ΕΛΛΗΝΙΚΟ ΜΕΣΟΓΕΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

ΤΟΜΕΑΣ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ & ΔΙΚΤΥΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Ανάπτυξη και Σύγκριση Μεθοδολογιών Μηχανικής Μάθησης για την Αυτόματη Ανίχνευση Πολύποδων του Παχέος Εντέρου από Ενδοσκοπικές Εικόνες»

"Development and comparison of machine learning methodologies for automatic detection of colorectal polyps from endoscopic images"

Ιωάννης Στυλιανίδης

ΤΛ20151

Επιβλέπων: Δρ. Ανάργυρος Μπακλέζος

Επικ. Καθηγητής ΕΛ.ΜΕ.ΠΑ.

Χανιά, Απρίλιος 2026

ΕΛΛΗΝΙΚΟ ΜΕΣΟΓΕΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

ΤΟΜΕΑΣ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ & ΔΙΚΤΥΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Ανάπτυξη και Σύγκριση Μεθοδολογιών Μηχανικής Μάθησης για την Αυτόματη Ανίχνευση Πολύποδων του Παχέος Εντέρου από Ενδοσκοπικές Εικόνες»

"Development and comparison of machine learning methodologies for automatic detection of colorectal polyps from endoscopic images"

Ιωάννης Στυλιανίδης

ΤΛ20151

Επιβλέπων: Δρ. Ανάργυρος Μπακλέζος

Επικ. Καθηγητής ΕΛ.ΜΕ.ΠΑ.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 08/04/2026.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Δρ. Ανάργυρος Μπακλέζος	Δρ. Χρήστος Νικολόπουλος	Δρ. Ιωάννης Βαρδιάμπασης
Επίκουρος Καθηγητής	Αναπληρωτής Καθηγητής	Καθηγητής Α' Βαθμίδας
ΕΛ.ΜΕ.ΠΑ.	ΕΛ.ΜΕ.ΠΑ.	ΕΛ.ΜΕ.ΠΑ.

Χανιά, Απρίλιος 2026

Copyright ©Ιωάννης Στυλιανίδης, 2026.
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

(Υπογραφή)

…….

Ιωάννης Στυλιανίδης

Ηλεκτρονικός Μηχανικός ΕΛ.ΜΕ.ΠΑ.

ΕΛΛΗΝΙΚΟ ΜΕΣΟΓΕΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

ΤΟΜΕΑΣ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ & ΔΙΚΤΥΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Ανάπτυξη και Σύγκριση Μεθοδολογιών Μηχανικής Μάθησης για την Αυτόματη Ανίχνευση Πολύποδων του Παχέος Εντέρου από Ενδοσκοπικές Εικόνες»

"Development and comparison of machine learning methodologies for automatic detection of colorectal polyps from endoscopic images"

Ιωάννης Στυλιανίδης

ΤΛ20151

Επιβλέπων: Δρ. Ανάργυρος Μπακλέζος

Επικ. Καθηγητής ΕΛ.ΜΕ.ΠΑ.

Χανιά, Απρίλιος 2026

Αφιερώνω την διπλωματική μου στην οικογένειά μου και σε μένα.

Π ρ ό λ ο γ ο ς

Για την παρούσα διπλωματική εργασία, η οποία εκπονήθηκε στο πλαίσιο των σπουδών μου στο Τμήμα Ηλεκτρονικών Μηχανικών του Ελληνικού Μεσογειακού Πανεπιστημίου, θα ήθελα αρχικά να ευχαριστήσω τον επιβλέποντα καθηγητή μου, Δρ. Ανάργυρο Μπακλέζο, για τη συνεργασία και την καθοδήγησή του κατά τη διάρκεια της εκπόνησης της εργασίας.

Πάνω απ΄όλα θέλω να ευχαριστήσω τους γονείς μου, οι οποίοι με στήριξαν σε όλη τη διάρκεια των σπουδών μου. Η συμπαράσταση, η υπομονή και η συνεχής ενθάρρυνσή τους αποτέλεσαν για εμένα πολύ σημαντικό στήριγμα σε κάθε στάδιο αυτής της διαδρομής.

Ευχαριστώ επίσης τους φίλους μου και τους ανθρώπους που στάθηκαν δίπλα μου όλο αυτό το διάστημα, για τη στήριξη, την κατανόηση και την παρουσία τους. Η συμβολή τους, με τον δικό της τρόπο, ήταν πολύτιμη στην προσπάθειά μου να ολοκληρώσω αυτή την εργασία.

Τέλος, θα ήθελα να ευχαριστήσω τους καθηγητές του Ελληνικού Μεσογειακού Πανεπιστημίου για τις γνώσεις και τα εφόδια που μου προσέφεραν κατά τη διάρκεια των σπουδών μου.

Χανιά 08/04/ 2026

Π ε ρ ί λ η ψ η

Η παρούσα διπλωματική εργασία εστιάζει στην ανάπτυξη, αξιολόγηση και σύγκριση μεθοδολογιών μηχανικής και βαθιάς μάθησης για την αυτόματη ανίχνευση πολυπόδων του παχέος εντέρου από ενδοσκοπικές εικόνες. Στόχος είναι η δημιουργία ενός αξιόπιστου συστήματος υποβοήθησης διαγνώσεων (CAD) ικανό να ενισχύσει την ακρίβεια των κολονοσκοπήσεων, να μειώσει το ποσοστό παράλειψης πολυπόδων και να συμβάλει αποτελεσματικά στην πρόληψη του καρκίνου του παχέος εντέρου, ο οποίος παραμένει μια από τις κύριες αιτίες θανάτου παγκοσμίως.

