορισμοί και προκλήσεις

Παρά την υψηλή απόδοση των Vision Transformers, μέσα από τη βιβλιογραφία αναδεικνύονται και σημαντικοί περιορισμοί. Για παράδειγμα, η μελέτη των Ji κ.ά., Automated Lung and Colon Cancer Classification using Histopathological Images, αναφέρεται στο γεγονός ότι τα αποτελέσματα που επιτυγχάνουν τα μοντέλα σε ελεγχόμενα datasets δεν μπορούν να γενικευτούν άμεσα σε πραγματικά κλινικά δεδομένα [25], λόγω της μεγαλύτερης διακύμανσης ανάμεσα στις εικόνες.

Επιπλέον, στη μελέτη Vision Transformers for Computational Histopathology αναφέρεται πως τα transformer μοντέλα απαιτούν μεγάλο όγκο δεδομένων για αποτελεσματική εκπαίδευση, έχουν

υψηλό υπολογιστικό κόστος και παρουσιάζουν δυσκολίες κατά την εφαρμογή τους σε Whole Slide Images (WSI) [24].

Τέλος, τα transformers εξαρτώνται από υπερπαραμέτρους όπως το patch size, στο οποίο αναφέρεται η μελέτη Vision Transformer Based Effective Model for Early Detection and Classification of Lung Cancer των Kumar κ.ά., γεγονός που αυξάνει τον βαθμό πολυπλοκότητας κατά τη διαδικασία βελτιστοποίησης [23].

3.5 Συγκριτική Ανάλυση Σχετικών Ερευνών

Η συγκριτική αποτίμηση των ερευνητικών προσεγγίσεων για την ταξινόμηση καρκίνου του πνεύμονα και του παχέος εντέρου μέσω της βαθιάς μάθησης αναδεικνύει σαφείς διαφοροποιήσεις ως προς την απόδοση των μοντέλων, τη σταθερότητα των αποτελεσμάτων και την ικανότητα γενίκευσης σε διαφορετικά datasets. Η βιβλιογραφία που αναφέρθηκε και συλλέχθηκε για την εκπόνηση αυτής της διπλωματικής εργασίας διαχωρίζεται κυρίως σε τρεις βασικές κατηγορίες: τα συνελικτικά νευρωνικά δίκτυα, τις μεθόδους μεταφοράς μάθησης (Transfer Learning) και τα μοντέλα Vision Transformer, με κάθε κατηγορία να παρουσιάζει διαφορετικά χαρακτηριστικά ως προς την αποτελεσματικότητα και την πολυπλοκότητα.

Οι CNN-based προσεγγίσεις αποτελούν μία από τις πρώτες και πιο καθιερωμένες κατηγορίες μοντέλων στον τομέα της ιατρικής απεικόνισης και εμφανίζουν σταθερά υψηλά ποσοστά απόδοσης. Στη μελέτη της Goldy Verma [2], στην οποία προτείνεται μια προσαρμοσμένη συνελικτική αρχιτεκτονική για την ταξινόμηση εικόνων καρκίνου του παχέος εντέρου και του πνεύμονα, επιτυγχάνεται ποσοστό ακρίβειας που φτάνει το 99%, γεγονός το οποίο αποδεικνύει πως τα CNNs είναι ικανά να εξάγουν μορφολογικά χαρακτηριστικά από ιστοπαθολογικές εικόνες σε πολύ ικανοποιητικό βαθμό. Παρόμοια επίπεδα καταγράφονται και σε άλλες μελέτες, στις οποίες η ακρίβεια κυμαίνεται από 97% έως 99% (πιο συγκεκριμένα στις έρευνες Deep Learning for Lung Cancer Detection [5], Deep Learning and Texture Analysis for Lung and Colon Cancer Prediction [17], Resource Efficient Attention Guided Deep Learning [22]), τόσο σε δεδομένα αξονικής τομογραφίας (Computed Tomography – CT) όσο και σε ιστοπαθολογικές εικόνες. Ωστόσο, στις περισσότερες περιπτώσεις δεν παρέχονται πλήρεις μετρικές αξιολόγησης, όπως το F1-score, γεγονός που περιορίζει την πλήρη εκτίμηση της ισορροπίας μεταξύ precision και recall. Επιπλέον, οι αρχιτεκτονικές CNNs, παρά την υψηλή απόδοση που επιτυγχάνουν, παρουσιάζουν περιορισμούς ως προς τη δυνατότητα κατανόησης μακρινών χωρικών σχέσεων και προβλήματα γενίκευσης όταν το πρόβλημα περιλαμβάνει πιο σύνθετα δεδομένα.

Οι εφαρμογές του Transfer Learning εμφανίζουν ακόμη υψηλότερα ποσοστά ακρίβειας και πιο σταθερά αποτελέσματα, αποτελώντας την πιο ευρέως χρησιμοποιούμενη προσέγγιση στη σύγχρονη βιβλιογραφία. Στην έρευνα των Mamoona Humayun κ.ά. [13], όπου συγκρίνονται τα μοντέλα VGG16, VGG19 και Xception, καταγράφονται ακρίβειες 98.83% (VGG16), 98.05% (VGG19) και 97.4% (Xception), επιβεβαιώνοντας ότι τα προεκπαιδευμένα μοντέλα μπορούν να προσαρμοστούν αποτελεσματικά σε ιατρικά δεδομένα. Ακόμη υψηλότερα ποσοστά ακρίβειας παρατηρούνται σε μελέτες όπου χρησιμοποιήθηκαν τα μοντέλα MobileNetV2 και InceptionResNetV2, με την ακρίβεια να φτάνει έως και το 99.98% [18], υποδεικνύοντας ότι η αξιοποίηση προεκπαιδευμένων χαρακτηριστικών σε συνδυασμό με κατάλληλο fine-tuning οδηγεί σε εξαιρετικά αποτελέσματα. Στη μελέτη των Shahid Mehmood κ.ά. [16], η εφαρμογή της μεθόδου CSIP σε συνδυασμό με Transfer Learning αύξησε την ακρίβεια του baseline AlexNet από 89% σε 98.4%, γεγονός που αναδεικνύει τη σημασία της προεπεξεργασίας και της ενίσχυσης των χαρακτηριστικών. Επιπλέον, στη μελέτη Prediction of Lung and Colon Cancer using Pre-trained Visualization [8] αναφέρεται ακρίβεια περίπου 97–98%,

ενισχύοντας περαιτέρω τη σταθερότητα της κατηγορίας αυτής. Παρά τις υψηλές αποδόσεις, όπως και στα CNNs, υπάρχει περιορισμένη αναφορά σε μετρικές όπως το F1-score, γεγονός που αποτελεί κοινό χαρακτηριστικό πολλών σχετικών ερευνών.

Η χρήση των Vision Transformers επιτρέπει μια διαφορετική συγκριτική ανάλυση, διότι τα μοντέλα αυτά εμφανίζουν σημαντικές διαφοροποιήσεις τόσο ως προς την απόδοση όσο και ως προς τον τρόπο συμπεριφοράς τους. Στη μελέτη των Arvind Kumar κ.ά. [21], η χρήση Vision Transformer με patch size 16×16 οδηγεί σε ακρίβεια 98.84%, επιβεβαιώνοντας ότι τα transformer-based μοντέλα μπορούν να ανταγωνιστούν άμεσα τις CNN και Transfer Learning προσεγγίσεις. Όμως, σε πιο σύνθετα προβλήματα, όπως στη μελέτη των Zeid κ.ά. [4], που αφορά πολυκατηγορική ταξινόμηση με οκτώ κλάσεις, η απόδοση μειώνεται στο 93.3%, ενώ μια βελτιωμένη εκδοχή, το Compact Convolutional Transformer, φτάνει περίπου το 95%. Το γεγονός αυτό καθιστά εμφανές ότι η απόδοση των transformers εξαρτάται από τη δομή του προβλήματος και την αρχιτεκτονική υλοποίηση. Αντίθετα, στη μελέτη των Ji κ.ά. [23], όπου χρησιμοποιείται ο Swin Transformer V2, καταγράφεται ακρίβεια 100%, καθώς και precision, recall και F1-score ίσα με 100%, αποτελώντας το μοναδικό παράδειγμα στη βιβλιογραφία όπου παρέχονται πλήρη metrics σε τόσο υψηλό επίπεδο απόδοσης. Το αποτέλεσμα αυτό υποδηλώνει ότι οι εξελιγμένες μορφές Vision Transformers, ιδιαίτερα οι υβριδικές αρχιτεκτονικές, έχουν τη δυνατότητα να υπερκεράσουν τις παραδοσιακές προσεγγίσεις, αν και η γενικευσιμότητα των αποτελεσμάτων αυτών παραμένει υπό διερεύνηση.

Η σύγκριση και των τριών κατηγοριών αναδεικνύει το γεγονός ότι τα μοντέλα Transfer Learning παρουσιάζουν τη μεγαλύτερη σταθερότητα, με την ακρίβεια στα περισσότερα μοντέλα να κυμαίνεται άνω του 98% και με μικρές διακυμάνσεις ανάμεσα στις διαφορετικές μελέτες (Humayun κ.ά. [13], Mehmood κ.ά. [16], Transfer Learning Colon Cancer Study [18]). Τα μοντέλα CNNs εμφανίζουν επίσης υψηλή απόδοση και αξιοπιστία, αλλά πολλές φορές δεν παρουσιάζουν ούτε την ίδια σταθερότητα ούτε την κορυφαία ακρίβεια που επιτυγχάνουν τα μοντέλα Transfer Learning (Verma [2], Resource Efficient Attention Guided Deep Learning [24]). Αντιθέτως, τα Vision Transformers παρουσιάζουν μεγάλη διακύμανση, με την ακρίβεια να κυμαίνεται από 93% έως και 100% (Kumar κ.ά., Zeid κ.ά., Ji κ.ά.), γεγονός που δείχνει τόσο την ικανότητά τους για εξαιρετικά υψηλή απόδοση όσο και την εξάρτησή τους από συγκεκριμένες αρχιτεκτονικές και τα χαρακτηριστικά των δεδομένων. Ιδιαίτερης σημασίας αποτελεί το γεγονός ότι τα μοντέλα Vision Transformers είναι κατά κύριο λόγο τα μόνα που υποβάλλονται σε πιο πλήρη αξιολόγηση μέσω μετρικών όπως precision, recall και F1-score, γεγονός που επιτρέπει πιο ολοκληρωμένη εκτίμηση της απόδοσης [4][21][23].

Συνολικά, μέσα από τη συγκριτική ανάλυση γίνεται εμφανές ότι, ενώ το Transfer Learning παραμένει η πιο αξιόπιστη και πρακτική επιλογή για εφαρμογές ταξινόμησης σε ιστοπαθολογικά δεδομένα, τα Vision Transformers και ιδιαίτερα οι εξελιγμένες υβριδικές μορφές τους αποτελούν την πιο υποσχόμενη κατεύθυνση για μελλοντική έρευνα. Ταυτόχρονα, η υψηλή ακρίβεια που αναφέρεται σε ορισμένες μελέτες, όπως το 100% του Swin Transformer (Ji κ.ά. [23]), θα πρέπει να αξιολογείται με προσοχή, καθώς η γενικευσιμότητα σε πραγματικά κλινικά δεδομένα δεν είναι πάντα εξασφαλισμένη. Ως εκ τούτου, η επιλογή της κατάλληλης μεθοδολογίας εξαρτάται σε μεγάλο βαθμό από τη φύση του dataset, την πολυπλοκότητα του προβλήματος και τους υπολογιστικούς περιορισμούς, ενώ η μελλοντική έρευνα φαίνεται να κατευθύνεται προς τη σύγκλιση των CNN και transformer-based προσεγγίσεων σε υβριδικά μοντέλα υψηλής απόδοσης.

Από τον συγκριτικό Πίνακα 3.1 που παρατίθεται παρακάτω παρατηρείται ότι οι περισσότερες μελέτες χρησιμοποιούν το LC25000, ένα δημόσια διαθέσιμο dataset με 25.000 ιστοπαθολογικές εικόνες πνεύμονα και παχέος εντέρου, κατανεμημένες σε πέντε ισορροπημένες κλάσεις. Η κοινή χρήση του

ίδιου dataset επιτρέπει πιο άμεση σύγκριση μεταξύ CNN, Transfer Learning και Vision Transformer προσεγγίσεων. Συνολικά, τα μοντέλα Transfer Learning εμφανίζουν τη μεγαλύτερη σταθερότητα, ενώ οι transformer-based αρχιτεκτονικές παρουσιάζουν πολύ υψηλές επιδόσεις και σημαντική προοπτική για μελλοντική έρευνα. Ωστόσο, η σύγκριση πρέπει να γίνεται με προσοχή, καθώς ορισμένες μελέτες, όπως η Zeid et al. [4], χρησιμοποιούν διαφορετικό dataset, συγκεκριμένα το Kather Colorectal Cancer Histology Dataset, το οποίο αφορά μόνο καρκίνο του παχέος εντέρου.

Από τη βιβλιογραφική ανασκόπηση προκύπτει ότι τα CNN και οι τεχνικές transfer learning έχουν εφαρμοστεί εκτενώς στην ιστοπαθολογική ταξινόμηση, ενώ τα Vision Transformers αποτελούν πιο πρόσφατη προσέγγιση με αυξημένες απαιτήσεις σε δεδομένα και υπολογιστικούς πόρους. Η παρούσα εργασία διαφοροποιείται μέσω της συγκριτικής αξιολόγησης πολλαπλών αρχιτεκτονικών στο ίδιο dataset και με ενιαίο πειραματικό πρωτόκολλο.

Πίνακας 3.1: Πίνακας Σύγκρισης Ερευνών

Ἔρευνα	Dataset	Τεχνική	Ειδος εικόνας	Μετρικές	Βασικό συμπέρασμα
Verma [2]	LC25000	CNN	Ιστοπαθολογικές εικόνες H&E πνεύμονα και παχέος εντέρου	Lung cancer: 97%, Colon cancer: 96%	Η μελέτη δείχνει ότι τα CNNs μπορούν να επιτύχουν υψηλή ακρίβεια στην ταξινόμηση ιστοπαθολογικών εικόνων πνεύμονα και παχέος εντέρου, ακόμη και με σχετικά απλούστερη αρχιτεκτονική.
Zeid, El-Bahnasy & Abo-youssef [4]	Kather	Vision Transformer και Compact Convolutional Transformer	Ιστοπαθολογικές RGB εικόνες H&E καρκίνου παχέος εντέρου	ViT: Accuracy 93.3%, Precision 93.44%, Recall 93.33%, F1-score 93.3%. CCT: Accuracy 94.75%, Precision 94.8%, Recall 94.75%, F1-score 94.74%	Το CCT υπερέχει του απλού ViT, επειδή συνδυάζει transformer αρχιτεκτονική με convolutional tokenization, διατηρώντας καλύτερα την τοπική πληροφορία των ιστοπαθολογικών εικόνων.
Garg & Garg [8]	LC25000	Pre-trained CNN models με visualization activation	Ιστοπαθολογικές εικόνες H&E πνεύμονα και παχέος εντέρου	Accuracy περίπου 97%–98%	Η χρήση προεκπαιδευμένων CNN μοντέλων σε συνδυασμό με οπτικοποιήσεις ενεργοποίησης ενισχύει τόσο την απόδοση όσο και την ερμηνευσιμότητα της ταξινόμησης.

Amitbhave, Maheswari, James & Remya [11]	LC25000	Transfer Learning EfficientNetB3	Ιστοπαθολογικές εικόνες H&E πνεύμονα και παχέος εντέρου	Accuracy: 99%	Το EfficientNetB3 μέσω transfer learning μπορεί να επιτύχει πολύ υψηλή ακρίβεια στην ταξινόμηση ιστοπαθολογικών εικόνων μειώνοντας παράλληλα την ανάγκη εκπαίδευσης μοντέλου από την αρχή.
Humayun et al. [13]	LC25000	Transfer Learning με VGG16, VGG19 και Xception	Ιστοπαθολογικές εικόνες H&E πνεύμονα και παχέος εντέρου	VGG16: 98.83%, VGG19: 98.05%, Xception: 97.4%	Το Transfer Learning παρουσιάζει υψηλή και σταθερή απόδοση, καθώς αξιοποιεί προεκπαιδευμένα χαρακτηριστικά και τα προσαρμόζει αποτελεσματικά στα ιστοπαθολογικά δεδομένα.
Mehmood et al. [16]	LC25000	AlexNet + CSIP + Transfer Learning	Ιστοπαθολογικές εικόνες H&E πνεύμονα και παχέος εντέρου	Baseline AlexNet: 89%, με CSIP: 98.4%	Η ενίσχυση της ποιότητας της εικόνας και η κατάλληλη προεπεξεργασία μπορούν να αυξήσουν σημαντικά την απόδοση ενός προεκπαιδευμένου μοντέλου.
Neamah, Al-Ani & George [17]	LC25000	GLCM + AM-EfficientNet B2	Ιστοπαθολογικές εικόνες H&E πνεύμονα και παχέος εντέρου	GLCM: Accuracy 92.28%, Precision 92.57%, Recall 92.28%, F1-score 92.19% AM-EfficientNet B2: Accuracy 99.96%, Precision 100%, Recall 100%, F1-score 100%	Η μελέτη δείχνει ότι τα deep features του AM-EfficientNet B2 είναι σημαντικά πιο αποτελεσματικά από τα αποκλειστικά texture-based χαρακτηριστικά του GLCM.
Kumar et al. [21]	LC25000	Vision Transformer	Ιστοπαθολογικές εικόνες H&E πνεύμονα και παχέος εντέρου	Accuracy 98.84%	Τα Vision Transformers μπορούν να ανταγωνιστούν άμεσα τις CNN και Transfer Learning
Hossain & Shaban [22]	LC25000, με εξωτερική επαλήθευση σε NCT-CRC-HE-100K και CRC-VAL-HE-7K	Lightweight CNN με DPCA blocks, channel attention και spatial attention	Ιστοπαθολογικές εικόνες H&E πνεύμονα και παχέος εντέρου	Colon 99.97%, Lung 3-class: έως 99.13%	Η μελέτη αναδεικνύει ότι μπορεί να επιτευχθεί πολύ υψηλή ακρίβεια με εξαιρετικά μικρό αριθμό παραμέτρων, γεγονός που καθιστά το μοντέλο κατάλληλο για περιβάλλοντα περιορισμένων υπολογιστικών πόρων.

Jie et al. [23]	LC25000	Swin Transformer V2	Ιστοπαθολογικές εικόνες πνεύμονα και παχέος εντέρου	Accuracy: 100%, Precision: 100%, Recall: 100%, F1-score: 100%	Το Swin Transformer V2 εμφανίζει την υψηλότερη αναφερόμενη απόδοση, απαιτείται προσοχή ως προς τη γενικευσιμότητα σε πραγματικά κλινικά δεδομένα.

Κεφάλαιο 4ο: Δεδομένα και Προεπεξεργασία

Ένα μοντέλο Μηχανικής και Βαθιάς Μάθησης, για να πετύχει γενικευσιμότητα, ειδικά σε έναν ευαίσθητο τομέα όπως η ιατρική απεικόνιση και η ογκολογία μέσω υπολογιστικών μέσων, εξαρτάται άμεσα από την ποιότητα, την ποσότητα και τη σωστή διαχείριση των δεδομένων που χρησιμοποιούνται για την εκπαίδευσή του. Το παρόν κεφάλαιο εστιάζει στην ανάλυση του συνόλου δεδομένων (dataset) που χρησιμοποιήθηκε για την ανάπτυξη των συγκριθέντων μοντέλων, στη φύση των ιστοπαθολογικών εικόνων, καθώς και στα στάδια προεπεξεργασίας και επαύξησης δεδομένων (data augmentation) που εφαρμόστηκαν. Τέλος, παρατίθεται το πειραματικό πρωτόκολλο που ακολουθήθηκε, διασφαλίζοντας την αξιοπιστία, την αμεροληψία και την αναπαραγωγιμότητα των αποτελεσμάτων της έρευνας.

4.1 Περιγραφή του Συνόλου Δεδομένων LC25000

Για τους σκοπούς της παρούσας διπλωματικής εργασίας, επιλέχθηκε το αναγνωρισμένο σύνολο δεδομένων LC25000 (Lung and Colon Cancer Histopathological Images). Πρόκειται για μια εξαιρετικά σημαντική συλλογή δεδομένων στον τομέα της ψηφιακής παθολογίας, η οποία δημιουργήθηκε με σκοπό να διευκολύνει την ανάπτυξη αλγορίθμων τεχνητής νοημοσύνης για την αυτόματη ανίχνευση και ταξινόμηση καρκινικών ιστών στον πνεύμονα και το παχύ έντερο.

Το dataset αποτελείται από συνολικά 25.000 έγχρωμες ιστοπαθολογικές εικόνες (RGB), οι οποίες εξήχθησαν από ψηφιοποιημένες αντικειμενοφόρους πλάκες (whole-slide images) υπό μικροσκοπική μεγέθυνση. Αξίζει να σημειωθεί πως το αρχικό σύνολο δεδομένων αποτελείται από 750 αρχικές εικόνες (500 για τις κατηγορίες του εντέρου και 250 για τις κατηγορίες του πνεύμονα) οι οποίες αυξήθηκαν στις 25.000 εικόνες μέσω data augmentation. Το κύριο πλεονέκτημα χρήσης του συγκεκριμένου συνόλου δεδομένων είναι το γεγονός ότι υπάρχει απόλυτη ισορροπία στις κλάσεις. Πιο συγκεκριμένα, κάθε μία από τις πέντε κλάσεις περιλαμβάνει 5.000 εικόνες. Η ισορροπία αυτή είναι ζωτικής σημασίας, διότι εξαλείφεται από την αρχή το πρόβλημα της ανισορροπίας δεδομένων, που συχνά οδηγεί τα μοντέλα στη δημιουργία μεροληπτικών συμπεριφορών προς τις κλάσεις στις οποίες οι εικόνες πλειοψηφούν.

Οι εικόνες του συγκεκριμένου dataset έχουν αρχική ανάλυση 768×768 pixels και συλλέχθηκαν από κλινικά δείγματα ασθενών, τηρώντας όλους τους κανόνες ανωνυμοποίησης και προστασίας ιατρικών δεδομένων (π.χ. πρότυπα HIPAA). Η ψηφιοποίησή του πραγματοποιήθηκε με σύγχρονους σαρωτές ιστοπαθολογίας, διασφαλίζοντας υψηλή ευκρίνεια στα μορφολογικά χαρακτηριστικά των κυττάρων, όπως ο πυρήνας, το κυτταρόπλασμα και η δομή του ιστού. Με βάση τους παραπάνω λόγους, το dataset είναι ιδανικό για την αξιολόγηση αρχιτεκτονικών Deep Learning, προσφέροντας επαρκή όγκο δεδομένων για την εκπαίδευση βαθιών δικτύων χωρίς τον κίνδυνο άμεσης υπερεκπαίδευσης λόγω έλλειψης δειγμάτων.

4.2 Κατηγορίες ιστοπαθολογικών εικόνων

Η αρχιτεκτονική του προβλήματος ταξινόμησης (classification problem) που εξετάζει η παρούσα εργασία είναι πολυταξική και περιλαμβάνει πέντε (5) διακριτές κατηγορίες-τάξεις , που αφορούν τόσο υγιείς όσο και παθολογικούς ιστούς από δύο διαφορετικά όργανα (παχύ έντερο και πνεύμονας). Η κατανόηση των βιολογικών χαρακτηριστικών κάθε κλάσης είναι απαραίτητη για την ερμηνεία των αποφάσεων του μοντέλου. Οι 5.000 εικόνες κάθε κλάσης κατανέμονται ως εξής:

• Καλοήθης ιστός πνεύμονα (Lung Benign Tissue - lung_n): Η κατηγορία αυτή περιλαμβάνει ιστοπαθολογικές εικόνες από υγιή πνευμονικό ιστό. Τα μορφολογικά χαρακτηριστικά παρουσιάζουν

κανονική κυτταρική οργάνωση, ομοιόμορφα μεγέθη πυρήνων και απουσία νεοπλασματικών αλλοιώσεων.

• Αδενοκαρκίνωμα πνεύμονα (Lung Adenocarcinoma - lung_aca): Πρόκειται για τον πιο συχνό τύπο καρκίνου του πνεύμονα (Μη Μικροκυτταρικός Καρκίνος του Πνεύμονα - NSCLC). Σε ιστολογικό επίπεδο, χαρακτηρίζεται από τον σχηματισμό αδενικών δομών ή την παραγωγή βλεννίνης. Τα κύτταρα παρουσιάζουν ατυπία, υπερχρωματικούς πυρήνες και αποδιοργάνωση του ιστού.

• Ακανθοκυτταρικό καρκίνωμα πνεύμονα (Lung Squamous Cell Carcinoma - lung_scc): Ένας άλλος υπότυπος NSCLC, ο οποίος προέρχεται από τα πλακώδη επιθηλιακά κύτταρα που επενδύουν τους αεραγωγούς. Χαρακτηρίζεται συχνά από την παρουσία κερατινοποίησης και διακυτταρικών γεφυρών. Η διάκριση μεταξύ αδενοκαρκινώματος και ακανθοκυτταρικού καρκινώματος είναι υψίστης κλινικής σημασίας για την επιλογή στοχευμένης θεραπείας.

• Καλοήθης ιστός παχέος εντέρου (Colon Benign Tissue - colon_n): Εικόνες υγιούς βλεννογόνου του παχέος εντέρου. Παρουσιάζουν φυσιολογικές κρύπτες (αδένες), που επενδύονται από κανονικά καλυκοειδή κύτταρα, με τακτική διάταξη και απουσία δυσπλασίας.

• Αδενοκαρκίνωμα παχέος εντέρου (Colon Adenocarcinoma - colon_aca): Κακοήθης όγκος που προέρχεται από τον αδενικό ιστό του εντέρου. Χαρακτηρίζεται από ακανόνιστους, διηθητικούς αδένες, νέκρωση, κυτταρική ατυπία και πλειομορφισμό.

Η χρήση του συγκεκριμένου dataset επιτρέπει την ταυτόχρονη εκπαίδευση των μοντέλων σε ιστούς δύο διαφορετικών οργάνων, προσθέτοντας ένα επιπλέον επίπεδο πολυπλοκότητας, καθώς το δίκτυο καλείται να μάθει τόσο την οργανο-ειδική μορφολογία (πνεύμονας vs έντερο) όσο και την παθολογία (καλοήθης vs κακοήθης ιστός).

4.3 Διαχωρισμός δεδομένων

Για να διασφαλιστεί η σωστή εκπαίδευση του μοντέλου και η ικανότητά του να γενικεύει σε νέα, άγνωστα δεδομένα (generalization), το σύνολο δεδομένων υποβλήθηκε σε αυστηρό διαχωρισμό. Η διαδικασία αυτή είναι θεμελιώδης για την αποφυγή της υπερπροσαρμογής, όπου το μοντέλο απλώς απομνημονεύει τα δεδομένα εκπαίδευσης χωρίς να μαθαίνει τα υποκείμενα μοτίβα.

Επιλέχθηκε μια προσέγγιση στρωματοποιημένης δειγματοληψίας (stratified sampling), προκειμένου να διατηρηθεί η αναλογία των κλάσεων (20% ανά κλάση) σε κάθε υποσύνολο. Η κατανομή πραγματοποιήθηκε με το σύνηθες πρότυπο 80% - 10% - 10% ως εξής:

• Σύνολο Εκπαίδευσης (Training Set - π.χ. 80%): Αποτελεί τον κύριο όγκο των δεδομένων (17.600 εικόνες) και χρησιμοποιείται αποκλειστικά για την ενημέρωση των βαρών (weights) του νευρωνικού δικτύου μέσω του αλγορίθμου οπισθοδιάδοσης του σφάλματος (backpropagation).

• Σύνολο Επικύρωσης (Validation Set - π.χ. 10%): Αποτελείται από εικόνες (4.400 εικόνες) που δεν συμμετέχουν στην ενημέρωση των βαρών. Αντιθέτως, χρησιμοποιείται στο τέλος κάθε εποχής (epoch) για την αξιολόγηση της απόδοσης του μοντέλου κατά τη διάρκεια της εκπαίδευσης. Αποτελεί κρίσιμο δείκτη για τον συντονισμό των υπερπαραμέτρων (hyperparameter tuning) και για την εφαρμογή τεχνικών όπως το Early Stopping, ώστε να σταματήσει η εκπαίδευση όταν το σφάλμα επικύρωσης αρχίσει να αυξάνεται.

• Σύνολο Ελέγχου (Test Set - π.χ. 10%): Ένα εντελώς ανεξάρτητο υποσύνολο (3.000 εικόνες), το οποίο παραμένει κλειδωμένο (unseen data) μέχρι την ολοκλήρωση της έρευνας και χρησιμοποιείται μόνο για την τελική αξιολόγηση. Η τελική αξιολόγηση και οι μετρικές (Accuracy, F1-Score κ.λπ.) που

παρουσιάζονται στα αποτελέσματα εξάγονται αποκλειστικά από αυτό το σύνολο, προσφέροντας μια αμερόληπτη και ρεαλιστική εκτίμηση της απόδοσης του μοντέλου σε πραγματικές συνθήκες.

4.4 Προεπεξεργασία εικόνων

Η χρήση ακατέργαστων εικόνων δεν είναι κατάλληλη για άμεση εισαγωγή σε Συνελικτικά Νευρωνικά Δίκτυα. Για τον λόγο αυτό απαιτούνται ορισμένα βήματα προεπεξεργασίας των εικόνων, με σκοπό τη βελτίωση της απόδοσης και τη διευκόλυνση της σύγκλισης του αλγορίθμου βελτιστοποίησης (Adam, SGD).

Το πρώτο βήμα προεπεξεργασίας είναι η αλλαγή μεγέθους των εικόνων. Επειδή η αρχική ανάλυση των εικόνων είναι 768×768 pixels και θεωρείται υπολογιστικά απαιτητική, όλες οι εικόνες σμικρύνθηκαν σε τυπικές διαστάσεις (128×128 ή 224×224 pixels) με τη χρήση αλγορίθμων παρεμβολής.

Ένα επιπλέον σημαντικό βήμα προεπεξεργασίας αποτελεί η κανονικοποίηση. Με δεδομένο ότι οι τιμές των pixels σε ψηφιακές εικόνες κυμαίνονται από 0 έως 255 (για τα χρωματικά κανάλια R, G, B), η κανονικοποίηση συμβάλλει στην αποφυγή προβλημάτων εκρηγνυόμενων ή εξαφανιζόμενων κλίσεων (exploding/vanishing gradients). Για την επίτευξη καλύτερης σύγκλισης, οι τιμές κανονικοποιήθηκαν στο διάστημα [0,1] μέσω διαίρεσης με το 255.

Στη συνέχεια, λόγω της χρήσης Transfer Learning στην έρευνα, οι εικόνες τυποποιήθηκαν χρησιμοποιώντας τη μέση τιμή (mean) και την τυπική απόκλιση (standard deviation – std) του dataset ImageNet ανά κανάλι χρώματος (π.χ. mean = [0.485, 0.456, 0.406] και std = [0.229, 0.224, 0.225]).

Τέλος, ο χρωματικός χώρος παρέμεινε σταθερός, δηλαδή οι εικόνες διατηρήθηκαν στον χρωματικό χώρο RGB, καθώς οι ιστοχημικές χρώσεις (όπως η αιματοξυλίνη και ηωσίνη - H&E) βασίζονται στη χρωματική διαφοροποίηση για την ανάδειξη των πυρήνων (μπλε/μωβ) και του κυτταροπλάσματος (ροζ).

4.5 Τεχνικές Επαύξησης Δεδομένων

Αν και το dataset LC25000 είναι εκτενές, η χρήση data augmentation κρίνεται απαραίτητη στην ιατρική απεικόνιση. Με τη χρήση επαύξησης δεδομένων δημιουργούνται νέες συνθετικές αλλά ρεαλιστικές εκδοχές εικόνων κατά τη διάρκεια της εκπαίδευσης, αυξάνοντας τεχνητά την ποικιλομορφία του συνόλου εκπαίδευσης και προσδίδοντας στο μοντέλο ανθεκτικότητα σε αλλαγές οπτικής γωνίας και φωτισμού. Αξίζει να επιπλέον να αναφερθεί πως οι τεχνικές data augmentation εφαρμόστηκαν μόνο στο σύνολο εκπαίδευσης .

Η πρώτη τεχνική επαύξησης που χρησιμοποιήθηκε είναι η τυχαία περιστροφή. Η παράμετρος αυτή ορίζει ένα εύρος (σε μοίρες) μέσα στο οποίο η εικόνα εισόδου μπορεί να περιστραφεί τυχαία. Στην προκειμένη περίπτωση, για κάθε νέα εικόνα που παράγεται κατά την εκπαίδευση, επιλέγεται τυχαία μια γωνία περιστροφής μέσα στο εύρος που έχει ορίσει ο ερευνητής. Η περιστροφή είναι σημαντική, διότι στις συνθήκες λήψης ιστοπαθολογικών εικόνων η ευθυγράμμιση του δείγματος ιστού στο μικροσκόπιο μπορεί να ποικίλλει. Η τεχνική αυτή βοηθά το μοντέλο να αναγνωρίζει τα χαρακτηριστικά του ιστού ανεξάρτητα από τη γωνία λήψης.

Στη συνέχεια αξιοποιείται η τυχαία οριζόντια και κάθετη μετατόπιση (width/height shift range), η οποία μετατοπίζει τυχαία την εικόνα κατά μήκος του οριζόντιου άξονα (πλάτος) και του κάθετου άξονα (ύψος). Με τη χρήση της, τα αντικείμενα ενδιαφέροντος (π.χ. κυτταρικοί πυρήνες, αδενικές

δομές) δεν βρίσκονται πάντα τέλεια κεντραρισμένα στην εικόνα. Μέσω της μετατόπισης, το μοντέλο εκπαιδεύεται να αναγνωρίζει αυτά τα χαρακτηριστικά ακόμη και όταν εμφανίζονται εκτός κέντρου.

Επιπλέον, χρησιμοποιείται τυχαία διατμητική παραμόρφωση (shear range). Η διάτμηση είναι ένας γεωμετρικός μετασχηματισμός που «γέρνει» την εικόνα κατά μήκος ενός άξονα. Με αυτόν τον τρόπο προσομοιώνονται αλλαγές στην οπτική γωνία ή ελαφριές παραμορφώσεις που μπορεί να προκύψουν κατά τη διαδικασία προετοιμασίας του δείγματος ιστού. Η τεχνική αυτή βοηθά το μοντέλο να μην βασίζεται αποκλειστικά σε αυστηρά γεωμετρικά σχήματα, τα οποία μπορεί να μην είναι σταθερά σε όλα τα δείγματα.

Παράλληλα, τα δεδομένα έχουν υποστεί τυχαία μεγέθυνση και σμίκρυνση (zoom range). Η παράμετρος αυτή επιλέχθηκε διότι στην ιστοπαθολογία χρησιμοποιούνται διαφορετικές μεγεθύνσεις μικροσκοπίου για την εξέταση των δειγμάτων. Μέσω της τυχαίας εστίασης, το μοντέλο γίνεται πιο ανθεκτικό στις αλλαγές κλίμακας και μαθαίνει να αναγνωρίζει τα χαρακτηριστικά ανεξάρτητα από το μέγεθός τους.

Τέλος, οι εικόνες έχουν υποστεί τυχαία κάθετη και οριζόντια αναστροφή (vertical/horizontal flip). Οι αναστροφές βοηθούν σημαντικά τα μοντέλα, διότι στις ιστοπαθολογικές εικόνες ο προσανατολισμός του ιστού είναι αυθαίρετος. Με τη χρήση τους στον συγκεκριμένο τομέα πολλαπλασιάζονται οι παραλλαγές κάθε εικόνας, ενισχύοντας περαιτέρω τη δυνατότητα γενίκευσης του μοντέλου.

4.6 Πειραματικό Πρωτόκολλο

Για τη διασφάλιση της αξιοπιστίας και της επιστημονικής εγκυρότητας των αποτελεσμάτων, σχεδιάστηκε και εφαρμόστηκε ένα συνεπές πειραματικό πρωτόκολλο, το οποίο επικεντρώνεται στη δίκαιη σύγκριση των εξεταζόμενων μοντέλων υπό κοινές συνθήκες αξιολόγησης.

Αρχικά, όλα τα μοντέλα αξιολογήθηκαν χρησιμοποιώντας το ίδιο σύνολο δεδομένων και τον ίδιο διαχωρισμό σε training, validation και test sets. Για την αποφυγή πιθανής διαρροής πληροφορίας (data leakage) ο διαχωρισμός των εικόνων σε σύνολα εκπαίδευσης, επικύρωσης και ελέγχου πραγματοποιήθηκε πριν την εφαρμογή τεχνικών data augmentation. Οι τεχνικές ενίσχυσης δεδομένων εφαρμόστηκαν αποκλειστικά στο σύνολο εκπαίδευσης ενώ τα σύνολα επικύρωσης και ελέγχου διατηρήθηκαν αμετάβλητα. Το ίδιο test set χρησιμοποιήθηκε στην τελική αξιολόγηση όλων των μοντέλων ώστε να διασφαλιστεί η δίκαιη σύγκριση των αρχιτεκτονικών. Με τον τρόπο αυτό διασφαλίστηκε ότι οι διαφοροποιήσεις στην απόδοση οφείλονται αποκλειστικά στα χαρακτηριστικά και την αρχιτεκτονική των μοντέλων.

Στη συνέχεια, υιοθετήθηκε ενιαία στρατηγική εκπαίδευσης για τα convolutional μοντέλα, βασισμένη στη λογική του Transfer Learning. Συγκεκριμένα, η εκπαίδευση πραγματοποιήθηκε σε δύο στάδια για τα μοντέλα Transfer Learning: αρχικά εκπαιδεύτηκαν μόνο τα τελικά επίπεδα ταξινόμησης, ενώ στη συνέχεια πραγματοποιήθηκε fine-tuning επιλεγμένων βαθύτερων επιπέδων. Το σύνολο των εποχών εκπαίδευσης ήταν 10 (5 στην Α΄ φάση εκπαίδευσης και 5 στη Β΄ φάση). Η διαδικασία αυτή επιτρέπει την αποτελεσματική προσαρμογή των προεκπαιδευμένων μοντέλων στο συγκεκριμένο πρόβλημα, χωρίς να απαιτείται εκπαίδευση από την αρχή. Το μοντέλο CNN εκπαιδεύτηκε από την αρχή σε μία φάση, επίσης για 10 εποχές εκπαίδευσης.

Για το Vision Transformer εφαρμόστηκε διαφορετική προσέγγιση, καθώς το μοντέλο εκπαιδεύτηκε εξ ολοκλήρου από την αρχή. Η επιλογή αυτή κρίθηκε απαραίτητη λόγω της διαφορετικής φύσης της αρχιτεκτονικής, η οποία δεν βασίζεται σε προεκπαιδευμένα convolutional φίλτρα. Επιπλέον,

το μοντέλο αξιολογήθηκε σε πολλαπλά στάδια εκπαίδευσης, επιτρέποντας την ανάλυση της εξέλιξης της απόδοσής του.

Η διαδικασία εκπαίδευσης όλων των μοντέλων υποστηρίχθηκε από μηχανισμούς ελέγχου σύγκλισης και αποφυγής υπερπροσαρμογής, όπως η διακοπή εκπαίδευσης όταν δεν παρατηρείται βελτίωση (early stopping) και η προσαρμογή του ρυθμού μάθησης. Με τον τρόπο αυτό διασφαλίστηκε ότι τα μοντέλα δεν υπερεκπαιδεύονται και ότι τα τελικά αποτελέσματα αντανακλούν τη βέλτιστη δυνατή απόδοση.

Η αξιολόγηση των μοντέλων πραγματοποιήθηκε με τη χρήση πολλαπλών μετρικών, ώστε να αποτυπωθεί ολοκληρωμένα η απόδοσή τους. Πέρα από τη συνολική ακρίβεια, χρησιμοποιήθηκαν μετρικές όπως precision, recall και F1-score, οι οποίες παρέχουν πιο λεπτομερή εικόνα της συμπεριφοράς των μοντέλων. Παράλληλα, η ανάλυση των confusion matrices επέτρεψε την εις βάθος κατανόηση των σφαλμάτων και της συνέπειας των προβλέψεων. Τέλος, ιδιαίτερη έμφαση δόθηκε στην ομοιομορφία της διαδικασίας αξιολόγησης, καθώς όλα τα μοντέλα εξετάστηκαν υπό το ίδιο πλαίσιο και με τα ίδια κριτήρια. Αυτό αποτελεί κρίσιμο στοιχείο του πειραματικού πρωτοκόλλου, καθώς επιτρέπει την εξαγωγή αξιόπιστων και αντικειμενικών συμπερασμάτων.

Συνολικά, το πειραματικό πρωτόκολλο σχεδιάστηκε με στόχο τη μεγιστοποίηση της συγκρισιμότητας και της αξιοπιστίας των αποτελεσμάτων, διασφαλίζοντας ότι τα συμπεράσματα της εργασίας βασίζονται σε συνεπή και επιστημονικά τεκμηριωμένη διαδικασία.

Κεφάλαιο 5ο: Σχεδιασμός και Υλοποίηση Μοντέλων

5.1 Βασικό Συνελικτιό Νευρωνικό Δίκτυο

Στην έρευνα της ανάλυσης ιατρικής εικόνας, και ειδικότερα στην ιστοπαθολογία (όπου οι ιστολογικές τομές παρουσιάζουν υψηλή ενδο-ταξική ποικιλομορφία και δια-ταξική ομοιότητα), η ανάπτυξη ενός στιβαρού βασικού μοντέλου αποτελεί το θεμέλιο της πειραματικής διαδικασίας. Η προσέγγιση που υιοθετήθηκε στην παρούσα εργασία αποκλίνει από τις παραδοσιακές τεχνικές μηχανικής όρασης, όπου απαιτείται χειροκίνητος σχεδιασμός χαρακτηριστικών, και αξιοποιεί την ικανότητα των Συνελικτικών Νευρωνικών Δικτύων για ενδογενή και ιεραρχική αυτόματη εξαγωγή χαρακτηριστικών (end-to-end feature learning).

Στο πλαίσιο της παρούσας διπλωματικής εργασίας, η ανάπτυξη ενός βασικού Συνελικτικού Νευρωνικού Δικτύου αποτέλεσε το πρώτο και θεμελιώδες βήμα για την προσέγγιση του προβλήματος ταξινόμησης εικόνων. Το μοντέλο αυτό υλοποιήθηκε στο περιβάλλον του Jupyter Notebook και χρησιμοποιήθηκε ως baseline προσέγγιση, δηλαδή ως σημείο αναφοράς για την αξιολόγηση της αποτελεσματικότητας τόσο της αρχιτεκτονικής όσο και της συνολικής μεθοδολογίας.

Στην επιλογή του baseline CNN μοντέλου, η παρούσα εργασία δεν περιορίζεται σε μια απλή αρχική υλοποίηση, αλλά εντάσσεται σε μια διαδικασία πειραματισμού και σύγκρισης. Συγκεκριμένα, το baseline μοντέλο λειτουργεί ως σημείο αναφοράς που επιτρέπει την απομόνωση της επίδρασης της ίδιας της αρχιτεκτονικής από άλλους παράγοντες, όπως η χρήση προεκπαιδευμένων βαρών ή πιο σύνθετων μηχανισμών μάθησης. Με τον τρόπο αυτό καθίσταται δυνατή η αντικειμενική αξιολόγηση του κατά πόσο το dataset μπορεί να υποστηρίξει αποτελεσματική μάθηση χωρίς εξωτερική γνώση. Έτσι, καθίσταται εφικτή η αξιολόγηση της αναγκαιότητας πιο εξελιγμένων μοντέλων σε επόμενα στάδια του προβλήματος.

Η στρατηγική αυτή είναι κρίσιμη για τη σύγκριση με τα επόμενα μοντέλα που εξετάζονται στην εργασία. Τα μοντέλα Μεταφοράς Μάθησης (Transfer Learning), όπως τα VGG και ResNet, εισάγουν ήδη εκπαιδευμένες αναπαραστάσεις από μεγάλα datasets (π.χ. ImageNet), γεγονός που μπορεί να οδηγήσει σε σημαντική βελτίωση της απόδοσης. Ωστόσο, χωρίς ένα baseline που να έχει εκπαιδευτεί αποκλειστικά στο συγκεκριμένο dataset, δεν είναι εφικτό να ποσοτικοποιηθεί η πραγματική συνεισφορά της μεταφοράς γνώσης. Αντίστοιχα, οι αρχιτεκτονικές Vision Transformer, οι οποίες βασίζονται σε μηχανισμούς attention και όχι σε συνελικτικές πράξεις, διαφοροποιούνται πλήρως ως προς τον τρόπο επεξεργασίας της εικόνας. Το baseline CNN επιτρέπει τη σύγκριση όχι μόνο σε επίπεδο απόδοσης, αλλά και σε επίπεδο συμπεριφοράς ως προς την ικανότητα εξαγωγής τοπικών έναντι παγκόσμιων χαρακτηριστικών.

Η υλοποίηση του μοντέλου βασίστηκε σε μια ισορροπημένη αρχιτεκτονική, η οποία αποφεύγει τόσο την υπεραπλούστευση όσο και την υψηλή πολυπλοκότητα. Στόχος ήταν η δημιουργία ενός μοντέλου με επαρκή εκφραστική ικανότητα, το οποίο να μπορεί να μάθει ουσιαστικά χαρακτηριστικά χωρίς να απαιτεί υπερβολικό χρόνο εκπαίδευσης ή να οδηγείται εύκολα σε φαινόμενα υπερπροσαρμογής.

Ένα ακόμη κρίσιμο σημείο που αναδεικνύεται είναι η ανάγκη ισορροπίας μεταξύ εκφραστικής ικανότητας και υπολογιστικού κόστους. Στην παρούσα υλοποίηση, το baseline μοντέλο σχεδιάστηκε με τέτοιο τρόπο ώστε να είναι αρκετά βαθύ για να μπορεί να συλλάβει πολύπλοκα μοτίβα, αλλά όχι τόσο ώστε να καθίσταται δύσκολη η εκπαίδευσή του ή να οδηγεί εύκολα σε υπερπροσαρμογή. Η επιλογή

αυτή είναι άμεσα συνδεδεμένη με τα χαρακτηριστικά του dataset, το οποίο, αν και επαρκές σε μέγεθος, δεν συγκρίνεται με μεγάλα datasets γενικής χρήσης. Ένα υπερβολικά σύνθετο μοντέλο θα είχε αυξημένο κίνδυνο να μάθει θόρυβο αντί για ουσιαστικά χαρακτηριστικά, ενώ ένα υπερβολικά απλό μοντέλο δεν θα μπορούσε να αποτυπώσει την πολυπλοκότητα των ιστολογικών δομών.