Αρχικά παρουσιάζεται το κλινικό και επιδημιολογικό υπόβαθρο των πολυπόδων, αναλύονται οι υφιστάμενες τεχνικές τεχνητής νοημοσύνης στο πεδίο της ιατρικής απεικόνισης και τα διαθέσιμα σύνολα δεδομένων. Στη συνέχεια, εξετάζονται βασικές αρχιτεκτονικές βαθιάς μάθησης, με ιδιαίτερη έμφαση στα Συνελικτικά Νευρωνικά Δίκτυα (CNNs) και τη στρατηγική της μεταφορά μάθησης. Για την υλοποίηση του συστήματος, επιλέχθηκε η αποδοτική αρχιτεκτονική EfficientNetB3, η οποία προσαρμόστηκε και βελτιστοποιήθηκε μέσω fine-tuning σε ένα ευρύ σύνολο δεδομένων πραγματικών κολονοσκοπικών εικόνων από το σύνολο δεδομένων REAL-colon.

Η μεθοδολογία περιλαμβάνει σχολαστική προεπεξεργασία, επαύξηση δεδομένων (data augmentation) και στρατηγικές για τη διαχείριση της έντονης ανισορροπίας μεταξύ θετικών και αρνητικών καρέ. Το τελικό μοντέλο αξιολογήθηκε σε ανεξάρτητο σύνολο ελέγχου, επιδεικνύοντας ανώτερη απόδοση με ακρίβεια (accuracy) 91,4%, ευαισθησία (recall) 92,3%, ακρίβεια προβλέψεων (precision) 89,6% και F1-score 90,9%. Η καμπύλη ROC (AUC=0,945) επιβεβαίωσε την υψηλή διακριτική ικανότητα του συστήματος. Επιπλέον, με την εφαρμογή της τεχνικής Grad-CAM, παρέχεται οπτική επισήμανση των περιοχών της εικόνας που επηρέασαν κρίσιμα την πρόβλεψη του μοντέλου, προσφέροντας ένα επίπεδο διαφάνειας και ερμηνευσιμότητας (Explainable AI).

Η εργασία ολοκληρώνεται με την αναγνώριση περιορισμών και με προτάσεις για μελλοντική έρευνα, συμπεριλαμβανομένης της χρήσης πιο εξελιγμένων αρχιτεκτονικών (π.χ., transformers) και της ενσωμάτωσης χρονικής πληροφορίας από ακολουθίες βίντεο. Το προτεινόμενο σύστημα αποτελεί ένα ισχυρό θεμέλιο για την ανάπτυξη κλινικά χρήσιμων εργαλείων που θα υποστηρίξουν τους ενδοσκόπους στην επίτευξη υψηλότερων ποσοστών ανίχνευσης και, εν τέλει, στην αποτελεσματικότερη πρόληψη του καρκίνου του παχέος εντέρου.

Λέξεις Κλειδιά: Ανίχνευση πολυπόδων, βαθιά μάθηση, συνελικτικά νευρωνικά δίκτυα (CNN), EfficientNet, μεταφορά μάθησης, ενδοσκοπική απεικόνιση, διάγνωση με υποβοήθεια υπολογιστή (CAD), καρκίνος παχέος εντέρου, κολονοσκόπηση, REAL-colon dataset.

A B S T R A C T

This thesis focuses on the development, evaluation, and comparison of machine and deep learning methodologies for the automatic detection of colorectal polyps from endoscopic images. The goal is to create a reliable computer-aided diagnosis (CAD) system capable of enhancing the accuracy of colonoscopies, reducing the polyp miss rate, and effectively contributing to the prevention of colorectal cancer, which remains a leading cause of death globally.

First, the clinical and epidemiological background of polyps is presented, existing artificial intelligence techniques in the field of medical imaging are analyzed, and available datasets are reviewed. Subsequently, fundamental deep learning architectures are examined, with particular emphasis on Convolutional Neural Networks (CNNs) and the transfer learning strategy. For the system's implementation, the efficient EfficientNetB3 architecture was chosen, adapted, and optimized via fine-tuning on a large dataset of real colonoscopy images from the REAL-colon dataset.

The methodology includes meticulous preprocessing, data augmentation, and strategies to handle the severe imbalance between positive and negative frames. The final model was evaluated on an independent test set, demonstrating superior performance with an accuracy of 91.4%, sensitivity (recall) of 92.3%, precision of 89.6%, and an F1-score of 90.9%. The ROC curve (AUC=0.945) confirmed the system's high discriminative ability. Furthermore, by applying the Grad-CAM technique, visual highlighting of the image regions that critically influenced the model's prediction is provided, offering a level of transparency and interpretability (Explainable AI).

The work concludes by acknowledging limitations—such as the difficulty in detecting very small or flat polyps—and by proposing directions for future research, including the use of more advanced architectures (e.g., transformers) and the integration of temporal information from video sequences. The proposed system constitutes a strong foundation for the development of clinically useful tools that will support endoscopists in achieving higher detection rates and, ultimately, in the more effective prevention of colorectal cancer.

Keywords: Polyp detection, deep learning, convolutional neural networks (CNN), EfficientNet, transfer learning, endoscopic imaging, computer-aided diagnosis (CAD), colorectal cancer, colonoscopy, REAL-colon dataset.

Π ε ρ ι ε χ ό μ ε ν α

1 Εισαγωγή 1

1.1 Σκοπός και σημασία της εργασίας 1

2 Πολύποδες Παχέος Εντέρου 3

3 Η Τεχνητή Νοημοσύνη και η Χρήση της στην Ανίχνευση Πολυπόδων του Παχέος Εντέρου 9

3.1 Διάγνωση με τη βοήθεια υπολογιστή (CAD) στην κολονοσκόπηση 11

3.2 Πρώιμες τεχνικές επεξεργασίας εικόνας (ανίχνευση ακμών, μορφολογία) 11

3.3 Μετάβαση σε προσεγγίσεις μηχανικής μάθησης και βαθιάς μάθησης 12

3.4 Η κατάσταση της τεχνολογίας στην ανίχνευση πολυπόδων 13

3.5 Ερευνητικά κενά 31

4 Συνελικτικά Νευρωνικά Δίκτυα 35

4.1 Η βαθιά μάθηση στην ιατρική απεικόνιση 35

4.2 Επισκόπηση των CNNs, της μεταφοράς μάθησης και της ανίχνευσης αντικειμένων 35

4.3 Σχετικές αρχιτεκτονικές: VGG, ResNet, U-Net, YOLO, EfficientNet κ.λπ. 37

4.4 CNNs 39

4.5 Μετρικές αξιολόγησης 41

5 EfficientNetBX 45

5.1 Η αρχιτεκτονική EfficientNetBx 45

5.2 EfficientNetB3 47

6 Δεδομένα 49

6.1 Κατασκευή και οργάνωση του dataset 50

7 Υλοποίηση και Αξιολόγηση του Μοντέλου 53

7.1 Δημιουργία και ανάπτυξη μοντέλου 53

7.2 Διαχείριση ανισορροπίας και στρατηγικές δειγματοληψίας 59

7.3 Προεπεξεργασία και επαύξηση δεδομένων 59

7.4 Επιλογή και εξέλιξη αρχιτεκτονικής 61

7.5 Στρατηγική εκπαίδευσης και fine-tuning 62

7.6 Πειραματισμοί, αποτυχίες και σταθεροποίηση του συστήματος 63

7.7 Τελική ρύθμιση και αναπαραγωγιμότητα 64

7.8 Περιγραφή υλοποίησης και παραπομπή στον κώδικα 65

7.8.1 Σύνοψη κώδικα εκπαίδευσης μοντέλου 65

7.8.2 Σύνοψη κώδικα αξιολόγησης μοντέλου 65

7.9 Αποτελέσματα 66

7.9.1 Συνοπτικά αποτελέσματα τελικού μοντέλου 66
7.9.2 Συγκριτικά αποτελέσματα αρχιτεκτονικών 66
7.9.3 Οπτικοποίηση αποτελεσμάτων 67

8 eXplainable AI 71

8.1 Εφαρμογή Grad-CAM στο προτεινόμενο σύστημα 72

9 Συμπεράσματα και Μελλοντικές Προτάσεις 75

9.1 Συμπεράσματα 75

9.2 Περίληψη της εργασίας 75

9.3 Βασικές συνεισφορές 77

9.4 Τελικές παρατηρήσεις 78

9.5 Μελλοντικές προτάσεις 79
9.5.1 Βελτίωση της αρχιτεκτονικής και των τεχνικών μάθησης 79
9.5.2 Επέκταση και βελτίωση συνόλων δεδομένων 80
9.5.3 Ενσωμάτωση πληροφορίας από βίντεο και πολυτροπικά δεδομένα 80
9.5.4 Κλινική εφαρμογή και επικύρωση 80
9.5.5 Εξηγήσιμη ΤΝ και υποστήριξη απόφασης 81
9.5.6 Διεπιστημονική συνεργασία και ηθικά πλαίσια 81

10 Βιβλιογραφία 83

Παράρτημα Α: Κώδικας Εκπαίδευσης Μοντέλου 89

Παράρτημα Β: Κώδικας Αξιολόγησης Μοντέλου 96

Ευρετήριο Εικόνων

Εικόνα 1 – Ανίχνευση καρκίνου και αντίστοιχοι θάνατοι το 2020 [1] 1
Εικόνα 2 – Πολύποδες μέσα σε έντερο ασθενούς [5] 3
Εικόνα 3 – Καυτηρίαση πολύποδα ενδοσκοπικά με χρήση λειζερ [7]. 4
Εικόνα 4 – Αιτίες ανάπτυξης πολυπόδων [4] 5
Εικόνα 5 – Αφαίρεση πολύποδα ενδοσκοπικά [7] 6
Εικόνα 6 – Εξέλιξη ακρίβειας αλλά και υπολογιστικής πολυπλοκότητας συστημάτων τεχνητής νοημοσύνης [13] 10
Εικόνα 7 – Ανίχνευση πολυπόδων με χρήση CAD [21] 15
Εικόνα 8 – Αναπαράσταση αρχιτεκτονικής Convolutional Neural Network [49] 40
Εικόνα 9 – Αξιολόγηση μοντέλου με τη χρήση της καμπύλης ROC [54] 43
Εικόνα 10 – Αναπαράσταση αρχιτεκτονικής EfficientNet [13] 45
Εικόνα 11 – Πρόβλημα overfitting στα αρχικά στάδια του μοντέλου (το μοντέλο έχει μάθει τα δεδομένα εκπαίδευσης και εμφανίζει accuracy: 100%) 54
Εικόνα 12 – Αποτελέσματα μοντέλου που παρουσίασε overfitting (ενώ το μοντέλο έχει μάθει τα δεδομένα εκπαίδευσης, στο validation set η ακρίβεια ήταν κάτω από 30%) 54
Εικόνα 13 – Αποτελέσματα Confusion Matrix 68
Εικόνα 14 – Αποτελέσματα ROC Curve 68
Εικόνα 15 – Αποτελέσματα Sample Predictions 69
Εικόνα 16 – Αναπαράσταση αποτελεσμάτων Grad-CAM ως Heatmaps. 73

Ευρετήριο Πινάκων

Πίνακας 1- Αποτελέσματα τελικού μοντέλου στο test set 66
Πίνακας 2 - Συγκριτική αξιολόγηση αρχιτεκτονικών στο test set 67