Η πρακτική σημασία αυτής της ισορροπίας φαίνεται και κατά την εκπαίδευση στο notebook, όπου η εξέλιξη των μετρικών απόδοσης αντικατοπτρίζει την ικανότητα του μοντέλου να γενικεύει. Η παρακολούθηση της διαφοράς μεταξύ training και validation απόδοσης επιτρέπει την αξιολόγηση του κατά πόσο το μοντέλο έχει επαρκή χωρητικότητα χωρίς να υπερπροσαρμόζεται. Η πληροφορία αυτή είναι καθοριστική για τη μετάβαση σε πιο σύνθετα μοντέλα, καθώς καθορίζει αν η βελτίωση που επιτυγχάνεται οφείλεται στην αρχιτεκτονική ή απλώς στην αύξηση των παραμέτρων.

Ιδιαίτερη σημασία έχει επίσης η ικανότητα του CNN να αντιμετωπίζει το πρόβλημα των «λεπτών διαφοροποιήσεων» στις ιστοπαθολογικές εικόνες. Στις συγκεκριμένες εικόνες, οι διαφορές μεταξύ κατηγοριών δεν είναι εμφανείς σε μακροσκοπικό επίπεδο, αλλά εντοπίζονται σε μικρές αλλαγές στη δομή και την υφή των κυττάρων. Το baseline CNN, μέσω της τοπικής επεξεργασίας που υλοποιείται στα convolutional layers, επιτρέπει την ανίχνευση αυτών των μικρής κλίμακας μοτίβων. Η σταδιακή αύξηση του βάθους της αρχιτεκτονικής επιτρέπει τον συνδυασμό αυτών των τοπικών πληροφοριών σε πιο σύνθετες αναπαραστάσεις, οι οποίες είναι κρίσιμες για τη διάκριση μεταξύ παρόμοιων κατηγοριών.

Σε αντίθεση με πιο εξελιγμένα μοντέλα που ενσωματώνουν attention ή παγκόσμια επεξεργασία της εικόνας, το baseline CNN βασίζεται αποκλειστικά σε τοπικά φίλτρα. Το χαρακτηριστικό αυτό το καθιστά ιδιαίτερα κατάλληλο για την ανίχνευση μικροσκοπικών διαφοροποιήσεων, αλλά ενδέχεται να περιορίζει την ικανότητά του να συλλαμβάνει μακροσκοπικές συσχετίσεις. Η παρατήρηση αυτή αποτελεί βασικό σημείο σύνδεσης με τα επόμενα στάδια της εργασίας, όπου εξετάζεται κατά πόσο αρχιτεκτονικές όπως τα ViTs μπορούν να βελτιώσουν την απόδοση μέσω παγκόσμιας κατανόησης της εικόνας.

Όσον αφορά το επίπεδο υλοποίησης, το μοντέλο αξιοποιεί διαδοχικά convolutional blocks με φίλτρα μικρού μεγέθους, τα οποία εφαρμόζονται τοπικά στις εικόνες και επιτρέπουν την εξαγωγή χωρικών χαρακτηριστικών. Η χρήση pooling layers μειώνει τις διαστάσεις των ενδιάμεσων αναπαραστάσεων και περιορίζει την υπολογιστική επιβάρυνση, ενώ παράλληλα συμβάλλει στη γενίκευση του μοντέλου. Στο τελικό στάδιο, τα χαρακτηριστικά που έχουν εξαχθεί μετατρέπονται μέσω flattening σε ένα διάνυσμα και οδηγούνται σε πλήρως συνδεδεμένα επίπεδα για την τελική ταξινόμηση. Η στρατηγική αυτή επιτρέπει στο μοντέλο να μαθαίνει προοδευτικά χρήσιμες αναπαραστάσεις, χωρίς να απαιτείται ξεκάθαρος καθορισμός χαρακτηριστικών από τον χρήστη. Ωστόσο, σε αντίθεση με μια καθαρά θεωρητική προσέγγιση της ιεραρχικής μάθησης, η παρούσα εργασία εστιάζει κυρίως στο πώς αυτή η διαδικασία εκφράζεται πρακτικά μέσω της αρχιτεκτονικής που υλοποιήθηκε και στον τρόπο που επηρεάζει τα αποτελέσματα. Η απόδοση του baseline μοντέλου παρέχει άμεση ένδειξη για το κατά πόσο τα δεδομένα μπορούν να διαχωριστούν αποτελεσματικά με μια σχετικά απλή συνελικτική προσέγγιση.

Συνολικά, η διαδικασία μετασχηματισμού των δεδομένων από απλές σε σύνθετες αναπαραστάσεις αποτελεί το θεμέλιο της επιτυχίας των συνελικτικών νευρωνικών δικτύων και εξηγεί την υψηλή τους απόδοση σε προβλήματα ανάλυσης εικόνας, όπως αυτό που εξετάζεται στην παρούσα εργασία, λειτουργώντας παράλληλα ως το βασικό μέτρο σύγκρισης (benchmark) για την αποτίμηση πιθανών μελλοντικών μετασχηματισμών στην αρχιτεκτονική, όπως η εισαγωγή μηχανισμών προσοχής ή αλγορίθμων μεταφοράς μάθησης , οι οποίοι χρησιμοποιήθηκαν στα επόμενα στάδια της εργασίας.

5.2 Σχεδιασμός αρχιτεκτονικής CNN

Ο σχεδιασμός της αρχιτεκτονικής του Συνελικτικού Νευρωνικού Δικτύου που υλοποιήθηκε στο πλαίσιο της εργασίας είναι μια στοχευμένη διαδικασία, η οποία συνδυάζει καθιερωμένες αρχές της βαθιάς μάθησης με προσαρμογές που ανταποκρίνονται στις ιδιαιτερότητες του συγκεκριμένου προβλήματος. Η αρχιτεκτονική CNN δεν επιλέχθηκε αυθαίρετα αλλά με κριτήριο την ικανότητά της να εξάγει ιεραρχικά χαρακτηριστικά από τις εικόνες, διατηρώντας ταυτόχρονα την ισορροπία μεταξύ απόδοσης και γενίκευσης.

Η συνολική δομή του μοντέλου ακολουθεί τη λογική των σύγχρονων CNNs, όπου το δίκτυο χωρίζεται λειτουργικά σε δύο βασικά υποσυστήματα: το τμήμα εξαγωγής χαρακτηριστικών (Feature Extraction Backbone) και το τμήμα ταξινόμησης (Classification Head). Η αρχιτεκτονική του προτεινόμενου συστήματος ανήκει στην κατηγορία των μοντέλων ακολουθιακής ροής (Sequential) και βελτιστοποιήθηκε για την επεξεργασία έγχρωμων εικόνων (με RGB) με χωρική ανάλυση 128x128 εικονοστοιχείων (pixels). Ο σχεδιασμός ακολουθεί την αρχή της «πυραμιδικής» συμπίεσης: καθώς αυξάνεται το βάθος του δικτύου, η χωρική διάσταση των δεδομένων ελαττώνεται, ενώ η διάσταση των καναλιών αυξάνεται.

Στο στάδιο της εξαγωγής χαρακτηριστικών, το μοντέλο αξιοποιεί διαδοχικά συνελικτικά επίπεδα, στα οποία εφαρμόζονται φίλτρα μικρών διαστάσεων. Τα φίλτρα αυτά λειτουργούν ως ανιχνευτές τοπικών προτύπων, επιτρέποντας στο δίκτυο να αναγνωρίζει βασικά μορφολογικά στοιχεία των εικόνων. Το στάδιο αυτό αποτελείται από τέσσερα συνελικτικά blocks, στα οποία εκτελείται μια αλληλουχία γραμμικών και μη γραμμικών μετασχηματισμών. Πρώτα απ’ όλα, στα επίπεδα συνέλιξης, που αποτελούν τον πυρήνα του δικτύου, χρησιμοποιήθηκαν πυρήνες συνέλιξης μεγέθους 3x3 σε όλα τα επίπεδα, ελαχιστοποιώντας τον αριθμό των παραμέτρων και διατηρώντας το ίδιο πεδίο υποδοχής με μεγαλύτερα φίλτρα. Στο πρώτο block βρίσκονται 32 φίλτρα, τα οποία ανιχνεύουν χαμηλού επιπέδου μοτίβα, όπως οι ακμές κυτταρικών μεμβρανών. Σταδιακά, στα επόμενα blocks, ο αριθμός των φίλτρων αυξάνεται εκθετικά (από 64 στα 128 και στα 256), επιτρέποντας στα βαθύτερα επίπεδα να συνθέτουν τα απλά μοτίβα σε πολύπλοκες, αφηρημένες αναπαραστάσεις υψηλού επιπέδου. Ως συνάρτηση ενεργοποίησης έχει επιλεγεί η ReLU, επειδή η γραμμικότητά της για θετικές τιμές αποτρέπει το φαινόμενο της εξαφάνισης της κλίσης (vanishing gradient problem) κατά την οπισθοδιάδοση, επιταχύνοντας σημαντικά τη σύγκλιση του μοντέλου.

Η στρατηγική αυτή τονίζει την ανάγκη για μεγαλύτερη εκφραστική ικανότητα στα βαθύτερα επίπεδα, όπου το μοντέλο καλείται να αναγνωρίσει πιο σύνθετα και εξειδικευμένα χαρακτηριστικά. Στο πλαίσιο της εργασίας, η επιλογή αυτή συμβάλλει στην αποτελεσματική ανάλυση των δεδομένων, επιτρέποντας την ανίχνευση διαφορών που δεν είναι άμεσα εμφανείς.

Μετά από κάθε συνέλιξη και πριν από την υποδειγματοληψία, παρεμβάλλεται ένα επίπεδο Κανονικοποίησης Παρτίδας. Επειδή στις ιστολογικές εικόνες παρατηρείται συχνά μεγάλη διαφοροποίηση στους χρωματικούς τόνους λόγω διαφορετικών πρωτοκόλλων χρώσης στα εργαστήρια, η κανονικοποίηση παρτίδας χρησιμοποιείται για να ελαττώσει την εσωτερική μετατόπιση συμμεταβλητών, βοηθώντας στη σταθεροποίηση της κατανομής των ενεργοποιήσεων και καθιστώντας το μοντέλο πιο ανθεκτικό σε «θορύβους».

Παράλληλα, ενσωματώνονται επίπεδα υποδειγματοληψίας μεγέθους 2x2, που υποδιπλασιάζουν τη διάσταση των feature maps σε κάθε βήμα. Η διαδικασία αυτή δεν μειώνει μόνο δραματικά τον υπολογιστικό φόρτο μέσω της μείωσης των παραμέτρων στα επόμενα επίπεδα, αλλά συμβάλλει και στη δημιουργία πιο ανθεκτικών αναπαραστάσεων, προσδίδοντας στο δίκτυο τη χαρακτηριστική ιδιότητα της «χωρικής αναλλοιωτότητας». Αυτό σημαίνει ότι το δίκτυο μαθαίνει να αναγνωρίζει έναν καρκινικό πυρήνα ανεξάρτητα από το πού ακριβώς βρίσκεται τοποθετημένος μέσα στο πλαίσιο της εικόνας, δηλαδή το μοντέλο γίνεται λιγότερο ευαίσθητο σε μικρές μετατοπίσεις ή παραμορφώσεις της εικόνας.

Στο πρόβλημα που αναλύεται στην παρούσα εργασία, όπου οι εικόνες εμφανίζουν διαφοροποιήσεις, η ιδιότητα αυτή είναι ζωτικής σημασίας.

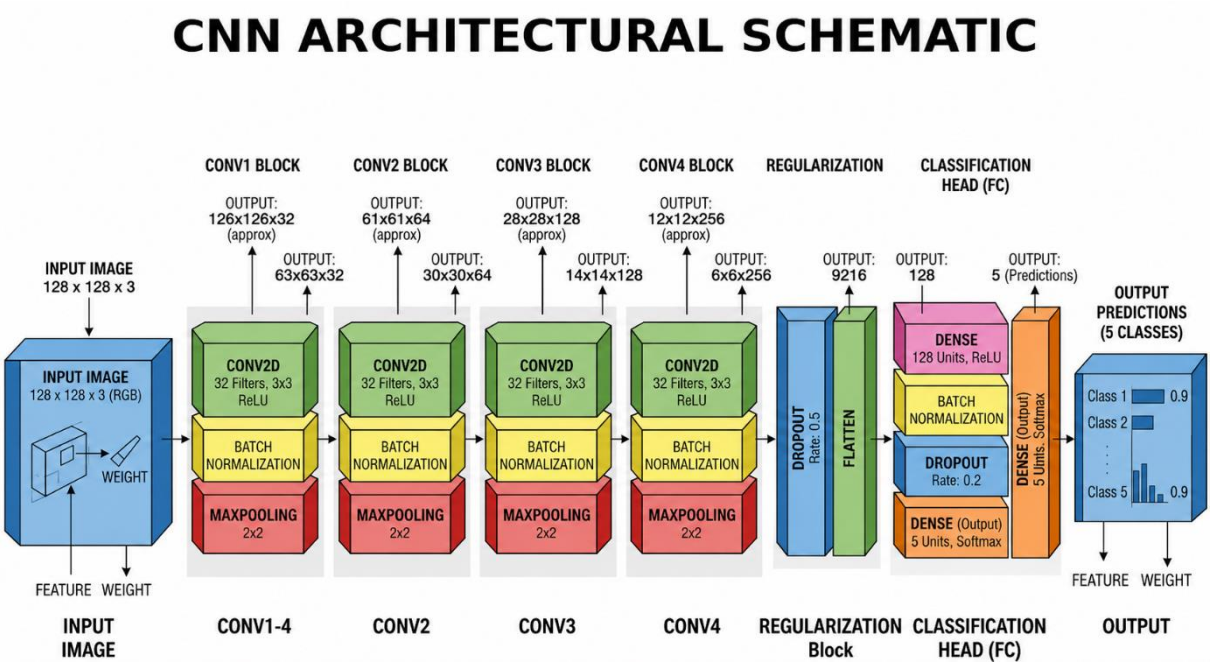
Κρίσιμο σημείο επίσης αποτελεί η μετάβαση από τον δισδιάστατο χώρο των feature maps στον μονοδιάστατο χώρο των πλήρως συνδεδεμένων επιπέδων. Μετά το τελευταίο συνελικτικό block, οι πολυδιάστατοι χάρτες χαρακτηριστικών μετατρέπονται σε ένα μονοδιάστατο διάνυσμα μέσω της συνάρτησης επιπέδωσης (Flatten). Η διαδικασία αυτή λειτουργεί ως ενδιάμεσο στάδιο, μετατρέποντας τη χωρική πληροφορία σε κατάλληλη μορφή για την τελική ταξινόμηση.

Το διάνυσμα που προαναφέρθηκε τροφοδοτείται σε ένα πλήρως συνδεδεμένο επίπεδο (Dense Layer) 128 νευρώνων όπως φαίνεται και στο σχήμα 5.1, το οποίο λειτουργεί ως ταξινομητής, συνδυάζοντας τα χαρακτηριστικά που έχουν εξαχθεί και παράγοντας την τελική πρόβλεψη. Η ύπαρξη ενός ή περισσότερων τέτοιων επιπέδων επιτρέπει στο μοντέλο να μάθει μη γραμμικούς συνδυασμούς χαρακτηριστικών, αυξάνοντας την ακρίβεια της ταξινόμησης.

Έμφαση επιπλέον δόθηκε και στην αποτροπή της υπερεκπαίδευσης, που είναι ένα συχνό πρόβλημα, ειδικά όταν τα ιατρικά δεδομένα είναι πεπερασμένα, μέσω της ενσωμάτωσης επιπέδου Dropout. Συνολικά, το Dropout εφαρμόστηκε με ποσοστό 30% πριν από το πρώτο Dense επίπεδο και 20% πριν από το επίπεδο εξόδου. Με την τεχνική αυτή απενεργοποιείται τυχαία ένα ποσοστό νευρώνων σε κάθε εποχή κατά την εκπαίδευση, αναγκάζοντας το δίκτυο να αναπτύξει κατανεμημένες αναπαραστάσεις. Ως αποτέλεσμα, το μοντέλο δεν βασίζεται υπερβολικά σε μεμονωμένα χαρακτηριστικά της εικόνας, αυξάνοντας την ικανότητα γενίκευσης σε άγνωστα δεδομένα.

Τέλος, το επίπεδο εξόδου διαθέτει πέντε νευρώνες, όσες και οι διαγνωστικές κλάσεις του προβλήματος. Ιδιαίτερη σημασία έχει και η επιλογή της τελικής συνάρτησης ενεργοποίησης. Ανάλογα με τη φύση του προβλήματος, χρησιμοποιείται sigmoid ή Softmax, επιτρέποντας την ερμηνεία της εξόδου ως πιθανότητας. Στην παρούσα εργασία χρησιμοποιήθηκε η Softmax, η οποία μετατρέπει τις πραγματικές τιμές εξόδου (logits) σε μια κανονικοποιημένη κατανομή πιθανοτήτων, επιτρέποντας την ξεκάθαρη και ποσοτικοποιημένη ταξινόμηση κάθε εικόνας στην πιθανότερη παθολογική ή φυσιολογική κατηγορία.

Στο Σχήμα 5.1 φαίνεται ο συνολικός αριθμός επιπέδων και η δομή της αρχιτεκτονικής VGG16 που υλοποιήθηκε στην Δ.Ε.



Σχήμα 5.1: Αναπαράσταση αναπτυγμένης αρχιτεκτονικής CNN

5.3 Μεταφορά Μάθησης με Προεκπαιδευμένα Μοντέλα

Η αξιοποίηση τεχνικών Μεταφοράς Μάθησης (Transfer Learning) αποτελεί μία από τις πλέον αποδοτικές στρατηγικές στη σύγχρονη υπολογιστική όραση, ιδιαίτερα σε προβλήματα όπου το διαθέσιμο σύνολο δεδομένων είναι περιορισμένο ή παρουσιάζει υψηλή πολυπλοκότητα. Στο πλαίσιο της παρούσας εργασίας, εφαρμόστηκε Transfer Learning σε τέσσερις διαφορετικές αρχιτεκτονικές συνελικτικών νευρωνικών δικτύων, συγκεκριμένα στα μοντέλα VGG16, VGG19, ResNet50 και InceptionResNetV2, τα οποία έχουν προεκπαιδευτεί στο ευρέως χρησιμοποιούμενο dataset ImageNet.

Η βασική αρχή του Transfer Learning έγκειται στη μεταφορά γνώσης από ένα πρόβλημα γενικής φύσεως σε ένα πιο εξειδικευμένο. Τα προεκπαιδευμένα μοντέλα έχουν ήδη μάθει να αναγνωρίζουν βασικά οπτικά χαρακτηριστικά, όπως ακμές, υφές και δομές, μέσα από την εκπαίδευσή τους σε εκατομμύρια εικόνες. Η γνώση αυτή μπορεί να επαναχρησιμοποιηθεί, επιτρέποντας στο μοντέλο να προσαρμοστεί ταχύτερα και αποτελεσματικότερα στο νέο πρόβλημα.

Στο περιβάλλον των notebooks που χρησιμοποιούνται στην εργασία, η διαδικασία εφαρμογής του Transfer Learning ακολουθεί μια κοινή μεθοδολογία. Αρχικά, φορτώνεται το προεκπαιδευμένο μοντέλο χωρίς τα τελικά πλήρως συνδεδεμένα επίπεδα ταξινόμησης. Με τον τρόπο αυτό, το μοντέλο χρησιμοποιείται ως μηχανισμός εξαγωγής χαρακτηριστικών. Στη συνέχεια, προστίθεται ένα νέο classification head, το οποίο είναι προσαρμοσμένο στις ανάγκες του συγκεκριμένου προβλήματος.

Κεντρικό στοιχείο της διαδικασίας αποτελεί η επιλογή των επιπέδων που θα παραμείνουν παγωμένα και αυτών που θα επανεκπαιδευτούν. Τα πρώτα επίπεδα των CNNs περιέχουν γενικά χαρακτηριστικά και για τον λόγο αυτό διατηρούνται αμετάβλητα, ενώ τα βαθύτερα επίπεδα προσαρμόζονται μέσω fine-tuning. Η στρατηγική αυτή επιτρέπει στο μοντέλο να διατηρεί τη γενική γνώση που έχει αποκτήσει, ενώ παράλληλα εξειδικεύεται στο νέο dataset.

5.3.1 Αναλυτική Εφαρμογή VGG16

Η αρχιτεκτονική VGG16 αποτελείται από 19 επίπεδα με εκπαιδεύσιμα σταθμά (1 επίπεδο εισόδου, 13 συνελικτικά και, στην αρχική της μορφή, 5 πλήρως συνδεδεμένα επίπεδα). Η βασική σχεδιαστική φιλοσοφία της είναι η αντικατάσταση των μεγάλων συνελικτικών φίλτρων (π.χ. 11x11) με αλλεπάλληλα μικρότερα φίλτρα διαστάσεων 3x3, σε συνδυασμό με βήμα 1 και χρήση padding. Στο συγκεκριμένο ανεπτυγμένο μοντέλο VGG16 υλοποιούνται συνολικά 22 επίπεδα, καθώς υπάρχουν και 3 επιπλέον επίπεδα: ένα επίπεδο Average Pooling 2D, που μετατρέπει τα δεδομένα σε μονοδιάστατο διάνυσμα, ένα επίπεδο Dropout για τη μείωση της υπερεκπαίδευσης και το τελικό επίπεδο εξόδου.

Η εφαρμογή του μοντέλου VGG16 στο πλαίσιο της παρούσας εργασίας βασίζεται στην αξιοποίηση της προεκπαιδευμένης αρχιτεκτονικής ως μηχανισμού εξαγωγής χαρακτηριστικών, προσαρμοσμένου στις απαιτήσεις του συγκεκριμένου προβλήματος ταξινόμησης. Το VGG16, έχοντας εκπαιδευτεί σε μεγάλης κλίμακας σύνολα δεδομένων, διαθέτει την ικανότητα ανίχνευσης γενικών οπτικών χαρακτηριστικών, τα οποία μεταφέρονται και επαναχρησιμοποιούνται στο νέο περιβάλλον.

Κατά το στάδιο της αρχικοποίησης, το μοντέλο φορτώνεται χωρίς τα τελικά επίπεδα ταξινόμησης, επιτρέποντας την απομόνωση του τμήματος εξαγωγής χαρακτηριστικών. Η επιλογή αυτή είναι κρίσιμη, καθώς τα αρχικά convolutional blocks περιέχουν φίλτρα τα οποία έχουν μάθει να ανιχνεύουν βασικές δομές εικόνας, όπως ακμές και υφές. Τα επίπεδα αυτά διατηρούνται παγωμένα, ώστε να αποφευχθεί η αλλοίωση της ήδη εκπαιδευμένης γνώσης.

Στη συνέχεια, προστίθεται ένα νέο classification head, το οποίο αποτελείται από πλήρως συνδεδεμένα επίπεδα. Το τμήμα αυτό εκπαιδεύεται από την αρχή και είναι υπεύθυνο για τη

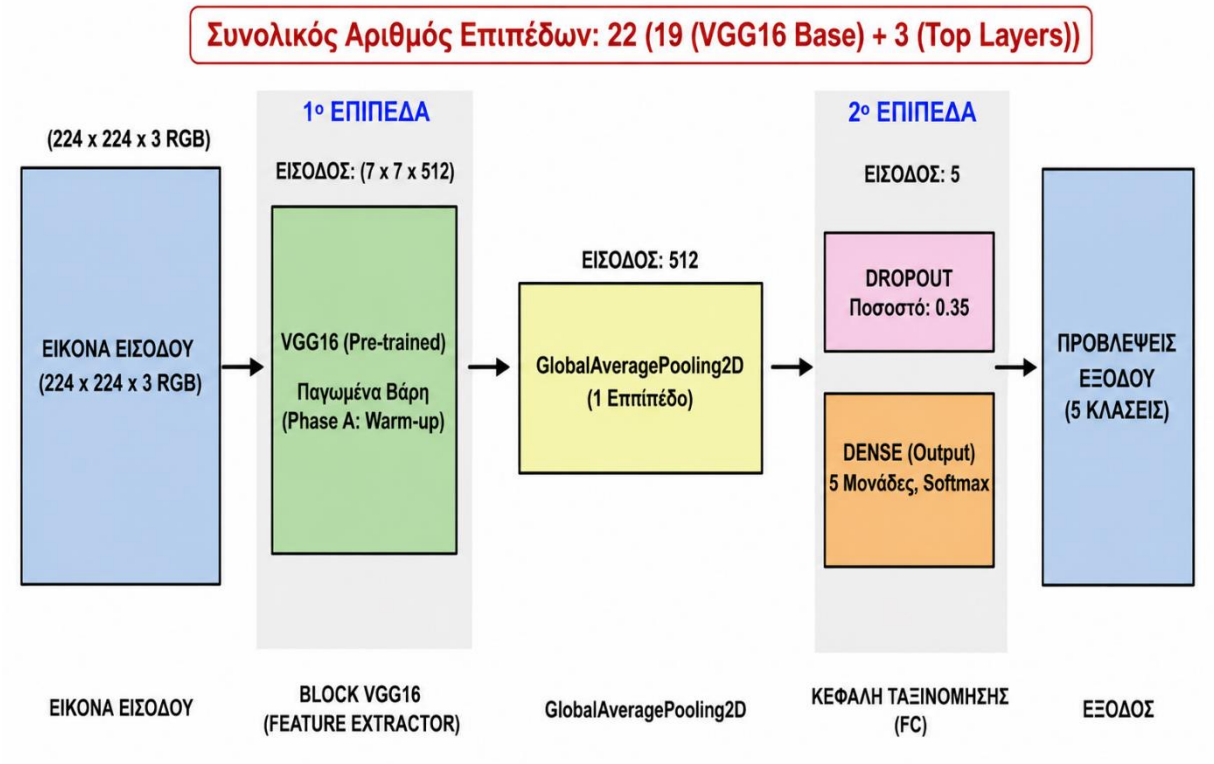
χαρτογράφηση των εξαγόμενων χαρακτηριστικών στις κατηγορίες του προβλήματος. Η μετάβαση από το convolutional στο fully connected μέρος γίνεται μέσω της διαδικασίας flattening, η οποία μετατρέπει τα feature maps σε μονοδιάστατα διανύσματα, όπως και στα κλασικά CNNs.

Σε επίπεδο λειτουργίας, το VGG16 μετασχηματίζει την αρχική εικόνα σε μια ακολουθία feature maps, τα οποία διατηρούν τη χωρική δομή της εισόδου, ενώ παράλληλα εμπλουτίζονται με σημασιολογική πληροφορία. Τα αρχικά επίπεδα του δικτύου λειτουργούν ως φίλτρα χαμηλού επιπέδου, τα οποία αποκρίνονται σε βασικά μορφολογικά στοιχεία. Η επαναληπτική εφαρμογή των συνελικτικών πράξεων οδηγεί σε σταδιακή ενίσχυση των ενεργοποιήσεων που αντιστοιχούν σε δομές με μεγαλύτερη πληροφοριακή αξία.

Καθώς η πληροφορία προωθείται βαθύτερα στο δίκτυο, τα feature maps μετασχηματίζονται σε πιο αφαιρετικές αναπαραστάσεις. Η αύξηση του αριθμού των καναλιών, σε συνδυασμό με τη μείωση της χωρικής διάστασης, επιτρέπει τη συγκέντρωση πληροφορίας σε έναν πιο συμπαγή και σημασιολογικά πλούσιο χώρο. Το αποτέλεσμα είναι η δημιουργία ενός representation space, στο οποίο διαφορετικές κατηγορίες εικόνων καταλαμβάνουν διακριτές περιοχές.

Η δομή του VGG16 ευνοεί τη σταθερότητα των αναπαραστάσεων, καθώς η απουσία σύνθετων μηχανισμών, όπως skip connections ή πολυδιαδρομική επεξεργασία, οδηγεί σε προβλέψιμη ροή πληροφορίας. Η ιδιότητα αυτή καθιστά το μοντέλο ιδιαίτερα κατάλληλο για περιπτώσεις όπου απαιτείται σαφής και ερμηνεύσιμη εξαγωγή χαρακτηριστικών.

Στο Σχήμα 5.2 φαίνεται ο συνολικός αριθμός επιπέδων και η δομή της αρχιτεκτονικής VGG16 που υλοποιήθηκε στην εργασία.



Σχήμα 5.2: Αναπαράσταση αναπτυγμένης αρχιτεκτονικής VGG16

5.3.2 Αναλυτική Εφαρμογή VGG19

Η αρχιτεκτονική του VGG19 επεκτείνεται πάνω στη φιλοσοφία του VGG16 μέσω της αύξησης του βάθους του δικτύου, γεγονός που επιτρέπει την περαιτέρω ανάπτυξη της ιεραρχίας των χαρακτηριστικών. Το VGG19 αποτελεί τη βαθύτερη και πιο εκτεταμένη παραλλαγή της οικογένειας VGG. Περιλαμβάνει συνολικά 22 επίπεδα με σταθμά, εκ των οποίων τα 16 είναι συνελικτικά, 5 επίπεδα Max Pooling, 1 επίπεδο Average Pooling και 2 επίπεδα στην κεφαλή ταξινόμησης όπως φαίνεται και στο Σχήμα 5.3. Ο σχεδιασμός ακολουθεί τον ίδιο κανόνα των φίλτρων 3x3, ωστόσο η προσθήκη τριών επιπλέον συνελικτικών επιπέδων στα βαθύτερα blocks του δικτύου προσδίδει αυξημένη χωρητικότητα. Η προσθήκη επιπλέον συνελικτικών επιπέδων οδηγεί σε μεγαλύτερη ικανότητα διάκρισης, καθώς το μοντέλο μπορεί να συνθέτει πιο σύνθετες και εξειδικευμένες αναπαραστάσεις, γεγονός που επηρεάζει άμεσα τη διαδικασία του fine-tuning.

Κατά τη φόρτωση του μοντέλου, απομονώνεται το convolutional τμήμα, ενώ τα fully connected layers αντικαθίστανται από νέα επίπεδα που προσαρμόζονται στο συγκεκριμένο πρόβλημα. Τα πρώτα επίπεδα παραμένουν παγωμένα, ενώ τα βαθύτερα μπορούν να συμμετέχουν στη διαδικασία fine-tuning, ανάλογα με τη στρατηγική εκπαίδευσης.

Κατά τη διαδικασία fine-tuning, αποδεσμεύονται τα τελευταία convolutional blocks, επιτρέποντας στο μοντέλο να προσαρμόσει τις υψηλού επιπέδου αναπαραστάσεις. Η εκπαίδευση σε αυτό το στάδιο είναι πιο ευαίσθητη, καθώς η ύπαρξη μεγάλου αριθμού παραμέτρων αυξάνει τον κίνδυνο αστάθειας. Για τον λόγο αυτό, η εκπαίδευση πραγματοποιείται με μικρό learning rate και συχνή παρακολούθηση της απώλειας.

Σε επίπεδο λειτουργίας, το VGG19 παρουσιάζει πιο εκτεταμένη επεξεργασία των feature maps πριν από τη μετάβαση σε πιο αφαιρετικές μορφές. Η μεγαλύτερη ακολουθία συνελικτικών επιπέδων επιτρέπει την ενίσχυση συγκεκριμένων ενεργοποιήσεων, οδηγώντας σε πιο «καθαρά» και διακριτικά χαρακτηριστικά.

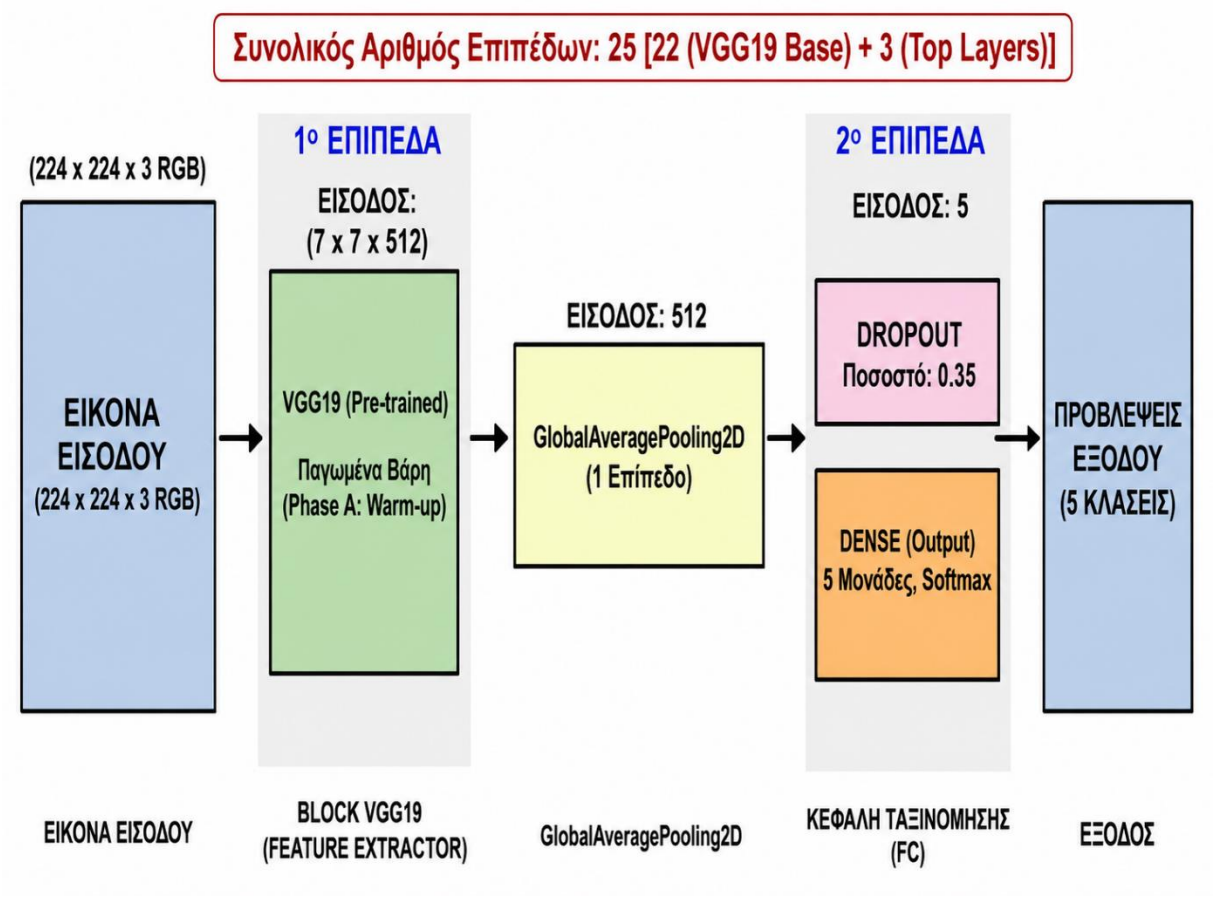
Η αυξημένη χωρητικότητα του μοντέλου αντικατοπτρίζεται στη δυνατότητα αποτύπωσης λεπτών διαφορών μεταξύ εικόνων. Τα βαθύτερα επίπεδα λειτουργούν ως εξειδικευμένοι ανιχνευτές μοτίβων, οι οποίοι αποκρίνονται σε πολύπλοκες δομές και συσχετίσεις. Η ιδιότητα αυτή είναι ιδιαίτερα σημαντική σε προβλήματα όπου οι κατηγορίες παρουσιάζουν υψηλό βαθμό ομοιότητας.

Ένα σημαντικό χαρακτηριστικό της εφαρμογής στην έρευνα είναι η σταδιακή προσαρμογή των gradients στα βαθύτερα επίπεδα. Οι αλλαγές στα weights των convolutional filters είναι μικρές αλλά ουσιαστικές, οδηγώντας σε βελτίωση της εξειδίκευσης χωρίς απώλεια της γενικής γνώσης.

Η χρήση Dropout στα fully connected layers συμβάλλει στον περιορισμό της υπερπροσαρμογής με ποσοστό 0.35 όπως φαίνεται και στο Σχήμα 5.3, ενώ η παρουσία βαθύτερης αρχιτεκτονικής επιτρέπει στο μοντέλο να διαχωρίζει πιο λεπτές διαφορές μεταξύ των κατηγοριών.

Η ροή πληροφορίας στο VGG19 παραμένει γραμμική, ωστόσο η αυξημένη πολυπλοκότητα οδηγεί σε πιο πλούσιο representation space. Τα feature maps στα τελικά επίπεδα εμφανίζουν υψηλή συμπύκνωση πληροφορίας, ενώ διατηρούν τη δυνατότητα διάκρισης μεταξύ διαφορετικών κατηγοριών.

Στο Σχήμα 5.3 φαίνεται ο συνολικός αριθμός επιπέδων και η δομή της αρχιτεκτονικής VGG19 που υλοποιήθηκε στην εργασία.



Σχήμα 5.3: Αναπαράσταση αναπτυγμένης αρχιτεκτονικής VGG19

5.3.3 Αναλυτική Εφαρμογή ResNet50

Η αρχιτεκτονική του ResNet50 διαφοροποιείται ουσιαστικά από τα VGG μοντέλα μέσω της εισαγωγής residual connections, οι οποίες επιτρέπουν την απευθείας μεταφορά πληροφορίας μεταξύ μη διαδοχικών επιπέδων. Σε παραδοσιακά σειριακά δίκτυα, όπως το VGG, η αύξηση του βάθους οδηγεί συχνά στο πρόβλημα της εξαφάνισης της κλίσης, καθιστώντας την εκπαίδευση ιδιαίτερα δύσκολη. Το ResNet λύνει το πρόβλημα αυτό χρησιμοποιώντας συνδέσεις παράκαμψης (skip connections). Οι συνδέσεις αυτές υλοποιούν μια μορφή ταυτοτικής χαρτογράφησης, επιτρέποντας στο δίκτυο να μαθαίνει μεταβολές αντί για πλήρεις μετασχηματισμούς.

Η βασική ιδέα της υπολειπόμενης μάθησης δεν είναι η εκμάθηση μιας πλήρους συνάρτησης μετασχηματισμού της εισόδου, αλλά η εκμάθηση του «υπολοίπου» σε σχέση με την ταυτότητα. Πιο συγκεκριμένα, αντί το δίκτυο να προσπαθεί να μάθει μια άμεση χαρτογράφηση $H(x)$, μαθαίνει μια συνάρτηση $F(x)=H(x)-x$, ώστε η τελική έξοδος να προκύπτει ως $H(x)=F(x)+x$. Η προσέγγιση αυτή επιτρέπει ευκολότερη βελτιστοποίηση, καθώς το δίκτυο μπορεί να διατηρεί την αρχική πληροφορία και να μαθαίνει μόνο τις απαραίτητες διορθώσεις.

Στην υλοποίηση του μοντέλου ResNet50, οι παραπάνω θεωρητικές αρχές εφαρμόζονται μέσω των residual blocks. Κάθε block αποτελείται από ένα σύνολο συνελικτικών επιπέδων, των οποίων η έξοδος προστίθεται απευθείας στην είσοδο του block μέσω μιας shortcut connection. Η διαδικασία αυτή δεν αποτελεί απλώς μια συγχώνευση, αλλά έναν μηχανισμό ενίσχυσης και διατήρησης της πληροφορίας, που επιτρέπει τη συνύπαρξη διαφορετικών επιπέδων αναπαράστασης.

Μια από τις σημαντικότερες καινοτομίες της αρχιτεκτονικής ResNet είναι η δημιουργία ενός μηχανισμού τύπου bottleneck, ο οποίος χρησιμοποιείται σε κάθε residual block. Ο σχεδιασμός αυτός περιλαμβάνει τρία διαδοχικά επίπεδα συνέλιξης με διαστάσεις πυρήνα 1x1, 3x3 και 1x1. Η πρώτη συνέλιξη 1x1 λειτουργεί ως μηχανισμός μείωσης διαστάσεων, περιορίζοντας τον αριθμό των καναλιών και συνεπώς το υπολογιστικό κόστος. Ακολουθεί η συνέλιξη 3x3, η οποία επεξεργάζεται τα χαρακτηριστικά στον μειωμένο χώρο, ενώ η τελική συνέλιξη 1x1 επαναφέρει τις διαστάσεις στο αρχικό τους μέγεθος. Η αρχιτεκτονική αυτή επιτρέπει στο δίκτυο να διατηρεί μεγάλο βάθος χωρίς εκθετική αύξηση των παραμέτρων, καθιστώντας το ιδιαίτερα αποδοτικό. Παράλληλα, ο συνδυασμός bottleneck δομής και residual συνδέσεων δημιουργεί ένα σύστημα όπου η πληροφορία συμπιέζεται, μετασχηματίζεται και επανασυντίθεται με ελεγχόμενο τρόπο.

Σε επίπεδο λειτουργίας, η ροή πληροφορίας δεν είναι πλέον αυστηρά γραμμική, αλλά περιλαμβάνει διακλαδώσεις. Με την ύπαρξη των skip connections επιτρέπεται η διατήρηση της αρχικής εισόδου και ο συνδυασμός της με τις εξόδους των συνελικτικών επιπέδων. Με τη διαδικασία αυτή, τα feature maps που δημιουργούνται περιέχουν τόσο επεξεργασμένη όσο και πρωτογενή πληροφορία. Για τον λόγο αυτό, οι αναπαραστάσεις που προκύπτουν είναι πιο σταθερές και περισσότερο ανθεκτικές σε αλλοιώσεις καθώς το βάθος του δικτύου αυξάνεται.

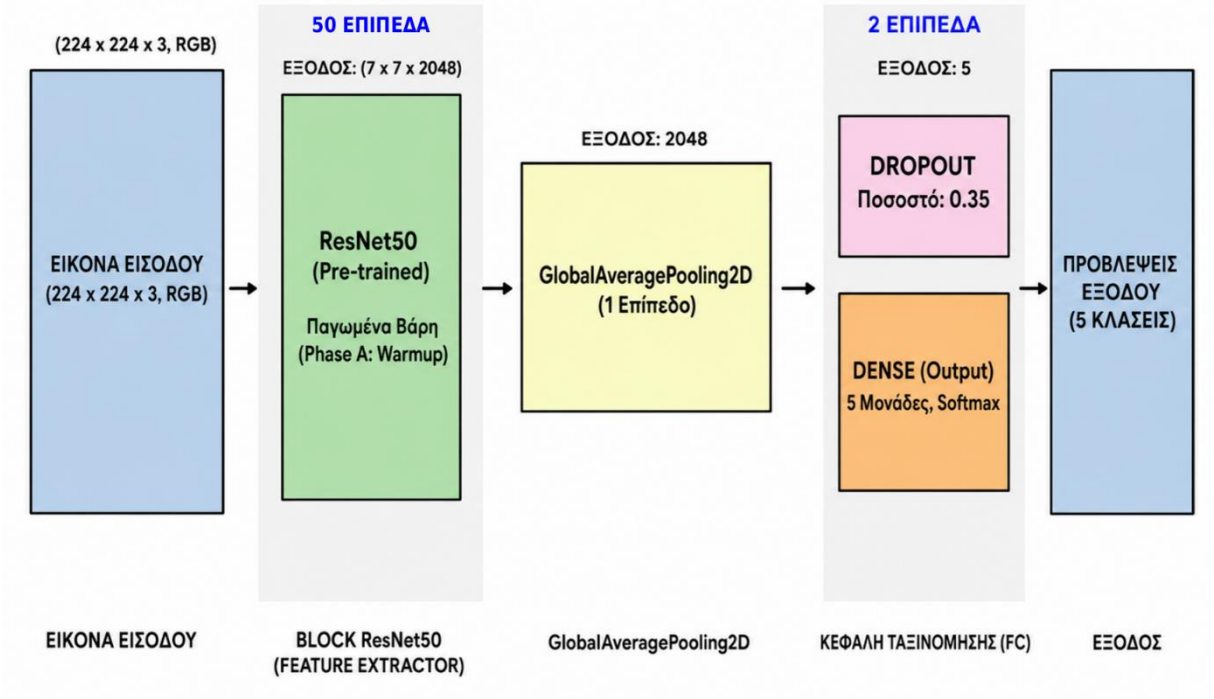
Σημαντικό επίσης είναι το γεγονός ότι τα residual blocks λειτουργούν ως μονάδες μετασχηματισμού μικρής κλίμακας. Δηλαδή, αντί να αντικαθιστούν σταδιακά την πληροφορία, προσθέτουν νέα στοιχεία πάνω στην ήδη υπάρχουσα αναπαράσταση. Αυτό οδηγεί στη συνεχή βελτίωση των χαρακτηριστικών, με κάθε επίπεδο να συμβάλλει με μικρές αλλά ουσιαστικές αλλαγές. Η προσέγγιση αυτή καθιστά το δίκτυο πιο ανθεκτικό και επιτρέπει την αποτελεσματική αξιοποίηση μεγάλου βάθους.

Στο πλαίσιο της εφαρμογής μέσω των notebooks, η αρχιτεκτονική του ResNet50 λειτουργεί ως ένας ισχυρός μηχανισμός εξαγωγής χαρακτηριστικών, ο οποίος διατηρεί πληροφορία από διαφορετικά επίπεδα αφαίρεσης. Τα αρχικά επίπεδα συλλαμβάνουν βασικά χαρακτηριστικά, ενώ τα βαθύτερα επίπεδα ενσωματώνουν πιο σύνθετες δομές, χωρίς όμως να χάνεται η σύνδεση με την αρχική είσοδο.

Ένα κρίσιμο στοιχείο κατά τη χρήση προεκπαιδευμένων αρχιτεκτονικών είναι η αυστηρή ευθυγράμμιση της προεπεξεργασίας των δεδομένων εισόδου με τη στρατηγική του εκάστοτε δικτύου. Ενώ τα μοντέλα της οικογένειας VGG και το βασικό ResNet50 βασίζονται στον ίδιο μετασχηματισμό των χρωματικών καναλιών από RGB σε BGR και στην αφαίρεση της μέσης τιμής του ImageNet (zero-centering), άλλες αρχιτεκτονικές της μελέτης, όπως το InceptionResNetV2, υιοθετούν διαφορετική προσέγγιση.

Συνολικά, η αρχιτεκτονική του ResNet50 διαμορφώνει έναν χώρο αναπαράστασης στον οποίο η πληροφορία δεν αντικαθίσταται, αλλά εμπλουτίζεται σταδιακά. Η συνύπαρξη πρωτογενών και επεξεργασμένων χαρακτηριστικών, σε συνδυασμό με τον υπολογιστικά αποδοτικό σχεδιασμό των bottleneck blocks, καθιστά το μοντέλο ιδιαίτερα ικανό να αποτυπώνει πολύπλοκες ιστοπαθολογικές δομές. Η λειτουργία του, όπως αναδεικνύεται μέσα από την πειραματική εφαρμογή στην παρούσα μελέτη, δεν βασίζεται σε μια απλή ιεραρχική εξαγωγή χαρακτηριστικών, αλλά σε μια πιο σύνθετη διαδικασία αθροιστικής επανασύνθεσης της πληροφορίας, η οποία ενισχύει σημαντικά τη διακριτική ικανότητα του μοντέλου.

Στο Σχήμα 5.4 φαίνεται ο συνολικός αριθμός επιπέδων και η δομή της αρχιτεκτονικής ResNet50 που υλοποιήθηκε στην εργασία.



Σχήμα 5.4: Αναπαράσταση αναπτυγμένης αρχιτεκτονικής ResNet50

5.3.4 Αναλυτική Εφαρμογή InceptionResNetV2

Η αρχιτεκτονική InceptionResNetV2 αποτελεί μία από τις πιο εξελιγμένες προσεγγίσεις στον σχεδιασμό συνελικτικών νευρωνικών δικτύων, διότι συνδυάζει δύο από τους σημαντικότερους πυλώνες της βαθιάς μάθησης: την πολυκλιμακική επεξεργασία μέσω των Inception modules και τη σταθεροποιημένη ροή πληροφορίας μέσω των υπολειπόμενων συνδέσεων. Η παράλληλη χρήση των δύο αυτών μηχανισμών δημιουργεί ένα σύνθετο, αλλά ταυτόχρονα εξαιρετικά αποδοτικό σύστημα.

Η βασική φιλοσοφία του μηχανισμού Inception πηγάζει από την ταυτόχρονη επεξεργασία της εισόδου σε πολλαπλά επίπεδα. Σε αντίθεση με τις σειριακές αρχιτεκτονικές, όπου κάθε επίπεδο εφαρμόζει έναν ενιαίο τύπο φίλτρου, τα Inception modules διακλαδώνουν τη ροή της πληροφορίας σε παράλληλα μονοπάτια. Κάθε μονοπάτι εφαρμόζει φίλτρα διαφορετικών διαστάσεων, όπως 1x1 και 3x3, επιτρέποντας στο δίκτυο να εξάγει χαρακτηριστικά που αντιστοιχούν σε διαφορετικά επίπεδα χωρικής ανάλυσης. Τα αποτελέσματα αυτών των διαδρομών συγχωνεύονται μέσω ενός μηχανισμού συνένωσης, δημιουργώντας feature maps υψηλής πληροφοριακής ποικιλίας.

Η προσέγγιση αυτή είναι κρίσιμη σε προβλήματα όπου η πληροφορία κατανέμεται σε διαφορετικά μεγέθη και επίπεδα λεπτομέρειας. Μέσα από την πειραματική εφαρμογή του μοντέλου στην παρούσα μελέτη, καθίσταται εμφανές ότι το InceptionResNetV2 δεν περιορίζεται στην ανίχνευση συγκεκριμένου τύπου χαρακτηριστικών, αλλά αναπτύσσει ένα πολυδιάστατο representation space, στο οποίο συνυπάρχουν τοπικές και παγκόσμιες αναπαραστάσεις.

Η ενσωμάτωση residual connections στα Inception blocks αποτελεί το στοιχείο που διαφοροποιεί ριζικά την αρχιτεκτονική από τα κλασικά Inception δίκτυα. Οι συνδέσεις αυτές επιτρέπουν την απευθείας πρόσθεση της εισόδου ενός block στην έξοδό του, δημιουργώντας μια μορφή identity mapping. Αντί το δίκτυο να μετασχηματίζει πλήρως την πληροφορία σε κάθε επίπεδο, διατηρεί ένα μέρος της αρχικής αναπαράστασης και προσθέτει σε αυτήν τις απαραίτητες τροποποιήσεις.

Η λειτουργία αυτή έχει σημαντικές επιπτώσεις στη δομή των feature maps. Οι αναπαραστάσεις που προκύπτουν δεν αποτελούν απλώς προϊόν επεξεργασίας των συνελικτικών φίλτρων, αλλά συνδυασμό πρωτογενούς και επεξεργασμένης πληροφορίας. Αυτό οδηγεί σε μεγαλύτερη σταθερότητα των χαρακτηριστικών και επιτρέπει τη διατήρηση κρίσιμων στοιχείων της εισόδου σε όλα τα επίπεδα του δικτύου.

Η αρχιτεκτονική του InceptionResNetV2 χαρακτηρίζεται από έντονη μη γραμμικότητα και πολυδιαδρομική ροή πληροφορίας. Κάθε επίπεδο δεν λειτουργεί ως απλή συνέχεια του προηγούμενου, αλλά ως κόμβος όπου συνδυάζονται πολλαπλές αναπαραστάσεις. Τα Inception modules παράγουν ένα σύνολο feature maps που αντιστοιχούν σε διαφορετικές οπτικές «προσεγγίσεις» της ίδιας εικόνας, ενώ οι residual connections διασφαλίζουν τη συνοχή αυτών των αναπαραστάσεων.

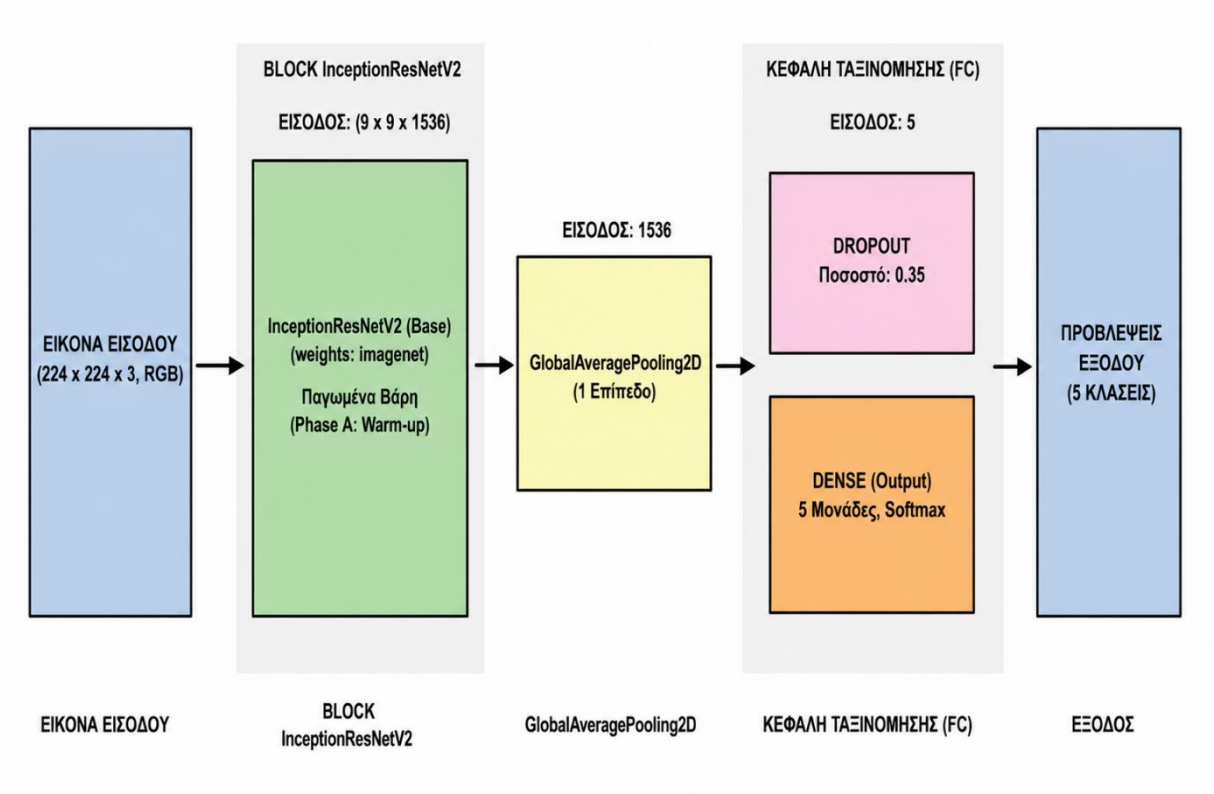
Σε επίπεδο αναπαράστασης, το μοντέλο δημιουργεί εξαιρετικά πυκνούς (dense) χώρους χαρακτηριστικών, στους οποίους η πληροφορία δεν κατανέμεται ομοιόμορφα, αλλά συγκεντρώνεται σε περιοχές υψηλής σημασιολογικής αξίας. Οι τελικοί feature maps εμφανίζουν υψηλό βαθμό αφαιρετικότητας, ενώ ταυτόχρονα διατηρούν στοιχεία από διαφορετικές κλίμακες ανάλυσης.

Η προεπεξεργασία των δεδομένων εισόδου αποτελεί επίσης κρίσιμο στοιχείο της λειτουργίας του μοντέλου. Σε αντίθεση με άλλες αρχιτεκτονικές, το InceptionResNetV2 χρησιμοποιεί κανονικοποίηση των τιμών των pixel στο διάστημα [-1, 1], γεγονός που επηρεάζει άμεσα την κατανομή των ενεργοποιήσεων στα αρχικά επίπεδα. Η επιλογή αυτή συνδέεται με τον τρόπο σχεδίασης του μοντέλου και συμβάλλει στη σταθερότητα της ροής πληροφορίας.

Μέσα από την ανάλυση της συμπεριφοράς του κατά την εκπαίδευση, καθίσταται σαφές ότι το InceptionResNetV2 λειτουργεί ως ένας μηχανισμός σύνθεσης πολλαπλών επιπέδων πληροφορίας. Τα χαρακτηριστικά που εξάγονται δεν προκύπτουν από μία ενιαία διαδικασία μετασχηματισμού, αλλά από τον συνδυασμό διαφορετικών διαδρομών επεξεργασίας. Η ιδιότητα αυτή επιτρέπει στο μοντέλο να αποτυπώνει πολύπλοκες και πολυδιάστατες δομές, οι οποίες δεν μπορούν να αναπαρασταθούν επαρκώς μέσω απλών σειριακών αρχιτεκτονικών.

Η συνολική λειτουργία του μοντέλου διαμορφώνει ένα representation space υψηλής εκφραστικής ικανότητας, όπου κάθε είσοδος αναπαρίσταται μέσω ενός συνδυασμού χαρακτηριστικών που προέρχονται από διαφορετικές κλίμακες και επίπεδα αφαίρεσης. Η συνύπαρξη πολυκλιμακικής ανάλυσης και υπολειπόμενης μάθησης καθιστά το InceptionResNetV2 ένα από τα πλέον ισχυρά εργαλεία για την εξαγωγή πλούσιων και πολυδιάστατων αναπαραστάσεων.

Στο Σχήμα 5.5 φαίνεται ο συνολικός αριθμός επιπέδων και η δομή της αρχιτεκτονικής InceptionResNetV2 που υλοποιήθηκε στην εργασία. Όπως γίνεται αντιληπτό από το Σχήμα η εικόνα εισόδου έχει μέγεθος 224×224×3 και περνά από το προεκπαιδευμένο block InceptionResNetV2 με παγωμένα βάρη. Στη συνέχεια εφαρμόζεται GlobalAveragePooling2D, dropout 0.35 και τελικό Dense layer με softmax για πρόβλεψη 5 κλάσεων.



Σχήμα 5.5: Αναπαράσταση αναπτυγμένης αρχιτεκτονικής InceptionResNetV2

5.4 Σχεδιασμός αρχιτεκτονικής Vision Transformer

Η αρχιτεκτονική των Vision Transformers δεν αποτελεί απλώς μια εναλλακτική ή εξελικτική προσέγγιση στην επεξεργασία εικόνας. Αποτελεί μια θεμελιώδη αλλαγή στον τρόπο τον οποίο ορίζεται, δομείται και αναπαρίσταται η οπτική πληροφορία. Η ανάλυση εικόνας στην οποία ιστορικά κυριαρχούσαν τα Συνελικτικά Νευρωνικά Δίκτυα, βασίζονται σε μια παραδοχή ότι τα pixels που βρίσκονται το ένα κοντά στο άλλο σχετίζονται πιο ισχυρά με pixels που βρίσκονται μακριά, καθώς και ότι το μοτίβο αυτό έχει σημασία ανεξάρτητα από το που εμφανίζεται. Στο Σχήμα 5.6 της αρχιτεκτονικής του Vision Transformer αποτυπώνεται αυτή η διαφορετική λογική, καθώς η εικόνα εισόδου διαστάσεων 224 × 224 × 3 RGB δεν επεξεργάζεται άμεσα με συνελικτικά φίλτρα, αλλά διασπάται αρχικά σε επιμέρους patches.

Σε αντίθεση με τα CNNs τα οποία καταλήγουν στην παραγωγή απαντήσεων μέσα από τοπική και ιεραρχική επεξεργασία (από πιο απλά σε πιο σύνθετα χαρακτηριστικά), τα ViTs υιοθετούν μια πιο καθολική θεώρηση. Στα Transformers η σημασία κάθε στοιχείου δεν καθορίζεται από την θέση του στην εικόνα αλλά από τις αλληλεπιδράσεις του με τα υπόλοιπα στοιχεία της εικόνας. Η αρχιτεκτονική αυτή ξεκινά σαν άγραφος πίνακας μαθαίνοντας τα χαρακτηριστικά την χωρική κανονικότητα και κανόνες της οπτικής φύσης αποκλειστικά μέσα από τα δεδομένα.

5.4.1 Δημιουργία τμημάτων και Γραμμική προβολή

Η υλοποίηση ξεκινά με την εισαγωγή των εικόνων σε μορφή tensors τεσσάρων διαστάσεων, όπου κάθε batch έχει διαστάσεις που αντιστοιχούν στον αριθμό των δειγμάτων, το ύψος, το πλάτος και τα κανάλια της εικόνας. Το πρώτο υπολογιστικό στάδιο του μοντέλου είναι η διάσπαση της εικόνας σε σταθερού μεγέθους μη επικαλυπτόμενα τμήματα (Patchification), διαδικασία που υλοποιείται μέσω της συνάρτησης image.extract_patches. Σύμφωνα με το Σχήμα 5.6, το μέγεθος κάθε patch είναι 16 × 16,

γεγονός που καθορίζει τον τρόπο με τον οποίο η αρχική εικόνα μετατρέπεται σε ακολουθία οπτικών tokens. Με την επιλογή ίδιου μεγέθους παραθύρου και βήματος μετακίνησης, εξασφαλίζεται ότι η εικόνα καλύπτεται πλήρως από patches χωρίς επαναλήψεις.

Το αποτέλεσμα της διαδικασίας είναι ένα tensor στο οποίο κάθε θέση περιέχει το περιεχόμενο ενός patch σε μορφή ήδη «ισοπεδωμένου» διανύσματος. Το tensor αυτό αναδιαμορφώνεται με χρήση της συνάρτησης reshape σε μία δομή όπου κάθε εικόνα του batch αντιστοιχεί σε μία ακολουθία από patches και κάθε patch σε ένα διάνυσμα χαρακτηριστικών. Στη συνέχεια εφαρμόζεται ένα πλήρως συνδεδεμένο επίπεδο (Dense layer), το οποίο λειτουργεί ως γραμμικός μετασχηματισμός (Linear projection) και προβάλλει κάθε patch σε έναν κοινό χώρο χαρακτηριστικών σταθερής διάστασης. Στην παρούσα αρχιτεκτονική, όπως φαίνεται και στο Σχήμα 5.6, η διάσταση προβολής είναι 64. Η πράξη αυτή εκτελείται ταυτόχρονα για όλα τα patches μέσω διανυσματοποιημένων πράξεων και αποτελεί το σημείο μετάβασης από τα αρχικά δεδομένα εικόνας σε αναπαραστάσεις κατάλληλες για transformer επεξεργασία.

5.4.2 Η Μετάβαση στον Σημασιολογικό Χώρο

Μετά τη γραμμική προβολή, το σύνολο των patches έχει μετατραπεί σε embeddings, ωστόσο η πληροφορία σχετικά με τη χωρική τους θέση δεν είναι πλέον διαθέσιμη. Για τον λόγο αυτό, στην εργασία χρησιμοποιείται η προσθήκη positional embeddings μέσω ενός Embedding layer, το οποίο αντιστοιχίζει κάθε θέση της ακολουθίας σε ένα trainable διάνυσμα. Η ύπαρξη του Position Embedding παρουσιάζεται και στο σχήμα της αρχιτεκτονικής, όπου εντάσσεται στο στάδιο εισαγωγής των patches, πριν από την είσοδο στα Transformer blocks. Η ακολουθία των θέσεων δημιουργείται με χρήση της range και τα παραγόμενα embeddings προστίθενται απευθείας στα patch embeddings μέσω στοιχειο-wise πρόσθεσης.

Παράλληλα, δημιουργείται ένα επιπλέον trainable διάνυσμα, το οποίο αντιστοιχεί στο classification token. Το διάνυσμα αυτό αρχικοποιείται ως μεταβλητή και επεκτείνεται σε όλα τα στοιχεία του batch, ώστε να αποκτήσει συμβατές διαστάσεις με την υπόλοιπη ακολουθία. Στη συνέχεια τοποθετείται στην αρχή της ακολουθίας μέσω συνένωσης, με αποτέλεσμα το τελικό tensor εισόδου στους transformer blocks να περιλαμβάνει τόσο τα patches όσο και το CLS token. Ωστόσο, στην απεικόνιση και στον Πίνακα 5.1 της συγκεκριμένης αρχιτεκτονικής, η τελική ταξινόμηση παρουσιάζεται μέσω Global GAP στην κεφαλή ταξινόμησης, κάτι που δείχνει ότι η τελική αναπαράσταση προκύπτει από συγκεντρωτική επεξεργασία των embeddings πριν από το Dense layer εξόδου.

5.4.3 Η σημασία του Μηχανισμού Αυτοπροσοχής

Ο μηχανισμός αυτοπροσοχής αποτελεί τον πυρήνα του μηχανισμού των Vision Transformers. Ο μηχανισμός αυτός είναι μία διαδικασία υπολογισμού ενός δυναμικού γράφου σχέσεων μεταξύ των tokens. Κάθε token συνδέεται με τα υπόλοιπα μέσω βαρών προσοχής, τα οποία καθορίζουν τη σημασία του στην εικόνα. Σε αντίθεση με τους μηχανισμούς που χρησιμοποιούν τα CNNs, στους οποίους τα βάρη είναι σταθερά μετά την εκπαίδευση και εφαρμόζονται τοπικά, οι συνδέσεις στο self-attention υπολογίζονται δυναμικά τη στιγμή της εκτέλεσης βάσει του περιεχομένου των ίδιων των patches.

Η μαθηματική δομή του attention επιτρέπει τη χαρτογράφηση των tokens σε έναν χώρο όπου η πληροφορία ανακατανέμεται βάσει συμβατότητας. Τα patches μετασχηματίζονται σε τρεις διαφορετικές οντότητες: Queries (Q), Keys (K) και Values (V). Ο μαθηματικός τύπος του μηχανισμού προσοχής, όπως έχει αναφερθεί και στο Κεφάλαιο (2.4.2), ο τύπος (2.6).

Ο υπολογισμός του dot product μεταξύ Q και K, που δείχνει τη σχετικότητα ανάμεσα σε δύο patches, δημιουργεί έναν πίνακα συσχετίσεων. Η διαδικασία αυτή μπορεί να θεωρηθεί ως ένας ανώτερος μηχανισμός «συσχέτισης προτύπων». Το μοντέλο εντοπίζει ποια patches παρουσιάζουν κοινά χαρακτηριστικά ή συμπληρωματικές λειτουργικές σχέσεις, επιτρέποντας την επικοινωνία μεταξύ στοιχείων που βρίσκονται στα δύο άκρα της εικόνας (long-range dependencies) ήδη από το πρώτο επίπεδο του δικτύου. Στο Σχήμα 5.6, η λειτουργία αυτή αποτυπώνεται στο εσωτερικό του Transformer block μέσω του επιπέδου Multi-Head Attention.

Το αποτέλεσμα κανονικοποιείται μέσω διαίρεσης με τη ρίζα της διάστασης και μετατρέπεται σε κατανομή πιθανοτήτων με εφαρμογή της softmax. Οι τιμές αυτές χρησιμοποιούνται ως βάρη για τον συνδυασμό των values, οδηγώντας σε μία νέα αναπαράσταση της ακολουθίας, όπου κάθε token περιέχει πληροφορία από όλα τα υπόλοιπα. Η διαδικασία αυτή υλοποιείται πλήρως μέσω πράξεων γραμμικής άλγεβρας σε επίπεδο tensors.

5.4.4 Πολλαπλές Κεφαλές Προσοχής και πολυπλοκότητα

Επίσης σημαντική είναι η ύπαρξη πολλαπλών κεφαλών προσοχής (Multi-Head Attention). Οι πολλαπλές κεφαλές ενισχύουν σε σημαντικό βαθμό τη διαδικασία του μηχανισμού προσοχής, επιτρέποντας στο μοντέλο να έχει την απαιτούμενη εκφραστικότητα. Αν το μοντέλο λειτουργούσε με έναν μόνο μηχανισμό προσοχής, τότε το δίκτυο θα μπορούσε να εστιάσει μόνο σε έναν τύπο σχέσης (π.χ. μόνο στο χρώμα ή μόνο στο περίγραμμα). Χρησιμοποιώντας πολλαπλά heads, το ViT προβάλλει τα Q, K, V σε διαφορετικούς γραμμικούς υποχώρους. Στη συγκεκριμένη αρχιτεκτονική, όπως φαίνεται στο Σχήμα 5.6 και συνοψίζεται στον Πίνακα 5.1, χρησιμοποιούνται 4 κεφαλές προσοχής.

Στο πλαίσιο της αρχιτεκτονικής Vision Transformer που αξιοποιήθηκε στην παρούσα εργασία, ο μηχανισμός multi-head self-attention διαδραματίζει καθοριστικό ρόλο στη διαδικασία εξαγωγής χαρακτηριστικών από τις ιατρικές εικόνες. Κάθε attention head μαθαίνει να εστιάζει σε διαφορετικές πτυχές των δεδομένων, λειτουργώντας ως εξειδικευμένος «ανιχνευτής» προτύπων.

Στην υλοποίηση της εργασίας, η multi-head attention αποτελεί επέκταση του βασικού self-attention, όπου η ίδια διαδικασία εφαρμόζεται παράλληλα σε πολλαπλά υποσύνολα του embedding χώρου. Το αρχικό tensor εισόδου έχει μορφή (B, N+1, E), όπου E είναι το συνολικό embedding dimension. Το πρώτο βήμα είναι ο διαχωρισμός αυτού του χώρου σε h επιμέρους κεφαλές, με κάθε κεφαλή να επεξεργάζεται διανύσματα διάστασης head_dim = E / h. Αυτό υλοποιείται μέσω συνδυασμού Dense layers και πράξεων reshape/transpose. Συγκεκριμένα, μετά τον υπολογισμό των Q, K και V, τα tensors αναδιαμορφώνονται σε (B, N+1, h, head_dim) και στη συνέχεια μετατίθενται σε (B, h, N+1, head_dim), ώστε η διάσταση των heads να έρθει δεύτερη και να επιτρέψει παράλληλη επεξεργασία. Σε αυτό το σημείο, η διαδικασία attention εφαρμόζεται ανεξάρτητα σε κάθε head, με τον υπολογισμό των attention scores να γίνεται σε επίπεδο (N+1 x N+1) για κάθε κεφαλή.

Η παράλληλη εκτέλεση επιτυγχάνεται χωρίς ρητούς βρόχους, αξιοποιώντας πλήρως διανυσματοποιημένες πράξεις TensorFlow. Τα αποτελέσματα κάθε head διατηρούνται ξεχωριστά μέχρι το στάδιο της συγχώνευσης, όπου πραγματοποιείται αντίστροφη μετατόπιση και reshape, ώστε να επανέλθουν στη μορφή (B, N+1, E). Στη συνέχεια εφαρμόζεται ένα τελικό Dense layer, το οποίο λειτουργεί ως γραμμικός συνδυασμός των heads και επιτρέπει στο μοντέλο να μάθει πώς να ενοποιεί τις διαφορετικές αναπαραστάσεις.

Από υπολογιστικής απόψεως, το κρίσιμο σημείο της υλοποίησης είναι ο υπολογισμός του attention matrix, ο οποίος έχει πολυπλοκότητα τετραγωνική ως προς τον αριθμό των tokens. Δεδομένου ότι το N προκύπτει από το πλήθος των patches, η αύξηση της ανάλυσης της εικόνας οδηγεί σε

σημαντική αύξηση της υπολογιστικής απαίτησης, κάτι που αντανακλάται άμεσα στον χρόνο εκπαίδευσης και στη χρήση μνήμης GPU.

5.4.5 Σταδιακή Ομογενοποίηση και Ροή Πληροφορίας

Κατά τη διέλευση των δεδομένων μέσω των διαδοχικών Transformer blocks της αρχιτεκτονικής Vision Transformer που αξιοποιήθηκε στην παρούσα εργασία, παρατηρείται το φαινόμενο της σταδιακής ομογενοποίησης και συμπύκνωσης των embeddings. . Όπως παρουσιάζεται στο Σχήμα 5.6, το Transformer block επαναλαμβάνεται 4 φορές, γεγονός που επιτρέπει την επαναληπτική επεξεργασία των tokens και τη σταδιακή ενίσχυση των αναπαραστάσεων. Στα αρχικά επίπεδα του δικτύου, οι αναπαραστάσεις των patches διατηρούν έντονα τα τοπικά χαρακτηριστικά της εικόνας, όπως ακμές, υφές και χρωματικές διαφοροποιήσεις, τα οποία αντανακλούν την άμεση πληροφορία της εισόδου. Ωστόσο, καθώς η πληροφορία προωθείται σε βαθύτερα επίπεδα, τα embeddings εξελίσσονται σε πιο αφηρημένες και παγκόσμιες αναπαραστάσεις, ενσωματώνοντας πληροφορία από ολόκληρη την εικόνα.

Η ροή πληροφορίας μέσα στο μοντέλο υλοποιείται μέσω επαναλαμβανόμενων Transformer blocks, τα οποία έχουν πανομοιότυπη δομή και εφαρμόζονται σειριακά. Κάθε block δέχεται είσοδο tensor μορφής (B, N+1, E) και διατηρεί ακριβώς την ίδια διάσταση στην έξοδο, γεγονός που επιτρέπει την εύκολη στοίβαξη πολλαπλών blocks. Η πρώτη λειτουργία σε κάθε block είναι το LayerNormalization, το οποίο εφαρμόζεται ανεξάρτητα σε κάθε token και κανονικοποιεί τα χαρακτηριστικά κατά μήκος της διάστασης E. Αυτό μειώνει τη μεταβλητότητα των ενεργοποιήσεων και σταθεροποιεί την εκπαίδευση.

Ακολουθεί η multi-head attention, η οποία επιστρέφει tensor ίδιας μορφής. Το αποτέλεσμα προστίθεται στο αρχικό input μέσω residual σύνδεσης (x + attention_output). Η πρόσθεση αυτή υλοποιείται στοιχειο-wise και δεν απαιτεί αλλαγή διαστάσεων, εξασφαλίζοντας ότι η αρχική πληροφορία διατηρείται. Στη συνέχεια εφαρμόζεται δεύτερο LayerNormalization και ακολουθεί το MLP block. Το MLP αποτελείται από δύο Dense layers, με μονάδες [128, 64] σύμφωνα με τον Πίνακα 5.1, ενώ στη συνέχεια η πληροφορία επιστρέφει στη βασική ροή του μοντέλου. Και εδώ εφαρμόζεται residual σύνδεση, δηλαδή το output του MLP προστίθεται στο input του block.

Η επανάληψη αυτής της δομής οδηγεί σε σταδιακή «ομογενοποίηση» της πληροφορίας, καθώς κάθε token εμπλουτίζεται επαναληπτικά με πληροφορία από όλα τα υπόλοιπα tokens μέσω attention, ενώ το MLP ενισχύει τις σημαντικές συνιστώσες. Σημαντικό στοιχείο της υλοποίησης είναι ότι σε κανένα σημείο δεν αλλάζει το σχήμα των δεδομένων, κάτι που απλοποιεί τη ροή και μειώνει τον κίνδυνο σφαλμάτων συμβατότητας.

5.4.6 Ο Κομβικός Ρόλος του Classification Token

Στο πλαίσιο της αρχιτεκτονικής Vision Transformer, ένα από τα πλέον κρίσιμα και εννοιολογικά ισχυρά στοιχεία είναι το λεγόμενο Classification token (CLS token), το οποίο εισάγεται από την αρχή της ακολουθίας εισόδου και αποτελεί άμεση μεταφορά ιδεών από τα γλωσσικά μοντέλα Transformer, όπως το BERT (Bidirectional Encoder Representations from Transformers).

Το CLS token είναι ένα τεχνητό, μαθητεύσιμο διάνυσμα, το οποίο δεν αντιστοιχεί σε κάποιο συγκεκριμένο τμήμα της εικόνας, σε αντίθεση με τα υπόλοιπα tokens που προκύπτουν από τον διαχωρισμό της εικόνας σε patches. Αντί να κωδικοποιεί τοπική οπτική πληροφορία, λειτουργεί ως ένας καθολικός φορέας συγκέντρωσης πληροφορίας, με στόχο τη δημιουργία μίας ενιαίας αναπαράστασης υψηλού επιπέδου για ολόκληρη την εικόνα.

Στην υλοποίηση του notebook, το classification token εισάγεται πριν από τα transformer blocks και παραμένει στην πρώτη θέση της ακολουθίας σε όλη τη διάρκεια της επεξεργασίας. Καθώς περνά από κάθε attention layer, το CLS token συμμετέχει πλήρως στον υπολογισμό των attention weights, τόσο ως query όσο και ως key/value, επιτρέποντάς του να συγκεντρώνει πληροφορία από όλα τα patches. Ωστόσο, με βάση το σχήμα και τον Πίνακα 5.1 της αρχιτεκτονικής που παρουσιάζονται στην εργασία, η τελική κεφαλή ταξινόμησης περιλαμβάνει Global GAP, Dropout με ποσοστό 0.3, Dense layer με 5 εξόδους και Softmax. Επομένως, στο επίπεδο της απεικόνισης δίνεται έμφαση στη συγκεντρωτική αναπαράσταση των embeddings μέσω Global GAP και όχι αποκλειστικά στη χρήση του CLS token.

Το τελικό στάδιο της ταξινόμησης μετατρέπει την παραγόμενη αναπαράσταση σε πιθανότητες για τις επιμέρους κατηγορίες. Όπως φαίνεται στο σχήμα, η κεφαλή ταξινόμησης οδηγεί σε έξοδο 5 κλάσεων, ενώ η συνάρτηση Softmax χρησιμοποιείται για την παραγωγή της τελικής κατανομής πιθανοτήτων. Από υλοποιητικής πλευράς, το classification head παραμένει σχετικά απλό σε σχέση με την υπόλοιπη αρχιτεκτονική, καθώς αποτελεί το τελικό στάδιο μετατροπής των αναπαραστάσεων σε προβλέψεις. Παρατίθεται παρακάτω και ο Πίνακας της Αρχιτεκτονικής του Vision Transformer που υλοποιήθηκε στην εργασία.

Πίνακας 5.1: Πίνακας Αρχιτεκτονικής Vision Transformer

Στοιχείο Αρχιτεκτονικής	Τι χρησιμοποιείται
Είσοδος	Εικόνα 224x224x3 RGB
Patch size	16x16
Embedding Dimention	64
Transformer Blocks	4
Attention Heads	4
Διαστάσεις MLP	[128,64]
Classification Head	Global GAP, 0.3 Dropout, 5 Dense Layers, Softmax Activation
Έξοδος	5 κλάσεις

5.4.7 Ερμηνευσιμότητα και Οπτικοποίηση στο Vision Transformer

Ένα από τα σημαντικότερα πλεονεκτήματα της αρχιτεκτονικής Vision Transformer είναι η εγγενής δυνατότητα ερμηνείας της διαδικασίας λήψης απόφασης, η οποία διαφοροποιεί ουσιαστικά τα μοντέλα αυτά από τις κλασικές αρχιτεκτονικές Συνελικτικών Νευρωνικών Δικτύων (CNNs). Συγκεκριμένα, μέσω της εξαγωγής των attention weights από τα επιμέρους heads κάθε Transformer block, καθίσταται εφικτή η λεπτομερής ανάλυση του τρόπου με τον οποίο το μοντέλο κατανέμει τη σημασία στα διάφορα patches της εικόνας. Η ιδιότητα αυτή επιτρέπει την οπτικοποίηση των περιοχών στις οποίες εστιάζει το μοντέλο κατά τη διαδικασία της ταξινόμησης, παρέχοντας ένα επίπεδο διαφάνειας που είναι κρίσιμο για εφαρμογές υψηλής αξιοπιστίας, όπως η ιατρική διάγνωση.

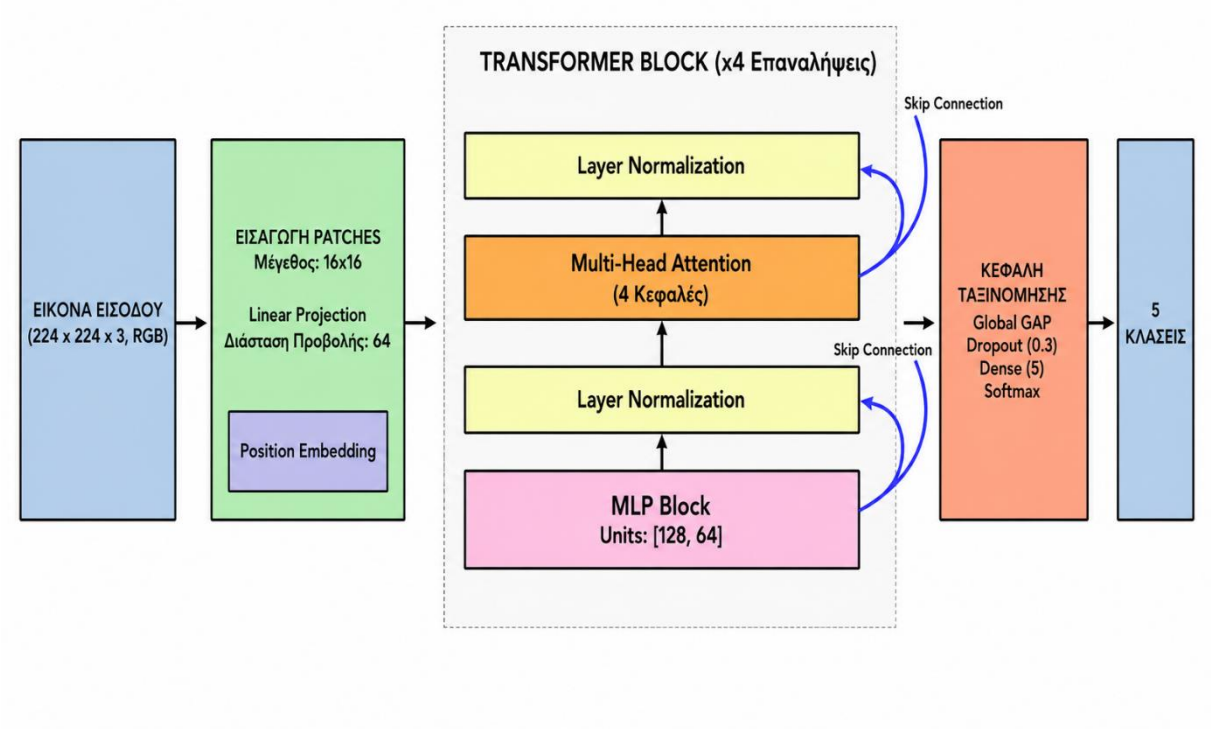
Στο πλαίσιο της παρούσας διπλωματικής εργασίας, η ερμηνευσιμότητα του ViT ενισχύεται περαιτέρω μέσω της αξιοποίησης τεχνικών οπτικοποίησης, όπως το Grad-CAM και οι χάρτες συμβολής patches. Οι μέθοδοι αυτές επιτρέπουν την αποτύπωση της ροής των gradients προς τα attention layers, δημιουργώντας heatmaps τα οποία αναδεικνύουν τις περιοχές της εικόνας που συνέβαλαν περισσότερο στην τελική πρόβλεψη.

Η εφαρμογή τεχνικών ερμηνευσιμότητας στην έρευνα βασίζεται σε gradient-based προσέγγιση, η οποία επιτρέπει την ανάλυση της συμβολής των patches στην τελική πρόβλεψη. Η διαδικασία ξεκινά με τη χρήση του GradientTape, το οποίο καταγράφει τις πράξεις του forward pass και επιτρέπει τον υπολογισμό των gradients. Κατά τη διάρκεια του forward pass, υπολογίζεται η έξοδος του μοντέλου για ένα συγκεκριμένο δείγμα και επιλέγεται η πιθανότητα της κλάσης ενδιαφέροντος. Στη συνέχεια, μέσω του gradient tape υπολογίζονται τα gradients αυτής της εξόδου ως προς ένα ενδιάμεσο tensor, το οποίο συνήθως αντιστοιχεί στις ενεργοποιήσεις των patches πριν από το classification στάδιο.

Τα gradients αυτά έχουν την ίδια δομή με το επιλεγμένο tensor και χρησιμοποιούνται για την εκτίμηση της σημαντικότητας κάθε patch. Συγκεκριμένα, εφαρμόζεται μέσος όρος κατά μήκος της διάστασης των tokens ή των χαρακτηριστικών, παράγοντας ένα σύνολο βαρών. Τα βάρη αυτά πολλαπλασιάζονται με τις αντίστοιχες ενεργοποιήσεις και το αποτέλεσμα αθροίζεται, ώστε να παραχθεί ένας μονοδιάστατος χάρτης σημαντικότητας. Ο χάρτης αυτός αναδιαμορφώνεται σε δισδιάστατο πλέγμα, με διαστάσεις που αντιστοιχούν στη διάταξη των patches (π.χ. 14x14). Στη συνέχεια, εφαρμόζεται επαναδειγματοληψία μέσω image.resize, ώστε να αποκτήσει τις ίδιες διαστάσεις με την αρχική εικόνα. Ακολουθεί κανονικοποίηση των τιμών στο εύρος [0,1], ώστε να είναι δυνατή η απεικόνισή του.

Τέλος, το heatmap προβάλλεται πάνω στην αρχική εικόνα μέσω blending, επιτρέποντας την οπτική απεικόνιση των περιοχών που επηρέασαν περισσότερο την απόφαση του μοντέλου. Όλη η διαδικασία υλοποιείται με tensor operations και αξιοποιεί πλήρως τις δυνατότητες του TensorFlow, το οποίο χρησιμοποιείται για την ανάπτυξη του Vision Transformer, τόσο για αυτόματη παραγώγιση όσο και για τη διαχείριση γραφημάτων υπολογισμού.

Στο σχήμα 5.6 φαίνεται ο συνολικός αριθμός επιπέδων και η δομή της αρχιτεκτονικής Vision Transformer που υλοποιήθηκε στην εργασία.



Σχήμα 5.6: Αναπαράσταση αναπτυγμένης αρχιτεκτονικής Vision Transformer

5.5 Διαδικασία Εκπαίδευσης Μοντέλων

Η διαδικασία εκπαίδευσης των μοντέλων στην παρούσα εργασία υλοποιήθηκε σε περιβάλλον TensorFlow/Keras και βασίστηκε σε μια ενιαία, επαναχρησιμοποιήσιμη pipeline, η οποία εφαρμόστηκε σε διαφορετικές αρχιτεκτονικές, όπως ένα custom CNN, τα προεκπαιδευμένα μοντέλα VGG16, VGG19, ResNet50, InceptionResNetV2, καθώς και ένα Vision Transformer.

Αρχικά, πραγματοποιείται η φόρτωση και προεπεξεργασία των δεδομένων εικόνας μέσω των κατάλληλων εργαλείων του Keras (ImageDataGenerator ή αντίστοιχων pipelines). Οι εικόνες μετατρέπονται σε σταθερό μέγεθος εισόδου, συμβατό με τις απαιτήσεις των προεκπαιδευμένων μοντέλων, ενώ εφαρμόζεται κανονικοποίηση των τιμών των pixels. Επιπλέον, γίνεται χρήση τεχνικών data augmentation, όπως περιστροφές, μετατοπίσεις και αναστροφές, με στόχο τη βελτίωση της γενίκευσης των μοντέλων.

Στη συνέχεια, τα δεδομένα διαχωρίζονται σε σύνολα εκπαίδευσης και επικύρωσης, διασφαλίζοντας ότι η αξιολόγηση κατά τη διάρκεια της εκπαίδευσης πραγματοποιείται σε μη παρατηρημένα δεδομένα. Η εκπαίδευση των μοντέλων υλοποιείται μέσω της μεθόδου model.fit(), με χρήση batches δεδομένων και παρακολούθηση των μετρικών απόδοσης ανά εποχή.

Για τα προεκπαιδευμένα μοντέλα (VGG16, VGG19, ResNet50, InceptionResNetV2), εφαρμόζεται η τεχνική του transfer learning. Συγκεκριμένα, το βασικό convolutional σώμα του μοντέλου φορτώνεται με προεκπαιδευμένα βάρη (π.χ. από το ImageNet), ενώ τα αρχικά layers παραμένουν παγωμένα. Στην κορυφή του μοντέλου προστίθενται πλήρως συνδεδεμένα επίπεδα, τα οποία εκπαιδεύονται εκ νέου για το συγκεκριμένο πρόβλημα ταξινόμησης. Σε επόμενη φάση, σε ορισμένα μοντέλα εφαρμόζεται fine-tuning, όπου αποδεσμεύονται επιλεγμένα βαθύτερα layers για περαιτέρω προσαρμογή. Αντίθετα, στο custom CNN η εκπαίδευση ξεκινά από τυχαία αρχικοποιημένα βάρη και το μοντέλο μαθαίνει εξ αρχής τα χαρακτηριστικά των εικόνων μέσω διαδοχικών convolutional και pooling layers.

Για το Vision Transformer, η διαδικασία εκπαίδευσης διαφοροποιείται σημαντικά, καθώς η εικόνα διασπάται σε patches και μετατρέπεται σε embeddings, τα οποία επεξεργάζονται μέσω transformer blocks και μηχανισμών self-attention. Το ViT εκπαιδεύεται είτε εξ ολοκλήρου από την αρχή είτε με βάση προεκπαιδευμένα embeddings, ακολουθώντας παρόμοια διαδικασία βελτιστοποίησης.

Κατά τη διάρκεια της εκπαίδευσης χρησιμοποιούνται callbacks, όπως το EarlyStopping για την αποφυγή υπερπροσαρμογής και το ModelCheckpoint για την αποθήκευση του καλύτερου μοντέλου. Η παρακολούθηση της εκπαίδευσης γίνεται μέσω των καμπυλών loss και accuracy, τόσο για το training όσο και για το validation set. Επιπλέον, σε ορισμένες περιπτώσεις εφαρμόζεται δυναμική προσαρμογή του learning rate μέσω scheduler callbacks, με στόχο τη σταθεροποίηση της σύγκλισης και τη βελτίωση της τελικής απόδοσης.

Η συνολική διαδικασία σχεδιάστηκε ώστε να είναι συνεπής μεταξύ των διαφορετικών μοντέλων, επιτρέποντας την άμεση σύγκριση των αρχιτεκτονικών υπό παρόμοιες συνθήκες εκπαίδευσης. Με τον τρόπο αυτό, οι διαφορές στα αποτελέσματα αποδίδονται κυρίως στα ιδιαίτερα χαρακτηριστικά κάθε αρχιτεκτονικής και όχι σε διαφοροποιήσεις του πειραματικού πρωτοκόλλου.

5.6 Ρυθμίσεις υπερπαραμέτρων

Η ρύθμιση των υπερπαραμέτρων στην παρούσα εργασία πραγματοποιήθηκε με στόχο τη βελτιστοποίηση της απόδοσης των μοντέλων, διατηρώντας παράλληλα τη σταθερότητα της εκπαίδευσης.

Κεντρικό ρόλο κατέχει ο optimizer, όπου χρησιμοποιείται ο Adam λόγω της ικανότητάς του να προσαρμόζει δυναμικά το learning rate κατά τη διάρκεια της εκπαίδευσης. Η αρχική τιμή του learning rate (ξεκινώντας από 10−4) επιλέγεται ώστε να επιτυγχάνεται ισορροπία μεταξύ ταχύτητας σύγκλισης και σταθερότητας, ενώ σε ορισμένες περιπτώσεις γίνεται χρήση learning rate scheduling ή μείωσης του learning rate κατά τη διάρκεια της εκπαίδευσης.

Το batch size καθορίζεται με βάση τους υπολογιστικούς περιορισμούς του συστήματος, ιδιαίτερα τη διαθέσιμη μνήμη της GPU, ενώ επηρεάζει και τη συμπεριφορά σύγκλισης του μοντέλου. Στη συγκεκριμένη εργασία, το batch size ορίστηκε ίσο με 32. Μικρότερα batch sizes χρησιμοποιούνται για πιο σταθερή γενίκευση, ενώ αποφεύγονται πολύ μεγάλα batches λόγω αυξημένων απαιτήσεων μνήμης.

Ο αριθμός των epochs επιλέγεται έτσι ώστε να επιτρέπεται επαρκής εκπαίδευση των μοντέλων χωρίς να οδηγεί σε overfitting. Για τον λόγο αυτό αξιοποιείται το EarlyStopping, το οποίο παρακολουθεί το validation loss και διακόπτει την εκπαίδευση όταν δεν παρατηρείται περαιτέρω βελτίωση. Τα μοντέλα της παρούσας έρευνας εκπαιδεύτηκαν συνολικά σε 10 εποχές, καθώς το dataset θεωρείται σχετικά περιορισμένο σε σύγκριση με μεγάλης κλίμακας σύνολα δεδομένων γενικής χρήσης. Το μόνο μοντέλο που εκπαιδεύτηκε παραπάνω εποχές αποτελεί το Vision Transformer το οποίο από την «φύση» του χρειάζεται παραπάνω εκπαίδευση.

Στα προεκπαιδευμένα μοντέλα, σημαντική υπερπαράμετρος αποτελεί το ποσοστό των layers που παραμένουν παγωμένα. Στα αρχικά στάδια, το μεγαλύτερο μέρος του μοντέλου παραμένει frozen, ενώ σταδιακά αποδεσμεύονται επιλεγμένα layers για fine-tuning. Η επιλογή αυτή επηρεάζει σημαντικά την ικανότητα του μοντέλου να προσαρμοστεί στο dataset, καθώς καθορίζει την ισορροπία μεταξύ γενικής γνώσης και εξειδίκευσης.

Στο Vision Transformer, ιδιαίτερη σημασία δίνεται σε υπερπαραμέτρους όπως ο αριθμός των transformer blocks και η διάσταση των embeddings. Παρατηρείται ότι η επιλογή μικρότερου βάθους σε ορισμένες περιπτώσεις οδηγεί σε πιο σταθερή εκπαίδευση, ειδικά όταν το dataset είναι περιορισμένο. Επιπλέον, η επιλογή του patch size επηρεάζει άμεσα τον αριθμό των tokens και, κατά συνέπεια, την υπολογιστική πολυπλοκότητα του attention mechanism.

Επιπλέον, χρησιμοποιούνται τεχνικές κανονικοποίησης όπως το Dropout, κυρίως στα fully connected layers, για την αποφυγή υπερπροσαρμογής. Στην παρούσα εργασία, οι τιμές του Dropout κυμάνθηκαν από 0.2 έως 0.3, ανάλογα με την αρχιτεκτονική του μοντέλου. Η τεχνική αυτή μειώνει την εξάρτηση του μοντέλου από συγκεκριμένα μοτίβα, ενισχύοντας τη δυνατότητα γενίκευσης σε άγνωστα δεδομένα.