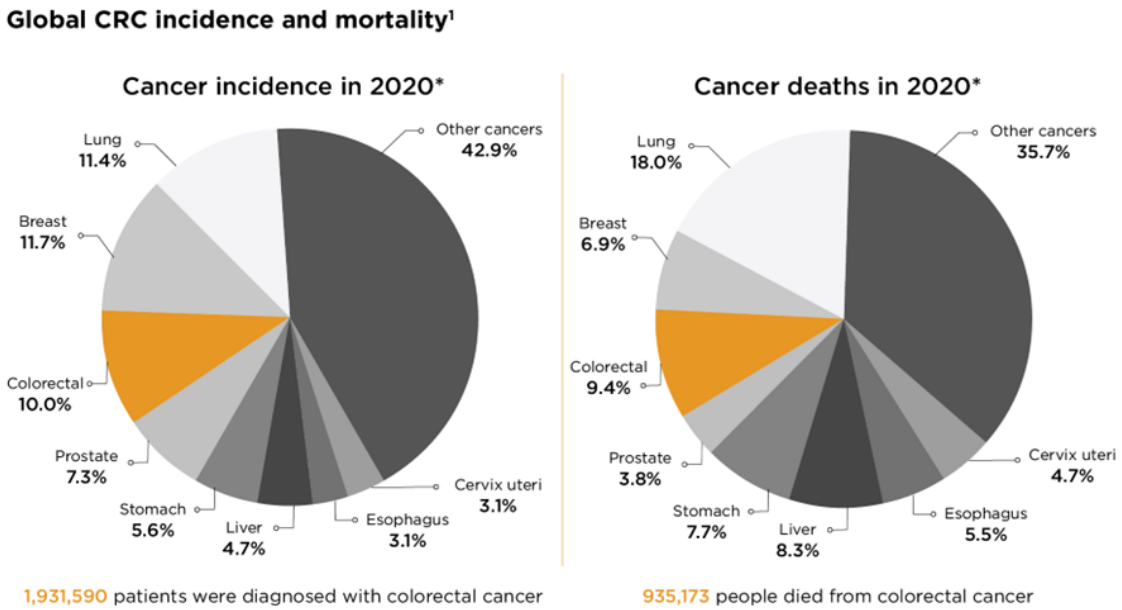
ΚΕΦΑΛΑΙΟ 1

1 Εισαγωγή

1.1 Σκοπός και σημασία της εργασίας

Η πρόληψη και έγκαιρη διάγνωση του καρκίνου του παχέος εντέρου αποτελεί ένα από τα σημαντικότερα ζητήματα δημόσιας υγείας του 21ου αιώνα. Σύμφωνα με στοιχεία της Διεθνούς Οργάνωσης για την Έρευνα στον Καρκίνο (IARC), ο καρκίνος του παχέος εντέρου αποτελεί τη δεύτερη συχνότερη αιτία θανάτου από καρκίνο παγκοσμίως, με περισσότερους από 1,9 εκατομμύρια νέους διαγνωσμένους ασθενείς ετησίως [1]. Η συντριπτική πλειοψηφία αυτών των περιστατικών ξεκινά με την εμφάνιση πολύποδων, οι οποίοι αν δεν ανιχνευθούν και αφαιρεθούν έγκαιρα, μπορεί να εξελιχθούν σε αδενωματώδεις βλάβες και στη συνέχεια σε καρκίνο.

Οι πολύποδες του παχέος εντέρου αποτελούν μορφώματα του βλεννογόνου, τα οποία στην πλειονότητα των περιπτώσεων είναι καλοήθη, όμως οι λεγόμενοι προκαρκινικοί αδενωματώδεις πολύποδες ενέχουν αυξημένο κίνδυνο κακοήθους εξαλλαγής. Η κολονοσκόπηση παραμένει μέχρι σήμερα η πιο αξιόπιστη μέθοδος εντοπισμού και αφαίρεσης των πολύποδων, όμως ακόμη και οι πιο έμπειροι γαστρεντερολόγοι ενδέχεται να παραλείψουν έως και 22% των μικρών ή επίπεδων πολύποδων κατά την εξέταση [2].



Εικόνα 1 – Ανίχνευση καρκίνου και αντίστοιχοι θάνατοι το 2020 [1]

Η ανθρώπινη κόπωση, η εμπειρία του ιατρού, αλλά και η πολυπλοκότητα του εντερικού τοπίου κατά την κολονοσκόπηση, αποτελούν βασικούς παράγοντες σφαλμάτων. Αυτή η κατάσταση

καθιστά σαφές ότι η εισαγωγή αυτόματων μεθόδων ανίχνευσης πολύποδων, με τη βοήθεια της τεχνητής νοημοσύνης και της μηχανικής μάθησης, αποτελεί ένα κρίσιμο τεχνολογικό άλμα.

Κατά την τελευταία δεκαετία, οι τεχνικές βαθιάς μάθησης (deep learning) και ιδιαίτερα τα συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNNs) έχουν φέρει επανάσταση στην ανάλυση ιατρικών εικόνων. Εφαρμογές τους περιλαμβάνουν από την αναγνώριση όγκων σε μαγνητικές τομογραφίες, μέχρι την αυτόματη ταξινόμηση δερματικών βλαβών. Στο πεδίο της γαστρεντερολογίας, το 2020 και μετά παρατηρείται ραγδαία ανάπτυξη αλγορίθμων για την αυτόματη ανίχνευση πολύποδων σε ενδοσκοπικά βίντεο και εικόνες, με ακρίβειες που σε ορισμένες περιπτώσεις ξεπερνούν το 94% [3].

Η παρούσα διπλωματική εργασία εστιάζει στην ανάπτυξη, αξιολόγηση και σύγκριση διαφορετικών μοντέλων μηχανικής μάθησης για τον εντοπισμό πολύποδων σε ενδοσκοπικές εικόνες παχέος εντέρου. Ο βασικός στόχος είναι η αυτοματοποίηση της διαδικασίας εντοπισμού, με σκοπό να μειωθεί το ποσοστό μη ανιχνευμένων αλλοιώσεων και να ενισχυθεί η αποτελεσματικότητα της προληπτικής ιατρικής.

Για τον σκοπό αυτό, εκπαιδεύτηκε και βελτιστοποιήθηκε ένα νευρωνικό δίκτυο τύπου CNN, με χρήση προεπεξεργασμένων εικόνων από ανοικτά και επικυρωμένα σύνολα δεδομένων. Επιπλέον, δοκιμάστηκαν μοντέλα transfer learning με χρήση προκαταρτισμένων αρχιτεκτονικών όπως ResNet50, EfficientNet και YOLOv5, συγκρίνοντας την απόδοσή τους με βάση κλασικούς δείκτες (accuracy, recall, precision, F1-score).