Η διαδικασία επιλογής των υπερπαραμέτρων βασίστηκε σε επαναληπτικά πειράματα, όπου δοκιμάστηκαν διαφορετικές ρυθμίσεις και επιλέχθηκαν εκείνες που παρείχαν την καλύτερη ισορροπία μεταξύ απόδοσης, σταθερότητας και υπολογιστικού κόστους. Με τον τρόπο αυτό διασφαλίστηκε ότι τα τελικά μοντέλα δεν βελτιστοποιήθηκαν αποκλειστικά ως προς την ακρίβεια, αλλά και ως προς τη δυνατότητα γενίκευσης και τη σταθερότητα της εκπαίδευσης.

5.7 Προβλήματα που εμφανίστηκαν κατά την υλοποίηση

Κατά την υλοποίηση των notebooks και την εκπαίδευση των μοντέλων προέκυψαν διάφορα προβλήματα, τα οποία σχετίζονται τόσο με το υπολογιστικό περιβάλλον όσο και με τη συμπεριφορά των ίδιων των μοντέλων.

Ένα από τα κύρια ζητήματα ήταν η καθυστέρηση στην εκπαίδευση, ιδιαίτερα όταν τα μοντέλα εκτελούνταν σε CPU αντί για GPU. Τα πιο σύνθετα μοντέλα, όπως το InceptionResNetV2 και το Vision Transformer, παρουσίασαν σημαντικά αυξημένο χρόνο ανά epoch, καθιστώντας τη διαδικασία εκπαίδευσης ιδιαίτερα χρονοβόρα. Το πρόβλημα αυτό έγινε ακόμη πιο εμφανές στα transformer-based μοντέλα, όπου ο μηχανισμός self-attention αυξάνει σημαντικά την υπολογιστική πολυπλοκότητα λόγω του μεγάλου αριθμού πράξεων μεταξύ των tokens.

Επιπλέον, παρουσιάστηκαν προβλήματα συμβατότητας μεταξύ διαφορετικών εκδόσεων TensorFlow και Keras. Σε ορισμένες περιπτώσεις, η χρήση παραμέτρων που δεν υποστηρίζονταν από τη συγκεκριμένη έκδοση του optimizer οδήγησε σε σφάλματα εκτέλεσης, απαιτώντας προσαρμογή του κώδικα ή αλλαγές στις βιβλιοθήκες του περιβάλλοντος. Παράλληλα, διαφορές στον τρόπο διαχείρισης των callbacks και των layers μεταξύ εκδόσεων δημιούργησαν προβλήματα κατά τη φόρτωση ή επαναχρησιμοποίηση προεκπαιδευμένων μοντέλων.

Ένα σημαντικό πρόβλημα που παρατηρήθηκε ήταν το overfitting, ιδιαίτερα στα προεκπαιδευμένα μοντέλα transfer learning. Τα μοντέλα εμφάνιζαν πολύ υψηλή απόδοση στο training set, αλλά αισθητά χαμηλότερη στο validation set, γεγονός που υποδεικνύει περιορισμένη ικανότητα γενίκευσης. Η αντιμετώπιση πραγματοποιήθηκε μέσω τεχνικών data augmentation, χρήσης Dropout και αξιοποίησης του EarlyStopping, ώστε να αποφεύγεται η υπερεκπαίδευση των δικτύων. Επιπλέον, σε ορισμένες περιπτώσεις περιορίστηκε το ποσοστό των layers που αποδεσμεύονταν κατά το fine-tuning, μειώνοντας έτσι τον αριθμό των εκπαιδεύσιμων παραμέτρων.

Αντίθετα, σε απλούστερα μοντέλα όπως το custom CNN παρατηρήθηκε σε ορισμένες περιπτώσεις underfitting, όπου το μοντέλο δεν μπορούσε να μάθει επαρκώς τα χαρακτηριστικά των δεδομένων. Το φαινόμενο αυτό συνδέεται κυρίως με τη μικρότερη εκφραστική ικανότητα της αρχιτεκτονικής, καθώς και με τον περιορισμένο αριθμό παραμέτρων σε σύγκριση με βαθύτερα δίκτυα. Για την αντιμετώπισή του πραγματοποιήθηκαν δοκιμές με διαφορετικό αριθμό filters, μεγαλύτερο βάθος δικτύου και προσαρμογές στις υπερπαραμέτρους της εκπαίδευσης.

Στο Vision Transformer εμφανίστηκε αυξημένη ευαισθησία στην επιλογή της αρχιτεκτονικής, όπως στον αριθμό των transformer blocks και στη διάσταση των embeddings. Παρατηρήθηκε ότι βαθύτερα μοντέλα δεν οδηγούσαν απαραίτητα σε καλύτερη απόδοση, ειδικά σε datasets περιορισμένου μεγέθους. Επιπλέον, η αύξηση του αριθμού των patches αύξανε σημαντικά τη χρήση μνήμης και τον χρόνο εκπαίδευσης, γεγονός που περιόρισε την επιλογή μεγαλύτερης ανάλυσης εικόνας ή μικρότερου patch size.

Παράλληλα, παρουσιάστηκαν προβλήματα διαχείρισης μνήμης, ιδιαίτερα κατά τη χρήση μεγάλων batch sizes ή πολύπλοκων μοντέλων. Σε αρκετές περιπτώσεις εμφανίστηκαν σφάλματα εξάντλησης μνήμης GPU (Out Of Memory – OOM), κυρίως στα transformer-based μοντέλα και στο InceptionResNetV2. Η μείωση του batch size αποτέλεσε βασική λύση για την αντιμετώπιση αυτών των περιορισμών, ενώ σε ορισμένες περιπτώσεις απαιτήθηκε και μείωση της ανάλυσης εισόδου ή απλοποίηση της αρχιτεκτονικής.

Τέλος, η αναπαραγωγιμότητα των αποτελεσμάτων αποτέλεσε σημαντική πρόκληση, καθώς η στοχαστική φύση της εκπαίδευσης, όπως η τυχαία αρχικοποίηση βαρών (random initialization), το shuffling των δεδομένων και οι στοχαστικές λειτουργίες των optimizers, οδηγεί σε μικρές διαφοροποιήσεις μεταξύ διαφορετικών εκτελέσεων. Για τον λόγο αυτό κρίθηκε απαραίτητη η χρήση σταθερών random seeds, όπου αυτό ήταν εφικτό, καθώς και η διατήρηση σταθερών συνθηκών εκπαίδευσης σε όλα τα πειράματα.

Συνολικά, τα προβλήματα που εμφανίστηκαν αντιμετωπίστηκαν μέσω συνδυασμού τεχνικών βελτιστοποίησης, προσαρμογής του υπολογιστικού περιβάλλοντος και επαναληπτικής πειραματικής διαδικασίας. Η διαδικασία αυτή συνέβαλε ουσιαστικά στη βελτίωση της σταθερότητας των μοντέλων και στην ενίσχυση της αξιοπιστίας των τελικών αποτελεσμάτων της έρευνας.

Κεφάλαιο 6ο: Πειραματική Αξιολόγηση

6.1 Μετρικές αξιολόγησης

Ένα μοντέλο μηχανικής μάθησης, μετά το πέρας της εκπαίδευσής του, είναι αναγκαίο να αξιολογηθεί σε δεδομένα τα οποία είναι άγνωστα. Μέσω της αξιολόγησης ενός μοντέλου ML επιτρέπεται η αντικειμενική εκτίμηση της ποιότητας των αποτελεσμάτων — και συνολικότερα της απόδοσής του — αλλά και η σύγκρισή του με εναλλακτικές προσεγγίσεις. Η διαδικασία αυτή καθίσταται δυνατή μέσω της αξιοποίησης των μετρικών αξιολόγησης των αλγορίθμων μηχανικής μάθησης. Γενικότερα, οι μετρικές λειτουργούν ως ποσοτικοί δείκτες, οι οποίοι συνδέουν τον θεωρητικό τρόπο λειτουργίας των αλγορίθμων με το πόσο αποτελεσματικοί είναι στην πράξη, διευκολύνοντας τον ερευνητή τόσο στη ρύθμιση των υπερπαραμέτρων όσο και στην ερμηνεία των αποτελεσμάτων.

Στην παρούσα έρευνα χρησιμοποιήθηκαν μετρικές ταξινόμησης (classification metrics) για την αξιολόγηση των μοντέλων. Η επιλογή αυτή σχετίζεται άμεσα με τη φύση του προβλήματος που εξετάζεται, καθώς η εργασία επικεντρώνεται στην ταξινόμηση ιστοπαθολογικών εικόνων σε πολλαπλές κατηγορίες. Οι μετρικές ταξινόμησης επιτρέπουν την ποσοτική αποτίμηση της ικανότητας των μοντέλων να διαχωρίζουν σωστά τις διαφορετικές κατηγορίες ιστών και παρέχουν πιο ολοκληρωμένη εικόνα της συμπεριφοράς τους σε σχέση με μία μόνο συνολική μέτρηση.

Οι βασικές μετρικές που χρησιμοποιήθηκαν στην αξιολόγηση των μοντέλων είναι οι εξής: Accuracy, Precision, Recall, F1-score, καθώς και ο Πίνακας Σύγχυσης (Confusion Matrix). Η χρήση πολλαπλών μετρικών κρίθηκε απαραίτητη, καθώς κάθε μία αποτυπώνει διαφορετική πτυχή της απόδοσης του μοντέλου. Με τον τρόπο αυτό καθίσταται δυνατή η πιο ολοκληρωμένη και αξιόπιστη αξιολόγηση των αρχιτεκτονικών που εξετάζονται στην παρούσα εργασία.

Accuracy

Η μετρική ακρίβειας (Accuracy) εκφράζει το συνολικό ποσοστό των σωστών προβλέψεων σε σχέση με το σύνολο των δειγμάτων, παρέχοντας μια γενική εικόνα της απόδοσης του μοντέλου. Αποτελεί τη θεμελιώδη και πιο εύκολα ερμηνεύσιμη μετρική ταξινόμησης και χρησιμοποιείται ευρέως για την αξιολόγηση μοντέλων μηχανικής μάθησης. Ωστόσο, ανάλογα με τη φύση του προβλήματος, τα αποτελέσματά της μπορεί να αποδειχθούν παραπλανητικά ως προς την πραγματική ικανότητα γενίκευσης του μοντέλου.1

Η περίπτωση αυτή εμφανίζεται κυρίως όταν το μοντέλο έχει εκπαιδευτεί σε σύνολα δεδομένων με ανισόρροπες κλάσεις, όπου το μοντέλο τείνει να εμφανίζει υψηλή απόδοση στις κλάσεις με μεγάλο αριθμό δειγμάτων και χαμηλότερη απόδοση στις κλάσεις με μικρότερο αριθμό δειγμάτων. Στην παρούσα έρευνα, το πρόβλημα αυτό δεν υφίσταται, καθώς το σύνολο δεδομένων που χρησιμοποιείται είναι πλήρως ισορροπημένο (25.000 εικόνες σε 5 κλάσεις, δηλαδή 5.000 εικόνες ανά κλάση).

Precision

Η μετρική Precision εκφράζει το ποσοστό των σωστών θετικών προβλέψεων σε σχέση με το σύνολο των προβλέψεων που χαρακτηρίστηκαν ως θετικές. Η μετρική αυτή είναι ιδιαίτερα χρήσιμη για την αξιολόγηση προβλημάτων στα οποία τα μοντέλα εμφανίζουν αυξημένο ποσοστό ψευδώς θετικών αποτελεσμάτων.

Το φαινόμενο αυτό είναι ιδιαίτερα συχνό σε προβλήματα που σχετίζονται με την ιατρική επιστήμη, όπως και στο αντικείμενο που εξετάζει η παρούσα διπλωματική εργασία, γεγονός που

καθιστά τη συγκεκριμένη μετρική ιδιαίτερα σημαντική για την αξιολόγηση των αλγορίθμων που αναπτύχθηκαν στο πλαίσιο της έρευνας.

Recall

Η μετρική ευαισθησίας (Recall) εκφράζει το ποσοστό των πραγματικά θετικών περιπτώσεων που αναγνωρίστηκαν σωστά από το μοντέλο. Αποτελεί μία ιδιαίτερα σημαντική μετρική αξιολόγησης, κυρίως σε εφαρμογές όπου η αποτυχία ανίχνευσης θετικών περιστατικών μπορεί να έχει σοβαρές συνέπειες, όπως σε προβλήματα που σχετίζονται με την υγεία ή την ασφάλεια.

Όπως και η μετρική Precision, έτσι και η ευαισθησία διαδραματίζει καθοριστικό ρόλο στην παρούσα έρευνα, καθώς το πρόβλημα που εξετάζεται αφορά την ανίχνευση καρκινικών ιστοπαθολογικών εικόνων. Μια λανθασμένη πρόβλεψη αρνητικού αποτελέσματος σε πραγματικά καρκινικό δείγμα παράγοντας ένα ψευδώς αρνητικό αποτέλεσμα το οποίο μπορεί να επηρεάσει σημαντικά τη διαδικασία διάγνωσης και, κατ' επέκταση, τη θεραπευτική αντιμετώπιση του ασθενούς.

F1-Score

Η μετρική F1-Score αποτελεί έναν συνδυαστικό δείκτη αξιολόγησης, ο οποίος προκύπτει από τον συνδυασμό των μετρικών Precision και Recall μέσω του υπολογισμού του αρμονικού μέσου τους. Ο αρμονικός μέσος είναι ένας ειδικός τύπος μέσου όρου, ο οποίος χρησιμοποιείται σε περιπτώσεις όπου υπάρχουν σημαντικές ανισορροπίες μεταξύ των τιμών, δίνοντας μεγαλύτερη βαρύτητα στις χαμηλές επιδόσεις. Με τον τρόπο αυτό, αποφεύγεται η παραπλανητική αξιολόγηση ενός μοντέλου όταν μία από τις δύο μετρικές παρουσιάζει ιδιαίτερα χαμηλή τιμή.

Το F1-Score αποτελεί ιδιαίτερα σημαντική μετρική σε πολυταξικά προβλήματα ταξινόμησης, καθώς και σε εφαρμογές όπου απαιτείται ισορροπημένη αξιολόγηση ανάμεσα στην ικανότητα του μοντέλου να εντοπίζει σωστά τα θετικά δείγματα και στην αξιοπιστία των προβλέψεών του. Για τον λόγο αυτό χρησιμοποιείται ευρέως σε προβλήματα ιατρικής διάγνωσης και ανάλυσης εικόνας, όπου τόσο τα ψευδώς θετικά όσο και τα ψευδώς αρνητικά αποτελέσματα έχουν ιδιαίτερη σημασία.

Το F1-Score υπολογίζεται με τον παρακάτω τύπο:

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{6.1}$$

Σε πολυταξικά προβλήματα, όπως αυτό της παρούσας διπλωματικής εργασίας, η μετρική F1-Score μπορεί να υπολογιστεί με διαφορετικούς τρόπους, ανάλογα με τη φύση και τις απαιτήσεις του προβλήματος. Οι πιο συνηθισμένες προσεγγίσεις είναι οι macro και weighted μέθοδοι υπολογισμού.

Η μέθοδος macro υπολογίζει αρχικά το F1-Score ξεχωριστά για κάθε κλάση και στη συνέχεια υπολογίζεται ο μέσος όρος των επιμέρους τιμών, αποδίδοντας ίσο βάρος σε όλες τις κλάσεις. Η προσέγγιση αυτή χρησιμοποιείται κυρίως σε περιπτώσεις όπου όλες οι κλάσεις θεωρούνται εξίσου σημαντικές, ανεξάρτητα από τον αριθμό των δειγμάτων που περιέχουν.

Αντίθετα, η μέθοδος weighted λαμβάνει υπόψη το μέγεθος κάθε κλάσης κατά τον υπολογισμό του τελικού μέσου όρου. Πιο συγκεκριμένα, το βάρος κάθε κλάσης είναι ανάλογο με το ποσοστό των δειγμάτων της στο dataset. Η συγκεκριμένη εκδοχή του F1-Score χρησιμοποιείται κυρίως σε προβλήματα όπου τα σύνολα δεδομένων είναι ανισόρροπα, καθώς επιτρέπει πιο ρεαλιστική αξιολόγηση της συνολικής απόδοσης του μοντέλου. Ωστόσο, υπάρχει πιθανότητα εμφάνισης bias προς τις κλάσεις με μεγαλύτερο αριθμό δειγμάτων, καθώς αυτές επηρεάζουν περισσότερο το τελικό αποτέλεσμα.

Η επιλογή του κατάλληλου τρόπου υπολογισμού του F1-Score εξαρτάται από τη φύση του προβλήματος και από την ύπαρξη ή μη ανισορροπίας μεταξύ των κλάσεων του dataset. Στην παρούσα διπλωματική εργασία, όλες οι κλάσεις αντιμετωπίζονται με το ίδιο βάρος, καθώς το dataset που χρησιμοποιείται για την εκπαίδευση των μοντέλων είναι πλήρως ισορροπημένο.

Πίνακας Σύγχυσης

Ο Πίνακας Σύγχυσης (Confusion Matrix) αποτελεί ένα από τα σημαντικότερα εργαλεία για την ανάλυση της απόδοσης ενός μοντέλου ταξινόμησης μηχανικής μάθησης, καθώς αποτυπώνει τη σχέση μεταξύ των πραγματικών κλάσεων των δεδομένων και των προβλέψεων του μοντέλου. Μέσω του πίνακα σύγχυσης καθίσταται δυνατή η λεπτομερής κατανόηση των σφαλμάτων του μοντέλου, καθώς παρουσιάζεται ο τρόπος με τον οποίο οι προβλέψεις κατανέμονται στις επιμέρους κατηγορίες.

Στον πίνακα σύγχυσης συνδυάζονται πληροφορίες που σχετίζονται με τις μετρικές Accuracy, Precision και Recall, οι οποίες αναλύθηκαν στις προηγούμενες ενότητες. Με τον τρόπο αυτό παρέχεται πιο ολοκληρωμένη εικόνα της συμπεριφοράς του μοντέλου, τόσο ως προς τις σωστές όσο και ως προς τις λανθασμένες προβλέψεις.

Επιπλέον, ο πίνακας σύγχυσης προσφέρει αναλυτική εικόνα της απόδοσης ανά κλάση, γεγονός ιδιαίτερα σημαντικό σε πολυταξικά προβλήματα ταξινόμησης, όπου το μοντέλο καλείται να διαχωρίσει πολλαπλές κατηγορίες με πιθανές μορφολογικές ομοιότητες μεταξύ τους.

6.2 Αποτελέσματα Βασικού CNN

Μέσα από την αξιολόγηση του baseline CNN αναδεικνύεται μια ιδιαίτερα υψηλή απόδοση στο συγκεκριμένο πειραματικό πλαίσιο της πολυταξικής ταξινόμησης ιστοπαθολογικών εικόνων, γεγονός το οποίο αναδεικνύει την αποτελεσματικότητα του συγκεκριμένου CNN στο παρόν πειραματικό σύνολο δεδομένων στην εξαγωγή χωρικών και μορφολογικών χαρακτηριστικών μέσω της χρήσης τους σε ιατρικά δεδομένα. Το μοντέλο επιτυγχάνει συνολική ακρίβεια (test accuracy) της τάξης του 98.3% όπως είναι φανερό από τον Πίνακα 6.1, γεγονός που υποδηλώνει ότι μέσω της διαδικασίας εκπαίδευσης επιτεύχθηκε ισχυρή γενίκευση του μοντέλου στο συγκεκριμένο test set, χωρίς να υπάρχουν εμφανή σημάδια υποεκπαίδευσης.

Πίνακας 6.1: Αποτελέσματα Baseline CNN

Class Type	Precision	Recall	F1-score	Support
Colon_aca	0.99	0.99	0.99	600
Colon_n	0.99	0.99	0.99	600
Lung_aca	0.94	0.99	0.96	600
Lung_n	1	1	1	600
Lung_Scc	0.99	0.93	0.96	600
Accuracy	-	-	0.983	3000
Macro_avg	0.98	0.98	0.98	3000
Weighted_avg	0.98	0.98	0.98	3000

Ένας ακόμη παράγοντας που υποστηρίζει το παραπάνω συμπέρασμα είναι η χαμηλή τιμή της συνάρτησης κόστους (loss = 0.0577), η οποία δείχνει ότι το μοντέλο φαίνεται να έχει αποτυπώσει αποτελεσματικά βασικά μοτίβα των δεδομένων, περιορίζοντας σημαντικά το σφάλμα πρόβλεψης. Ο συνδυασμός υψηλής ακρίβειας και χαμηλού loss αποτελεί ένδειξη ότι το μοντέλο δεν αποτυγχάνει συστηματικά σε συγκεκριμένες κατηγορίες, αλλά παρουσιάζει συνολικά συνεπή συμπεριφορά.

Σε επίπεδο αναφοράς ανά κλάση, τα αποτελέσματα εμφανίζουν σημαντικές διαφοροποιήσεις, οι οποίες σχετίζονται με τη φύση των δεδομένων καθώς και με την πολυπλοκότητα των μορφολογικών χαρακτηριστικών κάθε κατηγορίας. Όσον αφορά τις κατηγορίες του παχέος εντέρου (Colon_aca και Colon_n), το μοντέλο εμφανίζει σχεδόν ιδανική απόδοση (precision = recall = F1-score ≈ 0.99), γεγονός που υποδηλώνει ότι έχει καταφέρει να μάθει με μεγάλη σαφήνεια τα διακριτικά χαρακτηριστικά μεταξύ καρκινικών και υγιών ιστών στο παχύ έντερο. Η ισορροπία μεταξύ precision και recall είναι ιδιαίτερα σημαντική, καθώς υποδεικνύει ότι το μοντέλο δεν ευνοεί ούτε την υπερταξινόμηση ούτε την υποταξινόμηση. Αυτό αποτελεί ένδειξη ότι τα χαρακτηριστικά των συγκεκριμένων κλάσεων είναι επαρκώς διαχωρίσιμα στον χώρο χαρακτηριστικών που έχει μάθει το CNN.

Στην κατηγορία lung_n, το μοντέλο επιτυγχάνει τέλεια επίδοση στο συγκεκριμένο πειραματικό περιβάλλον (precision = recall = F1-score = 1.00), γεγονός που αποτελεί ισχυρή ένδειξη ότι τα φυσιολογικά δείγματα πνεύμονα διαθέτουν σαφή και συνεπή μορφολογικά πρότυπα. Το αποτέλεσμα αυτό πιθανώς οφείλεται στη χαμηλή ενδοκλασική διακύμανση (intra-class variance), επιτρέποντας στο μοντέλο να δημιουργήσει ένα σταθερό και διακριτό decision boundary.

Ωστόσο, η απόδοση του μοντέλου στις κατηγορίες lung_aca και lung_scc παρουσιάζει ελαφρώς μειωμένα επίπεδα, με F1-score ίσο με 0.96. Η βασική αιτία αυτής της μείωσης εντοπίζεται στην πτώση του recall (0.94 και 0.93 αντίστοιχα), γεγονός που υποδηλώνει ότι το μοντέλο δεν καταφέρνει να εντοπίσει το σύνολο των θετικών περιπτώσεων. Με άλλα λόγια, παρατηρείται η ύπαρξη ψευδώς αρνητικών προβλέψεων, οι οποίες σε προβλήματα ιατρικής ταξινόμησης έχουν ιδιαίτερη σημασία, καθώς σχετίζονται με μη ανίχνευση παθολογικών καταστάσεων.

Η συγκεκριμένη συμπεριφορά μπορεί να αποδοθεί σε διάφορους παράγοντες:

- μορφολογική ομοιότητα μεταξύ υποτύπων καρκίνου του πνεύμονα, η οποία καθιστά δυσκολότερο τον διαχωρισμό τους από το μοντέλο,
- υψηλή ενδοκλασική ποικιλότητα (intra-class variability), όπου δείγματα της ίδιας κατηγορίας εμφανίζουν σημαντικές διαφοροποιήσεις,
- επικάλυψη χαρακτηριστικών (feature overlap) μεταξύ διαφορετικών κατηγοριών στον embedding space.

Η ανάλυση των συνολικών μετρικών (macro average και weighted average) καταδεικνύει τιμές ίσες με 0.98 για precision, recall και F1-score, επιβεβαιώνοντας ότι η απόδοση του μοντέλου είναι ομοιόμορφη σε όλες τις κλάσεις. Η ταύτιση macro και weighted average αποτελεί ένδειξη ότι το dataset είναι ισορροπημένο (ίσο support = 600 ανά κλάση), γεγονός που εξαλείφει πιθανές μεροληψίες λόγω ανισορροπίας δεδομένων.

Από πλευράς γενίκευσης, τα αποτελέσματα υποδηλώνουν ότι το μοντέλο έχει μάθει ουσιαστικά και όχι επιφανειακά χαρακτηριστικά των δεδομένων. Η ύπαρξη σαφούς διαχωρισμού μεταξύ training και test συνόλων ενισχύει την αξιοπιστία των μετρικών απόδοσης, καθώς διασφαλίζεται ότι η αξιολόγηση πραγματοποιείται σε μη παρατηρημένα δεδομένα. Συνεπώς, η υψηλή ακρίβεια του μοντέλου δεν μπορεί να αποδοθεί σε απλή απομνημόνευση (memorization), αλλά υποδηλώνει καλύτερη ικανότητα απόδοσης σε μη παρατηρημένα δείγματα του ίδιου dataset. Το γεγονός αυτό είναι ιδιαίτερα σημαντικό σε ιατρικές εφαρμογές, όπου η δυνατότητα μεταφοράς της γνώσης σε νέα, άγνωστα δεδομένα βασική προϋπόθεση για μελλοντική διερεύνηση σε πιο ρεαλιστικά και ετερογενή δεδομένα .

Συνολικά, το baseline CNN παρουσιάζει ισχυρή πειραματική συμπεριφορά, αποτελώντας ένα ιδιαίτερα ανταγωνιστικό σημείο αναφοράς για σύγκριση με πιο προηγμένα μοντέλα, όπως βαθύτερα CNNs (π.χ. ResNet, DenseNet) ή αρχιτεκτονικές βασισμένες σε attention μηχανισμούς (π.χ. Vision Transformers). Παρά την υψηλή του απόδοση, τα ευρήματα υποδεικνύουν ότι υπάρχει περιθώριο

βελτίωσης, κυρίως στη μείωση των ψευδώς αρνητικών προβλέψεων στις κατηγορίες καρκίνου του πνεύμονα, κάτι που θα μπορούσε να επιτευχθεί μέσω τεχνικών όπως data augmentation, fine-tuning βαθύτερων μοντέλων ή ενσωμάτωσης attention μηχανισμών.

6.3 Αποτελέσματα Τεχνικών Μεταφοράς Μάθησης

6.3.1 Αποτελέσματα InceptionResNetV2

Πρώτα αποτελέσματα InceptionResNetV2

Πίνακας 6.2: Πρώτα αποτελέσματα InceptionResNetV2

Class Type	Precision	Recall	F1-score	Support
Colon_aca	0.96	0.95	0.95	600
Colon_n	1	0.98	0.99	600
Lung_aca	0.88	0.88	0.88	600
Lung_n	1	0.99	0.99	600
Lung_Scc	0.90	0.92	0.91	600
Accuracy	-	-	0.945	3000
Macro_avg	0.95	0.95	0.95	3000
Weighted_avg	0.95	0.95	0.95	3000

Η αξιολόγηση του μοντέλου InceptionResNetV2 αναδεικνύει υψηλή, αλλά συγκριτικά χαμηλότερη απόδοση σε σχέση με το baseline CNN, με συνολική ακρίβεια 94.53% και τιμή συνάρτησης κόστους loss = 0.1412 όπως αναδικνύεται και από τον Πίνακα 6.2. Είναι σημαντικό να σημειωθεί ότι τα συγκεκριμένα αποτελέσματα αντιστοιχούν σε πρώτη εκτέλεση του μοντέλου, χωρίς εκτεταμένη διαδικασία βελτιστοποίησης ή συστηματική ρύθμιση υπερπαραμέτρων.

Η επισήμανση αυτή είναι κρίσιμη, καθώς το InceptionResNetV2 αποτελεί μια ιδιαίτερα βαθιά και σύνθετη αρχιτεκτονική, η οποία συνήθως απαιτεί προσεκτική προσαρμογή, επαρκή αριθμό εποχών εκπαίδευσης και κατάλληλη επιλογή learning rate, προκειμένου να αποδώσει στο μέγιστο των δυνατοτήτων της. Συνεπώς, η παρούσα απόδοση δεν αντικατοπτρίζει απαραίτητα το πλήρες δυναμικό του μοντέλου.

Η τιμή του loss είναι αισθητά υψηλότερη σε σχέση με το baseline CNN, γεγονός που υποδηλώνει ότι το μοντέλο δεν έχει ακόμη συγκλίνει πλήρως σε ένα βέλτιστο σημείο λύσης. Το εύρημα αυτό ενισχύει την υπόθεση ότι απαιτείται περαιτέρω βελτιστοποίηση.

Η λεπτομερής εξέταση των μετρικών precision, recall και F1-score, όπως παρουσιάζονται στον Πίνακα 6.2, αποκαλύπτει διαφοροποιήσεις μεταξύ των κλάσεων, οι οποίες είναι ενδεικτικές της συμπεριφοράς του μοντέλου σε αυτό το αρχικό στάδιο χωρίς βελτιστοποιήσεις.

Η κατηγορία colon_n εμφανίζει εξαιρετική απόδοση (precision = 1.00, recall = 0.98, F1-score = 0.99), όπως φαίνεται στον Πίνακα 6.2, γεγονός που υποδηλώνει ότι ακόμη και στην αρχική φάση εκπαίδευσης το μοντέλο μπορεί να αναγνωρίσει με μεγάλη ακρίβεια τα φυσιολογικά δείγματα του παχέος εντέρου. Τα ελάχιστα σφάλματα, κυρίως προς την κατηγορία colon_aca, θεωρούνται αναμενόμενα, καθώς σε ορισμένες περιπτώσεις η μετάβαση από φυσιολογικό σε καρκινικό ιστό συνοδεύεται από σταδιακές μορφολογικές μεταβολές. Η πολύ υψηλή τιμή precision δείχνει ότι το μοντέλο σπάνια συγχέει άλλες κατηγορίες με την colon_n, ενώ το ελαφρώς μειωμένο recall υποδηλώνει ότι ένα πολύ μικρό ποσοστό πραγματικών δειγμάτων δεν αναγνωρίζεται σωστά.

Αντίστοιχα, η κατηγορία lung_n παρουσιάζει σχεδόν τέλεια απόδοση (precision = 1.00, recall = 0.99, F1-score = 0.99), επιβεβαιώνοντας ότι οι φυσιολογικές κλάσεις αποτελούν ευκολότερο πρόβλημα ταξινόμησης. Η σχεδόν ιδανική επίδοση υποδεικνύει ότι τα χαρακτηριστικά της κατηγορίας αυτής καταλαμβάνουν μια σαφώς διακριτή περιοχή στον χώρο χαρακτηριστικών (feature space), με ελάχιστη επικάλυψη με άλλες κατηγορίες. Τα περιορισμένα σφάλματα προς άλλες κλάσεις επιβεβαιώνουν ότι το μοντέλο έχει μάθει αποτελεσματικά τα βασικά μορφολογικά πρότυπα των φυσιολογικών ιστών.

Η κατηγορία colon_aca παρουσιάζει επίσης υψηλή απόδοση, με F1-score = 0.95. Ωστόσο, η μικρή μείωση στο recall (0.95) σε σχέση με το precision (0.96) υποδηλώνει την ύπαρξη ψευδώς αρνητικών προβλέψεων. Από τον πίνακα σύγχυσης παρατηρείται ότι ένα μέρος των δειγμάτων ταξινομείται λανθασμένα κυρίως ως lung_aca, γεγονός ιδιαίτερα ενδιαφέρον, καθώς πρόκειται για κατηγορίες διαφορετικών οργάνων. Η συμπεριφορά αυτή υποδηλώνει ότι το μοντέλο, στο παρόν στάδιο εκπαίδευσης, δεν έχει ακόμη διαμορφώσει πλήρως διαχωριστικά χαρακτηριστικά υψηλού επιπέδου, ικανά να διαφοροποιούν επαρκώς τις ιστολογικές δομές μεταξύ παχέος εντέρου και πνεύμονα. Αντίθετα, φαίνεται να βασίζεται εν μέρει σε πιο γενικά μορφολογικά μοτίβα, όπως υφές ή κυτταρική πυκνότητα, τα οποία ενδέχεται να εμφανίζονται σε περισσότερες από μία κατηγορίες.

Η κατηγορία lung_aca αποτελεί την πιο απαιτητική περίπτωση για το μοντέλο, με F1-score = 0.88 και σχετικά χαμηλές τιμές precision και recall σύμφωνα με τον Πίνακα 6.2. Η ανάλυση του confusion matrix αποκαλύπτει σημαντικό αριθμό λανθασμένων ταξινομήσεων προς την κατηγορία lung_scc, γεγονός που υποδηλώνει ισχυρή επικάλυψη χαρακτηριστικών μεταξύ των δύο υποτύπων καρκίνου. Το εύρημα αυτό είναι ιδιαίτερα σημαντικό, καθώς δείχνει ότι το embedding space που έχει μάθει το μοντέλο δεν επιτυγχάνει ακόμη σαφή διαχωρισμό μεταξύ των δύο κατηγοριών. Πιθανές αιτίες περιλαμβάνουν τη μορφολογική ομοιότητα των ιστολογικών δομών, την υψηλή ενδοκλασική μεταβλητότητα ή την ανεπαρκή εκμάθηση λεπτών διαφορών. Η ύπαρξη τόσο ψευδώς θετικών όσο και ψευδώς αρνητικών προβλέψεων καθιστά τη συγκεκριμένη κατηγορία κρίσιμο σημείο βελτίωσης του μοντέλου.

Η κατηγορία lung_scc παρουσιάζει μέτρια προς υψηλή απόδοση (F1-score = 0.91), ωστόσο εμφανίζει επίσης σημαντική σύγχυση με την κατηγορία lung_aca. Συγκεκριμένα, ένας αριθμός δειγμάτων ταξινομείται λανθασμένα ως lung_aca, γεγονός που επιβεβαιώνει την ύπαρξη συμμετρικής σύγχυσης μεταξύ των δύο κατηγοριών. Η συμπεριφορά αυτή αποτελεί ένδειξη ότι το μοντέλο δεν έχει ακόμη μάθει τα λεπτά διακριτικά χαρακτηριστικά που διαφοροποιούν τους υποτύπους καρκίνου του πνεύμονα. Σε όρους representation learning, αυτό σημαίνει ότι οι δύο κατηγορίες καταλαμβάνουν γειτονικές περιοχές στον embedding space, με αποτέλεσμα την ύπαρξη ασαφών decision boundaries.

Οι τιμές των macro average και weighted average (0.95) υποδηλώνουν συνολικά σταθερή, αλλά όχι ακόμη βέλτιστη απόδοση. Το γεγονός ότι οι τιμές αυτές είναι χαμηλότερες από το baseline CNN δεν πρέπει να ερμηνευθεί ως μειονέκτημα της αρχιτεκτονικής, αλλά ως ένδειξη ότι το μοντέλο δεν έχει φτάσει ακόμη στο βέλτιστο σημείο εκπαίδευσης. Η συμπεριφορά αυτή είναι αναμενόμενη για πολύ βαθιά δίκτυα, ιδιαίτερα σε περιβάλλοντα Transfer Learning, όπου το fine-tuning και η σωστή ρύθμιση των υπερπαραμέτρων διαδραματίζουν καθοριστικό ρόλο.

Συνοψίζοντας, τα πρώτα αποτελέσματα του InceptionResNetV2 καταδεικνύουν ότι το μοντέλο διαθέτει ισχυρό δυναμικό, το οποίο όμως δεν έχει ακόμη αξιοποιηθεί πλήρως. Η μειωμένη απόδοση σε σχέση με το baseline CNN δεν αποτελεί ένδειξη ανωτερότητας του τελευταίου, αλλά αντανακλά τη διαφορετική φάση εκπαίδευσης και την αυξημένη πολυπλοκότητα της αρχιτεκτονικής.

Πίνακας 6.3: Τελικά αποτελέσματα InceptionResNetV2

Class Type	Precision	Recall	F1-score	Support
Colon_aca	1	1	1	600
Colon_n	1	1	1	600
Lung_aca	0.96	1	0.98	600
Lung_n	1	0.99	1	600
Lung_Scc	1	0.97	0.98	600
Accuracy	-	-	0.992	3000
Macro_avg	0.99	0.99	0.99	3000
Weighted_avg	0.99	0.99	0.99	3000

Στην επαναξιολόγηση του μοντέλου InceptionResNetV2 όπως φανερώνεται και από τον Πίνακα 6.3, μετά από διαδικασία fine-tuning και περαιτέρω βελτιστοποίησης, αναδεικνύει μια σημαντική αύξηση της απόδοσης, με συνολική ακρίβεια 99.20% και ιδιαίτερα χαμηλή τιμή συνάρτησης κόστους (loss = 0.0194). Η σημαντική αυτή βελτίωση επιβεβαιώνει ότι το μοντέλο, λόγω της αυξημένης αρχιτεκτονικής του πολυπλοκότητας, απαιτεί κατάλληλη προσαρμογή ώστε να αξιοποιήσει πλήρως τις δυνατότητές του. Σε αντίθεση με τα αρχικά αποτελέσματα, όπου παρατηρούνταν αδυναμίες κυρίως στις κατηγορίες καρκίνου του πνεύμονα, το βελτιστοποιημένο μοντέλο επιτυγχάνει πλέον εξαιρετικά υψηλές επιδόσεις σε όλες τις κλάσεις, υποδηλώνοντας σημαντική βελτίωση στη διαδικασία εκμάθησης και διαχωρισμού χαρακτηριστικών.

Από τη λεπτομερή ανάλυση των επιδόσεων ανά κλάση αποκαλύπτεται ότι το βελτιστοποιημένο μοντέλο InceptionResNetV2 έχει επιτύχει σχεδόν πλήρη διαχωρισμό των κατηγοριών στον χώρο χαρακτηριστικών, γεγονός που αποτυπώνεται στις εξαιρετικά υψηλές τιμές precision, recall και F1-score όπως φαίνεται στον Πίνακα 6.3. Παρά τη συνολική σχεδόν τέλεια απόδοση, η διερεύνηση σε επίπεδο κλάσεων παρέχει σημαντικές πληροφορίες σχετικά με τα ελάχιστα υπολειπόμενα σφάλματα και τη φύση της μάθησης του μοντέλου.

Στην κατηγορία colon_aca παρουσιάζεται απόλυτη απόδοση (precision = recall = F1-score = 1.00), γεγονός που υποδηλώνει ότι το μοντέλο έχει μάθει πλήρως τα χαρακτηριστικά των καρκινικών ιστών του παχέος εντέρου. Από πλευράς representation learning, αυτό σημαίνει ότι τα δείγματα της κατηγορίας αυτής σχηματίζουν μια σαφώς διακριτή και συμπαγή περιοχή στον embedding space, με ελάχιστη ή μηδενική επικάλυψη με άλλες κατηγορίες. Η πλήρης απουσία σφαλμάτων υποδηλώνει επίσης ότι το μοντέλο δεν βασίζεται πλέον σε επιφανειακά χαρακτηριστικά, αλλά έχει μάθει βαθύτερες και σημασιολογικά πλούσιες αναπαραστάσεις, οι οποίες επιτρέπουν επιτρέπει σαφή πειραματική διάκριση .

Η κατηγορία colon_n εμφανίζει επίσης τέλεια επίδοση (F1-score = 1.00), γεγονός που επιβεβαιώνει ότι τα φυσιολογικά δείγματα παχέος εντέρου διαθέτουν ισχυρά και συνεπή μορφολογικά χαρακτηριστικά. Το μοντέλο φαίνεται να έχει μάθει ένα σαφές decision boundary μεταξύ φυσιολογικού και καρκινικού ιστού στο ίδιο όργανο. Η τέλεια απόδοση τόσο στο precision όσο και στο recall φανερώνει ότι δεν υπάρχουν ούτε ψευδώς θετικές ούτε ψευδώς αρνητικές προβλέψεις, στοιχείο που αποτελεί ένδειξη υψηλής ενδοκλασικής συνοχής και επιτυχούς εκμάθησης των βασικών δομικών μοτίβων.

Η κατηγορία lung_aca παρουσιάζει εντυπωσιακή βελτίωση (F1-score = 0.98), ωστόσο εξακολουθεί να αποτελεί τη μοναδική κατηγορία με μικρή απόκλιση από την τέλεια επίδοση. Το recall = 1.00 υποδηλώνει ότι το μοντέλο εντοπίζει όλες τις πραγματικές περιπτώσεις, στο συγκεκριμένο test set δεν εμφανίζει ψευδώς αρνητικά για την κατηγορία αυτή . Η μικρή μείωση στο precision (0.96)

υποδηλώνει ότι ένα περιορισμένο πλήθος δειγμάτων άλλων κατηγοριών, κυρίως lung_scc, ταξινομούνται λανθασμένα ως lung_aca. Αυτό σημαίνει ότι, παρά τη σημαντική βελτίωση, εξακολουθεί να υπάρχει μια μικρή επικάλυψη μεταξύ των δύο κατηγοριών στον χώρο χαρακτηριστικών όπως γίνεται αντιληπτό από τον Πίνακα 6.3. Από άποψη feature learning, το αποτέλεσμα αυτό υποδηλώνει ότι το μοντέλο έχει μάθει τα βασικά διακριτικά χαρακτηριστικά της κατηγορίας, αλλά ορισμένα οριακά δείγματα εξακολουθούν να βρίσκονται κοντά στο decision boundary.

Η κατηγορία lung_n διατηρεί τέλεια απόδοση (F1-score = 1.00), γεγονός που επιβεβαιώνει ότι τα φυσιολογικά δείγματα πνεύμονα είναι πλήρως διαχωρίσιμα από τις καρκινικές κατηγορίες. Το μοντέλο έχει καταφέρει να δημιουργήσει ένα ισχυρό και σταθερό αναπαραστατικό πρότυπο για τη συγκεκριμένη κατηγορία. Η συμπεριφορά αυτή είναι ιδιαίτερα σημαντική, καθώς υποδηλώνει ότι τα φυσιολογικά δείγματα αποτελούν μια «εύκολη» κλάση, με χαμηλή επικάλυψη με άλλες κατηγορίες, γεγονός που επιτρέπει στο μοντέλο να επιτύχει σχεδόν τέλεια ταξινόμηση.

Η κατηγορία lung_scc παρουσιάζει επίσης σημαντική βελτίωση (F1-score = 0.98), με precision = 1.00 και recall = 0.97. Το γεγονός ότι το precision είναι τέλειο υποδηλώνει ότι το μοντέλο δεν συγχέει άλλες κατηγορίες με την lung_scc, ενώ το ελαφρώς μειωμένο recall δείχνει ότι ένα πολύ μικρό ποσοστό πραγματικών δειγμάτων ταξινομείται λανθασμένα ως lung_aca. Η ασυμμετρία αυτή (precision > recall) υποδηλώνει ότι το μοντέλο έχει μετατοπίσει το decision boundary με τρόπο που ελαχιστοποιεί τα ψευδώς θετικά, εις βάρος ενός εξαιρετικά μικρού αριθμού ψευδώς αρνητικών.

Από τη συνολική εικόνα του μοντέλου προκύπτει ότι επιτυγχάνεται πλήρης διαχωρισμός των κατηγοριών του παχέος εντέρου, σχεδόν τέλεια απόδοση στις φυσιολογικές κατηγορίες και ουσιαστική επίλυση της σύγχυσης μεταξύ των κατηγοριών καρκίνου του πνεύμονα, η οποία αποτελούσε το βασικό πρόβλημα στην αρχική εκπαίδευση. Η δραματική μείωση των σφαλμάτων μεταξύ lung_aca και lung_scc αποτελεί ένδειξη ότι το μοντέλο, μετά το fine-tuning, έχει μάθει λεπτομερή και διακριτικά χαρακτηριστικά, τα οποία επιτρέπουν σαφέστερο διαχωρισμό στον embedding space.

Οι τιμές των macro average και weighted average (0.99 σε όλες τις μετρικές) υποδηλώνουν σχεδόν ιδανική συνολική απόδοση όπως φαίνεται στον Πίνακα 6.3. Η ταύτιση των δύο μετρικών επιβεβαιώνει την ισορροπημένη φύση του dataset, ενώ η υψηλή τιμή τους υποδηλώνει ότι το μοντέλο αποδίδει ομοιόμορφα σε όλες τις κατηγορίες. Η δραματική βελτίωση σε σχέση με τα αρχικά αποτελέσματα καταδεικνύει ότι το μοντέλο, όταν εκπαιδευτεί επαρκώς, μπορεί να εκμεταλλευτεί πλήρως τη βαθιά αρχιτεκτονική του και να μάθει εξαιρετικά διακριτικές αναπαραστάσεις.

Η υψηλή απόδοση στο test set, σε συνδυασμό με τον σαφή διαχωρισμό training και test δεδομένων, παρουσιάζει ενδείξεις ικανότητας γενίκευσης στο συγκεκριμένο test set. Η χαμηλή τιμή του loss και η υψηλή ακρίβεια αποτελούν ενδείξεις ότι το μοντέλο έχει συγκλίνει σε μια σταθερή και αποδοτική λύση.

Συνοψίζοντας, το βελτιστοποιημένο μοντέλο InceptionResNetV2 παρουσιάζει εξαιρετική απόδοση, ξεπερνώντας τόσο τα αρχικά του αποτελέσματα όσο και το baseline CNN. Η εξέλιξη αυτή αναδεικνύει τη σημασία της σωστής εκπαίδευσης και προσαρμογής των βαθιών αρχιτεκτονικών, καθώς και τη δυναμική που διαθέτουν όταν αξιοποιηθούν πλήρως. Το μοντέλο καταφέρνει πλέον να μειώνει σημαντικά τη σύγχυση στις δυσκολότερες περιπτώσεις ταξινόμησης, όπως είναι η διάκριση μεταξύ υποτύπων καρκίνου του πνεύμονα, γεγονός που το αναδεικνύει ως υποσχόμενη προσέγγιση για ερευνητικές εφαρμογές ιατρικής απεικόνισης. Τα αποτελέσματα αυτά καταδεικνύουν ότι η αρχική υστέρηση του InceptionResNetV2 δεν οφειλόταν σε περιορισμούς της αρχιτεκτονικής, αλλά στη μη βελτιστοποιημένη αρχική εκπαίδευση.

6.3.2 Αποτελέσματα VGG16

Πίνακας 6.4: Πρώτα αποτελέσματα VGG16

Class Type	Precision	Recall	F1-score	Support
Colon_aca	0.96	0.96	0.96	600
Colon_n	0.99	0.98	0.99	600
Lung_aca	0.92	0.88	0.9	600
Lung_n	0.99	0.99	0.99	600
Lung_Scc	0.89	0.94	0.92	600
Accuracy	-	-	0.9503	3000
Macro_avg	0.95	0.95	0.95	3000
Weighted_avg	0.95	0.95	0.95	3000

Η αξιολόγηση του μοντέλου VGG16 στα αρχικά στάδια εκπαίδευσης καταδεικνύει ικανοποιητική αλλά όχι βέλτιστη απόδοση, επιτυγχάνοντας συνολική ακρίβεια 95.03% και τιμή συνάρτησης κόστους loss = 0.1466. Είναι σημαντικό να επισημανθεί ότι τα αποτελέσματα αυτά αντιστοιχούν σε πρώτη εκτέλεση του μοντέλου, χωρίς εκτεταμένη διαδικασία fine-tuning ή συστηματικής ρύθμισης υπερπαραμέτρων. Το VGG16, ως βαθύ συνελικτικό δίκτυο με απλή και ομοιογενή αρχιτεκτονική, παρουσιάζει σταθερή αλλά λιγότερο εκφραστική ικανότητα σε σχέση με πιο σύγχρονες αρχιτεκτονικές, όπως το InceptionResNetV2. Ωστόσο, η σχετικά υψηλή ακρίβεια δείχνει ότι το μοντέλο μπορεί ήδη από τα πρώτα στάδια να μάθει σημαντικά μορφολογικά χαρακτηριστικά των δεδομένων. Η υψηλότερη τιμή του loss, σε σύγκριση με πιο βελτιστοποιημένα μοντέλα, υποδηλώνει ότι η εκπαίδευση δεν έχει ακόμη συγκλίνει πλήρως, γεγονός που αφήνει περιθώρια σημαντικής βελτίωσης.

Η κατηγορία colon_aca παρουσιάζει ισορροπημένη απόδοση (precision = recall = F1-score = 0.96), γεγονός που υποδηλώνει ότι το μοντέλο έχει μάθει ικανοποιητικά τα βασικά χαρακτηριστικά των καρκινικών ιστών του παχέος εντέρου. Η συμμετρία μεταξύ precision και recall δείχνει ότι το decision boundary που έχει διαμορφώσει το μοντέλο είναι σχετικά σταθερό, χωρίς έντονη μεροληψία. Ωστόσο, η μη τέλεια απόδοση υποδηλώνει ότι ορισμένα δείγματα βρίσκονται κοντά στα όρια απόφασης, οδηγώντας σε αμφίσημες προβλέψεις. Αυτό αποτελεί ένδειξη ότι το μοντέλο δεν έχει ακόμη καταφέρει να μάθει πλήρως υψηλού επιπέδου διακριτικά χαρακτηριστικά, αλλά βασίζεται εν μέρει σε πιο γενικευμένα μορφολογικά μοτίβα.

Η κατηγορία colon_n παρουσιάζει εξαιρετική απόδοση (F1-score = 0.99), γεγονός που υποδηλώνει ότι τα φυσιολογικά δείγματα παχέος εντέρου σχηματίζουν μια συμπαγή και καλά οριοθετημένη κατανομή στον χώρο χαρακτηριστικών όπως φαίνεται από τον Πίνακα 6.4. Τα ελάχιστα σφάλματα προς την κατηγορία colon_aca υποδηλώνουν ότι το μοντέλο αντιλαμβάνεται τη σχέση μεταξύ φυσιολογικού και καρκινικού ιστού ως ένα συνεχές φάσμα, γεγονός που είναι σύμφωνο με τη βιολογική πραγματικότητα. Η υψηλή τιμή precision υποδεικνύει ότι το μοντέλο αποφεύγει αποτελεσματικά την εσφαλμένη ταξινόμηση άλλων κατηγοριών ως colon_n.

Η κατηγορία lung_aca αποτελεί το πιο αδύναμο σημείο του μοντέλου, με F1-score = 0.90 και σημαντική μείωση στο recall (0.88). Το αποτέλεσμα αυτό υποδηλώνει ότι το μοντέλο αποτυγχάνει να εντοπίσει ένα ποσοστό πραγματικών θετικών περιπτώσεων, οδηγώντας σε αυξημένο αριθμό ψευδώς αρνητικών προβλέψεων. Η συμπεριφορά αυτή είναι ιδιαίτερα σημαντική, καθώς υποδηλώνει ότι το embedding space δεν παρέχει επαρκή διαχωρισιμότητα για τη συγκεκριμένη κατηγορία. Πιο συγκεκριμένα, τα δείγματα της lung_aca εμφανίζουν επικάλυψη με την lung_scc, ενώ το decision

boundary φαίνεται να είναι ασαφές ή μη επαρκώς προσαρμοσμένο στη λεπτή διαφοροποίηση των δύο κατηγοριών. Συνεπώς, το μοντέλο δεν έχει ακόμη μάθει τα λεπτά ιστολογικά χαρακτηριστικά που διαφοροποιούν τους δύο υποτύπους καρκίνου.

Η κατηγορία lung_n εμφανίζει εξαιρετική απόδοση (F1-score = 0.99), γεγονός που υποδηλώνει ότι τα φυσιολογικά δείγματα πνεύμονα αποτελούν σαφώς διακριτή κατηγορία. Το μοντέλο έχει καταφέρει να μάθει ισχυρά και γενικεύσιμα χαρακτηριστικά, τα οποία οδηγούν σε σταθερή και αξιόπιστη ταξινόμηση. Η σχεδόν πλήρης απουσία σφαλμάτων υποδεικνύει χαμηλή επικάλυψη με άλλες κατηγορίες και υψηλή ενδοκλασική συνοχή.

Η κατηγορία lung_scc παρουσιάζει F1-score = 0.92, με υψηλότερο recall (0.94) σε σχέση με το precision (0.89). Η ασυμμετρία αυτή είναι ιδιαίτερα σημαντική, καθώς υποδηλώνει ότι το μοντέλο εντοπίζει τις περισσότερες πραγματικές περιπτώσεις, αλλά παράγει σημαντικό αριθμό ψευδώς θετικών προβλέψεων. Η κύρια πηγή σφαλμάτων εντοπίζεται στη σύγχυση με την κατηγορία lung_aca, γεγονός που επιβεβαιώνει ότι οι δύο κατηγορίες καταλαμβάνουν γειτονικές περιοχές στον χώρο χαρακτηριστικών. Σε όρους μηχανικής μάθησης, αυτό σημαίνει ότι το μοντέλο δεν έχει ακόμη καταφέρει να δημιουργήσει επαρκώς διακριτά decision boundaries.

Από πλευράς γενίκευσης, το μοντέλο παρουσιάζει ικανοποιητική απόδοση σε μη παρατηρημένα δεδομένα, δεδομένου του σαφούς διαχωρισμού μεταξύ training και test συνόλων. Η επίδοση της τάξης του 95.03% υποδηλώνει ότι το μοντέλο έχει μάθει ουσιαστικά χαρακτηριστικά, χωρίς να βασίζεται σε απλή απομνημόνευση. Ωστόσο, η ύπαρξη σφαλμάτων σε κρίσιμες κατηγορίες υποδηλώνει ότι η ικανότητα γενίκευσης δεν είναι ακόμη βέλτιστη.

Συνολικά, το VGG16 φαίνεται ότι διαχωρίζει αποτελεσματικά τις φυσιολογικές κατηγορίες (colon_n, lung_n), επιτυγχάνει καλή απόδοση στις κατηγορίες παχέος εντέρου, αλλά παρουσιάζει σαφή αδυναμία στις κατηγορίες καρκίνου πνεύμονα, ιδιαίτερα στη μεταξύ τους διάκριση. Η συμπεριφορά αυτή είναι χαρακτηριστική των μοντέλων με περιορισμένη εκφραστική ικανότητα σε σχέση με πιο σύγχρονες αρχιτεκτονικές. Το VGG16 φαίνεται να μαθαίνει κυρίως γενικά χωρικά χαρακτηριστικά (edges, textures), χωρίς να καταφέρνει πλήρως να συλλάβει τις λεπτές ιστολογικές διαφοροποιήσεις.

Συνοψίζοντας, τα πρώτα αποτελέσματα του VGG16 καταδεικνύουν ότι το μοντέλο αποτελεί μια αξιόπιστη αλλά όχι κορυφαία επιλογή για το συγκεκριμένο πρόβλημα. Η σχετικά απλή αρχιτεκτονική του προσφέρει σταθερή απόδοση, αλλά δεν επαρκεί για την πλήρη διάκριση σύνθετων και μορφολογικά παρόμοιων κατηγοριών. Η περαιτέρω βελτιστοποίηση του μοντέλου μέσω fine-tuning οδηγήγει σε βελτίωση της απόδοσης, ιδιαίτερα στις δύσκολες κατηγορίες όπως γίνεται αντιληπτό μέσω του Πίνακα 6.5 που παρατίθεται.

Πίνακας 6.5: Τελικά αποτελέσματα VGG16

Class Type	Precision	Recall	F1-score	Support
Colon_aca	0.97	0.99	0.98	600
Colon_n	0.99	0.99	0.99	600
Lung_aca	0.95	0.89	0.92	600
Lung_n	1	1	1	600
Lung_Scc	0.91	0.95	0.93	600
Accuracy	-	-	0.965	3000
Macro_avg	0.96	0.96	0.96	3000
Weighted_avg	0.96	0.96	0.96	3000

Η επαναξιολόγηση του μοντέλου VGG16 μετά από περαιτέρω εκπαίδευση και βελτιστοποίηση αναδεικνύει σαφή βελτίωση της απόδοσής του, με συνολική ακρίβεια 96.47% και μειωμένη τιμή συνάρτησης κόστους (loss = 0.1124). Η βελτίωση αυτή επιβεβαιώνει ότι το μοντέλο ανταποκρίνεται θετικά στη διαδικασία εκπαίδευσης, ενισχύοντας την ικανότητά του να μαθαίνει πιο διακριτικά χαρακτηριστικά. Σε σύγκριση με τα αρχικά αποτελέσματα (~95%), παρατηρείται αύξηση της ακρίβειας κατά περίπου 1.5%, γεγονός που υποδηλώνει καλύτερη σύγκλιση και αποτελεσματικότερη εκμάθηση των δεδομένων. Παρά τη βελτίωση αυτή, το μοντέλο εξακολουθεί να υπολείπεται έναντι πιο σύνθετων αρχιτεκτονικών, γεγονός που συνδέεται με τους εγγενείς περιορισμούς της αρχιτεκτονικής του VGG16.

Η περαιτέρω ανάλυση των επιδόσεων ανά κλάση αποκαλύπτει ότι το VGG16, μετά τη βελτιστοποίηση, παρουσιάζει σημαντική βελτίωση στη δομή του χώρου χαρακτηριστικών, χωρίς ωστόσο να επιτυγχάνει πλήρη διαχωρισιμότητα μεταξύ όλων των κατηγοριών. Η μελέτη των precision–recall trade-offs και των υπολειπόμενων σφαλμάτων επιτρέπει μια πιο λεπτομερή κατανόηση της συμπεριφοράς του μοντέλου.

Η κατηγορία colon_aca (precision = 0.97, recall = 0.99, F1 = 0.98) παρουσιάζει σαφή βελτίωση κυρίως ως προς το recall όπως φαίνεται στον Πίνακα 6.5, γεγονός που υποδηλώνει σημαντική μείωση των false negatives. Αυτό σημαίνει ότι το μοντέλο έχει μετατοπίσει το decision boundary προς την κατεύθυνση της αυξημένης ευαισθησίας, δίνοντας προτεραιότητα στην ανίχνευση των καρκινικών περιπτώσεων. Η μικρή πτώση στο precision υποδηλώνει ότι αυτό επιτυγχάνεται εις βάρος ενός περιορισμένου αριθμού false positives, γεγονός που αποτελεί συνήθης συμπεριφορά σε πειραματικά προβλήματα ιατρικής ταξινόμησης , όπου η μείωση των ψευδώς αρνητικών έχει ιδιαίτερη σημασία σε τέτοιου τύπου προβλήματα.

Η κατηγορία colon_n (F1 = 0.99) συνεχίζει να παρουσιάζει εξαιρετική απόδοση, γεγονός που υποδηλώνει πολύ υψηλή ενδοκλασική συνοχή. Τα δείγματα της κατηγορίας αυτής φαίνεται να συγκεντρώνονται σε μια συμπαγή περιοχή στον embedding space, με σαφή απόσταση από άλλες κατηγορίες. Η μικρή απώλεια στο recall (0.98) υποδηλώνει την ύπαρξη οριακών περιπτώσεων που βρίσκονται κοντά στο decision boundary με την colon_aca, γεγονός που είναι σύμφωνο με τη βιολογική συνέχεια μεταξύ φυσιολογικού και παθολογικού ιστού.

Από τον Πίνακα 6.5 γίνεται αντιληπτό ότι, η κατηγορία lung_aca (precision = 0.95, recall = 0.89, F1 = 0.92) παραμένει η πιο προβληματική κατηγορία, παρά τη συνολική βελτίωση. Το χαμηλότερο recall υποδηλώνει ότι το μοντέλο συνεχίζει να παράγει false negatives, δηλαδή να αποτυγχάνει στην ανίχνευση ορισμένων καρκινικών περιπτώσεων. Το εύρημα αυτό είναι ιδιαίτερα σημαντικό, καθώς δείχνει ότι τα χαρακτηριστικά της κατηγορίας δεν έχουν αποτυπωθεί πλήρως και ότι ορισμένα δείγματα εξακολουθούν να βρίσκονται σε περιοχές υψηλής αβεβαιότητας. Παράλληλα, υποδηλώνει ότι το decision boundary μεταξύ των κατηγοριών lung_aca και lung_scc δεν είναι ακόμη επαρκώς διαμορφωμένο, με αποτέλεσμα τα embeddings των δύο κατηγοριών να παραμένουν γειτονικά στον χώρο χαρακτηριστικών.

Η κατηγορία lung_n (F1 = 1.00) παρουσιάζει πλήρη διαχωρισμό στο συγκεκριμένο test set. Αυτό σημαίνει ότι:

- το μοντέλο έχει μάθει πλήρως επαρκώς τα βασικά μορφολογικά χαρακτηριστικά,
- τα δείγματα σχηματίζουν ένα σαφές και απομονωμένο cluster,
- και τα decision boundaries είναι πλήρως καθορισμένα.

Η συμπεριφορά αυτή επιβεβαιώνει ότι οι φυσιολογικές κατηγορίες αποτελούν χαμηλής πολυπλοκότητας πρόβλημα για το μοντέλο.

Η κατηγορία lung_scc παρουσιάζει βελτίωση (F1-score = 0.93), με recall = 0.95 και precision = 0.91. Η αύξηση του recall υποδηλώνει καλύτερη ανίχνευση των πραγματικών περιπτώσεων, ενώ το χαμηλότερο precision υποδηλώνει ότι εξακολουθούν να υπάρχουν ψευδώς θετικά, κυρίως προς την κατηγορία lung_aca. Τα αποτελέσματα αυτά υποδηλώνουν ότι το μοντέλο εξακολουθεί να αντιμετωπίζει δυσκολίες στη σαφή διάκριση μεταξύ των δύο υποτύπων καρκίνου πνεύμονα.

Η συνολική εικόνα από τον Πίνακα 6.5 δείχνει ότι το μοντέλο VGG16, μετά τη βελτιστοποίηση, παρουσιάζει σημαντική βελτίωση της απόδοσης στις κατηγορίες καρκίνου πνεύμονα, διατηρώντας παράλληλα εξαιρετική συμπεριφορά στις φυσιολογικές κατηγορίες και επιτυγχάνοντας σχεδόν πλήρη διαχωρισμό στις κατηγορίες παχέος εντέρου. Ωστόσο, η σύγχυση μεταξύ lung_aca και lung_scc δεν έχει εξαλειφθεί πλήρως, γεγονός που υποδηλώνει ότι το μοντέλο δεν έχει ακόμη αναπτύξει πλήρως ικανότητα λεπτομερούς μορφολογικής διαφοροποίησης. Η βελτιωμένη απόδοση στο test set, σε συνδυασμό με τη μείωση του loss, υποδεικνύει ενίσχυση της ικανότητας γενίκευσης. Το μοντέλο δεν προκύπτουν ισχυρές ενδείξεις απομνημόνευσης από τα διαθέσιμα αποτελέσματα, αλλά φαίνεται να μαθαίνει πιο σταθερά και γενικεύσιμα χαρακτηριστικά.

Συνοψίζοντας, τα τελικά αποτελέσματα του VGG16 επιβεβαιώνουν ότι το μοντέλο μπορεί να βελτιωθεί σημαντικά μέσω περαιτέρω εκπαίδευσης και fine-tuning. Παρότι παρουσιάζει αξιόλογη απόδοση, εξακολουθεί να υπολείπεται σε σχέση με πιο προηγμένες αρχιτεκτονικές, κυρίως λόγω της περιορισμένης ικανότητάς του να διακρίνει λεπτές μορφολογικές διαφοροποιήσεις μεταξύ παρόμοιων κατηγοριών. Η εξέλιξη αυτή αναδεικνύει τη σημασία της σωστής εκπαίδευσης και προσαρμογής του μοντέλου, ενώ παράλληλα επιβεβαιώνει ότι η αρχιτεκτονική του VGG16, αν και σταθερή ως προς την πειραματική της συμπεριφορά, δεν αποτελεί την πλέον κατάλληλη επιλογή για προβλήματα υψηλής πολυπλοκότητας χωρίς επιπλέον μηχανισμούς βελτίωσης ή βαθύτερες αρχιτεκτονικές.

6.3.3 Αποτελέσματα VGG19

Η αξιολόγηση του μοντέλου VGG19 καταδεικνύει ιδιαίτερα υψηλή απόδοση στο πρόβλημα ταξινόμησης, επιτυγχάνοντας συνολική ακρίβεια 98.63% και χαμηλή τιμή συνάρτησης κόστους (loss = 0.0666). Τα αποτελέσματα αυτά υποδεικνύουν ότι το μοντέλο έχει επιτύχει υψηλό επίπεδο σύγκλισης και είναι σε θέση να εξάγει πλούσιες και διακριτικές αναπαραστάσεις από τα δεδομένα. Σε σύγκριση με το VGG16, παρατηρείται σημαντική βελτίωση της απόδοσης, γεγονός που αποδίδεται στο μεγαλύτερο βάθος του δικτύου (19 layers), το οποίο επιτρέπει την εκμάθηση πιο σύνθετων και ιεραρχικών χαρακτηριστικών. Το VGG19 λειτουργεί ως ενδιάμεση προσέγγιση μεταξύ απλούστερων CNN αρχιτεκτονικών και πιο σύγχρονων βαθιών μοντέλων.

Πίνακας 6.6: Αποτελέσματα VGG19

Class Type	Precision	Recall	F1-score	Support
Colon_aca	1	1	1	600
Colon_n	1	1	1	600
Lung_aca	1	0.93	0.96	600
Lung_n	1	1	1	600
Lung_Scc	0.94	1	0.97	600
Accuracy	-	-	0.986	3000
Macro_avg	0.99	0.99	0.99	3000
Weighted_avg	0.99	0.99	0.99	3000

Η λεπτομερής ανάλυση των αποτελεσμάτων του μοντέλου VGG19 σε επίπεδο κλάσεων, σύμφωνα με τον Πίνακα 6.6, αποκαλύπτει ότι το δίκτυο έχει επιτύχει έναν σχεδόν πλήρη και σταθερό

διαχωρισμό του χώρου χαρακτηριστικών στο συγκεκριμένο test set, με εξαιρετικά περιορισμένες περιοχές επικάλυψης μεταξύ των κατηγοριών. Η συνολική εικόνα υποδηλώνει ότι το μοντέλο έχει καταφέρει να οργανώσει τις αναπαραστάσεις του σε ένα ιδιαίτερα καλά δομημένο embedding space, όπου οι περισσότερες κλάσεις σχηματίζουν συμπαγή και σαφώς διακριτά clusters.

Η κατηγορία colon_aca παρουσιάζει ιδανική συμπεριφορά, με τιμές precision, recall και F1-score ίσες με 1.00, γεγονός που υποδηλώνει πλήρη απουσία τόσο ψευδώς θετικών όσο και ψευδώς αρνητικών προβλέψεων στο συγκεκριμένο test set . Το αποτέλεσμα αυτό φανερώνει ότι το μοντέλο φαίνεται να έχει αποτυπώσει επαρκώς βασικά διακριτικά μορφολογικά χαρακτηριστικά της συγκεκριμένης κατηγορίας, οδηγώντας στη δημιουργία ενός πλήρως απομονωμένου cluster στον χώρο χαρακτηριστικών. Το decision boundary που αντιστοιχεί στην κατηγορία αυτή είναι ιδιαίτερα σταθερό και χαρακτηρίζεται από μεγάλο περιθώριο διαχωρισμού σε σχέση με τις υπόλοιπες κατηγορίες.

Αντίστοιχη συμπεριφορά παρατηρείται και στην κατηγορία colon_n, όπου η επίδοση είναι επίσης τέλεια. Η πλήρης απουσία σφαλμάτων υποδηλώνει υψηλή ενδοκλασική συνοχή και πολύ μικρή διασπορά των χαρακτηριστικών εντός της κατηγορίας. Το γεγονός αυτό επιβεβαιώνει ότι τα φυσιολογικά δείγματα παχέος εντέρου διαθέτουν μορφολογικά χαρακτηριστικά τα οποία είναι σαφώς διακριτά από τις παθολογικές καταστάσεις, επιτρέποντας στο μοντέλο να επιτύχει πλήρη διαχωρισμό.

Η κατηγορία lung_n παρουσιάζει επίσης ιδανική απόδοση, με μηδενικά σφάλματα ταξινόμησης, όπως φαίνεται και από τις επιμέρους τιμές του Πίνακα 6.6. Το αποτέλεσμα αυτό υποδηλώνει ότι το μοντέλο έχει καταφέρει να μάθει ισχυρές αναπαραστάσεις για τα δεδομένα του συγκεκριμένου πειράματος για τα φυσιολογικά δείγματα πνεύμονα, τα οποία σχηματίζουν ένα απομονωμένο και συμπαγές cluster στον χώρο χαρακτηριστικών. Η πλήρης διαχωρισιμότητα της κατηγορίας αυτής αποτελεί ένδειξη ότι τα φυσιολογικά πρότυπα είναι λιγότερο πολύπλοκα και παρουσιάζουν μικρότερη ενδοκλασική ποικιλία.

Η κατηγορία lung_aca, αν και εμφανίζει πολύ υψηλή συνολική επίδοση, παρουσιάζει μικρή αλλά ουσιαστική μείωση στην τιμή του recall (0.93). Το αποτέλεσμα αυτό υποδηλώνει την ύπαρξη ψευδώς αρνητικών προβλέψεων, δηλαδή περιπτώσεων όπου το μοντέλο αποτυγχάνει να αναγνωρίσει σωστά καρκινικά δείγματα. Η ταυτόχρονη ύπαρξη τέλειου precision φανερώνει ότι το μοντέλο είναι ιδιαίτερα αυστηρό στην απόδοση της συγκεκριμένης ετικέτας, γεγονός που μεταφράζεται σε ένα ιδιαίτερα «σφιχτό» decision boundary γύρω από το cluster της κατηγορίας. Ως αποτέλεσμα, ορισμένα δείγματα που βρίσκονται στα όρια του cluster δεν ταξινομούνται σωστά, γεγονός που αποκαλύπτει περιορισμούς ως προς την πλήρη κάλυψη της ενδοκλασικής ποικιλότητας.

Η κατηγορία lung_scc παρουσιάζει διαφορετικό πρότυπο συμπεριφοράς, με recall ίσο με 1.00 και precision μειωμένο στο 0.94. Το μοτίβο αυτό υποδηλώνει ότι το μοντέλο καταφέρνει να εντοπίσει όλα τα πραγματικά θετικά δείγματα, ωστόσο παράγει αυξημένο αριθμό ψευδώς θετικών προβλέψεων. Η συμπεριφορά αυτή είναι ενδεικτική ενός decision boundary με αυξημένη περιοχή κάλυψης προς γειτονικές περιοχές του χώρου χαρακτηριστικών, οδηγώντας σε λανθασμένη ταξινόμηση δειγμάτων άλλων κατηγοριών ως lung_scc. Η κύρια πηγή σύγχυσης φαίνεται να εντοπίζεται στην κατηγορία lung_aca, γεγονός που υποδηλώνει ότι τα δύο clusters δεν είναι πλήρως διαχωρισμένα και παρουσιάζουν περιοχές μερικής επικάλυψης.

Συνολικά, η γεωμετρία του χώρου χαρακτηριστικών που προκύπτει από το VGG19 χαρακτηρίζεται από σαφώς διαχωρισμένα clusters στο πλαίσιο της παρούσας αξιολόγησης για τις κατηγορίες παχέος εντέρου και τα φυσιολογικά δείγματα, ενώ οι κατηγορίες καρκίνου πνεύμονα εμφανίζουν περιορισμένη αλλά υπαρκτή επικάλυψη. Η συμπεριφορά αυτή υποδηλώνει ότι, παρά το αυξημένο βάθος και την ενισχυμένη εκφραστική ικανότητα του μοντέλου, η διάκριση μεταξύ μορφολογικά παρόμοιων κατηγοριών εξακολουθεί να αποτελεί ένα εγγενώς δύσκολο πρόβλημα.

Συμπερασματικά, η ανάλυση σε επίπεδο κλάσεων δείχνει ότι το VGG19 επιτυγχάνει σχεδόν πλήρη διαχωρισμό των δεδομένων, με τις εναπομείνασες ασάφειες να περιορίζονται σε συγκεκριμένες

περιοχές υψηλής μορφολογικής ομοιότητας. Το γεγονός αυτό υποδηλώνει ότι το μοντέλο φαίνεται να πλησιάζει ένα υψηλό επίπεδο απόδοσης για τη συγκεκριμένη αρχιτεκτονική ως προς την εκμάθηση χαρακτηριστικών, όπου η περαιτέρω βελτίωση πιθανόν να απαιτεί είτε πιο εξελιγμένες αρχιτεκτονικές είτε διαφορετικές προσεγγίσεις αναπαράστασης της πληροφορίας.

Συνοψίζοντας, το μοντέλο VGG19 παρουσιάζει εξαιρετική απόδοση, πλησιάζοντας τα επίπεδα των πιο σύγχρονων αρχιτεκτονικών. Η αύξηση του βάθους σε σχέση με το VGG16 οδηγεί σε σημαντική βελτίωση της ικανότητας εκμάθησης και διαχωρισμού χαρακτηριστικών. Παρότι δεν επιτυγχάνει απόλυτη τελειότητα σε όλες τις κατηγορίες, όπως προκύπτει και από τον Πίνακα 6.6, η συνολική του απόδοση το καθιστά ιδιαίτερα αποδοτική ερευνητική προσέγγιση για το συγκεκριμένο πρόβλημα, με τις εναπομείνασες δυσκολίες να εντοπίζονται αποκλειστικά σε μορφολογικά παρόμοιες κατηγορίες.

6.3.3 Αποτελέσματα ResNet50

Στον Πίνακα 6.7 παρουσιάζονται οι βασικές μετρικές αξιολόγησης του ResNet50 ανά κατηγορία. Οι τιμές Precision, Recall και F1-score δίνουν συνοπτική εικόνα της απόδοσης του μοντέλου και της ικανότητάς του να διαχωρίζει σωστά τις επιμέρους ιστοπαθολογικές κλάσεις.

Πίνακας 6.7: Αποτελέσματα ResNet50

Class Type	Precision	Recall	F1-score	Support
Colon_aca	1	1	1	600
Colon_n	1	1	1	600
Lung_aca	0.97	0.97	0.97	600
Lung_n	0.99	1	0.99	600
Lung_Scc	0.99	0.97	0.98	600
Accuracy	-	-	0.987	3000
Macro_avg	0.99	0.99	0.99	3000
Weighted_avg	0.99	0.99	0.99	3000

Η αξιολόγηση του μοντέλου ResNet50 καταδεικνύει εξαιρετική απόδοση εξαιρετική απόδοση στο συγκεκριμένο πειραματικό περιβάλλον, με συνολική ακρίβεια στο σύνολο ελέγχου ίση με 98.73% και χαμηλή τιμή συνάρτησης κόστους (loss = 0.0346). Τα αποτελέσματα αυτά υποδηλώνουν ότι το μοντέλο έχει συγκλίνει σε μια ιδιαίτερα αποτελεσματική λύση και έχει καταφέρει να μάθει διακριτικές και διακριτικές αναπαραστάσεις των δεδομένων των ιστοπαθολογικών εικόνων. Η χαμηλή τιμή του loss, σε συνδυασμό με την πολύ υψηλή ακρίβεια, αποτελεί ένδειξη σταθερής εκπαίδευσης και ενδείξεις ικανότητας γενίκευσης στο συγκεκριμένο σύνολο ελέγχου.

Η υψηλή επίδοση του ResNet50 μπορεί να ερμηνευθεί και με βάση τη δομή της αρχιτεκτονικής του. Οι residual συνδέσεις επιτρέπουν την αποτελεσματική ροή της πληροφορίας μέσα από βαθύτερα επίπεδα του δικτύου, περιορίζοντας προβλήματα όπως η υποβάθμιση της απόδοσης σε πολύ βαθιά μοντέλα. Με τον τρόπο αυτό, το δίκτυο είναι σε θέση να μαθαίνει όχι μόνο απλά τοπικά μοτίβα, αλλά και πιο σύνθετες ιεραρχικές αναπαραστάσεις, γεγονός που είναι ιδιαίτερα σημαντικό στην ιστοπαθολογική ανάλυση, όπου οι μορφολογικές διαφορές μεταξύ των κλάσεων μπορεί να είναι λεπτές και πολυπαραγοντικές.

Η λεπτομερής εξέταση των μετρικών precision, recall και F1-score, σύμφωνα με τον Πίνακα 6.7, δείχνει ότι το μοντέλο παρουσιάζει εξαιρετική και σε μεγάλο βαθμό ομοιόμορφη συμπεριφορά στις περισσότερες κατηγορίες, με περιορισμένες αποκλίσεις μόνο σε ορισμένες περιπτώσεις.

Η κατηγορία colon_aca παρουσιάζει τέλεια επίδοση στο συγκεκριμένο test set, με precision, recall και F1-score ίσα με 1.00. Το αποτέλεσμα αυτό υποδηλώνει ότι το μοντέλο αναγνωρίζει όλα τα δείγματα της κατηγορίας χωρίς να παράγει ούτε ψευδώς θετικά ούτε ψευδώς αρνητικά. Από πλευράς αναπαράστασης, αυτό σημαίνει ότι το ResNet50 έχει καταφέρει να οργανώσει τα δείγματα της συγκεκριμένης κλάσης σε ένα σαφώς οριοθετημένο cluster στον χώρο χαρακτηριστικών, το οποίο είναι πλήρως διαχωρισμένο στο πλαίσιο της παρούσας αξιολόγησης. Η επίδοση αυτή υποδηλώνει ότι τα καρκινικά δείγματα παχέος εντέρου διαθέτουν μορφολογικά χαρακτηριστικά τα οποία το μοντέλο φαίνεται να έχει αποτυπώσει αποτελεσματικά .

Αντίστοιχα, η κατηγορία colon_n εμφανίζει επίσης τέλεια επίδοση στο συγκεκριμένο test set», με όλες τις μετρικές ίσες με 1.00. Η πλήρης απουσία σφαλμάτων υποδηλώνει ότι τα φυσιολογικά δείγματα παχέος εντέρου παρουσιάζουν υψηλή ενδοκλασική συνοχή και ότι το μοντέλο έχει μάθει ένα πολύ καθαρό decision boundary μεταξύ φυσιολογικού και παθολογικού ιστού στο συγκεκριμένο όργανο. Η τέλεια ταξινόμηση τόσο της colon_aca όσο και της colon_n δείχνει ότι το ResNet50 επιτυγχάνει πλήρη διαχωρισμό των κατηγοριών του παχέος εντέρου, στοιχείο που ενισχύει τη συνολική πειραματική εικόνα του μοντέλου .

Η κατηγορία lung_n παρουσιάζει επίσης σχεδόν ιδανική επίδοση, με precision 0.99, recall 1.00 και F1-score 0.99, όπως φαίνεται και στον Πίνακα 6.7. Η συμπεριφορά αυτή υποδηλώνει ότι το μοντέλο αναγνωρίζει πρακτικά όλα τα φυσιολογικά δείγματα πνεύμονα, ενώ παράγει ελάχιστα ψευδώς θετικά. Το γεγονός αυτό δείχνει ότι τα φυσιολογικά δείγματα πνεύμονα σχηματίζουν μια ιδιαίτερα συνεκτική και σαφώς διακριτή περιοχή στον χώρο χαρακτηριστικών. Η πολύ υψηλή απόδοση στη συγκεκριμένη κατηγορία ενισχύει την εικόνα ότι το μοντέλο έχει μάθει αποτελεσματικά τα βασικά μορφολογικά μοτίβα των φυσιολογικών ιστών.

Η κατηγορία lung_aca εμφανίζει πολύ υψηλή αλλά όχι τέλεια επίδοση στο συγκεκριμένο test set , με precision, recall και F1-score ίσα με 0.97. Η ισορροπία μεταξύ precision και recall υποδηλώνει ότι τα σφάλματα είναι συμμετρικά, δηλαδή το μοντέλο παρουσιάζει τόσο περιορισμένο αριθμό ψευδώς θετικών όσο και ψευδώς αρνητικών. Η συγκεκριμένη συμπεριφορά είναι ιδιαίτερα σημαντική, διότι υποδηλώνει ότι το decision boundary της κατηγορίας αυτής είναι σχετικά σταθερό, χωρίς έντονη μεροληψία είτε προς υπερταξινόμηση είτε προς υποταξινόμηση. Παρόλα αυτά, η μη τέλεια επίδοση στο πειραματικό περιβάλλον δείχνει ότι εξακολουθούν να υπάρχουν ορισμένες οριακές περιπτώσεις, πιθανότατα λόγω μορφολογικής επικάλυψης με άλλες κατηγορίες, κυρίως με την lung_scc.

Η κατηγορία lung_scc παρουσιάζει precision 0.99, recall 0.97 και F1-score 0.98, γεγονός που υποδηλώνει πολύ υψηλή συνολική επίδοση με ελάχιστη αλλά υπαρκτή απώλεια σε επίπεδο recall. Η μικρή μείωση του recall δείχνει ότι ένα περιορισμένο ποσοστό πραγματικών δειγμάτων της κατηγορίας δεν αναγνωρίζεται σωστά, οδηγώντας σε ψευδώς αρνητικές προβλέψεις. Το ταυτόχρονα πολύ υψηλό precision δείχνει ότι στο συγκεκριμένο test set, οι περισσότερες προβλέψεις lung_scc είναι ορθές. Η συγκεκριμένη συμπεριφορά είναι ιδιαίτερα σημαντική, καθώς φανερώνει ότι το decision boundary της κατηγορίας αυτής είναι σχετικά σταθερό, χωρίς έντονη μεροληψία είτε προς υπερταξινόμηση είτε προς υποταξινόμηση.

Οι μόνες περιοχές όπου εξακολουθεί να παρατηρείται περιορισμένη ασάφεια εντοπίζονται στις δύο κατηγορίες καρκίνου πνεύμονα, δηλαδή στις lung_aca και lung_scc. Η παρατήρηση αυτή είναι ιδιαίτερα σημαντική, καθώς υποδηλώνει ότι η δυσκολία του προβλήματος δεν έγκειται πλέον στη γενική ικανότητα μάθησης του μοντέλου, αλλά στη λεπτομερή διαφοροποίηση μεταξύ μορφολογικά παρόμοιων υποτύπων. Με άλλα λόγια, το ResNet50 έχει μειώσει σε πολύ μεγάλο βαθμό τα σφάλματα

ταξινόμησης στο συγκεκριμένο πείραμα, αλλά εξακολουθεί να αντιμετωπίζει ελάχιστες περιοχές επικάλυψης εκεί όπου οι ιστολογικές διαφορές είναι πιο λεπτές και πιο δύσκολο να μοντελοποιηθούν.

Από πλευράς γενίκευσης, η απόδοση του ResNet50 είναι ιδιαίτερα ισχυρή. Η πολύ χαμηλή τιμή loss, σε συνδυασμό με την υψηλή ακρίβεια και τις ομοιόμορφες μετρικές ανά κλάση, όπως συνοψίζονται στον Πίνακα 6.7, υποδηλώνουν ότι το μοντέλο έχει μάθει ουσιαστικά χαρακτηριστικά των δεδομένων και όχι αποκλειστικά επιφανειακά μοτίβα, σύμφωνα με τις διαθέσιμες μετρικές.

Συνοψίζοντας, το μοντέλο ResNet50 παρουσιάζει πολύ υψηλή απόδοση και συγκαταλέγεται ανάμεσα στις ισχυρότερες αρχιτεκτονικές που αξιολογούνται στο συγκεκριμένο πρόβλημα, όπως φαίνεται μέσα από τον Πίνακα 6.7. Η αρχιτεκτονική του, μέσω των μηχανισμών υπολειπόμενης μάθησης, επιτρέπει αποτελεσματική εκπαίδευση σε μεγάλο βάθος και οδηγεί σε πλούσιες και διακριτικές αναπαραστάσεις. Το μοντέλο επιτυγχάνει πλήρη διαχωρισμό στις κατηγορίες του παχέος εντέρου και σχεδόν τέλεια ταξινόμηση στις κατηγορίες του πνεύμονα, με τις ελάχιστες εναπομείνασες ασάφειες να περιορίζονται αποκλειστικά στη διάκριση μεταξύ μορφολογικά συγγενών υποτύπων καρκίνου πνεύμονα.

6.4 Αποτελέσματα Vision Transformer

Αποτελέσματα Vision Transformer(10 εποχές εκπαίδευσης)

Πίνακας 6.8: Αποτελέσματα ViT(10 epochs)

Class Type	Precision	Recall	F1-score	Support
Colon_aca	0.86	0.93	0.89	600
Colon_n	0.92	0.87	0.90	600
Lung_aca	0.88	0.93	0.9	600
Lung_n	0.99	1	1	600
Lung_Scc	0.94	0.86	0.9	600
Accuracy	-	-	0.916	3000
Macro_avg	0.92	0.92	0.92	3000
Weighted_avg	0.92	0.92	0.92	3000

Η αξιολόγηση του μοντέλου Vision Transformer μετά από εκπαίδευση για 10 εποχές καταδεικνύει χαμηλότερη απόδοση σε σχέση με τα CNN-based μοντέλα, με συνολική ακρίβεια ίση με 91.60% και σχετικά υψηλή τιμή συνάρτησης κόστους (loss = 0.2346). Τα αποτελέσματα αυτά δείχνουν ότι το μοντέλο με βάση τις διαθέσιμες μετρικές, δεν φαίνεται να έχει συγκλίνει πλήρως και ότι η διαδικασία εκπαίδευσης δεν ήταν επαρκής ώστε να μάθει επαρκώς σταθερές και διακριτικές αναπαραστάσεις.

Η συμπεριφορά αυτή είναι αναμενόμενη για τα Vision Transformers, καθώς τα μοντέλα αυτού του τύπου απαιτούν συνήθως μεγαλύτερο όγκο δεδομένων και περισσότερες εποχές εκπαίδευσης για να επιτύχουν υψηλή απόδοση. Σε αντίθεση με τα CNNs, τα οποία ενσωματώνουν ισχυρές επαγωγικές προκαταλήψεις όπως η τοπικότητα και η μεταθετικότητα, το ViT βασίζεται αποκλειστικά στον μηχανισμό Αυτοπροσοχής, γεγονός που το καθιστά ιδιαίτερα απαιτητικό ως προς τον όγκο των δεδομένων και δυσκολότερο στην εκπαίδευση σε περιορισμένα datasets.

Η ανάλυση των μετρικών precision, recall και F1-score μέσω του Πίνακα 6.8 αποκαλύπτει σημαντικές διαφοροποιήσεις στην απόδοση μεταξύ των κατηγοριών, υποδηλώνοντας ότι το μοντέλο δεν έχει επιτύχει ομοιόμορφη εκμάθηση όλων των κλάσεων.

Η κατηγορία colon_aca παρουσιάζει precision 0.86, recall 0.93 και F1-score 0.89. Η υψηλότερη τιμή του recall σε σχέση με το precision υποδηλώνει ότι το μοντέλο εντοπίζει τα περισσότερα πραγματικά θετικά δείγματα, αλλά παράγει σημαντικό αριθμό ψευδώς θετικών. Η συμπεριφορά αυτή δείχνει ότι το decision boundary της κατηγορίας είναι σχετικά "χαλαρό", οδηγώντας σε υπερταξινόμηση.

Η κατηγορία colon_n εμφανίζει precision 0.92, recall 0.87 και F1-score 0.90. Σε αυτή την περίπτωση παρατηρείται το αντίστροφο μοτίβο, όπου το precision είναι υψηλότερο από το recall, γεγονός που υποδηλώνει ότι το μοντέλο είναι πιο συντηρητικό στην αναγνώριση της κατηγορίας. Η ύπαρξη ψευδώς αρνητικών δείχνει ότι ορισμένα φυσιολογικά δείγματα συγχέονται με καρκινικά, γεγονός που υποδηλώνει ατελή διαχωρισμό στον χώρο χαρακτηριστικών.

Η κατηγορία lung_aca παρουσιάζει precision 0.88, recall 0.93 και F1-score 0.90, με συμπεριφορά παρόμοια με την colon_aca, όπως φαίνεται και στον Πίνακα 6.8. Το μοντέλο εμφανίζει αυξημένη ευαισθησία, αλλά εις βάρος της ακρίβειας, γεγονός που υποδηλώνει ότι τα decision boundaries είναι σχετικά εκτεταμένα και περιλαμβάνουν δείγματα άλλων κατηγοριών.

Η κατηγορία lung_n αποτελεί την καλύτερα αποδιδόμενη κατηγορία, με precision 0.99, recall 1.00 και F1-score 1.00. Η σχεδόν τέλεια επίδοση στο συγκεκριμένο test set υποδηλώνει ότι το μοντέλο φαίνεται να έχει αποτυπώσει αποτελεσματικά βασικά χαρακτηριστικά των φυσιολογικών δειγμάτων πνεύμονα, τα οποία πιθανώς σχηματίζουν μια πιο διακριτή περιοχή στον χώρο χαρακτηριστικών.

Η κατηγορία lung_scc παρουσιάζει precision 0.94, recall 0.86 και F1-score 0.90. Η χαμηλότερη τιμή recall υποδηλώνει ότι το μοντέλο αποτυγχάνει να εντοπίσει ένα σημαντικό ποσοστό πραγματικών θετικών δειγμάτων, οδηγώντας σε ψευδώς αρνητικές προβλέψεις. Η συμπεριφορά αυτή υποδεικνύει ότι το decision boundary της κατηγορίας είναι σχετικά περιορισμένο και δεν καλύπτει πλήρως την ενδοκλασική ποικιλία.

Η συνολική εικόνα που προκύπτει από την ανάλυση σε επίπεδο κλάσεων δείχνει ότι το ViT, μετά από 10 εποχές εκπαίδευσης, δεν φαίνεται να έχει οργανώσει τον χώρο χαρακτηριστικών με την ίδια αποτελεσματικότητα με την ίδια αποτελεσματικότητα όπως τα CNN-based μοντέλα. Παρατηρείται αυξημένη επικάλυψη μεταξύ των κατηγοριών, ιδιαίτερα μεταξύ φυσιολογικών και καρκινικών ιστών, γεγονός που οδηγεί σε μεγαλύτερο αριθμό σφαλμάτων.

Οι κατηγορίες του παχέος εντέρου (colon_aca και colon_n) δεν παρουσιάζουν τον καθαρό διαχωρισμό που παρατηρείται στα CNNs, γεγονός που δείχνει ότι το μοντέλο πιθανώς δεν έχει αποτυπώσει επαρκώς λεπτομερή τοπικά μορφολογικά χαρακτηριστικά που να διαφοροποιούν αποτελεσματικά τις δύο καταστάσεις. Αντίστοιχα, στις κατηγορίες του πνεύμονα, παρατηρείται σημαντική ασάφεια μεταξύ των υποτύπων καρκίνου, γεγονός που υποδηλώνει περιορισμένη ικανότητα του μοντέλου να αποτυπώσει λεπτές διαφοροποιήσεις στην συγκεκριμένη φάση εκπαίδευσης.

Συνοψίζοντας, το Vision Transformer με εκπαίδευση 10 εποχών παρουσιάζει σαφώς κατώτερη απόδοση σε σχέση με τα CNN-based μοντέλα, γεγονός που αποδίδεται κυρίως στην ανεπαρκή εκπαίδευση και στην απουσία ισχυρών επαγωγικών προκαταλήψεων. Η ανάλυση σε επίπεδο κλάσεων, όπως συνοψίζεται στον Πίνακα 6.8, δείχνει ότι το μοντέλο δεν έχει καταφέρει να επιτύχει πλήρη διαχωρισμό των κατηγοριών, με σημαντική επικάλυψη στον χώρο χαρακτηριστικών και αυξημένο αριθμό σφαλμάτων.

Αποτελέσματα Vision Transformer(15 εποχές εκπαίδευσης)

Πίνακας 6.9: Αποτελέσματα ViT(15 epochs)

Class Type	Precision	Recall	F1-score	Support
Colon_aca	0.86	0.93	0.89	600
Colon_n	0.92	0.87	0.90	600
Lung_aca	0.88	0.93	0.9	600
Lung_n	0.99	1	1	600
Lung_Scc	0.94	0.86	0.9	600
Accuracy	-	-	0.916	3000
Macro_avg	0.92	0.92	0.92	3000
Weighted_avg	0.92	0.92	0.92	3000

Η αξιολόγηση του μοντέλου Vision Transformer μετά από εκπαίδευση για 15 εποχές παρουσιάζει σαφή βελτίωση σε σχέση με το προηγούμενο στάδιο (10 εποχές), με συνολική ακρίβεια ίση με 93.97% και σημαντικά μειωμένη τιμή συνάρτησης κόστους (loss = 0.1567). Η βελτίωση αυτή υποδηλώνει ότι το μοντέλο φαίνεται να μαθαίνει πιο ουσιαστικές αναπαραστάσεις και να φαίνεται οργανώνει σταδιακά τον χώρο χαρακτηριστικών με πιο συνεκτικό τρόπο. Σε αντίθεση με το αρχικό στάδιο εκπαίδευσης, η βελτίωση των μετρικών υποδηλώνει πιθανή αποτελεσματικότερη αξιοποίηση του μηχανισμού self-attention, ενισχύοντας τη δυνατότητα σύλληψης μακρινών εξαρτήσεων μεταξύ των patches της εικόνας. Παρόλα αυτά, εξακολουθεί να παρουσιάζει χαμηλότερη απόδοση σε σχέση με τα CNN-based μοντέλα, τα οποία έχουν εκπαιδευτεί στις ίδιες ή και λιγότερες εποχές εκπαίδευσης.

Η ανάλυση των μετρικών precision, recall και F1-score του Πίνακα 6.9 δείχνει βελτιωμένη αλλά όχι πλήρως ισορροπημένη συμπεριφορά μεταξύ των κατηγοριών, με σαφή μείωση των σφαλμάτων σε σχέση με τις 10 εποχές.