Η μεθοδολογία περιλαμβάνει αναλυτική επεξεργασία εικόνας, στατιστική αξιολόγηση αποτελεσμάτων, και χρήση κατάλληλων εργαλείων λογισμικού όπως TensorFlow, PyTorch και OpenCV.

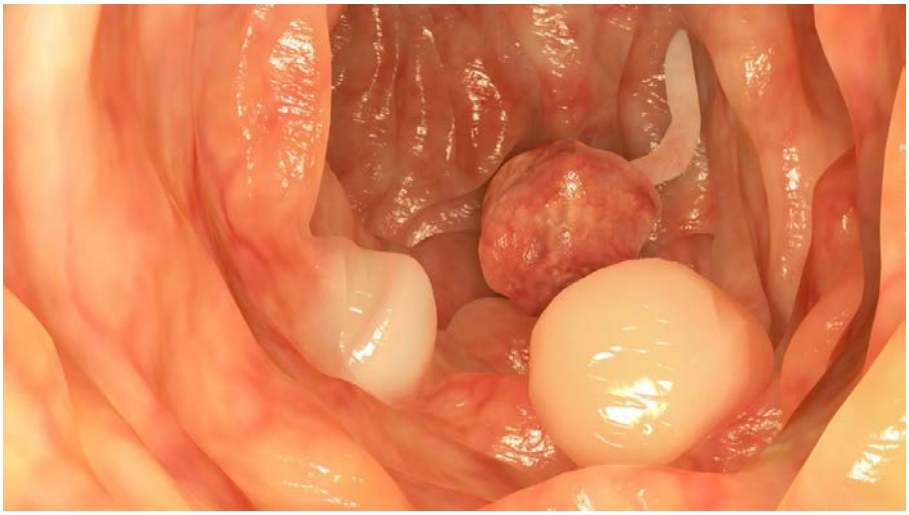
Η παρούσα εργασία φιλοδοξεί να συνεισφέρει στη διεύρυνση της κατανόησης και αξιοποίησης της τεχνητής νοημοσύνης στην ιατρική πρακτική, αποδεικνύοντας έμπρακτα τη σημασία της διεπιστημονικής προσέγγισης στο πεδίο της Υγείας.

ΚΕΦΑΛΑΙΟ 2

2 Πολύποδες Παχέος Εντέρου

Οι πολύποδες του παχέος εντέρου αποτελούν ευρήματα υψίστης σημασίας στο πλαίσιο της προληπτικής γαστρεντερολογικής πρακτικής. Πρόκειται για ανώμαλες εκβλαστήσεις του εντερικού βλεννογόνου που αναπτύσσονται ενδοαυλικά, κυρίως στο παχύ έντερο και στο ορθό. Αν και οι περισσότεροι είναι καλοήθεις, συγκεκριμένοι τύποι, όπως οι αδενωματώδεις και οι οδοντωτοί πολύποδες, παρουσιάζουν δυνητική κακοήθη εξαλλαγή, καθιστώντας την έγκαιρη ανίχνευση και αφαίρεσή τους κεντρική προτεραιότητα στην πρόληψη του καρκίνου του παχέος εντέρου.

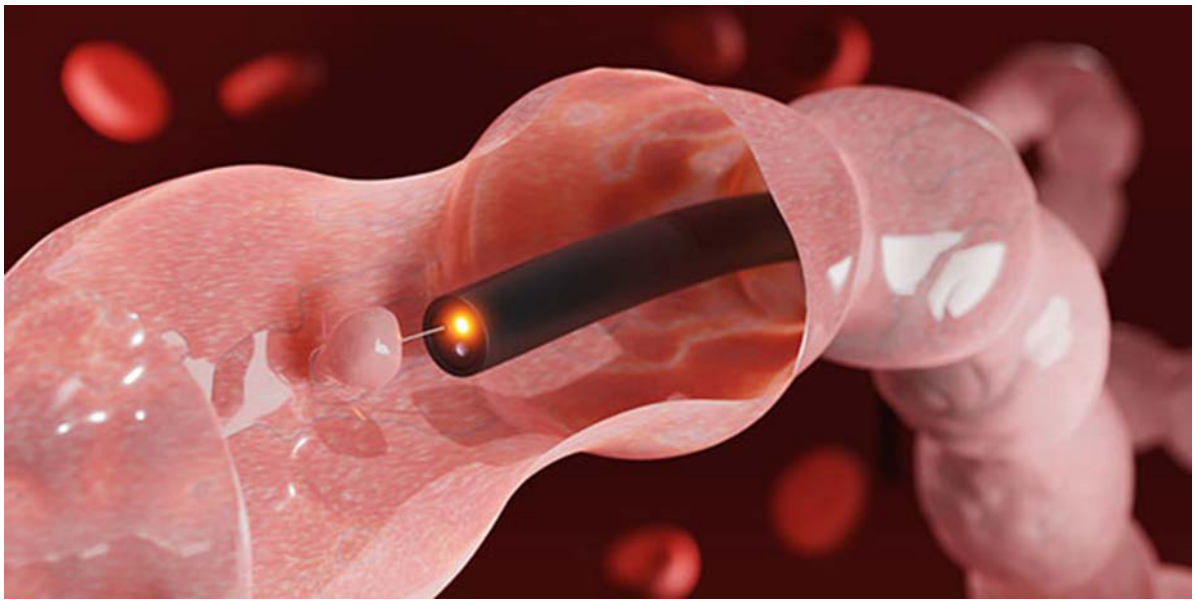
Το ενδιαφέρον της ιατρικής κοινότητας για τους πολύποδες έχει ενταθεί ιδιαίτερα τις τελευταίες δεκαετίες, καθώς αυξάνεται η κατανόηση της βιολογίας τους και της συμβολής τους στην αλληλουχία γεγονότων που οδηγεί στην καρκινογένεση. Ο λεγόμενος «αδενωματώδης-καρκινικός» μετασχηματισμός θεωρείται πλέον τεκμηριωμένη βιολογική οδός, γεγονός που επιβάλλει την κλινική επαγρύπνηση και τη συστηματική επιτήρηση των ατόμων που παρουσιάζουν τέτοιου είδους αλλοιώσεις [4].