Η κατηγορία colon_aca παρουσιάζει precision 0.98, recall 0.90 και F1-score 0.94. Η υψηλή τιμή precision υποδηλώνει ότι το μοντέλο αποδίδει την κατηγορία με υψηλή ακρίβεια στο συγκεκριμένο test set, ωστόσο η χαμηλότερη τιμή recall δείχνει ότι εξακολουθούν να υπάρχουν ψευδώς αρνητικές προβλέψεις. Η συμπεριφορά αυτή φανερώνει ότι το decision boundary της κατηγορίας είναι σχετικά αυστηρό, περιορίζοντας την κάλυψη της ενδοκλασικής ποικιλίας.

Η κατηγορία colon_n εμφανίζει precision 0.91, recall 0.98 και F1-score 0.95. Σε αυτή την περίπτωση παρατηρείται αυξημένη ευαισθησία, με το μοντέλο να αναγνωρίζει σχεδόν όλα τα φυσιολογικά δείγματα στο συγκεκριμένο test set, αλλά συνεχίζει να παράγει ορισμένα ψευδώς θετικά. Η ασυμμετρία αυτή υποδηλώνει ότι το decision boundary είναι πιο εκτεταμένο προς την πλευρά της κατηγορίας, γεγονός που οδηγεί σε ελαφρά υπερταξινόμηση.

Η κατηγορία lung_aca παρουσιάζει precision 0.94, recall 0.86 και F1-score 0.90, εμφανίζοντας μειωμένο recall σε σχέση με άλλες κατηγορίες, όπως φαίνεται και στον Πίνακα 6.9. Η συμπεριφορά αυτή δείχνει ότι το μοντέλο εξακολουθεί να δυσκολεύεται στην πλήρη κάλυψη της κατηγορίας, οδηγώντας σε ψευδώς αρνητικές προβλέψεις. Η δυσκολία αυτή πιθανότατα σχετίζεται με τη μορφολογική ομοιότητα με την κατηγορία lung_scc και υποδηλώνει ότι το μοντέλο δεν φαίνεται ακόμη να έχει διαμορφώσει επαρκώς διακριτές αναπαραστάσεις για τους δύο υποτύπους καρκίνου.

Η κατηγορία lung_n παραμένει η πιο σταθερή και εύκολη προς ταξινόμηση κατηγορία, με precision 0.99, recall 1.00 και F1-score 1.00. Η σχεδόν τέλεια επίδοση στο πειραματικό περιβάλλον υποδηλώνει ότι το μοντέλο φαίνεται να έχει αποτυπώσει αποτελεσματικά βασικά χαρακτηριστικά των

φυσιολογικών δειγμάτων πνεύμονα, τα οποία πιθανώς σχηματίζουν μια πιο διακριτή περιοχή στον χώρο χαρακτηριστικών.

Η κατηγορία lung_scc εμφανίζει precision 0.88, recall 0.95 και F1-score 0.91. Η υψηλή τιμή recall υποδηλώνει ότι το μοντέλο εντοπίζει τα περισσότερα πραγματικά θετικά δείγματα, αλλά η χαμηλότερη τιμή precision δείχνει ότι εξακολουθούν να υπάρχουν ψευδώς θετικά. Η συμπεριφορά αυτή είναι ενδεικτική ενός decision boundary που έχει επεκταθεί προς γειτονικές περιοχές του χώρου χαρακτηριστικών, οδηγώντας σε σύγχυση κυρίως με την κατηγορία lung_aca.

Η ανάλυση σε βάθος δείχνει ότι το ViT στις 15 εποχές φαίνεται να διαμορφώνει πιο σαφή clusters στον χώρο χαρακτηριστικών σε σχέση με τις 10 εποχές, χωρίς όμως να έχει επιτύχει ακόμη πλήρη διαχωρισμό μεταξύ όλων των κατηγοριών. Παρατηρείται σημαντική βελτίωση στη συνοχή των κατηγοριών και μείωση της επικάλυψης, ιδιαίτερα στις κατηγορίες του παχέος εντέρου.

Παρά τη βελτίωση αυτή, οι κατηγορίες καρκίνου πνεύμονα (lung_aca και lung_scc) εξακολουθούν να παρουσιάζουν περιοχές επικάλυψης στον embedding space. Το γεγονός αυτό υποδηλώνει ότι το μοντέλο δεν φαίνεται ακόμη να αποτυπώνει πλήρως τις λεπτές μορφολογικές διαφοροποιήσεις μεταξύ των υποτύπων, κάτι που αποτελεί βασική πρόκληση για τις αρχιτεκτονικές Vision Transformer όταν εκπαιδεύονται σε περιορισμένα datasets.

Συνοψίζοντας, το Vision Transformer μετά από 15 εποχές παρουσιάζει σαφή βελτίωση σε σχέση με το αρχικό στάδιο εκπαίδευσης, με καλύτερη οργάνωση του χώρου χαρακτηριστικών, μειωμένο αριθμό σφαλμάτων και πιο σταθερές αναπαραστάσεις, όπως προκύπτει από τον Πίνακα 6.9. Παρόλα αυτά, εξακολουθεί να υπολείπεται των CNN-based μοντέλων, κυρίως λόγω της ατελούς διάκρισης μεταξύ μορφολογικά παρόμοιων κατηγοριών και της αυξημένης απαίτησης του μοντέλου για μεγαλύτερο όγκο δεδομένων και εκτενέστερη εκπαίδευση.

Αποτελέσματα Vision Transformer(20 εποχές εκπαίδευσης)

Η τελική αξιολόγηση του μοντέλου Vision Transformer μετά από εκπαίδευση για 20 20 εποχές καταδεικνύει περαιτέρω βελτίωση της απόδοσης σε σχέση με τα προηγούμενα στάδια εκπαίδευσης, με συνολική ακρίβεια ίση με 94.80% και χαμηλότερη τιμή συνάρτησης κόστους (loss = 0.1350). Η εξέλιξη αυτή υποδηλώνει ότι το μοντέλο συνεχίζει να συγκλίνει, παρουσιάζοντας ενδείξεις σταδιακής βελτίωσης των αναπαραστάσεων που μαθαίνει. Σε αυτό το στάδιο, το ViT αρχίζει να προσεγγίζει περισσότερο τις επιδόσεις των CNN-based μοντέλων, αν και εξακολουθεί να υπολείπεται σε επίπεδο συνολικής ακρίβειας.

Πίνακας 6.10: Αποτελέσματα ViT(20 epochs)

Class Type	Precision	Recall	F1-score	Support
Colon_aca	1	0.92	0.96	600
Colon_n	0.93	0.99	0.96	600
Lung_aca	0.92	0.90	0.91	600
Lung_n	1	1	1	600
Lung_Scc	0.91	0.92	0.91	600
Accuracy	-	-	0.948	3000
Macro_avg	0.95	0.95	0.95	3000
Weighted_avg	0.95	0.95	0.95	3000

Η σταδιακή βελτίωση από τις 10 στις 20 εποχές εκπαίδευσης καθώς και η ανάλυση των πινάκων 6.8 έως 6.10 αναδεικνύει τη φύση των Vision Transformers ως αρχιτεκτονικών που απαιτούν

εκτεταμένη εκπαίδευση προκειμένου να αξιοποιήσουν αποτελεσματικότερα τον μηχανισμό self-attention και να οργανώσουν πιο αποτελεσματικά τον χώρο χαρακτηριστικών, όπως υποδηλώνεται από τις μετρικές. Σε αντίθεση με τα CNNs, τα οποία διαθέτουν ενσωματωμένες επαγωγικές προκαταλήψεις σχετικές με τη χωρική δομή της εικόνας, το ViT βασίζεται κυρίως στην εκμάθηση σχέσεων μεταξύ των patches μέσω attention μηχανισμών, γεγονός που απαιτεί μεγαλύτερο όγκο εκπαίδευσης ώστε να διαμορφωθούν σταθερές και διακριτές αναπαραστάσεις.

Η ανάλυση των μετρικών precision, recall και F1-score αποκαλύπτει σημαντική βελτίωση στη συνοχή των κατηγοριών, καθώς και πιο ισορροπημένη συμπεριφορά σε σχέση με τις προηγούμενες φάσεις εκπαίδευσης.

Η κατηγορία colon_aca παρουσιάζει precision 1.00, recall 0.92 και F1-score 0.96. Η τιμή precision 1.00 στο συγκεκριμένο test set υποδηλώνει ότι το μοντέλο δεν παράγει ψευδώς θετικά για τη συγκεκριμένη κατηγορία, ωστόσο η ελαφρώς μειωμένη τιμή recall δείχνει ότι εξακολουθούν να υπάρχουν ορισμένες ψευδώς αρνητικές προβλέψεις. Η συμπεριφορά αυτή υποδηλώνει ένα σχετικά αυστηρό decision boundary, το οποίο περιορίζει την πλήρη κάλυψη της ενδοκλασικής ποικιλίας.

Η κατηγορία colon_n εμφανίζει precision 0.93, recall 0.99 και F1-score 0.96, παρουσιάζοντας υψηλή ευαισθησία και σχεδόν πλήρη αναγνώριση των φυσιολογικών δειγμάτων στο συγκεκριμένο test set. Η μικρή μείωση του precision υποδηλώνει περιορισμένο αριθμό ψευδώς θετικών, γεγονός που συνδέεται με ελαφρά επικάλυψη με την κατηγορία colon_aca.

Η κατηγορία lung_aca παρουσιάζει precision 0.92, recall 0.90 και F1-score 0.91. Η σχετική ισορροπία μεταξύ των μετρικών υποδηλώνει ότι το μοντέλο έχει αρχίσει να σταθεροποιεί το decision boundary της κατηγορίας, μειώνοντας τις ασυμμετρίες που παρατηρούνταν σε προηγούμενα στάδια. Παρόλα αυτά, η απόδοση παραμένει χαμηλότερη σε σχέση με άλλες κατηγορίες, γεγονός που δείχνει ότι η διάκριση της συγκεκριμένης κατηγορίας εξακολουθεί να αποτελεί πρόκληση.

Η κατηγορία lung_n παρουσιάζει και πάλι τέλεια επίδοση στο συγκεκριμένο test set (precision 1.00, recall 1.00, F1-score 1.00), γεγονός που υποδηλώνει ότι τα φυσιολογικά δείγματα πνεύμονα αποτελούν την πιο εύκολα διαχωρίσιμη κατηγορία για το μοντέλο και πιθανώς σχηματίζουν πιο διακριτή περιοχή στον χώρο χαρακτηριστικών.

Η κατηγορία lung_scc εμφανίζει precision 0.91, recall 0.92 και F1-score 0.91, παρουσιάζοντας σχετικά ισορροπημένη αλλά όχι υψηλή επίδοση. Η συμπεριφορά αυτή δείχνει ότι το μοντέλο εξακολουθεί να αντιμετωπίζει δυσκολίες στη διάκριση της συγκεκριμένης κατηγορίας, πιθανότατα λόγω της μορφολογικής ομοιότητας με την lung_aca.

Η ανάλυση σε βάθος δείχνει ότι το ViT στις 20 εποχές φαίνεται να παρουσιάζει καλύτερη οργάνωση του χώρου χαρακτηριστικών σε σχέση με τα προηγούμενα στάδια. Οι αναπαραστάσεις φαίνεται να είναι πιο συνεκτικές και καλύτερα διαχωρισμένες, ιδιαίτερα για τις κατηγορίες του παχέος εντέρου και τα φυσιολογικά δείγματα.

Παρόλα αυτά, οι κατηγορίες καρκίνου πνεύμονα εξακολουθούν να παρουσιάζουν περιοχές επικάλυψης στον embedding space. Η διάκριση μεταξύ των κατηγοριών lung_aca και lung_scc παραμένει το κύριο σημείο δυσκολίας, γεγονός που υποδηλώνει ότι το μοντέλο δεν φαίνεται ακόμη να έχει μάθει επαρκώς τα λεπτομερή μορφολογικά χαρακτηριστικά που διαφοροποιούν τους δύο υποτύπους.

Η περαιτέρω μείωση της τιμής loss και η αύξηση της ακρίβειας υποδηλώνουν καλύτερη απόδοση στο συγκεκριμένο σύνολο ελέγχου. Το μοντέλο φαίνεται να αποτυπώνει πιο σταθερά

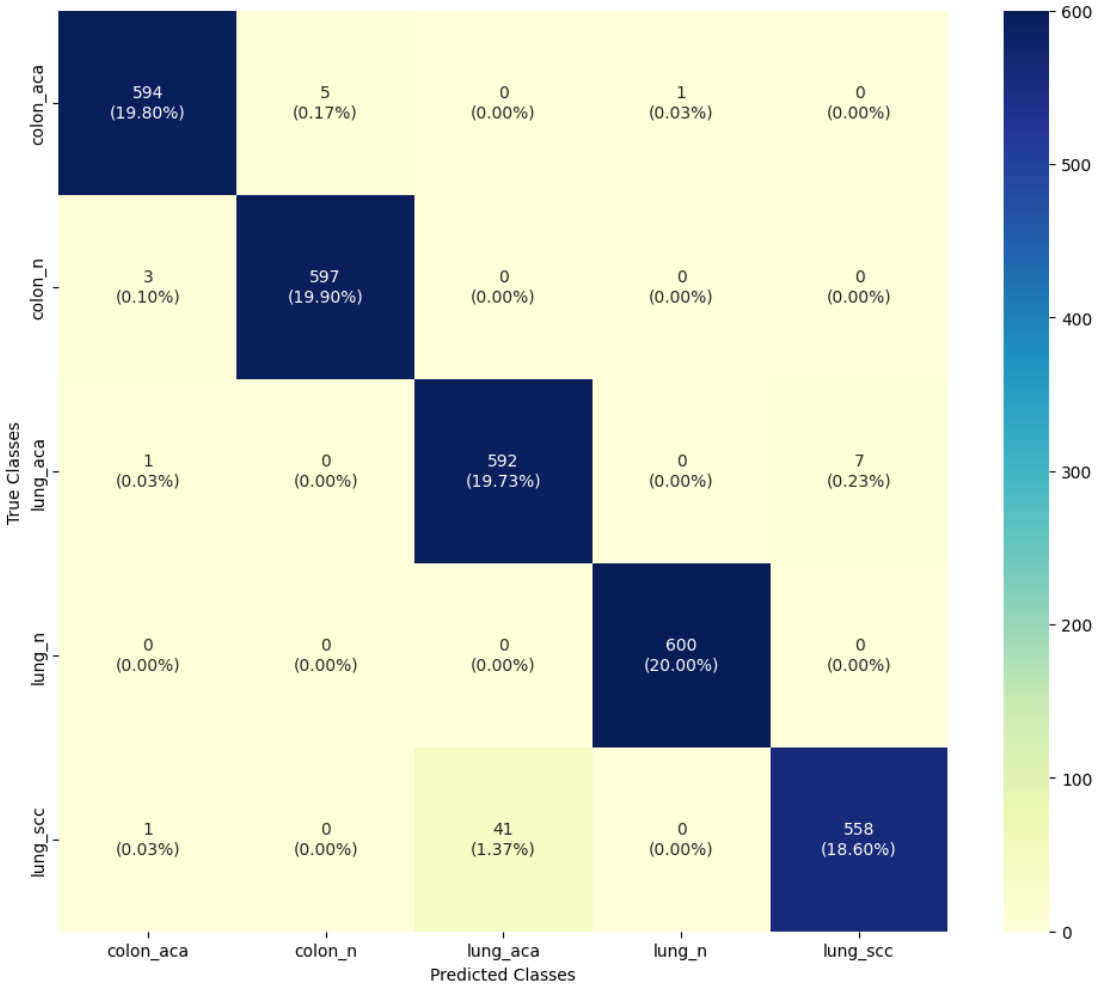
χαρακτηριστικά των δεδομένων, αν και η πραγματική ικανότητα γενίκευσης θα απαιτούσε αξιολόγηση σε ανεξάρτητα και ετερογενή δεδομένα. Παράλληλα, η βελτίωση της συνοχής των κατηγοριών μπορεί να υποδηλώνει αποτελεσματικότερη αξιοποίηση των attention μηχανισμών, οδηγώντας σε πιο σταθερές και διακριτικές αναπαραστάσεις.

Συνοψίζοντας, το Vision Transformer μετά από 20 εποχές παρουσιάζει σημαντική πρόοδο σε σχέση με τα προηγούμενα στάδια εκπαίδευσης, με βελτιωμένη συμπεριφορά ανά κλάση, καλύτερη απόδοση στο test set και μειωμένο αριθμό σφαλμάτων. Παρά τη βελτίωση αυτή, εξακολουθεί να υπολείπεται των CNN-based μοντέλων στο συγκεκριμένο πειραματικό περιβάλλον, κυρίως λόγω της δυσκολίας διάκρισης μεταξύ μορφολογικά παρόμοιων κατηγοριών.

Η εξέλιξη της απόδοσης από τις 10 στις 20 εποχές υποδηλώνει ότι το ViT επωφελείται από εκτενέστερη εκπαίδευση, γεγονός που συνδέεται με τη φύση της αρχιτεκτονικής του. Το συγκεκριμένο στάδιο μπορεί να θεωρηθεί ως ένα βελτιωμένο αλλά όχι πλήρως βελτιστοποιημένο σημείο της εκπαίδευσης, όπου το μοντέλο έχει παρουσιάσει σημαντική πρόοδο, αλλά εξακολουθεί να διαθέτει περιθώρια περαιτέρω βελτίωσης μέσω μεγαλύτερης εκπαίδευσης, προσεκτικότερου fine-tuning ή αξιοποίησης μεγαλύτερου όγκου δεδομένων.

6.5 Πίνακες Σύγχυσης και Ανάλυση Σφαλμάτων των τελικών αποτελεσμάτων

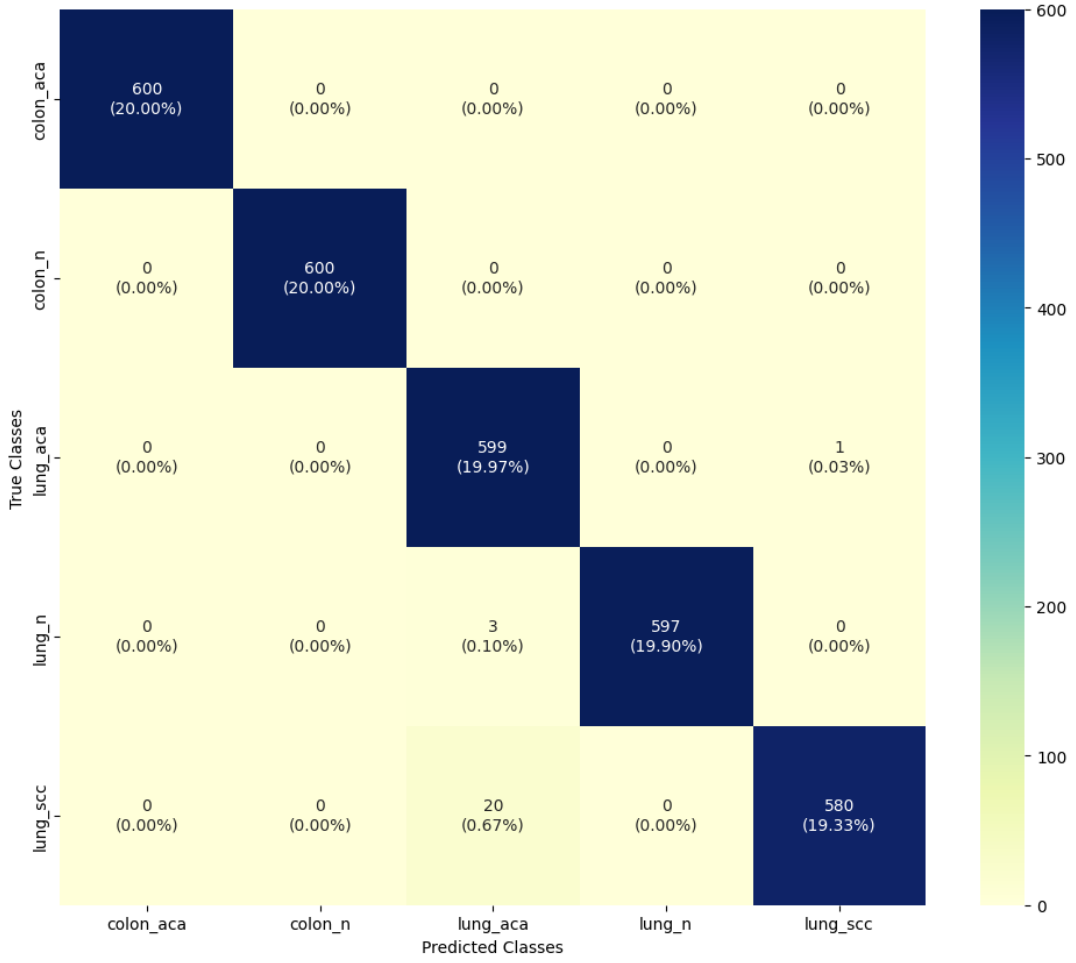
6.5.1 Πίνακας Σύγχυσης CNN



Σχήμα 6.1: Πίνακας Σύγχυσης CNN

Μέσα από την ανάλυση του Πίνακα Σύγχυσης 6.1 γίνεται φανερό ότι το μοντέλο CNN παρουσιάζει πολύ υψηλή συγκέντρωση τιμών στην διαγώνιο το οποίο αναδεικνύει ότι το μοντέλο έχει υψηλή ακρίβεια ταξινόμησης με πολύ σωστές προβλέψεις. Τα σφάλματα είναι λίγα σε αριθμό αλλά έχουν συγκεκριμένο μοτίβο. Συγκεκριμένα, καταγράφεται περιορισμένη σύγχυση μεταξύ φυσιολογικού και καρκινικού ιστού στο παχύ έντερο (colon_n και colon_aca) όπως φαίνεται και στο Σχήμα 6.1, γεγονός που υποδεικνύει ότι το μοντέλο δυσκολεύεται σε περιπτώσεις όπου οι μορφολογικές διαφοροποιήσεις είναι λεπτές. Επιπλέον, η πιο έντονη πηγή σφαλμάτων εντοπίζεται μεταξύ των υποτύπων καρκίνου του πνεύμονα, και ειδικότερα μεταξύ των κατηγοριών lung_aca και lung_scc, με σημαντικό αριθμό δειγμάτων της δεύτερης να ταξινομούνται εσφαλμένα ως lung_aca. Το εύρημα αυτό υποδηλώνει ότι, ενώ το μοντέλο καταφέρνει να αναγνωρίσει αποτελεσματικά το όργανο προέλευσης (organ-level classification), παρουσιάζει περιορισμένη ικανότητα στη λεπτομερή διάκριση μεταξύ ιστολογικών υποτύπων (fine-grained classification). Αντίθετα, η κατηγορία lung_n εμφανίζει τέλεια ταξινόμηση, γεγονός που υποδηλώνει ότι τα χαρακτηριστικά του φυσιολογικού ιστού του πνεύμονα είναι πιο διακριτά και ευκολότερα αναγνωρίσιμα από το μοντέλο. Συνολικά, τα αποτελέσματα που φαίνονται στο Σχήμα 6.1 καταδεικνύουν ότι το CNN μοντέλο αποδίδει εξαιρετικά σε επίπεδο γενικής κατηγοριοποίησης, αλλά εμφανίζει περιορισμούς στην αποτύπωση λεπτών διαφορών μεταξύ παρόμοιων κατηγοριών, κάτι που αποτελεί κρίσιμο σημείο για περαιτέρω βελτιστοποίηση και σύγκριση με πιο εξελιγμένες αρχιτεκτονικές.

6.5.2 Πίνακες Σύγχυσης μεθόδων Μεταφοράς Μάθησης



Σχήμα 6.2: Πίνακας Σύγχυσης InceptionResNetV2

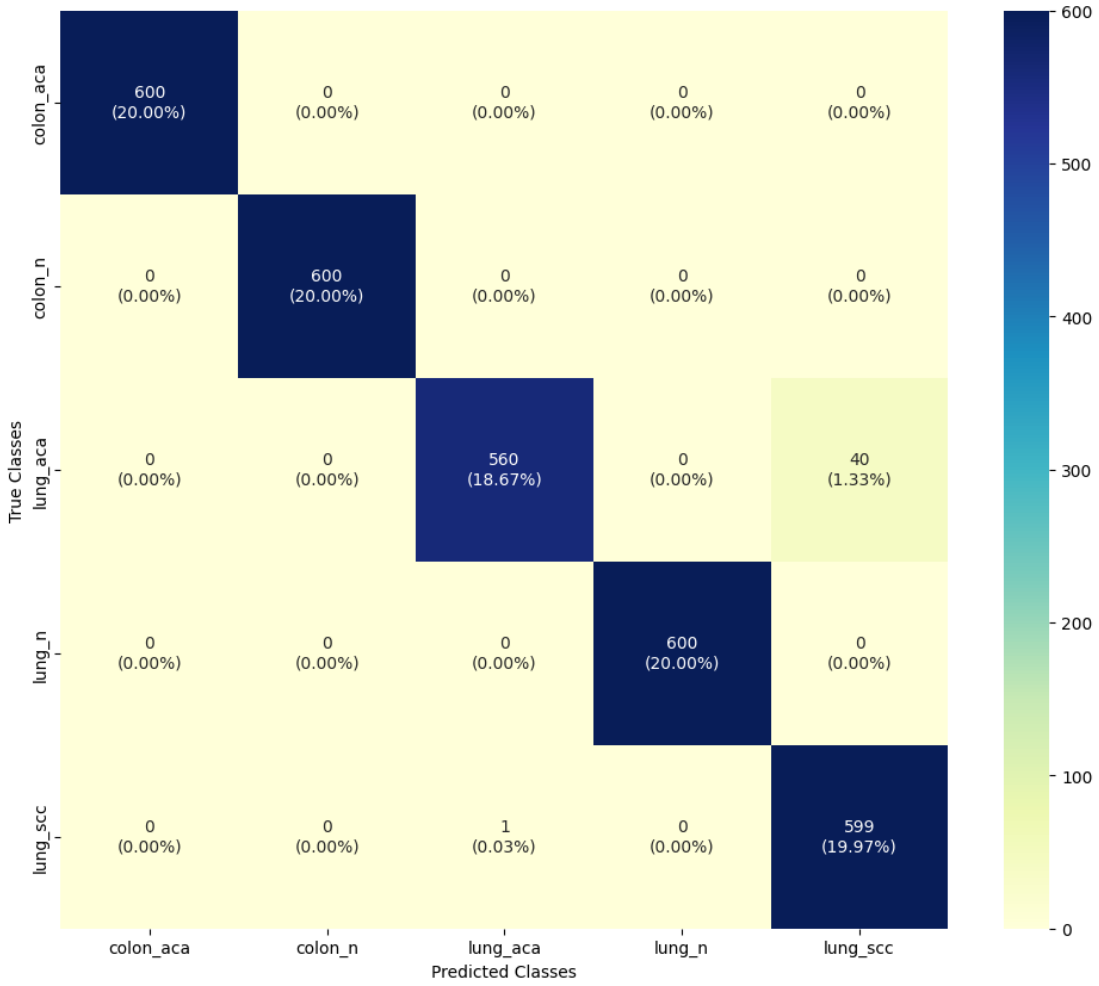
Η ανάλυση του Πίνακα Σύχγυσης του μοντέλου InceptionResNetV2 στο Σχήμα 6.2 καταδεικνύει υψηλή απόδοση ταξινόμησης, με σχεδόν πλήρη συγκέντρωση των προβλέψεων στη διαγώνιο του πίνακα, γεγονός που υποδηλώνει σχεδόν άψογη ικανότητα γενίκευσης. Ειδικότερα, οι κατηγορίες colon_aca και colon_n ταξινομούνται με απόλυτη ακρίβεια, χωρίς καμία λανθασμένη πρόβλεψη, ενώ και η κατηγορία lung_n παρουσιάζει εξαιρετικά υψηλή απόδοση με ελάχιστα σφάλματα. Παρά τη συνολική αυτή βελτίωση, μικρός αριθμός σφαλμάτων εξακολουθεί να παρατηρείται κυρίως μεταξύ των υποτύπων καρκίνου του πνεύμονα, και συγκεκριμένα μεταξύ των κατηγοριών lung_aca και lung_scc, όπου περιορισμένος αριθμός δειγμάτων της κατηγορίας lung_scc ταξινομείται ως lung_aca. Το φαινόμενο αυτό υποδηλώνει ότι, αν και το μοντέλο έχει βελτιώσει σημαντικά τη διακριτική του ικανότητα σε σχέση με απλούστερες αρχιτεκτονικές CNN, εξακολουθεί να αντιμετωπίζει δυσκολίες στη λεπτομερή διάκριση μεταξύ ιστολογικά παρόμοιων κατηγοριών (fine-grained classification). Συνολικά, το InceptionResNetV2 επιδεικνύει σαφή υπεροχή στην εξαγωγή πολύπλοκων χαρακτηριστικών και στη διαχείριση ενδοκατηγορικών διαφορών, μειώνοντας δραστικά τα σφάλματα που παρατηρήθηκαν σε προηγούμενα μοντέλα, γεγονός που το καθιστά ιδιαίτερα κατάλληλο για εφαρμογές ιατρικής ταξινόμησης υψηλής ακρίβειας.

VGG16

Σχήμα 6.3: Πίνακας Σύγχυσης VGG16

Η ανάλυση του confusion matrix του μοντέλου VGG16 στο Σχήμα 6.3 αποκαλύπτει υψηλή συνολική απόδοση ταξινόμησης, με την πλειοψηφία των προβλέψεων να συγκεντρώνεται στη διαγώνιο του πίνακα. Οι κατηγορίες του παχέος εντέρου (colon_aca και colon_n) ταξινομούνται με ιδιαίτερα υψηλή ακρίβεια, παρουσιάζοντας ελάχιστα σφάλματα μεταξύ τους, γεγονός που υποδηλώνει ικανοποιητική ικανότητα διάκρισης μεταξύ φυσιολογικού και καρκινικού ιστού στο συγκεκριμένο όργανο. Παρόμοια, η κατηγορία lung_n εμφανίζει σχεδόν τέλεια ταξινόμηση, επιβεβαιώνοντας ότι τα χαρακτηριστικά του φυσιολογικού ιστού του πνεύμονα είναι σαφώς διακριτά για το μοντέλο. Ωστόσο, σημαντικά σφάλματα εντοπίζονται στην κατηγορία lung_aca, όπου παρατηρείται έντονη σύγχυση με την κατηγορία lung_scc, με μεγάλο αριθμό δειγμάτων να ταξινομούνται λανθασμένα μεταξύ των δύο αυτών υποτύπων. Επιπλέον, παρατηρείται και αντίστροφη, αλλά μικρότερης έντασης, σύγχυση από lung_scc προς lung_aca. Το συγκεκριμένο μοτίβο του Σχήματος 6.3 υποδηλώνει ότι το μοντέλο δυσκολεύεται στη λεπτομερή διάκριση μεταξύ ιστολογικά παρόμοιων κατηγοριών καρκίνου του πνεύμονα, παρά την ικανοποιητική του απόδοση σε επίπεδο γενικής κατηγοριοποίησης. Συνολικά, το VGG16 εμφανίζει ισχυρή ικανότητα εξαγωγής βασικών μορφολογικών χαρακτηριστικών, ωστόσο παρουσιάζει περιορισμούς στην αποτύπωση πιο σύνθετων και λεπτών διαφορών μεταξύ συγγενών κατηγοριών, γεγονός που επηρεάζει την απόδοσή του σε σενάρια fine-grained classification.

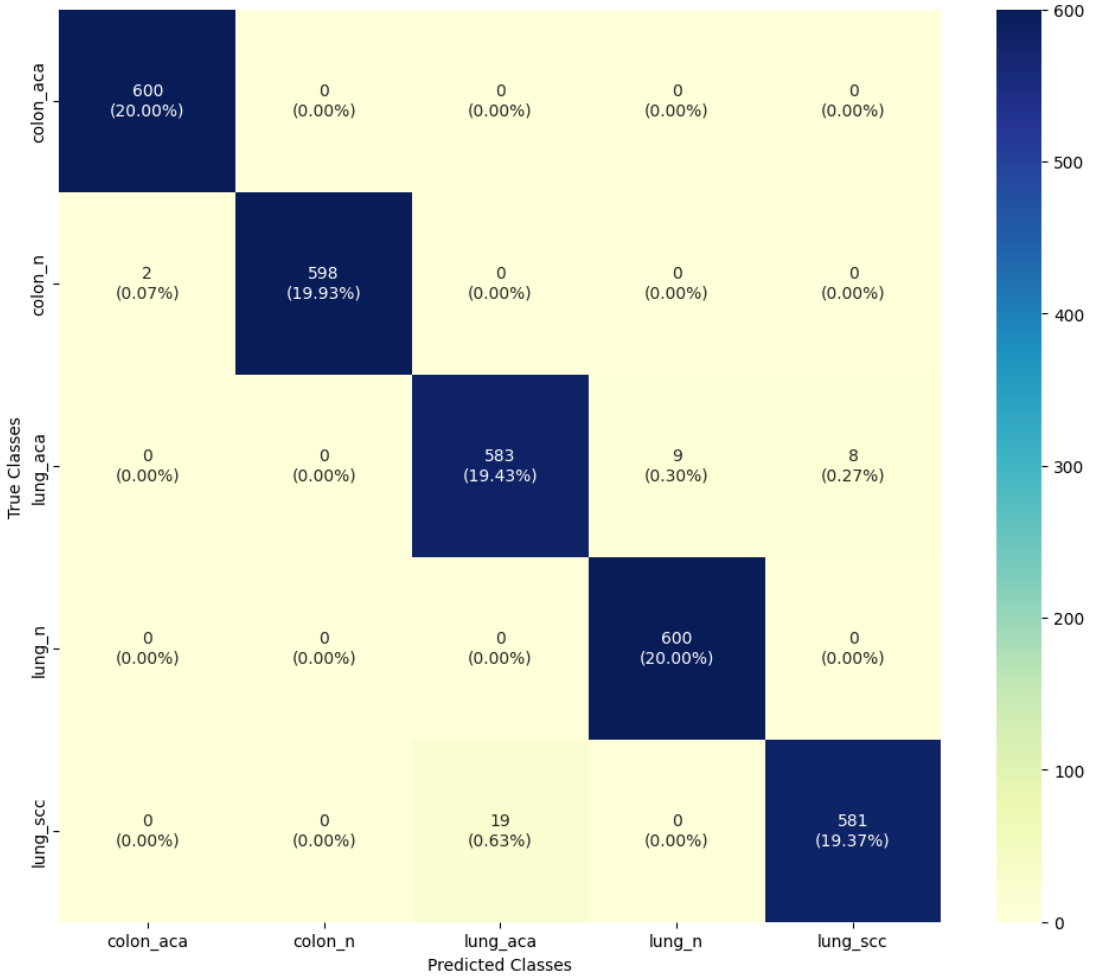
VGG19



Σχήμα 6.4: Πίνακας Σύγχυσης VGG19

Η ανάλυση του Σχήματος 6.4 του μοντέλου VGG19 καταδεικνύει υψηλή απόδοση ταξινόμησης, με πλήρη ακρίβεια στις κατηγορίες του παχέος εντέρου (colon_aca και colon_n), καθώς και στην κατηγορία lung_n, όπου δεν παρατηρούνται καθόλου σφάλματα. Το γεγονός αυτό υποδηλώνει ότι το μοντέλο είναι σε θέση να διακρίνει με απόλυτη σαφήνεια τόσο τον φυσιολογικό όσο και τον καρκινικό ιστό σε επίπεδο οργάνου, εκμεταλλευόμενο αποτελεσματικά τα μορφολογικά χαρακτηριστικά των εικόνων. Ωστόσο, μικρός αριθμός σφαλμάτων εντοπίζεται στην κατηγορία lung_aca, όπου παρατηρείται σύγχυση με την κατηγορία lung_scc, με μέρος των δειγμάτων να ταξινομούνται λανθασμένα ως πλακώδες καρκίνωμα. Επιπλέον, η αντίστροφη σύγχυση από lung_scc προς lung_aca είναι σχεδόν αμελητέα, γεγονός που υποδηλώνει ασυμμετρία στην κατανομή των σφαλμάτων και πιθανή μεροληψία του μοντέλου προς την κατηγορία lung_scc. Το συγκεκριμένο μοτίβο υποδεικνύει ότι, παρά την υψηλή συνολική απόδοση, το μοντέλο εξακολουθεί να αντιμετωπίζει περιορισμούς στη λεπτομερή διάκριση μεταξύ ιστολογικά συγγενών υποτύπων καρκίνου του πνεύμονα. Συνολικά, το VGG19 παρουσιάζει σημαντική βελτίωση σε σχέση με απλούστερες αρχιτεκτονικές, επιτυγχάνοντας σχεδόν τέλεια ταξινόμηση στις περισσότερες κατηγορίες, ωστόσο οι εναπομείνασες αστοχίες εντοπίζονται σε σενάρια υψηλής ομοιότητας χαρακτηριστικών, επιβεβαιώνοντας τη δυσκολία του προβλήματος σε επίπεδο fine-grained classification.

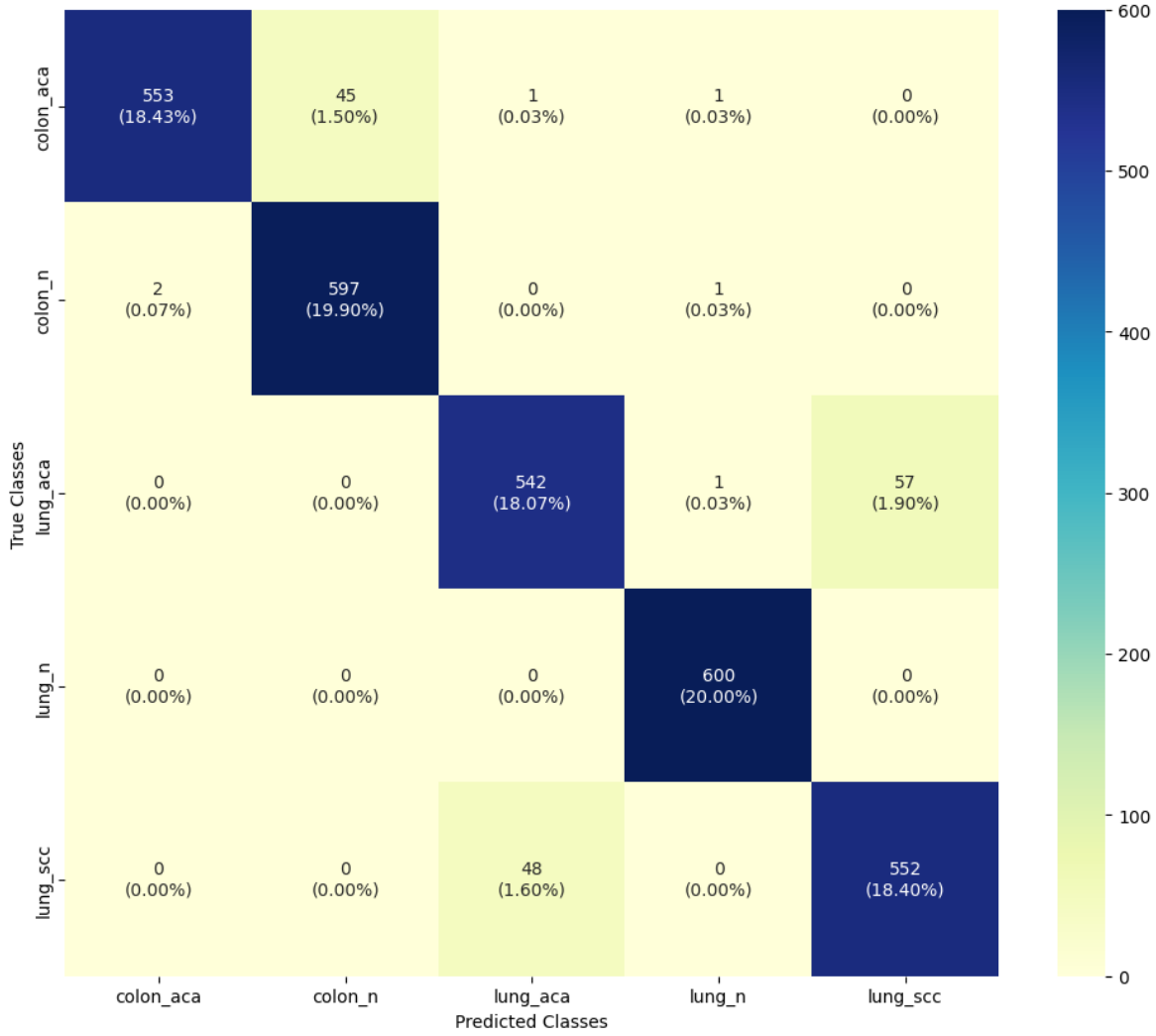
ResNet50



Σχήμα 6.5: Πίνακας Σύγχυσης ResNet50

Η ανάλυση του Σχήματος 6.5 του μοντέλου ResNet50 από το Σχήμα 6.5 το μοντέλο φαίνεται ότι, καταδεικνύει υψηλή απόδοση ταξινόμησης, με τις περισσότερες προβλέψεις να συγκεντρώνονται στη διαγώνιο του πίνακα, γεγονός που υποδηλώνει ισχυρή ικανότητα γενίκευσης και αποτελεσματική εξαγωγή χαρακτηριστικών. Οι κατηγορίες colon_aca και lung_n ταξινομούνται με απόλυτη ακρίβεια, ενώ και η κατηγορία colon_n παρουσιάζει σχεδόν τέλεια απόδοση με ελάχιστες λανθασμένες ταξινομήσεις. Παρόμοια, η κατηγορία lung_aca εμφανίζει υψηλή ακρίβεια, αν και παρατηρείται μικρός αριθμός σφαλμάτων προς τις κατηγορίες lung_n και lung_scc. Η σημαντικότερη πηγή σφαλμάτων εντοπίζεται, ωστόσο, στην κατηγορία lung_scc, όπου ένα μέρος των δειγμάτων ταξινομείται λανθασμένα ως lung_aca, υποδεικνύοντας σύγχυση μεταξύ των δύο υποτύπων καρκίνου του πνεύμονα. Το συγκεκριμένο μοτίβο επιβεβαιώνει ότι, παρά τη βελτιωμένη ικανότητα του ResNet50 στην εκμάθηση βαθύτερων και πιο σύνθετων χαρακτηριστικών μέσω των residual συνδέσεων, εξακολουθούν να υπάρχουν δυσκολίες στη λεπτομερή διάκριση μεταξύ ιστολογικά παρόμοιων κατηγοριών. Συνολικά από το Σχήμα 6.5 γίνεται αντιληπτό ότι, το ResNet50 επιτυγχάνει υψηλή απόδοση και εμφανίζει σαφή υπεροχή έναντι απλούστερων αρχιτεκτονικών, μειώνοντας σημαντικά τα σφάλματα, ωστόσο οι εναπομείνασες αστοχίες περιορίζονται κυρίως σε σενάρια fine-grained classification, τα οποία απαιτούν ακόμη πιο εξειδικευμένες αρχιτεκτονικές ή τεχνικές για περαιτέρω βελτίωση.

6.5.3 Πίνακας Σύγχυσης Vision Transformer(20 εποχές)



Σχήμα 6.6: Πίνακας Σύγχυσης Vision Transformer

Η ανάλυση του Σχήματος 6.6 του μοντέλου Vision Transformer μετά από 20 εποχές εκπαίδευσης αποκαλύπτει υψηλή αλλά όχι ομοιόμορφη απόδοση ταξινόμησης, με σαφή διαφοροποίηση μεταξύ των κατηγοριών. Ειδικότερα από το Σχήμα 6.6 φαίνεται πως, η κατηγορία lung_n ταξινομείται με απόλυτη ακρίβεια, γεγονός που υποδηλώνει ότι τα χαρακτηριστικά του φυσιολογικού ιστού του πνεύμονα είναι ιδιαίτερα διακριτά για το μοντέλο. Αντίστοιχα, η κατηγορία colon_n παρουσιάζει υψηλή ακρίβεια με ελάχιστα σφάλματα. Ωστόσο, σημαντικές αστοχίες παρατηρούνται στην κατηγορία colon_aca, όπου αρκετά δείγματα ταξινομούνται λανθασμένα ως colon_n, γεγονός που υποδεικνύει δυσκολία στη διάκριση μεταξύ φυσιολογικού και καρκινικού ιστού στο παχύ έντερο. Παράλληλα, έντονη σύγχυση εντοπίζεται μεταξύ των υποτύπων καρκίνου του πνεύμονα, και συγκεκριμένα μεταξύ lung_aca και lung_scc, με σημαντικό αριθμό δειγμάτων να ταξινομούνται λανθασμένα μεταξύ των δύο κατηγοριών. Το μοτίβο του Σχήματος 6.6 υποδηλώνει ότι, παρά την ικανότητα του ViT να συλλαμβάνει μακροπρόθεσμες συσχετίσεις μέσω του μηχανισμού self-attention, το μοντέλο εξακολουθεί να αντιμετωπίζει δυσκολίες στην αποτύπωση λεπτών ιστολογικών διαφορών (fine-grained features), ιδιαίτερα όταν οι κατηγορίες παρουσιάζουν υψηλή μορφολογική ομοιότητα. Συνολικά, τα αποτελέσματα δείχνουν ότι η απόδοση του ViT, αν και ικανοποιητική, υπολείπεται σε ορισμένα σημεία σε σχέση με βαθύτερες CNN αρχιτεκτονικές, γεγονός που ενδέχεται να οφείλεται είτε στο γεγονός ότι τα Vision Transformers χρειάζονται μεγαλύτερη διάρκεια εκπαίδευσης από τους μηχανισμούς που βασίζονται σε συνελικτικά νευρωνικά δίκτυα είτε στην ανάγκη για μεγαλύτερο όγκο δεδομένων, δεδομένης της φύσης των transformer μοντέλων.

6.6 Συγκριτική Αξιολόγηση των μοντέλων

Η αξιολόγηση των μοντέλων βασίστηκε όχι μόνο στην οπτική ανάλυση των confusion matrices, αλλά και στη συστηματική εξέταση βασικών μετρικών απόδοσης, όπως η ακρίβεια (Accuracy), η ακρίβεια θετικών προβλέψεων (Precision), η ανάκληση (Recall) και ο F1-score. Αν και όλα τα μοντέλα παρουσιάζουν ιδιαίτερα υψηλές τιμές Accuracy, η συγκεκριμένη μετρική από μόνη της δεν παρέχει πλήρη εικόνα της απόδοσης, καθώς δεν αποτυπώνει την κατανομή των σφαλμάτων μεταξύ των κατηγοριών. Συγκεκριμένα, η υψηλή συγκέντρωση σωστών προβλέψεων στη διαγώνιο των confusion matrices οδηγεί σε αυξημένο Accuracy, ακόμη και όταν υπάρχουν κρίσιμα σφάλματα μεταξύ μορφολογικά παρόμοιων κατηγοριών.

Η μετρική Recall αναδεικνύεται ως ιδιαίτερα κρίσιμη στο συγκεκριμένο πρόβλημα, καθώς σχετίζεται άμεσα με την ικανότητα των μοντέλων να αναγνωρίζουν σωστά τα δείγματα της εκάστοτε κατηγορίας, ιδιαίτερα στις κατηγορίες που αντιστοιχούν σε καρκινικούς ιστούς. Η ανάλυση δείχνει ότι τα απλούστερα μοντέλα, όπως το βασικό CNN και το VGG16, παρουσιάζουν μειωμένο Recall στις κατηγορίες lung_aca και lung_scc, λόγω αυξημένων False Negatives που προκύπτουν από τη σύγχυση μεταξύ των δύο υποτύπων καρκίνου. Αντίστοιχα, το μοντέλο Vision Transformer εμφανίζει χαμηλότερο Recall σε κατηγορίες όπως colon_aca και lung_aca, γεγονός που υποδηλώνει περιορισμένη ικανότητα εντοπισμού όλων των πραγματικών θετικών δειγμάτων. Αντίθετα, τα πιο σύνθετα μοντέλα, και ιδιαίτερα το InceptionResNetV2, επιτυγχάνουν σχεδόν τέλειες τιμές Recall, καταδεικνύοντας την ικανότητά τους να περιορίζουν σημαντικά τα False Negatives.

Η Precision εμφανίζεται γενικά υψηλή σε όλα τα μοντέλα, όπως φαίνεται και στον Πίνακα 6.11, γεγονός που υποδηλώνει ότι οι λανθασμένες θετικές προβλέψεις είναι περιορισμένες. Αυτό σημαίνει ότι τα μοντέλα σπάνια ταξινομούν λανθασμένα ένα δείγμα σε μια κατηγορία όταν αυτό ανήκει σε διαφορετική. Ωστόσο, η υψηλή Precision δεν επαρκεί από μόνη της για την αξιολόγηση της συνολικής απόδοσης, καθώς δεν λαμβάνει υπόψη τις περιπτώσεις όπου το μοντέλο αποτυγχάνει να ανιχνεύσει πραγματικά θετικά δείγματα, κάτι που είναι ιδιαίτερα κρίσιμο σε εφαρμογές ιατρικής ανάλυσης.

Ο F1-score, ως αρμονικός μέσος της Precision και του Recall, παρέχει μια πιο ισορροπημένη εκτίμηση της απόδοσης των μοντέλων. Η ανάλυση δείχνει ότι τα μοντέλα CNN και VGG16 εμφανίζουν χαμηλότερο F1-score στις πιο απαιτητικές κατηγορίες, λόγω της ασυμμετρίας μεταξύ Precision και Recall. Αντίθετα, το VGG19 και το ResNet50 παρουσιάζουν σημαντική βελτίωση, επιτυγχάνοντας πιο ισορροπημένες τιμές και υψηλότερο F1-score, γεγονός που αντανακλά τη μείωση τόσο των False Positives όσο και των False Negatives. Το InceptionResNetV2 επιτυγχάνει τις υψηλότερες τιμές F1-score, επιβεβαιώνοντας την ανωτερότητά του στη συνολική απόδοση και υποδηλώνοντας αποτελεσματικότερο διαχωρισμό των κατηγοριών στον χώρο χαρακτηριστικών. Από την άλλη πλευρά, το Vision Transformer παρουσιάζει χαμηλότερο F1-score σε συγκεκριμένες κατηγορίες, γεγονός που συνδέεται με τη μειωμένη ικανότητά του να αποτυπώνει λεπτές μορφολογικές διαφοροποιήσεις μεταξύ ιστολογικά παρόμοιων κατηγοριών.

Συνολικά, η συνδυαστική ανάλυση των μετρικών του Πίνακα 6.11 καταδεικνύει ότι, παρά την υψηλή συνολική ακρίβεια, η κύρια πρόκληση εντοπίζεται στη διάκριση μεταξύ ιστολογικά παρόμοιων κατηγοριών, όπως οι υποτύποι καρκίνου του πνεύμονα. Τα βαθύτερα και πιο σύνθετα CNN-based μοντέλα, όπως το ResNet50 και ιδιαίτερα το InceptionResNetV2, επιτυγχάνουν την καλύτερη ισορροπία μεταξύ Precision και Recall, οδηγώντας σε υψηλότερο F1-score και πιο αξιόπιστη ταξινόμηση. Αντίθετα, το Vision Transformer, αν και θεωρητικά ιδιαίτερα ισχυρό, εμφανίζει μειωμένη απόδοση στο συγκεκριμένο dataset, γεγονός που πιθανότατα σχετίζεται με την ανάγκη για μεγαλύτερο όγκο δεδομένων ή εκτενέστερη εκπαίδευση.

Τα αποτελέσματα αυτά υπογραμμίζουν τη σημασία επιλογής κατάλληλης αρχιτεκτονικής σε συνδυασμό με τα χαρακτηριστικά του dataset, ιδιαίτερα σε εφαρμογές ιατρικής ανάλυσης όπου η αξιόπιστη ανίχνευση και ο σωστός διαχωρισμός των παθολογικών κατηγοριών αποτελούν κρίσιμους παράγοντες. Στον Πίνακα 6.11 παρουσιάζεται συγκεντρωτικά η σύγκριση των βασικών μετρικών αξιολόγησης των μοντέλων.

Πίνακας 6.11: Πίνακας Σύγκρισης Μετρικών Μοντέλων

Μοντέλο	Precision	Recall	F1-score	Accuracy
CNN	0.98	0.98	0.98	98.03%
VGG16	0.96	0.96	0.96	96.47%
VGG19	0.99	0.99	0.99	98.63%
ResNet50	0.99	0.99	0.99	98.73%
InceptionResNetV2	0.99	0.99	0.99	99.2%
ViT(20 epochs)	0.95	0.95	0.95	94.8%

Κεφάλαιο 7ο: Ερμηνεία Αποτελεσμάτων

7.1 Εφαρμογή Grad-CAM

Η ενσωμάτωση της τεχνικής Grad-CAM αποτελεί κρίσιμο βήμα για την ερμηνεία της συμπεριφοράς του Vision Transformer, μετατοπίζοντας την ανάλυση από καθαρά ποσοτικές μετρικές σε ποιοτική κατανόηση των αποφάσεων του μοντέλου. Ειδικότερα, το Grad-CAM επιτρέπει τον εντοπισμό των περιοχών της εισόδου που επηρεάζουν περισσότερο την τελική πρόβλεψη, παρέχοντας μια μορφή «οπτικής εξήγησης» [7]. Στο πλαίσιο της συγκεκριμένης διπλωματικής εργασίας, όπου το μοντέλο εκπαιδεύεται για ταξινόμηση ιστοπαθολογικών εικόνων, η ανάγκη για ερμηνευσιμότητα είναι ιδιαίτερα έντονη. Οι μετρικές όπως Accuracy, Precision, Recall και F1-score προσφέρουν μια συνολική εικόνα της απόδοσης, δεν αποκαλύπτουν όμως τον τρόπο με τον οποίο το μοντέλο καταλήγει στις αποφάσεις του. Το Grad-CAM καλύπτει αυτό το κενό, επιτρέποντας την ανάλυση της εσωτερικής λειτουργίας του μοντέλου με τρόπο οπτικά κατανοητό [7][19].

Η σημασία της τεχνικής είναι ακόμη μεγαλύτερη στην περίπτωση των Vision Transformers, καθώς τα συγκεκριμένα μοντέλα δεν βασίζονται σε τοπικά φίλτρα συνέλιξης, όπως τα CNNs, αλλά σε μηχανισμούς αυτοπροσοχής, οι οποίοι κατανέμουν δυναμικά τη σημασία σε ολόκληρη την εικόνα [6][24]. Ως αποτέλεσμα, η διαδικασία λήψης απόφασης είναι λιγότερο διαισθητική και πιο δύσκολα ερμηνεύσιμη χωρίς τη χρήση εξειδικευμένων τεχνικών οπτικοποίησης

Ο μηχανισμός λειτουργίας του Grad-CAM βασίζεται στην αξιοποίηση των gradients της εξόδου ως προς τα ενδιάμεσα χαρακτηριστικά του μοντέλου, επιτρέποντας τον υπολογισμό της σχετικής σημαντικότητας κάθε περιοχής της εισόδου [7]. Σε αντίθεση με απλές τεχνικές οπτικοποίησης, το Grad-CAM δεν αποτυπώνει απλώς περιοχές προσοχής, αλλά αναδεικνύει τις περιοχές που έχουν ουσιαστική συμβολή στην τελική απόφαση του μοντέλου. Στην περίπτωση του Vision Transformer, η διαδικασία αυτή προσαρμόζεται ώστε να λειτουργεί σε επίπεδο patches, ακολουθώντας τη δομή της ίδιας της αρχιτεκτονικής [6][24]. Έτσι, η ερμηνεία δεν πραγματοποιείται σε επίπεδο μεμονωμένων pixels, αλλά σε επίπεδο τοπικών περιοχών της εικόνας, οι οποίες συμμετέχουν δυναμικά μέσω του μηχανισμού self-attention.

Η σημασία του Grad-CAM γίνεται ακόμη πιο εμφανής σε εφαρμογές υψηλής ευαισθησίας, όπως η ανάλυση ιατρικών εικόνων, όπου δεν αρκεί μόνο η επίτευξη υψηλής ακρίβειας, αλλά απαιτείται και η δυνατότητα αιτιολόγησης των αποφάσεων του μοντέλου. Μέσω των παραγόμενων heatmaps, καθίσταται δυνατή η επαλήθευση του κατά πόσο το μοντέλο βασίζεται πράγματι σε μορφολογικά χαρακτηριστικά που σχετίζονται με την παθολογία ή αν επηρεάζεται από άσχετες περιοχές, θόρυβο ή ανεπιθύμητα μοτίβα [19]. Η δυνατότητα αυτή επιτρέπει τον εντοπισμό λανθασμένων συσχετίσεων, οι οποίες διαφορετικά θα παρέμεναν κρυφές, μειώνοντας την αξιοπιστία του συστήματος σε πραγματικές συνθήκες.

Παράλληλα, το Grad-CAM λειτουργεί και ως εργαλείο διάγνωσης σφαλμάτων του μοντέλου. Μέσω της ανάλυσης των χαρτών ενεργοποίησης, μπορούν να εντοπιστούν περιπτώσεις όπου το μοντέλο αποτυγχάνει να εστιάσει στις κρίσιμες περιοχές της εικόνας, οδηγώντας σε ψευδώς αρνητικές προβλέψεις, ή αντίθετα περιπτώσεις όπου εμφανίζει ισχυρές ενεργοποιήσεις σε μη σχετικές περιοχές, προκαλώντας ψευδώς θετικά αποτελέσματα. Επιπλέον, σε περιπτώσεις χαμηλής βεβαιότητας παρατηρούνται διάχυτες και ασαφείς ενεργοποιήσεις, γεγονός που υποδηλώνει αδυναμία του μοντέλου να απομονώσει σαφή και διακριτικά χαρακτηριστικά. Η ιδιότητα αυτή καθιστά το Grad-CAM όχι μόνο εργαλείο ερμηνείας, αλλά και μηχανισμό εντοπισμού των αδυναμιών της αρχιτεκτονικής.

Τέλος, η συμβολή του Grad-CAM επεκτείνεται και στη διαδικασία βελτιστοποίησης του μοντέλου, λειτουργώντας ως μηχανισμός ανατροφοδότησης (feedback mechanism). Μέσα από την παρατήρηση των περιοχών στις οποίες εστιάζει το μοντέλο, μπορούν να εξαχθούν συμπεράσματα σχετικά με την ποιότητα των δεδομένων και την αποτελεσματικότητα της εκπαίδευσης. Για παράδειγμα, η συστηματική εστίαση σε περιοχές φόντου μπορεί να υποδηλώνει ανάγκη για καλύτερο data augmentation ή καθαρισμό του dataset, ενώ η αγνόηση κρίσιμων ιστολογικών χαρακτηριστικών μπορεί να σχετίζεται με ανεπαρκή εκπαίδευση ή ακατάλληλη ρύθμιση υπερπαραμέτρων. Συνεπώς, το Grad-CAM δεν αποτελεί μόνο τεχνική ερμηνείας, αλλά και εργαλείο αξιολόγησης της ποιότητας της μάθησης και καθοδήγησης της διαδικασίας βελτιστοποίησης του μοντέλου.

7.2 Οπτικοποίηση περιοχών ενδιαφέροντος

Η οπτικοποίηση μέσω της χρήσης Grad-CAM αποτελεί το τελικό στάδιο της διαδικασίας ερμηνείας του μοντέλου Vision Transformer που αναπτύχθηκε στο πλαίσιο της παρούσας έρευνας [7][19]. Μέσω αυτής της διαδικασίας, οι αφηρημένες αριθμητικές τιμές που προκύπτουν από τα gradients και τα embeddings μετατρέπονται σε οπτικές αναπαραστάσεις, οι οποίες αποτυπώνουν τη χωρική κατανομή της σημαντικότητας στην εικόνα. Η συγκεκριμένη απεικόνιση επιτρέπει την άμεση κατανόηση του τρόπου με τον οποίο το μοντέλο λαμβάνει τις αποφάσεις του, γεφυρώνοντας το χάσμα μεταξύ της υπολογιστικής λειτουργίας του δικτύου και της ανθρώπινης ερμηνείας.

Η διαδικασία ξεκινά με τη δημιουργία ενός χάρτη σημαντικότητας σε επίπεδο patches, ο οποίος προκύπτει από τη στάθμιση των embeddings με βάση τα gradients [7]. Ο χάρτης αυτός αποτελεί μια μονοδιάστατη ακολουθία τιμών, όπου κάθε τιμή αντιστοιχεί στη συμβολή ενός συγκεκριμένου patch στην τελική πρόβλεψη του μοντέλου. Για να αποκτήσει χωρική σημασία, η ακολουθία αυτή αναδιατάσσεται σε δισδιάστατη μορφή, σύμφωνα με τη γεωμετρική διάταξη των patches στην αρχική εικόνα. Στη συνέχεια εφαρμόζεται διαδικασία παρεμβολής, ώστε ο χάρτης να αποκτήσει τις ίδιες διαστάσεις με την εικόνα εισόδου, επιτρέποντας την ακριβή επικάλυψή του πάνω στην αρχική εικόνα.

Το αποτέλεσμα της διαδικασίας είναι ένα heatmap, στο οποίο οι περιοχές υψηλής σημαντικότητας απεικονίζονται με θερμά χρώματα, όπως κόκκινο και κίτρινο, ενώ οι περιοχές χαμηλής σημαντικότητας με ψυχρά χρώματα, όπως μπλε και σκούρο κυανό. Η επικάλυψη του heatmap πάνω στην αρχική εικόνα δημιουργεί μια σύνθετη αναπαράσταση, η οποία συνδυάζει την οπτική πληροφορία της εισόδου με την εσωτερική «προσοχή» του μοντέλου. Με τον τρόπο αυτό καθίσταται δυνατή η άμεση οπτική σύγκριση μεταξύ σημαντικών και μη σημαντικών περιοχών, επιτρέποντας στον ερευνητή να κατανοήσει ποια μορφολογικά χαρακτηριστικά επηρεάζουν περισσότερο την απόφαση του μοντέλου.

Ένα από τα βασικά χαρακτηριστικά της οπτικοποίησης στο Vision Transformer είναι η πιο διάχυτη κατανομή των ενεργοποιήσεων. Σε αντίθεση με τα CNNs, όπου οι ενεργοποιήσεις είναι συνήθως εντοπισμένες σε συγκεκριμένες περιοχές υψηλής έντασης, στο ViT οι περιοχές ενδιαφέροντος εμφανίζονται πιο εκτεταμένες και λιγότερο συγκεντρωμένες. Η συμπεριφορά αυτή οφείλεται στον μηχανισμό self-attention, ο οποίος επιτρέπει την αλληλεπίδραση μεταξύ όλων των patches της εικόνας και οδηγεί σε πιο ολιστική κατανομή της πληροφορίας [6][24]. Συνεπώς, τα heatmaps του Vision Transformer δεν πρέπει να ερμηνεύονται ως ακριβής εντοπισμός συγκεκριμένων δομών, αλλά ως ένδειξη της συνολικής σημασιολογικής συμβολής των επιμέρους περιοχών στην τελική απόφαση.

Η ανάλυση των οπτικοποιήσεων παρέχει ιδιαίτερα σημαντικές πληροφορίες για τη συμπεριφορά του μοντέλου. Σε περιπτώσεις σωστής ταξινόμησης, το Grad-CAM αναδεικνύει περιοχές που σχετίζονται άμεσα με τα μορφολογικά χαρακτηριστικά της εκάστοτε κατηγορίας, επιβεβαιώνοντας ότι το μοντέλο έχει μάθει ουσιαστικά και όχι επιφανειακά πρότυπα. Οι ενεργοποιήσεις εμφανίζονται

πιο συγκεντρωμένες σε ιστολογικές δομές υψηλής σημασίας, γεγονός που υποδηλώνει ότι το μοντέλο αξιοποιεί πραγματικά διαγνωστικά χαρακτηριστικά.

Αντίθετα, σε περιπτώσεις λανθασμένων προβλέψεων, τα heatmaps συχνά αποκαλύπτουν ότι το μοντέλο εστιάζει σε άσχετες ή παραπλανητικές περιοχές, όπως περιοχές φόντου, θόρυβο ή δευτερεύοντα μορφολογικά στοιχεία. Η συμπεριφορά αυτή αποτελεί ένδειξη ότι το μοντέλο δεν έχει καταφέρει να μάθει πλήρως τα διακριτικά χαρακτηριστικά της κατηγορίας ή ότι επηρεάζεται από ανεπιθύμητες συσχετίσεις. Σε αρκετές περιπτώσεις χαμηλής βεβαιότητας, οι ενεργοποιήσεις εμφανίζονται διάχυτες και ασαφείς, χωρίς σαφή περιοχή συγκέντρωσης, γεγονός που αντανακλά αδυναμία του μοντέλου να εντοπίσει σταθερά και αξιόπιστα πρότυπα

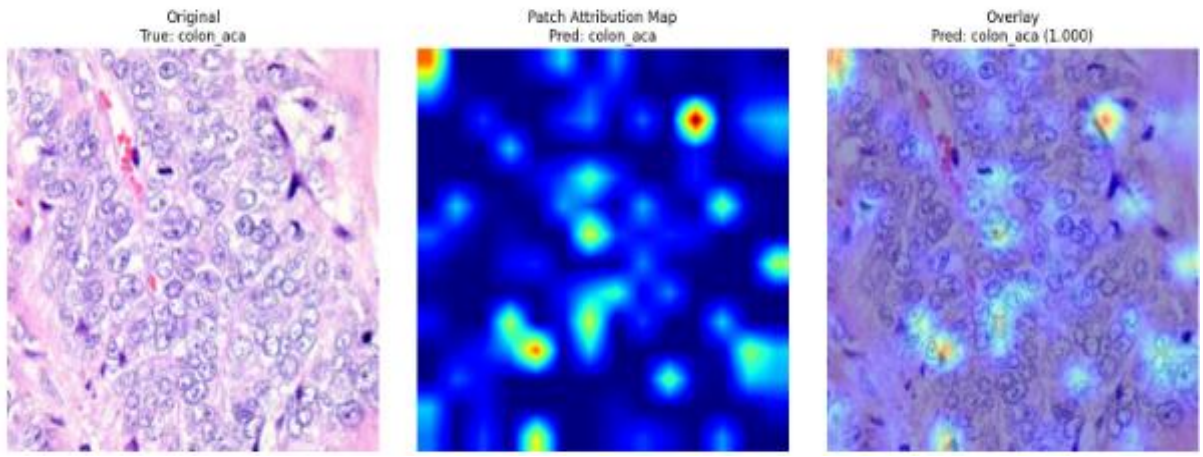
Επιπλέον, η οπτικοποίηση των περιοχών ενδιαφέροντος μπορεί να χρησιμοποιηθεί ως εργαλείο σύγκρισης μεταξύ διαφορετικών αρχιτεκτονικών. Σε σύγκριση με CNN-based προσεγγίσεις, το Vision Transformer εμφανίζει πιο “global” μορφή προσοχής, ενώ τα CNNs τείνουν να εστιάζουν περισσότερο σε τοπικά χαρακτηριστικά, όπως ακμές, υφές και μικρές περιοχές υψηλής αντίθεσης [6][24]. Η διαφορά αυτή αντικατοπτρίζεται άμεσα στα heatmaps και συμβάλλει στην ερμηνεία των διαφορών που παρατηρούνται στην απόδοση των μοντέλων.

Η σημασία της διαδικασίας οπτικοποίησης δεν περιορίζεται μόνο στην ερμηνεία της λειτουργίας του μοντέλου, αλλά επεκτείνεται και στη διαδικασία βελτιστοποίησής του. Μέσω της παρατήρησης των heatmaps μπορούν να εντοπιστούν συστηματικά προβλήματα, όπως υπερεξάρτηση από συγκεκριμένες περιοχές της εικόνας ή αδυναμία εστίασης σε κρίσιμα μορφολογικά χαρακτηριστικά. Οι πληροφορίες αυτές μπορούν να οδηγήσουν σε στοχευμένες παρεμβάσεις, όπως βελτίωση του dataset, εφαρμογή αποτελεσματικότερων τεχνικών data augmentation ή τροποποίηση της αρχιτεκτονικής και των υπερπαραμέτρων του μοντέλου.

Συνεπώς, η οπτικοποίηση μέσω Grad-CAM δεν αποτελεί απλώς ένα εργαλείο παρουσίασης αποτελεσμάτων, αλλά έναν ουσιαστικό μηχανισμό κατανόησης, αξιολόγησης και βελτίωσης της συμπεριφοράς του Vision Transformer. Μέσα από τη δυνατότητα ανάλυσης των περιοχών στις οποίες εστιάζει το μοντέλο, παρέχεται μια βαθύτερη κατανόηση της διαδικασίας λήψης αποφάσεων, ενισχύοντας τόσο την αξιοπιστία όσο και τη διαφάνεια της τεχνητής νοημοσύνης σε εφαρμογές ιατρικής απεικόνισης.

7.3 Ανάλυση αποφάσεων του μοντέλου και σχέση των ενεργοποιημένων περιοχών με μορφολογικά χαρακτηριστικά καρκίνου

Η ανάλυση των αποφάσεων του Vision Transformer μέσω των Grad-CAM heatmaps κατά τις 10, 15 και 20 εποχές εκπαίδευσης αποκαλύπτει με σαφήνεια τη δυναμική εξέλιξη της μαθησιακής διαδικασίας του μοντέλου, τόσο σε επίπεδο εστίασης της προσοχής όσο και σε επίπεδο σημασιολογικής κατανόησης της εικόνας. Η παρατήρηση της μεταβολής των heatmaps ανά στάδιο εκπαίδευσης επιτρέπει την εις βάθος μελέτη της εξέλιξης των εσωτερικών αναπαραστάσεων του μοντέλου και του τρόπου με τον οποίο αυτό οργανώνει σταδιακά τον χώρο χαρακτηριστικών.



Σχήμα 7.1: Grad-CAM αποφάσεων 10 epochs εκπαίδευση

Στο αρχικό στάδιο των 10 εποχών, το μοντέλο παρουσιάζει έντονα χαρακτηριστικά αστάθειας, καθώς οι ενεργοποιήσεις κατανέμονται διάχυτα σε ολόκληρη την εικόνα, χωρίς σαφή συγκέντρωση σε συγκεκριμένες περιοχές ενδιαφέροντος. Η συμπεριφορά αυτή δεν περιορίζεται απλώς σε μια «αδύναμη» εστίαση, αλλά αντανακλά μια βαθύτερη αδυναμία του μοντέλου να κατασκευάσει συνεκτικές εσωτερικές αναπαραστάσεις. Τα Grad-CAM heatmaps εμφανίζονται θορυβώδη και μη δομημένα, γεγονός που υποδηλώνει ότι το μοντέλο δεν έχει ακόμη αναπτύξει ικανότητα ιεράρχησης των χαρακτηριστικών της εικόνας. Αντί να αναγνωρίζει μορφολογικά μοτίβα που σχετίζονται με την εκάστοτε κλάση, φαίνεται να ανταποκρίνεται σε επιφανειακά στοιχεία, όπως έντονες χρωματικές αντιθέσεις ή γενικές υφές.

Ως αποτέλεσμα, οι αποφάσεις του χαρακτηρίζονται από χαμηλή αξιοπιστία και υψηλή μεταβλητότητα, με συχνή εμφάνιση τόσο ψευδώς θετικών όσο και ψευδώς αρνητικών προβλέψεων. Οι περιοχές ενεργοποίησης παρουσιάζουν μεγάλη διασπορά και δεν συνδέονται σταθερά με τις κρίσιμες ιστολογικές δομές της εικόνας, γεγονός που υποδηλώνει ότι το μοντέλο δεν έχει ακόμη καταφέρει να απομονώσει ουσιαστικά χαρακτηριστικά υψηλής σημασιολογικής αξίας. Η συνολική αυτή εικόνα επιβεβαιώνει την ύπαρξη φαινομένων underfitting, όπου το μοντέλο δεν έχει ακόμη αποκτήσει την απαιτούμενη ικανότητα γενίκευσης και κατανόησης της δομής των δεδομένων.

Παράλληλα, τα heatmaps των 10 εποχών αποκαλύπτουν ότι ο μηχανισμός self-attention δεν έχει ακόμη σταθεροποιηθεί. Η πληροφορία φαίνεται να κατανέμεται σχεδόν ομοιόμορφα μεταξύ των patches, χωρίς να αναδεικνύονται σαφείς περιοχές προτεραιότητας. Αυτό έχει ως αποτέλεσμα το μοντέλο να αντιμετωπίζει δυσκολία στη διάκριση μεταξύ σημαντικών και μη σημαντικών μορφολογικών στοιχείων, οδηγώντας σε ασαφή και μη διαχωρίσιμα decision boundaries. Η συμπεριφορά αυτή είναι ιδιαίτερα εμφανής στις κατηγορίες καρκίνου πνεύμονα, όπου οι μορφολογικές διαφορές είναι λεπτές και απαιτούν πιο εξειδικευμένη αναπαράσταση χαρακτηριστικών.

Επιπλέον, η απουσία συγκεντρωμένων ενεργοποιήσεων υποδηλώνει ότι το μοντέλο λειτουργεί ακόμη σε χαμηλό επίπεδο αφαιρετικότητας. Οι αναπαραστάσεις που δημιουργούνται παραμένουν κοντά στα αρχικά οπτικά χαρακτηριστικά της εικόνας και δεν έχουν εξελιχθεί σε υψηλού επιπέδου σημασιολογικά embeddings. Αυτό αντανακλάται άμεσα τόσο στην ποιοτική μορφή των heatmaps όσο και στις ποσοτικές μετρικές απόδοσης, οι οποίες στις 10 εποχές εμφανίζονται σημαντικά χαμηλότερες σε σχέση με τα μεταγενέστερα στάδια εκπαίδευσης.

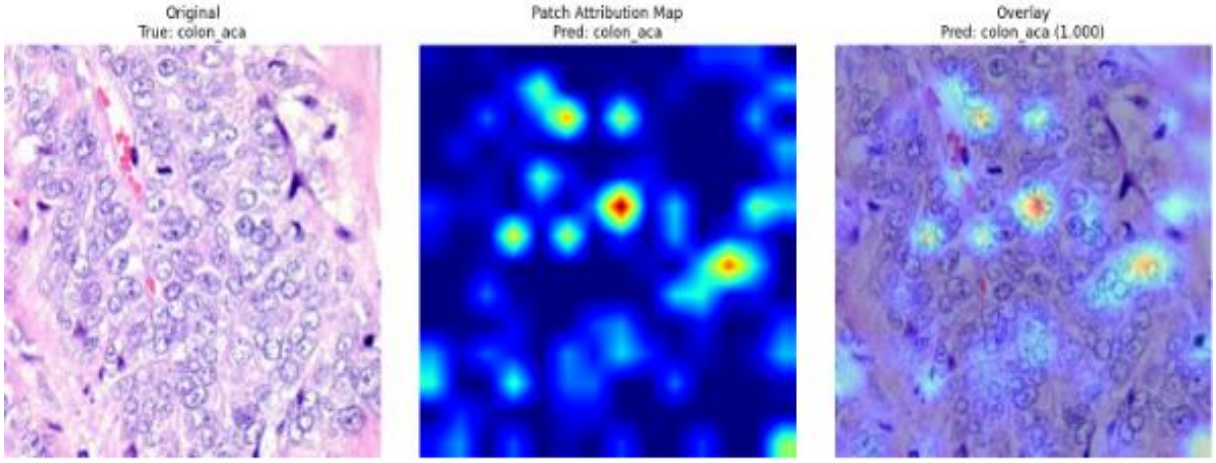
Στην περίπτωση του σχήματος 7.1, παρά τη συνολικά διάχυτη και μη πλήρως σταθεροποιημένη κατανομή των ενεργοποιήσεων, παρατηρούνται ορισμένες θερμές περιοχές του patch attribution map

που παρουσιάζουν μερική συσχέτιση με ιστοπαθολογικά χαρακτηριστικά του δείγματος. Ειδικότερα, οι ενεργοποιήσεις φαίνεται να εμφανίζονται σε περιοχές όπου παρατηρείται αυξημένη κυτταρική πυκνότητα, παρουσία πυρηνικών δομών και ανομοιογενής μορφολογία ιστού, στοιχεία που μπορούν να σχετίζονται με την κατηγορία colon_aca.

Ωστόσο, η εστίαση του μοντέλου δεν είναι ακόμη πλήρως συγκεντρωμένη ούτε αυστηρά περιορισμένη σε συγκεκριμένες διαγνωστικά κρίσιμες ιστολογικές δομές. Αντίθετα, οι θερμές περιοχές εμφανίζονται και σε διάσπαρτα σημεία της εικόνας, γεγονός που υποδηλώνει ότι το μοντέλο εξακολουθεί να αξιοποιεί συνδυασμό ουσιαστικών μορφολογικών μοτίβων, γενικών χαρακτηριστικών υφής και πιθανώς λιγότερο σχετικών οπτικών πληροφοριών.

Η συμπεριφορά αυτή είναι συμβατή με το πρώιμο στάδιο εκπαίδευσης των 10 εποχών, όπου το Vision Transformer έχει αρχίσει να εντοπίζει περιοχές με πιθανή διαγνωστική αξία, χωρίς όμως να έχει ακόμη αναπτύξει πλήρως σταθερές και σημασιολογικά εξειδικευμένες αναπαραστάσεις. Συνεπώς, το αποτέλεσμα μπορεί να θεωρηθεί εν μέρει ερμηνεύσιμο, αλλά η βιολογική σημασία των ενεργοποιήσεων παραμένει περιορισμένη και απαιτεί προσεκτική αξιολόγηση, ιδανικά από ειδικό παθολογοανατόμο.

Συνεπώς, τα Grad-CAM heatmaps των 10 εποχών λειτουργούν ως σαφής ένδειξη ότι το Vision Transformer βρίσκεται ακόμη σε πρώιμο στάδιο μάθησης, όπου η πληροφορία επεξεργάζεται αποσπασματικά και χωρίς επαρκή ιεραρχική οργάνωση. Η ανάλυση αυτή επιβεβαιώνει ότι η αρχιτεκτονική ViT απαιτεί μεγαλύτερη διάρκεια εκπαίδευσης προκειμένου να αξιοποιήσει αποτελεσματικά τον μηχανισμό self-attention και να διαμορφώσει σταθερές και σημασιολογικά πλούσιες αναπαραστάσεις.



Σχήμα 7.2: Grad-CAM αποφάσεων 15 epochs εκπαίδευση

Καθώς η εκπαίδευση προχωρά στις 15 εποχές, παρατηρείται μια ουσιαστική μεταβολή στη συμπεριφορά του μοντέλου, η οποία αντανακλάται άμεσα και στα Grad-CAM heatmaps. Οι ενεργοποιήσεις αρχίζουν να εμφανίζουν μεγαλύτερη συγκέντρωση και να διαμορφώνουν πρώιμες δομές προσοχής, υποδηλώνοντας ότι το μοντέλο αρχίζει να αναγνωρίζει περιοχές αυξημένης σημασίας μέσα στις ιστοπαθολογικές εικόνες. Η αλλαγή αυτή αποτελεί σαφή ένδειξη ότι το Vision Transformer μεταβαίνει σταδιακά από μια γενικευμένη και αποσπασματική επεξεργασία της εικόνας σε μια πιο οργανωμένη και σημασιολογικά καθοδηγούμενη ανάλυση.

Στο σχήμα 7.2 προκύπτει ότι, οι θερμές περιοχές του patch attribution map παρουσιάζουν εμφανή συσχέτιση με μορφολογικά και ιστοπαθολογικά χαρακτηριστικά του εξεταζόμενου δείγματος. Ειδικότερα, οι περιοχές υψηλής ενεργοποίησης εντοπίζονται κυρίως σε σημεία όπου παρατηρείται

αυξημένη κυτταρική πυκνότητα, έντονη παρουσία πυρηνικών δομών και σχετική ιστολογική ανομοιογένεια, χαρακτηριστικά τα οποία συνάδουν με τη μορφολογία αδενοκαρκινωματικού ιστού της κατηγορίας colon_aca.

Η οπτική επικάλυψη των ενεργοποιήσεων επί της αρχικής εικόνας καταδεικνύει ότι το μοντέλο δεν στηρίζεται σε τυχαία ή άσχετα τμήματα του ιστοτεμαχίου, αλλά εστιάζει σε περιοχές που φέρουν διαγνωστική πληροφορία. Παράλληλα, η χωρική κατανομή των θερμών περιοχών εμφανίζεται σχετικά διάχυτη και όχι αυστηρά περιορισμένη σε μία μόνο ιστολογική δομή, γεγονός που υποδηλώνει ότι το μοντέλο αξιοποιεί συνδυαστικά πολλαπλά τοπικά μορφολογικά πρότυπα και χαρακτηριστικά υφής.

Το εύρημα αυτό είναι ιδιαίτερα σημαντικό, καθώς δείχνει ότι ήδη από τις 15 εποχές εκπαίδευσης το μοντέλο έχει αρχίσει να αναπτύσσει έναν βαθμό ερμηνεύσιμης εστίασης σε βιολογικά συναφείς περιοχές. Ωστόσο, η παρουσία διάσπαρτων ενεργοποιήσεων υποδεικνύει ότι η προσοχή του μοντέλου δεν έχει ακόμη πλήρως εξειδικευθεί, γεγονός που σημαίνει ότι, παρά τη σαφή συσχέτιση με ουσιώδη ιστοπαθολογικά χαρακτηριστικά, η ερμηνευτική ακρίβεια των ενεργοποιήσεων παραμένει σε ένα ενδιάμεσο στάδιο ωρίμανσης.

Συνεπώς, το αποτέλεσμα μπορεί να θεωρηθεί ικανοποιητικά ερμηνεύσιμο, επιβεβαιώνοντας ότι το μοντέλο αντλεί την απόφασή του από περιοχές με μορφολογικό ενδιαφέρον, χωρίς όμως να έχει επιτύχει ακόμη πλήρως εντοπισμένη και απόλυτα εξειδικευμένη εστίαση.

Η ευθυγράμμιση αυτή, αν και ακόμη μερική, δείχνει ότι το μοντέλο παύει να βασίζεται αποκλειστικά σε επιφανειακά χαρακτηριστικά, όπως έντονες χρωματικές διαφοροποιήσεις ή τυχαίες υφές, και αρχίζει να αναγνωρίζει μορφολογικά μοτίβα που σχετίζονται περισσότερο με τη δομή των ιστών και τις παθολογικές αλλοιώσεις. Τα heatmaps εμφανίζονται πλέον λιγότερο θορυβώδη και περισσότερο συγκεντρωμένα γύρω από συγκεκριμένες περιοχές ενδιαφέροντος, γεγονός που υποδηλώνει βελτίωση της εσωτερικής οργάνωσης του χώρου χαρακτηριστικών.

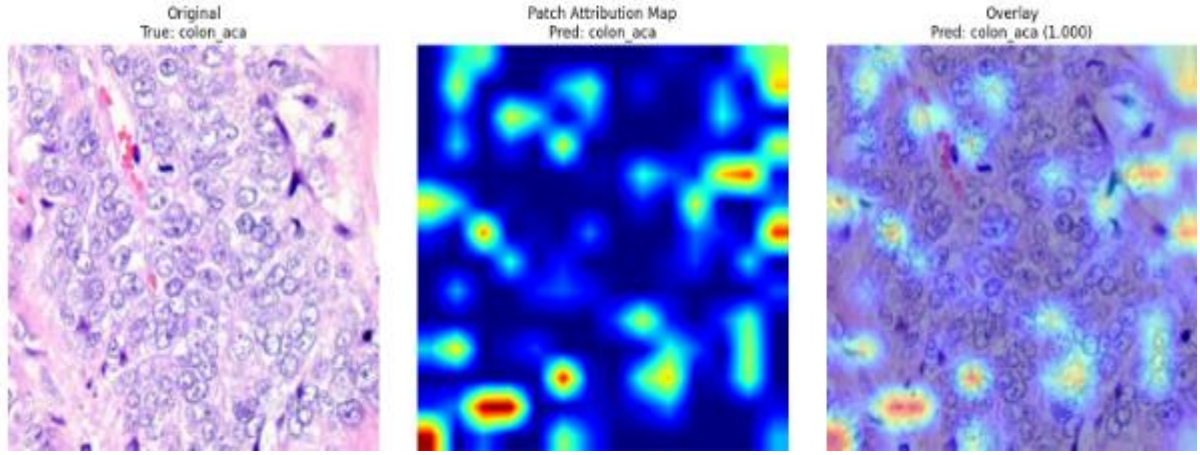
Παρόλα αυτά, η προσοχή του μοντέλου δεν έχει ακόμη πλήρως σταθεροποιηθεί. Παρατηρείται ταυτόχρονη ενεργοποίηση τόσο σε περιοχές υψηλής σημασιολογικής αξίας όσο και σε δευτερεύουσες ή λιγότερο σχετικές περιοχές της εικόνας. Η συμπεριφορά αυτή υποδηλώνει ότι το μοντέλο βρίσκεται ακόμη σε διαδικασία αναδιοργάνωσης των embeddings και των attention patterns, επιχειρώντας να διαχωρίσει τα κρίσιμα μορφολογικά χαρακτηριστικά από τον οπτικό θόρυβο. Με άλλα λόγια, το μοντέλο έχει αρχίσει να αναπτύσσει πιο ουσιαστικές αναπαραστάσεις, χωρίς όμως να έχει επιτύχει πλήρη συνέπεια και σταθερότητα στις αποφάσεις του.

Η μεταβολή αυτή είναι ιδιαίτερα εμφανής στις κατηγορίες που χαρακτηρίζονται από υψηλή μορφολογική ομοιότητα, όπως οι υποτύποι καρκίνου του πνεύμονα. Σε αρκετές περιπτώσεις, τα heatmaps αποκαλύπτουν ότι το μοντέλο αρχίζει να εστιάζει σε ιστολογικές δομές που σχετίζονται με τις καρκινικές αλλοιώσεις, ωστόσο μέρος της προσοχής εξακολουθεί να κατανέμεται σε γειτονικές ή μη κρίσιμες περιοχές. Το γεγονός αυτό αντανακλά τη μερική μόνο επίλυση του προβλήματος της επικάλυψης μεταξύ των κλάσεων στον χώρο χαρακτηριστικών.

Σε επίπεδο εσωτερικών αναπαραστάσεων, η φάση των 15 εποχών χαρακτηρίζεται από αυξημένη συμπύκνωση της πληροφορίας. Τα embeddings αρχίζουν να αποκτούν μεγαλύτερη σημασιολογική συνοχή, ενώ οι μηχανισμοί self-attention εμφανίζουν πιο σταθερά μοτίβα ενεργοποίησης. Οι attention maps παρουσιάζουν σαφέστερες περιοχές εστίασης, γεγονός που υποδηλώνει ότι το μοντέλο αρχίζει να αναπτύσσει ικανότητα ιεράρχησης των χαρακτηριστικών της εικόνας.

Η εξέλιξη αυτή αντικατοπτρίζεται άμεσα και στις ποσοτικές μετρικές απόδοσης. Παρατηρείται αισθητή βελτίωση στο Recall και στον F1-score, ιδιαίτερα στις πιο απαιτητικές κατηγορίες, γεγονός που υποδηλώνει μείωση των ψευδώς αρνητικών προβλέψεων. Παρόλα αυτά, η παρουσία ενεργοποιήσεων σε δευτερεύουσες περιοχές συνεχίζει να οδηγεί σε ορισμένα ψευδώς θετικά αποτελέσματα, αποκαλύπτοντας ότι το μοντέλο εξακολουθεί να εμφανίζει ευαισθησία σε μη κρίσιμα χαρακτηριστικά.

Η φάση αυτή μπορεί να χαρακτηριστεί ως μεταβατικό στάδιο της εκπαίδευσης, όπου το μοντέλο μετακινείται από την αδυναμία ουσιαστικής κατανόησης των δεδομένων προς μια πιο σταθερή και αξιόπιστη λειτουργία. Τα Grad-CAM heatmaps αποτυπώνουν με σαφήνεια αυτή τη μετάβαση, παρουσιάζοντας μια ενδιάμεση κατάσταση όπου η πληροφορία αρχίζει να οργανώνεται σημασιολογικά, χωρίς όμως να έχει ακόμη επιτευχθεί πλήρης διαχωρισμός και σταθεροποίηση των αναπαραστάσεων.



Σχήμα 7.3: Grad-CAM αποφάσεων 20 epochs εκπαίδευση

Στο στάδιο των 20 εποχών, το μοντέλο εμφανίζει σαφώς πιο ώριμη και σταθερή συμπεριφορά, με τα Grad-CAM heatmaps να παρουσιάζουν πιο καθαρά, έντονα και χωρικά οργανωμένα μοτίβα ενεργοποίησης. Η προσοχή του μοντέλου δεν εμφανίζεται πλέον στον ίδιο βαθμό διάχυτη ή ασταθής, αλλά τείνει να εστιάζεται περισσότερο σε συγκεκριμένα patches που αντιστοιχούν σε ουσιαστικά μορφολογικά χαρακτηριστικά της εικόνας. Η μεταβολή αυτή υποδηλώνει ότι το Vision Transformer έχει πλέον αναπτύξει πιο σταθερές και σημασιολογικά πλούσιες εσωτερικές αναπαραστάσεις, οι οποίες του επιτρέπουν να απομονώνει τις κρίσιμες περιοχές της εικόνας με μεγαλύτερη ακρίβεια.

Στην περίπτωση του σχήματος 7.3, το οποίο αντιστοιχεί στην εκπαίδευση του Vision Transformer για 20 εποχές, παρατηρείται περαιτέρω εξέλιξη της χωρικής κατανομής των ενεργοποιήσεων σε σχέση με τα προηγούμενα στάδια εκπαίδευσης. Οι θερμές περιοχές του patch attribution map εμφανίζονται περισσότερο οργανωμένες και παρουσιάζουν σαφέστερη σύνδεση με περιοχές του ιστοτεμαχίου που φέρουν μορφολογικό ενδιαφέρον. Η πρόβλεψη του μοντέλου παραμένει σωστή για την κλάση colon_aca, γεγονός που ενισχύει την ερμηνεία ότι οι ενεργοποιήσεις σχετίζονται με χαρακτηριστικά που συμβάλλουν στη διάκριση του αδενοκαρκινωματικού ιστού.

Ειδικότερα, στο overlay παρατηρείται ότι αρκετές θερμές περιοχές εντοπίζονται κοντά σε περιοχές με αυξημένη κυτταρική πυκνότητα, εμφανείς πυρηνικές δομές και ανομοιογενή αρχιτεκτονική ιστού. Τα χαρακτηριστικά αυτά μπορούν να θεωρηθούν ιστοπαθολογικά συναφή, καθώς η κλάση colon_aca συνδέεται με διαταραχή της φυσιολογικής ιστικής οργάνωσης, αυξημένη κυτταρικότητα και μορφολογική ετερογένεια. Σε σύγκριση με τις 10 εποχές, όπου οι ενεργοποιήσεις εμφανίζονταν

περισσότερο διάχυτες και λιγότερο σταθερές, στην περίπτωση των 20 εποχών η εστίαση του μοντέλου φαίνεται να έχει αποκτήσει μεγαλύτερο βαθμό οργάνωσης.

Παρά τη βελτιωμένη αυτή συμπεριφορά, οι θερμές περιοχές δεν περιορίζονται αποκλειστικά σε μία ενιαία και αυστηρά καθορισμένη ιστολογική δομή. Αντίθετα, παρατηρείται ακόμη παρουσία ενεργοποιήσεων σε διάφορα σημεία της εικόνας, γεγονός που δείχνει ότι το μοντέλο εξακολουθεί να αξιοποιεί συνδυασμό τοπικών μορφολογικών μοτίβων, χαρακτηριστικών υφής και χρωματικών διαφοροποιήσεων. Η συμπεριφορά αυτή είναι αναμενόμενη σε ιστοπαθολογικές εικόνες, όπου η διαγνωστική πληροφορία δεν εντοπίζεται πάντα σε ένα μοναδικό σημείο, αλλά κατανέμεται σε πολλαπλές μικροπεριοχές του ιστού.

Με βάση και την παρατήρηση του σχήματος 7.3, η διαφοροποίηση σε σχέση με τα προηγούμενα στάδια εκπαίδευσης γίνεται ακόμη πιο εμφανής. Ενώ στις 10 και 15 εποχές οι ενεργοποιήσεις παρουσίαζαν μεγαλύτερη διασπορά ή μερική μόνο συγκέντρωση, στα heatmaps των 20 εποχών παρατηρείται σαφέστερη εστίαση σε περιοχές που σχετίζονται με ιστολογικές αλλοιώσεις και μορφολογικές δομές της εκάστοτε κατηγορίας. Το μοντέλο φαίνεται πλέον να περιορίζει σε μεγαλύτερο βαθμό την επίδραση του οπτικού θορύβου και των δευτερευόντων χαρακτηριστικών, γεγονός που αποτελεί ένδειξη ότι οι attention μηχανισμοί έχουν σταθεροποιηθεί περισσότερο και ότι η πληροφορία ιεραρχείται αποτελεσματικότερα.

Η μετάβαση αυτή αντικατοπτρίζει τη δημιουργία υψηλού επιπέδου σημασιολογικών embeddings, μέσα από τα οποία το μοντέλο μπορεί να διακρίνει με μεγαλύτερη σαφήνεια μεταξύ διαφορετικών ιστολογικών κατηγοριών. Τα attention maps παρουσιάζουν πλέον περισσότερο συνεπή και επαναλήψιμα μοτίβα, γεγονός που υποδηλώνει ότι το Vision Transformer έχει καταφέρει να οργανώσει τον χώρο χαρακτηριστικών σε πιο διακριτές και συμπαγείς περιοχές. Οι ενεργοποιήσεις δεν κατανέμονται πλέον ομοιόμορφα στον ίδιο βαθμό, αλλά εμφανίζουν σαφέστερη ιεράρχηση ως προς τη σημασία των patches.