Εικόνα 2 – Πολύποδες μέσα σε έντερο ασθενούς [5]

Οι πολυποειδείς βλάβες εντοπίζονται συνήθως μέσω κολονοσκόπησης, η οποία αποτελεί την πιο αξιόπιστη διαγνωστική και θεραπευτική μέθοδο. Ωστόσο, δεν είναι όλες οι πολυποειδείς αλλοιώσεις ίδιες, ούτε παρουσιάζουν τον ίδιο βαθμό επικινδυνότητας. Η ιστολογική ταξινόμηση των πολύποδων περιλαμβάνει κυρίως τους υπερπλαστικούς, τους αδενωματώδεις και τους οδοντωτούς πολύποδες. Από αυτούς, οι υπερπλαστικοί θεωρούνται γενικά αβλαβείς, ενώ οι υπόλοιποι ενέχουν σημαντικό κίνδυνο κακοήθους εξέλιξης.

Η παρουσία πολύποδων είναι συχνή, ειδικά σε ηλικιωμένους ενήλικες. Σύμφωνα με μελέτες πληθυσμιακής βάσης, η ανεύρεση ενός ή περισσοτέρων πολύποδων σε άτομα ηλικίας άνω των 50 ετών αγγίζει το 30–40% [6]. Αυτή η υψηλή συχνότητα, σε συνδυασμό με τον χαμηλό βαθμό συμπτωματολογίας, καθιστά επιτακτική την ανάπτυξη αποτελεσματικών στρατηγικών μαζικού ελέγχου και αυτοματοποιημένης ανάλυσης ενδοσκοπικών εικόνων.



Εικόνα 3 – Καυτηρίαση πολύποδα ενδοσκοπικά με χρήση λειζερ [7].

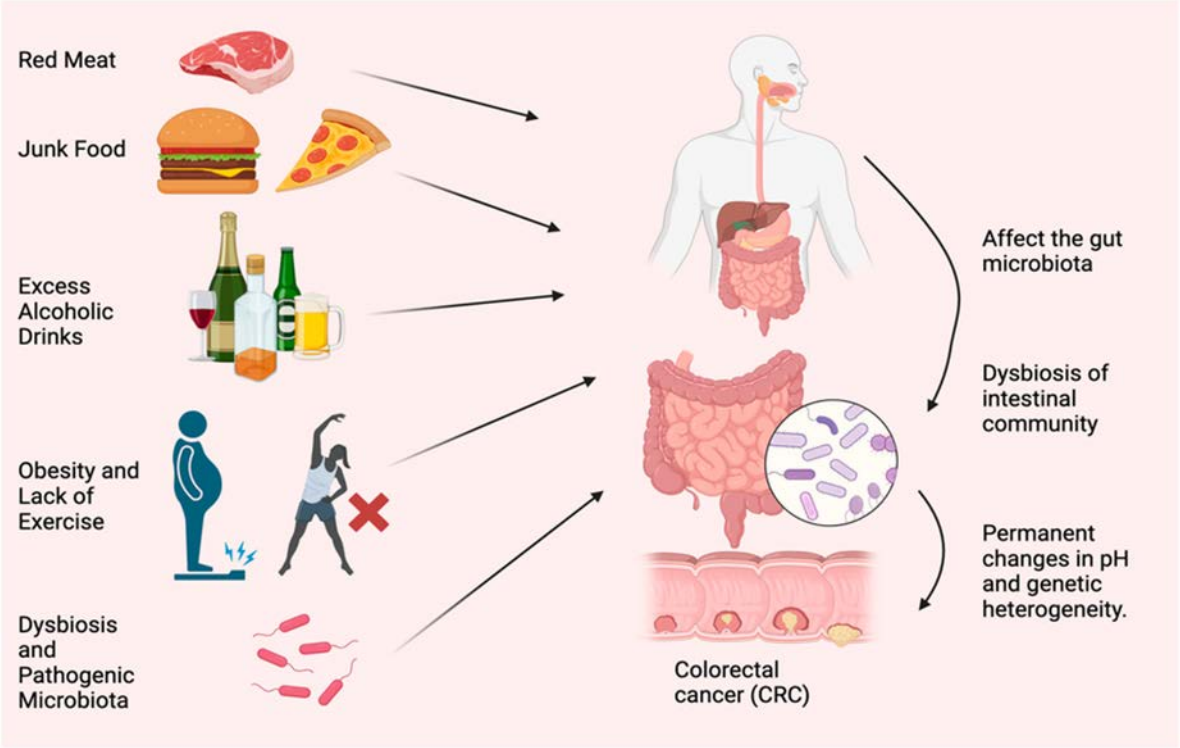
Πέραν της συχνότητας, εξίσου σημαντική είναι η παρακολούθηση της εξελικτικής δυναμικής των πολυπόδων. Ο αδενωματώδης πολύποδας, και ειδικότερα οι υποκατηγορίες του με λαχνωτό στοιχείο ή υψηλό βαθμό δυσπλασίας, θεωρείται πρόδρομη μορφή καρκινικής μετάλλαξης. Αυτή η εξελικτική πορεία, γνωστή ως “adenoma–carcinoma sequence”, μπορεί να διαρκέσει 5 έως 15 έτη, γεγονός που προσφέρει πολύτιμο διαγνωστικό παράθυρο ευκαιρίας, κατάλληλο για παρέμβαση [8].

Στον αντίποδα, οι οδοντωτοί πολύποδες, αν και ιστολογικά διακριτοί, ακολουθούν εναλλακτική καρκινογόνο πορεία (serrated pathway), η οποία συνδέεται με μικροδορυφορική αστάθεια και υπερμεθυλίωση DNA. Οι βλάβες αυτές, ειδικά όταν είναι επίπεδες και ανευρίσκονται στο εγγύς κόλον, συχνά διαφεύγουν της διάγνωσης, γεγονός που συντελεί σε καθυστερημένη ανίχνευση και υψηλότερη θνητότητα.