Η σταθερότητα αυτή είναι ιδιαίτερα εμφανής στις κατηγορίες όπου προηγουμένως παρατηρούνταν σημαντική μορφολογική επικάλυψη, όπως στις κατηγορίες lung_aca και lung_scc. Σε αρκετές περιπτώσεις, τα heatmaps αποκαλύπτουν ότι το μοντέλο έχει αρχίσει να απομονώνει πιο εξειδικευμένα ιστολογικά χαρακτηριστικά, τα οποία σχετίζονται άμεσα με τη διαφοροποίηση των δύο υποτύπων καρκίνου πνεύμονα. Παρόλο που εξακολουθούν να υπάρχουν ορισμένες περιοχές αβεβαιότητας, η συνολική συμπεριφορά είναι σαφώς πιο συνεπής και αξιόπιστη σε σχέση με τα προηγούμενα στάδια εκπαίδευσης.

Οι λιγότερο σχετικές περιοχές της εικόνας φαίνεται πλέον να επηρεάζουν σε μικρότερο βαθμό την τελική απόφαση, γεγονός που υποδηλώνει αποτελεσματικότερο φιλτράρισμα του θορύβου και καλύτερη εστίαση στα πραγματικά διαγνωστικά χαρακτηριστικά. Η εξέλιξη αυτή είναι ιδιαίτερα σημαντική σε εφαρμογές ιατρικής ανάλυσης, καθώς ενισχύει την αξιοπιστία του μοντέλου και μειώνει την πιθανότητα λανθασμένων συσχετίσεων. Τα heatmaps εμφανίζονται πλέον πιο «καθαρά» και περισσότερο ευθυγραμμισμένα με τις περιοχές ιστολογικού ενδιαφέροντος, γεγονός που υποδηλώνει βαθύτερη κατανόηση της δομής των δεδομένων.

Η σταθεροποίηση των ενεργοποιήσεων αντικατοπτρίζεται άμεσα και στις ποσοτικές μετρικές απόδοσης. Παρατηρείται σημαντική μείωση τόσο των ψευδώς θετικών όσο και των ψευδώς αρνητικών προβλέψεων, καθώς και βελτίωση της ισορροπίας μεταξύ Precision και Recall. Το γεγονός αυτό υποδηλώνει ότι τα decision boundaries έχουν γίνει πιο σαφή και πιο σταθερά, επιτρέποντας στο μοντέλο να λαμβάνει πιο αξιόπιστες αποφάσεις ακόμη και σε δύσκολες περιπτώσεις.

Σε επίπεδο representation learning, το στάδιο των 20 εποχών αντιστοιχεί σε μια πιο ώριμη φάση εκπαίδευσης, όπου τα embeddings εμφανίζουν αυξημένη σημασιολογική συνοχή και καλύτερη διαχωρισιμότητα. Το μοντέλο δεν φαίνεται να λειτουργεί πλέον κυρίως με βάση επιφανειακά οπτικά μοτίβα, αλλά αξιοποιεί πιο σύνθετες και αφαιρετικές αναπαραστάσεις της εικόνας. Η συμπεριφορά αυτή αποτυπώνεται ξεκάθαρα στα Grad-CAM heatmaps, τα οποία παρουσιάζουν πιο στοχευμένη και συνεπή κατανομή της προσοχής.

Συνολικά, το αποτέλεσμα των 20 εποχών μπορεί να θεωρηθεί περισσότερο ώριμο και ερμηνεύσιμο σε σχέση με τα προηγούμενα στάδια εκπαίδευσης. Το Vision Transformer φαίνεται να έχει αναπτύξει πιο σταθερές εσωτερικές αναπαραστάσεις και πιο συνεκτικά μοτίβα προσοχής, εστιάζοντας σε περιοχές που σχετίζονται με τη μορφολογία της κατηγορίας colon_aca. Ωστόσο, η παρουσία ορισμένων διάσπαρτων ενεργοποιήσεων δείχνει ότι η εστίαση του μοντέλου, αν και βελτιωμένη, δεν είναι απόλυτα εξειδικευμένη. Επομένως, η εικόνα των 20 εποχών υποδηλώνει ένα πιο προχωρημένο στάδιο μάθησης, στο οποίο το μοντέλο έχει αποκτήσει μεγαλύτερη ικανότητα εντοπισμού βιολογικά συναφών περιοχών, χωρίς όμως να εξαλείφεται πλήρως η ανάγκη προσεκτικής ερμηνείας των ενεργοποιήσεων.

Συνολικά, η σύγκριση των τριών σταδίων εκπαίδευσης αναδεικνύει μια σαφή εξελικτική πορεία, κατά την οποία το μοντέλο μεταβαίνει από μια ασαφή και μη δομημένη κατανομή προσοχής σε μια πιο στοχευμένη και σημασιολογικά συνεπή ερμηνεία της εικόνας. Η ανάλυση των Grad-CAM heatmaps καθιστά εμφανές ότι η αύξηση των εποχών εκπαίδευσης δεν επηρεάζει μόνο τις ποσοτικές μετρικές απόδοσης, αλλά και τον ίδιο τον τρόπο με τον οποίο το μοντέλο αντιλαμβάνεται και επεξεργάζεται την πληροφορία.

Με τον τρόπο αυτό, το Grad-CAM αναδεικνύεται σε ένα ιδιαίτερα ισχυρό εργαλείο κατανόησης της εσωτερικής λειτουργίας του Vision Transformer, παρέχοντας πολύτιμες πληροφορίες όχι μόνο για την ερμηνεία των αποφάσεων του μοντέλου, αλλά και για την αξιολόγηση της ποιότητας της μάθησης και της συνολικής διαδικασίας εκπαίδευσης.

Κεφάλαιο 8ο: Συζήτηση

8.1 Ερμηνεία πειραματικών αποτελεσμάτων

Μέσα από την ανάλυση των πειραματικών αποτελεσμάτων καθίστανται εμφανείς σημαντικές διαφοροποιήσεις στη συμπεριφορά των εξεταζόμενων μοντέλων, τόσο σε επίπεδο συνολικών μετρικών όσο και σε επίπεδο επιμέρους κλάσεων. Η αξιολόγηση πραγματοποιήθηκε σε ισορροπημένο dataset, γεγονός που επιτρέπει την άμεση και αξιόπιστη χρήση κλασικών μετρικών ταξινόμησης, όπως η ακρίβεια (Accuracy), η ακρίβεια θετικών προβλέψεων (Precision), η ανάκληση (Recall) και ο F1-score, χωρίς να απαιτούνται ειδικές τεχνικές αντιμετώπισης ανισορροπίας δεδομένων. Στο πλαίσιο αυτό, το Accuracy μπορεί να χρησιμοποιηθεί ως αντιπροσωπευτικός συνολικός δείκτης απόδοσης, ενώ οι υπόλοιπες μετρικές επιτρέπουν βαθύτερη κατανόηση της συμπεριφοράς των μοντέλων και της κατανομής των σφαλμάτων τους.

Από τη συγκριτική ανάλυση των αρχιτεκτονικών (Baseline CNN, VGG16, VGG19, ResNet50, InceptionResNetV2 και Vision Transformer) προκύπτει ότι τα βαθύτερα και πιο σύνθετα μοντέλα επιτυγχάνουν γενικά υψηλότερες επιδόσεις. Η συμπεριφορά αυτή σχετίζεται άμεσα με την αυξημένη ικανότητά τους να εξάγουν πολυεπίπεδες και πιο αφηρημένες αναπαραστάσεις των δεδομένων. Ειδικότερα, οι αρχιτεκτονικές ResNet50 και InceptionResNetV2 εμφανίζουν τις καλύτερες συνολικές επιδόσεις, καθώς οι residual συνδέσεις και οι μηχανισμοί multi-scale feature extraction επιτρέπουν αποτελεσματικότερη ροή πληροφορίας και βαθύτερη αναπαράσταση των ιστολογικών χαρακτηριστικών.

Το baseline CNN μοντέλο παρουσιάζει ιδιαίτερα υψηλή απόδοση, επιτυγχάνοντας Precision, Recall και F1-score ίσα με 0.98, καθώς και συνολική ακρίβεια 98.03%. Η επίδοση αυτή καταδεικνύει ότι ακόμη και μια σχετικά απλή συνελικτική αρχιτεκτονική μπορεί να εξαγάγει αποτελεσματικά μορφολογικά χαρακτηριστικά από ιστοπαθολογικές εικόνες, όταν το dataset είναι καλά οργανωμένο και ισορροπημένο. Παρόλα αυτά, το μοντέλο εμφανίζει ελαφρώς μειωμένη ικανότητα διάκρισης σε μορφολογικά παρόμοιες κατηγορίες, ιδιαίτερα στους υποτύπους καρκίνου πνεύμονα.

Το VGG16 παρουσιάζει χαμηλότερη συνολική απόδοση, με Precision, Recall και F1-score ίσα με 0.96 και συνολική ακρίβεια 96.47%. Η συμπεριφορά αυτή υποδηλώνει ότι, παρά τη σταθερότητα της αρχιτεκτονικής του, το μοντέλο διαθέτει περιορισμένη ικανότητα εξαγωγής σύνθετων χαρακτηριστικών σε σχέση με πιο σύγχρονες αρχιτεκτονικές. Αντίθετα, το VGG19 εμφανίζει σημαντική βελτίωση, επιτυγχάνοντας Precision, Recall και F1-score ίσα με 0.99 και Accuracy 98.63%. Η βελτίωση αυτή επιβεβαιώνει ότι η αύξηση του βάθους του δικτύου ενισχύει την ικανότητα του μοντέλου να μαθαίνει πιο αφηρημένες και διακριτικές αναπαραστάσεις.

Το ResNet50 παρουσιάζει από τις υψηλότερες συνολικές επιδόσεις, επιτυγχάνοντας Precision, Recall και F1-score ίσα με 0.99 και συνολική ακρίβεια 98.73%. Η ιδιαίτερα υψηλή απόδοση του μοντέλου αποδίδεται κυρίως στις residual συνδέσεις, οι οποίες επιτρέπουν αποτελεσματική εκπαίδευση βαθύτερων δικτύων χωρίς προβλήματα υποβάθμισης της πληροφορίας. Μέσω αυτής της αρχιτεκτονικής, το μοντέλο καταφέρνει να εξάγει πιο πλούσιες και σταθερές αναπαραστάσεις των ιστολογικών δομών, οδηγώντας σε πολύ υψηλή ικανότητα γενίκευσης.

Ακόμη υψηλότερη απόδοση παρουσιάζει το InceptionResNetV2, το οποίο επιτυγχάνει την καλύτερη συνολική επίδοση μεταξύ όλων των μοντέλων, με Accuracy 99.20% και αντίστοιχα πολύ υψηλές τιμές Precision, Recall και F1-score (0.99). Η υπεροχή του μοντέλου σχετίζεται τόσο με τη

χρήση residual learning όσο και με τους μηχανισμούς multi-scale feature extraction, οι οποίοι επιτρέπουν στο δίκτυο να αναλύει ταυτόχρονα χαρακτηριστικά διαφορετικής κλίμακας. Η δυνατότητα αυτή είναι ιδιαίτερα σημαντική στην ιστοπαθολογική ανάλυση, όπου τα μορφολογικά χαρακτηριστικά εμφανίζονται σε πολλαπλά επίπεδα λεπτομέρειας.

Αντίθετα, το Vision Transformer, ακόμη και μετά από 20 εποχές εκπαίδευσης, παρουσιάζει χαμηλότερη συνολική απόδοση σε σχέση με τα CNN-based μοντέλα, με Precision, Recall και F1-score περίπου ίσα με 0.95 και συνολική ακρίβεια 94.80%. Παρότι το μοντέλο παρουσιάζει σταδιακή βελτίωση κατά την εκπαίδευση, εξακολουθεί να υπολείπεται των convolutional αρχιτεκτονικών στο συγκεκριμένο dataset. Η συμπεριφορά αυτή συνδέεται με τη φύση των Vision Transformers, τα οποία δεν ενσωματώνουν εγγενείς χωρικές επαγωγικές προκαταλήψεις (inductive biases), όπως η τοπικότητα και η μεταθετικότητα που χαρακτηρίζουν τα CNNs. Ως αποτέλεσμα, τα ViTs απαιτούν συνήθως μεγαλύτερο όγκο δεδομένων και εκτενέστερη εκπαίδευση ώστε να επιτύχουν πλήρη εκμάθηση των χωρικών σχέσεων.

Παρά τη χαμηλότερη συνολική του απόδοση, το Vision Transformer παρουσιάζει ιδιαίτερο ερευνητικό ενδιαφέρον, καθώς μέσω του μηχανισμού self-attention μπορεί να αναπτύξει πιο ολιστικές και παγκόσμιες αναπαραστάσεις της εικόνας. Η σταδιακή βελτίωση που παρατηρείται από τις 10 έως τις 20 εποχές εκπαίδευσης υποδηλώνει ότι το μοντέλο διαθέτει σημαντικό δυναμικό περαιτέρω εξέλιξης, ιδιαίτερα σε περιβάλλοντα με μεγαλύτερα datasets ή εκτενέστερη διαδικασία fine-tuning.

Συνολικά, τα αποτελέσματα καταδεικνύουν ότι τα βαθύτερα και πιο σύνθετα CNN-based μοντέλα, όπως το ResNet50 και κυρίως το InceptionResNetV2, παρουσιάζουν την καλύτερη συνολική συμπεριφορά στο συγκεκριμένο πρόβλημα ταξινόμησης ιστοπαθολογικών εικόνων. Τα baseline CNN και VGG-based μοντέλα εμφανίζουν επίσης υψηλή απόδοση, αλλά υπολείπονται ελαφρώς στην ικανότητα διαχωρισμού μορφολογικά παρόμοιων κατηγοριών. Το Vision Transformer, παρότι θεωρητικά ιδιαίτερα ισχυρό, εμφανίζει μειωμένη αποτελεσματικότητα στο παρόν πειραματικό πλαίσιο, γεγονός που πιθανότατα σχετίζεται με τις αυξημένες απαιτήσεις του σε δεδομένα και διάρκεια εκπαίδευσης.

Η συνολική σύγκριση επιβεβαιώνει ότι η αρχιτεκτονική του μοντέλου διαδραματίζει καθοριστικό ρόλο στην τελική απόδοση, ακόμη και σε περιβάλλοντα όπου τα δεδομένα είναι πλήρως ισορροπημένα. Παράλληλα, αναδεικνύεται η σημασία της επιλογής κατάλληλης αρχιτεκτονικής ανάλογα με τη φύση του προβλήματος, ιδιαίτερα σε εφαρμογές ιατρικής ανάλυσης όπου η αξιοπιστία, η γενίκευση και η δυνατότητα ακριβούς διάκρισης μεταξύ μορφολογικά παρόμοιων κατηγοριών αποτελούν κρίσιμες απαιτήσεις.

8.2 Πλεονεκτήματα και περιορισμοί των διαφορετικών αρχιτεκτονικών

Η συγκριτική ανάλυση των μοντέλων που χρησιμοποιήθηκαν στο πλαίσιο της παρούσας εργασίας αναδεικνύει τόσο τα πλεονεκτήματα όσο και τους περιορισμούς κάθε αρχιτεκτονικής, επιτρέποντας μια πιο ολοκληρωμένη κατανόηση της συμπεριφοράς τους σε προβλήματα ταξινόμησης εικόνων.

Αρχικά, το βασικό μοντέλο CNN παρουσιάζει σημαντικά πλεονεκτήματα ως προς την απλότητα και την υπολογιστική αποδοτικότητα. Με ακρίβεια 98.03% και F1-score 0.98, αποδεικνύεται ότι μπορεί να επιτύχει υψηλή απόδοση χωρίς την ανάγκη ιδιαίτερα σύνθετων δομών. Το γεγονός αυτό το καθιστά κατάλληλο για εφαρμογές με περιορισμένους υπολογιστικούς πόρους ή όταν απαιτείται γρήγορη εκπαίδευση. Ωστόσο, ο βασικός του περιορισμός εντοπίζεται στην περιορισμένη ικανότητα

εξαγωγής βαθύτερων και πιο αφηρημένων χαρακτηριστικών, γεγονός που το καθιστά λιγότερο ανταγωνιστικό σε σύγκριση με πιο εξελιγμένα μοντέλα.

Τα μοντέλα VGG16 και VGG19 χαρακτηρίζονται από απλή και συνεπή αρχιτεκτονική, η οποία διευκολύνει την κατανόηση και την υλοποίηση. Το VGG16, με ακρίβεια 96.47% και F1-score 0.96, εμφανίζει σταθερή αλλά χαμηλότερη απόδοση, γεγονός που υποδηλώνει περιορισμένη ικανότητα μοντελοποίησης πολύπλοκων σχέσεων. Αντίθετα, το VGG19 παρουσιάζει βελτιωμένα αποτελέσματα (accuracy 98.63%, F1-score 0.99), επιβεβαιώνοντας ότι η αύξηση του βάθους μπορεί να ενισχύσει την αναπαραστατική ικανότητα του μοντέλου. Παρ' όλα αυτά, και τα δύο μοντέλα εμφανίζουν σημαντικό μειονέκτημα όσον αφορά τον μεγάλο αριθμό παραμέτρων, γεγονός που οδηγεί σε αυξημένες απαιτήσεις σε μνήμη και υπολογιστική ισχύ, καθώς και σε μεγαλύτερο χρόνο εκπαίδευσης.

Το ResNet50 αποτελεί ένα από τα πιο αποδοτικά μοντέλα της ανάλυσης, επιτυγχάνοντας ακρίβεια 98.73% και F1-score 0.99. Το βασικό του πλεονέκτημα έγκειται στη χρήση residual connections, τα οποία επιτρέπουν την εκπαίδευση πολύ βαθιών δικτύων χωρίς το πρόβλημα της εξαφάνισης των κλίσεων. Αυτό οδηγεί σε καλύτερη γενίκευση και πιο σταθερή εκπαίδευση. Ωστόσο, η αυξημένη πολυπλοκότητα του μοντέλου συνεπάγεται μεγαλύτερο υπολογιστικό κόστος, γεγονός που μπορεί να αποτελέσει περιοριστικό παράγοντα σε περιβάλλοντα με περιορισμένους πόρους.

Αντίστοιχα, το InceptionResNetV2 παρουσιάζει την υψηλότερη συνολική απόδοση (accuracy 99.2%, F1-score 0.99), συνδυάζοντας τα πλεονεκτήματα των Inception modules και των residual connections. Η δυνατότητα επεξεργασίας χαρακτηριστικών σε πολλαπλές κλίμακες επιτρέπει στο μοντέλο να συλλαμβάνει τόσο λεπτομερή όσο και γενικά μοτίβα, οδηγώντας σε εξαιρετική απόδοση. Παρ' όλα αυτά, το μοντέλο αυτό είναι το πιο απαιτητικό υπολογιστικά, τόσο σε επίπεδο εκπαίδευσης όσο και σε επίπεδο inference, γεγονός που περιορίζει την πρακτική του εφαρμογή σε συστήματα με υψηλή υπολογιστική ισχύ.

Το Vision Transformer, αν και αποτελεί μία σύγχρονη και πολλά υποσχόμενη αρχιτεκτονική, παρουσιάζει χαμηλότερη απόδοση στο συγκεκριμένο πείραμα (accuracy 94.8%, F1-score 0.95). Το βασικό του πλεονέκτημα είναι η ικανότητα μοντελοποίησης παγκόσμιων σχέσεων μέσω του μηχανισμού self-attention, χωρίς την ανάγκη τοπικών φίλτρων όπως στα CNNs. Ωστόσο, η απουσία ενσωματωμένης χωρικής επαγωγικής προκατάληψης το καθιστά ιδιαίτερα εξαρτώμενο από το μέγεθος του dataset και τον χρόνο εκπαίδευσης. Στην παρούσα εργασία, το σχετικά περιορισμένο πλήθος δεδομένων και οι 20 εποχές εκπαίδευσης δεν ήταν επαρκείς για την πλήρη αξιοποίηση των δυνατοτήτων του μοντέλου.

Συνοψίζοντας, τα αποτελέσματα καταδεικνύουν ότι τα CNN-based μοντέλα υπερέχουν στο συγκεκριμένο πρόβλημα, κυρίως λόγω της ενσωματωμένης χωρικής πληροφορίας και της αποδοτικής εξαγωγής χαρακτηριστικών. Τα πιο σύνθετα μοντέλα, όπως το ResNet50 και το InceptionResNetV2, προσφέρουν την καλύτερη ισορροπία μεταξύ απόδοσης και ικανότητας γενίκευσης, ενώ το ViT παρουσιάζει σημαντικό δυναμικό, το οποίο όμως απαιτεί μεγαλύτερη κλίμακα δεδομένων και εκπαίδευσης για να αναδειχθεί πλήρως.

8.3 Περιορισμοί του Συνόλου Δεδομένων και της Μεθοδολογίας

Παρά τα ιδιαίτερα ικανοποιητικά αποτελέσματα που προέκυψαν από την εκπαίδευση και αξιολόγηση των μοντέλων, είναι σημαντικό να επισημανθούν ορισμένοι περιορισμοί που σχετίζονται τόσο με το χρησιμοποιούμενο dataset όσο και με τη μεθοδολογία που ακολουθήθηκε. Η αναγνώριση αυτών των περιορισμών συμβάλλει στην ορθότερη ερμηνεία των αποτελεσμάτων και θέτει τα θεμέλια για μελλοντική βελτίωση.

Αρχικά, αν και το σύνολο δεδομένων είναι ισορροπημένο ως προς τις κλάσεις, γεγονός που διευκολύνει την εκπαίδευση και την αξιολόγηση των μοντέλων, ενδέχεται να μην αντικατοπτρίζει πλήρως τις συνθήκες πραγματικού κόσμου. Στην πράξη, πολλά προβλήματα ταξινόμησης εικόνων – και ιδιαίτερα σε ιατρικά ή κρίσιμα πεδία – χαρακτηρίζονται από έντονη ανισορροπία, όπου η μειοψηφική κλάση είναι και η πιο σημαντική. Επομένως, τα μοντέλα που αξιολογήθηκαν ενδέχεται να εμφανίζουν διαφορετική συμπεριφορά όταν εφαρμοστούν σε πραγματικά δεδομένα.

Ένας περιορισμός που προκύπτει είναι ότι μέρος του συνόλου δεδομένων LC25000 που χρησιμοποιείται προκύπτει από διαδικασίες data augmentation. Γι' τον λόγο αυτό, τα υψηλά ποσοστά ακρίβειας είναι αναγκαίο να ερμηνεύονται με προσοχή, διότι δεν ισοδυναμούν απαραίτητα με αντίστοιχη απόδοση σε πλήρως ανεξάρτητα δεδομένα.

Επιπλέον, ένας σημαντικός περιορισμός σχετίζεται με το μέγεθος και τη διαφοροποίηση του dataset. Εάν το πλήθος των εικόνων είναι περιορισμένο ή δεν καλύπτει επαρκώς όλες τις πιθανές παραλλαγές των δεδομένων (π.χ. διαφορετικές συνθήκες φωτισμού, γωνίες λήψης ή θόρυβος), τότε τα μοντέλα ενδέχεται να εμφανίζουν μειωμένη ικανότητα γενίκευσης. Αυτό είναι ιδιαίτερα εμφανές σε πιο σύνθετα μοντέλα, όπως το Vision Transformer, το οποίο απαιτεί μεγάλα σύνολα δεδομένων για να αποδώσει στο μέγιστο των δυνατοτήτων του.

Ένας ακόμη περιορισμός αφορά την πιθανή ύπαρξη μεροληψίας στο dataset. Εάν τα δεδομένα προέρχονται από συγκεκριμένη πηγή ή παρουσιάζουν ομοιομορφία ως προς ορισμένα χαρακτηριστικά, τα μοντέλα ενδέχεται να μάθουν μοτίβα που δεν είναι γενικεύσιμα. Αυτό μπορεί να οδηγήσει σε υπερεκτίμηση της απόδοσης κατά τη φάση της αξιολόγησης, ειδικά εάν το σύνολο εκπαίδευσης και το σύνολο ελέγχου παρουσιάζουν υψηλό βαθμό ομοιότητας.

Παράλληλα, ένας ακόμη περιορισμός της παρούσας μελέτης αφορά την ερμηνεία των αποτελεσμάτων μέσω των Grad-CAM οπτικοποιήσεων. Παρότι τα Grad-CAM heatmaps αποτελούν χρήσιμο εργαλείο κατανόησης της συμπεριφοράς του μοντέλου, δεν αποδεικνύουν ότι το μοντέλο αξιολογεί την εικόνα με τρόπο αντίστοιχο ενός ειδικού παθολογοανατόμου. Οι οπτικοποιήσεις αυτές παρέχουν κυρίως ενδείξεις συσχέτισης, αναδεικνύοντας τις περιοχές της εικόνας που επηρέασαν περισσότερο την τελική απόφαση.

Ωστόσο, οι θερμές περιοχές δεν πρέπει να ερμηνεύονται ως απόλυτη απόδειξη βιολογικής ή διαγνωστικής σημασίας. Είναι πιθανό το μοντέλο να βασίζεται σε χαρακτηριστικά όπως υφές, χρωματικές διαφοροποιήσεις ή τεχνουργήματα της εικόνας, τα οποία μπορεί να σχετίζονται με τα δεδομένα εκπαίδευσης χωρίς να αποτελούν ουσιαστικά ιστοπαθολογικά κριτήρια. Συνεπώς, τα Grad-CAM αποτελέσματα αξιοποιούνται στην παρούσα εργασία ως συμπληρωματικό εργαλείο ερμηνευσιμότητας και όχι ως μέθοδος διαγνωστικής επιβεβαίωσης. Η τελική αξιολόγηση της βιολογικής σημασίας των ενεργοποιήσεων απαιτεί επιβεβαίωση από ειδικό παθολογοανατόμο.

Σε επίπεδο μεθοδολογίας, ένας βασικός περιορισμός εντοπίζεται στη διαδικασία εκπαίδευσης των μοντέλων, και συγκεκριμένα στον αριθμό των εποχών και στις υπερπαραμέτρους που επιλέχθηκαν. Για παράδειγμα, το Vision Transformer εκπαιδεύτηκε για περιορισμένο αριθμό εποχών (έως 20), γεγονός που πιθανώς δεν επέτρεψε την πλήρη σύγκλιση του μοντέλου. Αντίστοιχα, η επιλογή υπερπαραμέτρων, όπως ο ρυθμός μάθησης, το batch size και οι τεχνικές κανονικοποίησης, μπορεί να επηρεάζουν σημαντικά την τελική απόδοση.

Τέλος, ένας ακόμη περιορισμός αφορά την έλλειψη εκτεταμένης διερεύνησης εναλλακτικών στρατηγικών εκπαίδευσης, όπως διαφορετικές αρχιτεκτονικές παραλλαγές, μεγαλύτερος αριθμός εποχών ή συνδυαστικές προσεγγίσεις. Αν και τα μοντέλα που επιλέχθηκαν καλύπτουν ένα ευρύ φάσμα

σύγχρονων τεχνικών, η περαιτέρω πειραματική διερεύνηση θα μπορούσε να οδηγήσει σε ακόμη καλύτερα αποτελέσματα.

Συνοψίζοντας, οι περιορισμοί του dataset και της μεθοδολογίας δεν αναιρούν την εγκυρότητα των αποτελεσμάτων, αλλά υπογραμμίζουν την ανάγκη προσεκτικής ερμηνείας τους. Παράλληλα, αναδεικνύουν σαφείς κατευθύνσεις για μελλοντική έρευνα, με στόχο τη βελτίωση της γενίκευσης και της αξιοπιστίας των μοντέλων.

8.4 Μελλοντικές βελτιώσεις του συστήματος

Παρά τα ιδιαίτερα ικανοποιητικά αποτελέσματα που επιτεύχθηκαν, υπάρχουν σημαντικά περιθώρια βελτίωσης του συστήματος, τόσο σε επίπεδο δεδομένων όσο και σε επίπεδο αρχιτεκτονικής και διαδικασίας εκπαίδευσης. Η διερεύνηση αυτών των κατευθύνσεων μπορεί να οδηγήσει σε περαιτέρω αύξηση της απόδοσης και της αξιοπιστίας των μοντέλων.

Αρχικά, μία από τις σημαντικότερες βελτιώσεις αφορά την επέκταση του dataset. Η αύξηση του πλήθους των διαθέσιμων δεδομένων, καθώς και η ενίσχυση της ποικιλίας τους (π.χ. διαφορετικές συνθήκες λήψης, παραλλαγές χαρακτηριστικών), θα μπορούσε να βελτιώσει σημαντικά την ικανότητα γενίκευσης των μοντέλων. Ιδιαίτερα για πιο σύνθετες αρχιτεκτονικές, όπως το Vision Transformer, η ύπαρξη μεγαλύτερου όγκου δεδομένων αποτελεί κρίσιμο παράγοντα για την πλήρη αξιοποίηση των δυνατοτήτων τους.

Επιπλέον, η βελτιστοποίηση της διαδικασίας εκπαίδευσης μπορεί να συμβάλει ουσιαστικά στην αύξηση της απόδοσης. Η διερεύνηση διαφορετικών υπερπαραμέτρων, όπως ο ρυθμός μάθησης (learning rate), το batch size και ο αριθμός εποχών, ενδέχεται να οδηγήσει σε καλύτερη σύγκλιση των μοντέλων. Στην περίπτωση του ViT, για παράδειγμα, η αύξηση των εποχών εκπαίδευσης πέρα από τις 20 ή η χρήση τεχνικών όπως learning rate scheduling θα μπορούσε να βελτιώσει σημαντικά την απόδοσή του

Επιπρόσθετα, η αξιοποίηση transfer learning σε μεγαλύτερο βαθμό αποτελεί μια πολλά υποσχόμενη προσέγγιση. Αν και τα περισσότερα μοντέλα βασίζονται ήδη σε προεκπαιδευμένα βάρη, η περαιτέρω προσαρμογή σε περισσότερα layers ή η χρήση πιο εξειδικευμένων pre-trained μοντέλων θα μπορούσε να οδηγήσει σε καλύτερη προσαρμογή στο συγκεκριμένο πρόβλημα.

Μια ιδιαίτερα ενδιαφέρουσα κατεύθυνση είναι η υλοποίηση ensemble μεθόδων, δηλαδή ο συνδυασμός πολλαπλών μοντέλων για τη λήψη τελικής απόφασης. Δεδομένου ότι τα διαφορετικά μοντέλα εμφανίζουν διαφορετικά πρότυπα σφαλμάτων, ο συνδυασμός τους μπορεί να μειώσει το συνολικό σφάλμα και να βελτιώσει τη συνολική απόδοση του συστήματος.

Τέλος, η περαιτέρω βελτίωση της ερμηνευσιμότητας του συστήματος αποτελεί σημαντικό στόχο, ιδιαίτερα σε εφαρμογές υψηλής σημασίας. Η εκτενέστερη χρήση τεχνικών οπτικοποίησης, όπως τα Grad-CAM και attention maps, μπορεί να προσφέρει βαθύτερη κατανόηση της διαδικασίας λήψης αποφάσεων των μοντέλων και να ενισχύσει την εμπιστοσύνη στον τρόπο λειτουργίας τους.

Κεφάλαιο 9ο: Συμπεράσματα και μελλοντική Έρευνα

9.1 Βασικά Συμπεράσματα της Εργασίας

Η παρούσα έρευνα επικεντρώθηκε στη συγκριτική αξιολόγηση διαφορετικών αρχιτεκτονικών βαθιάς μάθησης για την επίλυση προβλήματος ταξινόμησης εικόνων, με στόχο την ανάδειξη της επίδρασης της αρχιτεκτονικής δομής στην απόδοση των μοντέλων. Μέσα από συστηματική πειραματική διαδικασία, κατέστη δυνατή η εξαγωγή αξιόπιστων συμπερασμάτων σχετικά με τη συμπεριφορά τόσο των convolutional neural networks (CNNs) όσο και των transformer-based μοντέλων.

Τα αποτελέσματα κατέδειξαν ότι όλα τα μοντέλα πέτυχαν υψηλές επιδόσεις, γεγονός που επιβεβαιώνει την ωριμότητα των σύγχρονων τεχνικών βαθιάς μάθησης στον τομέα της οπτικής αναγνώρισης. Ωστόσο, παρατηρήθηκαν σαφείς διαφοροποιήσεις στην απόδοση, οι οποίες συνδέονται άμεσα με τη δομή και την πολυπλοκότητα κάθε αρχιτεκτονικής.

Συγκεκριμένα, τα CNN-based μοντέλα παρουσίασαν ιδιαίτερα ισχυρή και σταθερή απόδοση, με το InceptionResNetV2 να αναδεικνύεται ως το πλέον αποδοτικό μοντέλο, επιτυγχάνοντας ακρίβεια 99.2% και F1-score 0.99. Η επίδοση αυτή υποδηλώνει ότι ο συνδυασμός residual συνδέσεων και multi-scale feature extraction επιτρέπει στο μοντέλο να συλλαμβάνει αποτελεσματικά τόσο τοπικά όσο και παγκόσμια χαρακτηριστικά των εικόνων. Το ResNet50 ακολούθησε με πολύ υψηλή απόδοση (98.73%), επιβεβαιώνοντας την αξία των residual architectures στη βαθιά μάθηση. Παράλληλα, το VGG19 παρουσίασε εξαιρετικά αποτελέσματα (98.63%), καταδεικνύοντας ότι, παρά την απλότητά του σε σχέση με πιο σύγχρονες αρχιτεκτονικές, μπορεί να παραμείνει ανταγωνιστικό όταν το dataset είναι κατάλληλα διαμορφωμένο. Αντίθετα, το VGG16 εμφάνισε χαμηλότερη απόδοση (96.47%), γεγονός που υποδηλώνει ότι η αυξημένη χωρητικότητα του βαθύτερου VGG19 συμβάλλει ουσιαστικά στην καλύτερη αναπαράσταση των δεδομένων.

Το βασικό CNN, αν και απλούστερο σαν αρχιτεκτονική, πέτυχε ιδιαίτερα υψηλή ακρίβεια (98.03%), γεγονός που αναδεικνύει τη σημασία της σωστής προεπεξεργασίας και της δομής των δεδομένων. Η απόδοσή του αποδεικνύει ότι σε ορισμένες περιπτώσεις, η πολυπλοκότητα δεν αποτελεί απαραίτητα προϋπόθεση για υψηλή επίδοση.

Απεναντίας, το Vision Transformer παρουσίασε αισθητά χαμηλότερη απόδοση (94.8%), γεγονός που αποδίδεται κυρίως στην εξάρτησή του από μεγάλα datasets και εκτεταμένη εκπαίδευση. Η συμπεριφορά αυτή επιβεβαιώνει ότι τα transformers, αν και θεωρητικά ισχυρά, δεν υπερέχουν πάντα σε περιβάλλοντα περιορισμένων δεδομένων. Η απουσία ενσωματωμένης χωρικής επαγωγικής προκατάληψης καθιστά τη μάθηση πιο απαιτητική και λιγότερο αποδοτική σε τέτοιες συνθήκες.

Συνολικά, τα αποτελέσματα της εργασίας καταδεικνύουν ότι η επιλογή αρχιτεκτονικής πρέπει να γίνεται με βάση τα χαρακτηριστικά του προβλήματος, το μέγεθος του dataset και τους διαθέσιμους υπολογιστικούς πόρους. Τα CNNs παραμένουν ιδιαίτερα αποδοτικά και αξιόπιστα, ενώ τα transformer-based μοντέλα παρουσιάζουν σημαντικό δυναμικό, το οποίο όμως απαιτεί κατάλληλες συνθήκες για να αξιοποιηθεί πλήρως.

9.2 Συμβολή της Εργασίας

Η παρούσα εργασία συμβάλλει στον τομέα της βαθιάς μάθησης και της ανάλυσης εικόνας, παρέχοντας μια ολοκληρωμένη και συγκριτική προσέγγιση μεταξύ διαφορετικών αρχιτεκτονικών. Σε

αντίθεση με μελέτες που εστιάζουν αποκλειστικά σε ένα μοντέλο, η εργασία αυτή εξετάζει ένα ευρύ φάσμα προσεγγίσεων, επιτρέποντας την εξαγωγή πιο γενικευμένων και χρήσιμων συμπερασμάτων.

Μία από τις βασικές συνεισφορές είναι η εμπειρική επιβεβαίωση της υπεροχής των residual και hybrid αρχιτεκτονικών, όπως το ResNet50 και το InceptionResNetV2, σε σχέση με παραδοσιακά μοντέλα. Η εργασία αναδεικνύει τον ρόλο της αρχιτεκτονικής σχεδίασης στην απόδοση, προσφέροντας πρακτική τεκμηρίωση θεωρητικών εννοιών της βιβλιογραφίας.

Επιπλέον, η σύγκριση με το Vision Transformer παρέχει πολύτιμη γνώση σχετικά με τη συμπεριφορά των transformer-based μοντέλων σε πραγματικά σενάρια. Το εύρημα ότι το ViT υπολείπεται σε μικρότερα datasets αποτελεί σημαντική πρακτική παρατήρηση, η οποία μπορεί να καθοδηγήσει μελλοντικές επιλογές αρχιτεκτονικών σε παρόμοια προβλήματα. Η εργασία συμβάλλει επίσης στην ενίσχυση της ερμηνευσιμότητας των μοντέλων, μέσω της χρήσης τεχνικών οπτικοποίησης σε συνδιασμό με την υλοποίηση της αρχιτεκτονικής του ViT επιτρέποντας την καλύτερη κατανόηση της διαδικασίας λήψης αποφάσεων. Η προσέγγιση αυτή είναι ιδιαίτερα σημαντική σε εφαρμογές όπου απαιτείται διαφάνεια και αξιοπιστία.

Τέλος, η υλοποίηση όλων των μοντέλων υπό κοινές συνθήκες εκπαίδευσης και αξιολόγησης ενισχύει την εγκυρότητα της σύγκρισης, καθιστώντας τα συμπεράσματα της εργασίας άμεσα αξιοποιήσιμα τόσο σε ερευνητικό όσο και σε πρακτικό επίπεδο.

9.3 Προτάσεις για Μελλοντική Έρευνα

Οι προτάσεις για την μελλοντική επέκταση της παρούσας εργασίας επεκτείνονται σε πολλές κατευθύνσεις, με στόχο την βελτίωση της απόδοσης των αρχιτεκτονικών καθώς και της αξιοπιστίας των μοντέλων.

Την πιο βασική πρόταση επέκτασης της έρευνας αποτελεί η αξιοποίηση μεγαλύτερων και πιο σύνθετων datasets. Η αύξηση του όγκου και της ποικιλίας των δεδομένων αναμένεται να βελτιώσει σημαντικά την απόδοση, ιδιαίτερα για τα transformer-based μοντέλα, τα οποία απαιτούν μεγάλες ποσότητες δεδομένων για να αποδώσουν στο μέγιστο των δυνατοτήτων τους. Επιπλέον, η διερεύνηση πιο προηγμένων τεχνικών εκπαίδευσης μπορεί να οδηγήσει σε περαιτέρω βελτιώσεις. Η χρήση adaptive optimization algorithms και πιο εκτεταμένου fine-tuning αποτελεί σημαντική κατεύθυνση για τη βελτιστοποίηση της απόδοσης.

Ιδιαίτερο ενδιαφέρον παρουσιάζει η ανάπτυξη υβριδικών μοντέλων που συνδυάζουν τα πλεονεκτήματα των CNNs και των transformers. Τέτοιες προσεγγίσεις μπορούν να εκμεταλλευτούν τόσο την τοπική εξαγωγή χαρακτηριστικών όσο και τη μοντελοποίηση παγκόσμιων σχέσεων, οδηγώντας σε πιο αποδοτικά και ευέλικτα συστήματα. Παράλληλα, η αξιοποίηση ensemble τεχνικών μπορεί να συμβάλει στη μείωση του σφάλματος και στην αύξηση της αξιοπιστίας των προβλέψεων, μέσω του συνδυασμού διαφορετικών μοντέλων που εμφανίζουν συμπληρωματική συμπεριφορά.

Τέλος, ιδιαίτερη έμφαση πρέπει να δοθεί στην αξιολόγηση των μοντέλων σε πραγματικά δεδομένα και σε πιο απαιτητικά περιβάλλοντα. Η μεταφορά των μοντέλων από ελεγχόμενα πειραματικά πλαίσια σε πραγματικές συνθήκες αποτελεί κρίσιμο βήμα για την πρακτική αξιοποίησή τους.

9.4 Επίλογος

Η παρούσα εργασία κατέδειξε ότι η επιλογή της κατάλληλης αρχιτεκτονικής αποτελεί κρίσιμο παράγοντα για την επιτυχή ταξινόμηση εικόνων, με τα CNN-based μοντέλα να εμφανίζουν υψηλή και σταθερή απόδοση στο συγκεκριμένο πρόβλημα. Παράλληλα, αναδείχθηκε ότι πιο σύγχρονες

προσεγγίσεις, όπως τα Vision Transformers, διαθέτουν σημαντικό δυναμικό, το οποίο όμως εξαρτάται από τις συνθήκες εκπαίδευσης και το μέγεθος των δεδομένων. Τα αποτελέσματα επιβεβαιώνουν ότι δεν υπάρχει μία καθολικά βέλτιστη λύση, αλλά η απόδοση εξαρτάται από τον συνδυασμό δεδομένων, αρχιτεκτονικής και παραμετροποίησης. Συνολικά, η εργασία συμβάλλει στην κατανόηση των σύγχρονων μοντέλων βαθιάς μάθησης και θέτει τις βάσεις για την ανάπτυξη πιο αποδοτικών και αξιόπιστων συστημάτων στο μέλλον.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] S. M. Ali Shah, D. Jamil, M. N. Ali Khan and F. M. Mohammed Al-Jarwani, "Predicting Lung Cancer Risk Using Artificial Intelligence," *2024 2nd International Conference on Computing and Data Analytics (ICCDA)*, Shinas, Oman, 2024

[2] G. Verma, "Deep Learning-Based Classification of Lung and Colon Cancer using a Proposed CNN Architecture," *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, Bengaluru, India, 2024

[3] M. S. Naga Raju and D. B. Srinivasa Rao, "Classification of Colon Cancer through analysis of histopathology images using Transfer Learning," *2022 IEEE 2nd International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)*, Gunupur, Odisha, India, 2022

[4] M. A. -E. Zeid, K. El-Bahnasy and S. E. Abo-Youssef, "Multiclass Colorectal Cancer Histology Images Classification Using Vision Transformers," *2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)*, Cairo, Egypt, 2021

[5] Javed, Rabia, et al. "Deep learning for lungs cancer detection: a review." *Artificial Intelligence Review* 57.8 (2024): 197.

[6] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

[7] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *2017 IEEE International Conference on Computer Vision (ICCV)*

[8] Satvik Garg and Somya Garg. 2021. Prediction of lung and colon cancer through analysis of histopathological images by utilizing Pre-trained CNN models with visualization of class activation and saliency maps. In Proceedings of the 2020 3rd Artificial Intelligence and Cloud Computing Conference (AICCC '20). Association for Computing Machinery, New York, NY, USA

[9] J. Cañada, E. Cuello, L. Téllez, J. M. García, F. J. Velasco and J. Cabrera, "Assistance to lung cancer detection on histological images using Convolutional Neural Networks," *2022 E-Health and Bioengineering Conference (EHB)*, Iasi, Romania, 2022

[10] D. K. Soni and A. Taneja, "Enhancing Cancer Diagnosis: CNN-Based Classification of Lung and Colon Histopathological Images," *2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India, 2025

[11] G. Amirthayogam, S. S, G. Maheswari, M. James and K. Remya, "Lung and Colon Cancer Detection using Transfer Learning," *2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, Chennai, India, 2024

[12] A. R. Hidayah and U. N. Wisesty, "Lung Cancer Classification Based on Ensembling EfficientNet Using Histopathology Images," *2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, Bali, Indonesia, 2024

[13] Humayun, M.; Sujatha, R.; Almuayqil, S.N.; Jhanjhi, N.Z. A Transfer Learning Approach with a Convolutional Neural Network for the Classification of Lung Carcinoma. *Healthcare* 2022

[14] Coudray, N., Ocampo, P.S., Sakellaropoulos, T. *et al.* Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nat Med* **24**,(2018)

[15] Davri, A.; Birbas, E.; Kanavos, T.; Ntritsos, G.; Giannakeas, N.; Tzallas, A.T.; Batistatou, A. Deep Learning on Histopathological Images for Colorectal Cancer Diagnosis: A Systematic Review. *Diagnostics* **2022**

[16] S. Mehmood *et al*., "Malignancy Detection in Lung and Colon Histopathology Images Using Transfer Learning With Class Selective Image Processing," in *IEEE Access*, vol. 10, pp. 25657-25668, 2022,

[17] Neamah, Mohamed M., Laith A. Al-Ani, and Loay E. George. "Deep Learning and Texture Analysis for Lung and Colon Cancer Predicting." *Iraqi Journal for Computer Science and Mathematics* 6.3 (2025)

[18] Gessert, Nils, et al. "Deep transfer learning methods for colon cancer classification in confocal laser microscopy images." *International journal of computer assisted radiology and surgery* 14.11 (2019): 1837-1845.

[19] Jain, Tarun, and Andrew M. Lynn. "Interpretable self-supervised contrastive learning for colorectal cancer histopathology: GRADCAM visualization." *Bioinformation* 21.7 (2025): 1836.

[20] Adekunle, Olajumoke O., et al. "Lung Cancer Classification from CT Images Using ResNet." *arXiv preprint arXiv:2510.16310* (2025).

[21]Kumar, Arvind, et al. "Vision transformer based effective model for early detection and classification of lung cancer." *SN Computer Science* 5.7 (2024): 839.

[22] Hossain, M.B., Shaban, M. Resource efficient attention guided deep learning for histopathology based lung and colon cancer classification. *SIViP* **19**, 1311 (2025).

[23]Ji, Jie, et al. "Automated lung and colon cancer classification using histopathological images." *Biomedical Engineering and Computational Biology* 15 (2024): 11795972241271569.

[24]H. Xu *et al*., "Vision Transformers for Computational Histopathology," in *IEEE Reviews in Biomedical Engineering*, vol. 17, pp. 63-79, 2024

[25] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians.

[26] Zhang, XM., Gao, TH., Cai, QY. *et al.* Artificial intelligence in digital pathology diagnosis and analysis: technologies, challenges, and future prospects. *Military Med Res* **12**, 93 (2025)

[27] Inamura K. Update on Immunohistochemistry for the Diagnosis of Lung Cancer. Cancers (Basel). 2018

[28] Wang J, Wang T, Han R, Shi D, Chen B. Artificial intelligence in cancer pathology: Applications, challenges, and future directions 2025

[29] Klaver, C.E.L., Bulkmans, N., Drillenburg, P. *et al.* Interobserver, intraobserver, and interlaboratory variability in reporting pT4a colon cancer. *Virchows Arch* **476**, 219–230 (2020)

[30] World Health Organization. (2026). Lung cancer. World Health Organization.

[31] World Health Organization. (2026). Colorectal cancer. World Health Organization.