Οι πολυπαραγοντικές αιτίες που συνδέονται με την εμφάνιση πολύποδων περιλαμβάνουν τόσο γενετικούς όσο και περιβαλλοντικούς παράγοντες. Μεταξύ των τροποποιήσιμων παραγόντων κινδύνου περιλαμβάνονται:

- Η αυξημένη πρόσληψη επεξεργασμένου κρέατος και λιπών
- Η φτωχή σε φυτικές ίνες διατροφή
- Το κάπνισμα και η συστηματική κατανάλωση αλκοόλ
- Η καθιστική ζωή και η παχυσαρκία
- Η έλλειψη χρήσης μη στεροειδών αντιφλεγμονωδών φαρμάκων (NSAIDs), τα οποία έχει δειχθεί ότι μειώνουν τον κίνδυνο δυσπλασίας

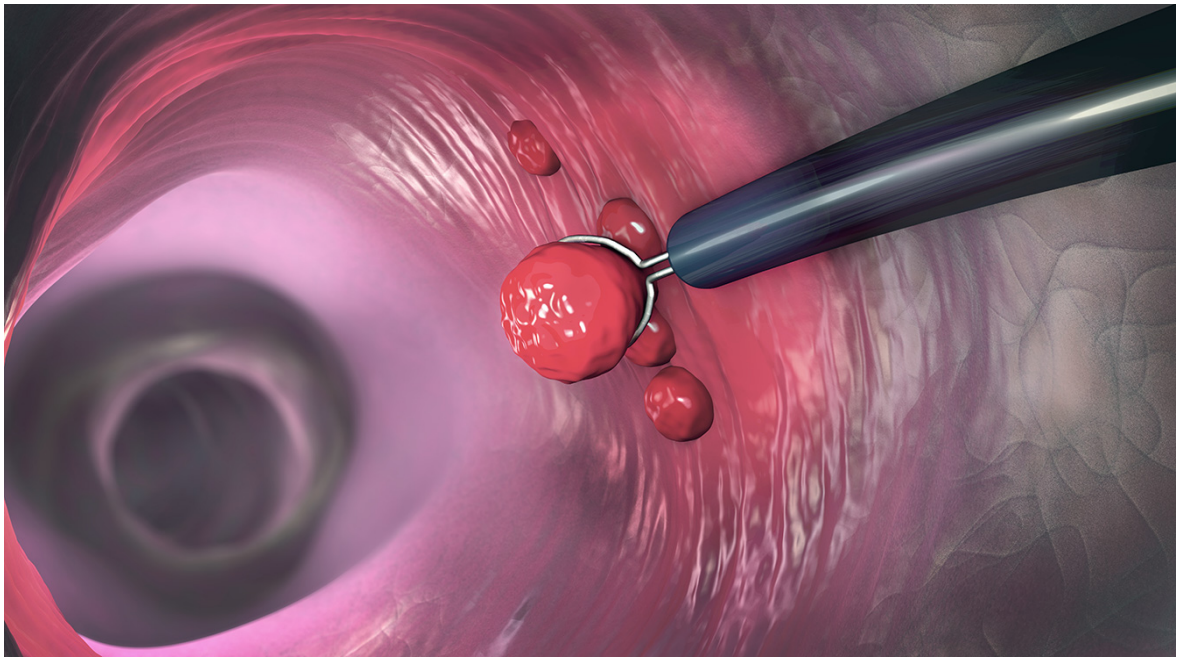
Οι παραπάνω παράγοντες δρουν συνεργικά με την ηλικία, το φύλο (με τους άνδρες να εμφανίζουν μεγαλύτερο κίνδυνο), καθώς και το οικογενειακό ιστορικό. Σε ορισμένες περιπτώσεις, η παρουσία συγκεκριμένων μεταλλάξεων στο DNA καθιστά τα άτομα ιδιαίτερα ευπαθή [9].



Εικόνα 4 – Αιτίες ανάπτυξης πολυπόδων [4]

Τα τελευταία χρόνια παρατηρείται σημαντική μετατόπιση του επιδημιολογικού προφίλ, με την εμφάνιση πολύποδων και καρκίνου σε άτομα νεότερης ηλικίας. Μελέτες που αναλύουν δεδομένα από την τελευταία 20ετία υποδεικνύουν αύξηση της επίπτωσης πολυπόδων σε άτομα ηλικίας 30–45 ετών, γεγονός που πιθανόν σχετίζεται με τις αλλαγές στη διατροφή και τον τρόπο ζωής [10]. Αυτή η νέα πραγματικότητα υπογραμμίζει την ανάγκη για επαναξιολόγηση των ηλικιακών ορίων έναρξης των ελέγχων.

Η πρόληψη και η πρώιμη παρέμβαση καθίστανται κρίσιμες, καθώς ο εντοπισμός και η αφαίρεση ενός πολυπόδα όχι μόνο περιορίζει τον κίνδυνο καρκίνου, αλλά μειώνει και την ανάγκη για επεμβατικές ή συστηματικές θεραπείες. Είναι ενδεικτικό ότι, σύμφωνα με μελέτες, η αφαίρεση αδενωματωδών πολυπόδων μειώνει τον κίνδυνο εμφάνισης καρκίνου του παχέος εντέρου έως και 76% [11].



Εικόνα 5 – Αφαίρεση πολύποδα ενδοσκοπικά [7]

Η κοινή κολονοσκόπηση, αν και αποτελεσματική, παρουσιάζει περιορισμούς που σχετίζονται κυρίως με την υποκειμενικότητα και την εμπειρία του εξεταστή. Υπολογίζεται ότι ποσοστό 15–22% των πολυπόδων μπορεί να διαφύγει κατά την πρώτη ενδοσκοπική εξέταση. Οι κύριοι λόγοι είναι η ύπαρξη επίπεδων βλαβών, η κακή προετοιμασία του εντέρου και η κόπωση του ιατρού [2].

Η ανάγκη ενίσχυσης της ακρίβειας της εξέτασης μέσω βοηθητικών τεχνολογιών είναι πλέον σαφής. Η ενσωμάτωση υπολογιστικών συστημάτων βασισμένων στην τεχνητή νοημοσύνη προσφέρει μια νέα διάσταση στην ενδοσκοπική απεικόνιση, βελτιώνοντας την ταχύτητα εντοπισμού και μειώνοντας τις πιθανότητες παράλειψης.

Πέρα από την κλινική διάσταση, το πρόβλημα των πολύποδων έχει και σημαντική οικονομική διάσταση. Το κόστος θεραπείας του καρκίνου του παχέος εντέρου σε προχωρημένα στάδια είναι κατά πολύ υψηλότερο σε σύγκριση με εκείνο της πρόληψης. Οι υπολογισμοί για το σύστημα υγείας των Ηνωμένων Πολιτειών καταγράφουν ετήσια δαπάνη που υπερβαίνει τα 10 δισεκατομμύρια δολάρια μόνο για θεραπείες σχετιζόμενες με τον καρκίνο του παχέος εντέρου [12].

Η πρώιμη ανίχνευση μέσω προληπτικών κολονοσκοπήσεων και αυτοματοποιημένων συστημάτων θα μπορούσε να περιορίσει αυτό το κόστος, καθιστώντας το σύστημα υγείας βιωσιμότερο και αποδοτικότερο.

Στην Ευρώπη, έχουν ξεκινήσει μαζικές προληπτικές καμπάνιες ελέγχου του πληθυσμού για πολύποδες, με έμφαση στην ηλικιακή ομάδα 50–74 ετών. Ωστόσο, ακόμα και στις ανεπτυγμένες χώρες, η συμμόρφωση στον προληπτικό έλεγχο παραμένει χαμηλή. Σε αυτό το πλαίσιο, η ενσωμάτωση τεχνολογιών μηχανικής μάθησης, που επιτρέπουν την ανίχνευση ακόμα και μικρών, ασαφών πολύποδων σε ενδοσκοπικές εικόνες, θα μπορούσε να ενισχύσει το επίπεδο εμπιστοσύνης και αποτελεσματικότητας του συστήματος.

Το επόμενο κεφάλαιο εξετάζει ακριβώς αυτό: πώς η τεχνητή νοημοσύνη εφαρμόζεται πλέον στον κλάδο της ιατρικής απεικόνισης και ειδικότερα στη γαστρεντερολογία.

Η βαθύτερη κατανόηση του προβλήματος των πολύποδων μάς οδηγεί σε τρία κρίσιμα συμπεράσματα:

1. Οι πολύποδες είναι εξαιρετικά συχνοί, πολλές φορές ασυμπτωματικοί, και εντοπίζονται καλύτερα με κολονοσκόπηση.

2. Ορισμένοι εξ αυτών φέρουν σημαντικό κακοήθη δυναμικό και η αφαίρεσή τους έχει αποδεδειγμένο προληπτικό όφελος.

3. Υπάρχει ανάγκη για προηγμένες μεθόδους εντοπισμού με χρήση τεχνητής νοημοσύνης και βαθιάς μάθησης, που θα ενισχύσουν το έργο του ιατρού και θα μειώσουν τα ποσοστά αποτυχίας ανίχνευσης.

ΚΕΦΑΛΑΙΟ 3

3 Η Τεχνητή Νοημοσύνη και η Χρήση της στην Ανίχνευση Πολυπόδων του Παχέος Εντέρου

Η ραγδαία πρόοδος της τεχνητής νοημοσύνης (ΤΝ) κατά τις δύο τελευταίες δεκαετίες έχει προκαλέσει μια βαθιά μεταμόρφωση σε πολλούς τομείς της επιστήμης και της τεχνολογίας, μεταξύ των οποίων η ιατρική κατέχει εξέχουσα θέση. Η ικανότητα των αλγορίθμων να μαθαίνουν από δεδομένα, να εντοπίζουν μοτίβα, να προβλέπουν εξελίξεις και να υποστηρίζουν τη λήψη κλινικών αποφάσεων έχει οδηγήσει στη δημιουργία νέων εργαλείων για την πρόληψη, τη διάγνωση και τη θεραπεία ασθενειών.

Ειδικότερα, στον τομέα της ιατρικής απεικόνισης, η ΤΝ έχει προσφέρει τη δυνατότητα επεξεργασίας και ερμηνείας μεγάλου όγκου οπτικών δεδομένων, επιτρέποντας τη βελτιστοποίηση της διαγνωστικής ακρίβειας. Σε εξετάσεις όπως οι ακτινογραφίες, οι τομογραφίες και οι ενδοσκοπήσεις, η συμβολή της βαθιάς μάθησης (deep learning – DL) έχει αποδειχθεί καθοριστική, μειώνοντας τα ποσοστά σφαλμάτων και ενισχύοντας την πρώιμη ανίχνευση ασθενειών.

Στο πλαίσιο της γαστρεντερολογίας, και ιδιαίτερα της ενδοσκόπησης του παχέος εντέρου (κολονοσκόπηση), η ΤΝ έχει ήδη αρχίσει να εφαρμόζεται ευρέως. Η διαδικασία της κολονοσκόπησης ενέχει πολυπλοκότητα και απαιτεί υψηλό επίπεδο προσοχής από τον ιατρό, καθώς η παράλειψη μικρών, επίπεδων ή δύσκολα ορατών πολυπόδων μπορεί να έχει σοβαρές κλινικές συνέπειες.

Η ανάπτυξη αλγορίθμων βαθιάς μάθησης, και ιδιαίτερα των συνελικτικών νευρωνικών δικτύων (Convolutional Neural Networks – CNNs), έχει καταστήσει εφικτή την αυτόματη αναγνώριση και επισήμανση πολύποδων σε πραγματικό χρόνο κατά τη διάρκεια της κολονοσκόπησης. Οι αλγόριθμοι αυτοί "εκπαιδεύονται" σε μεγάλες βάσεις ενδοσκοπικών εικόνων και μαθαίνουν να διακρίνουν μορφολογικά πρότυπα που συνδέονται με την παρουσία πολυπόδων.

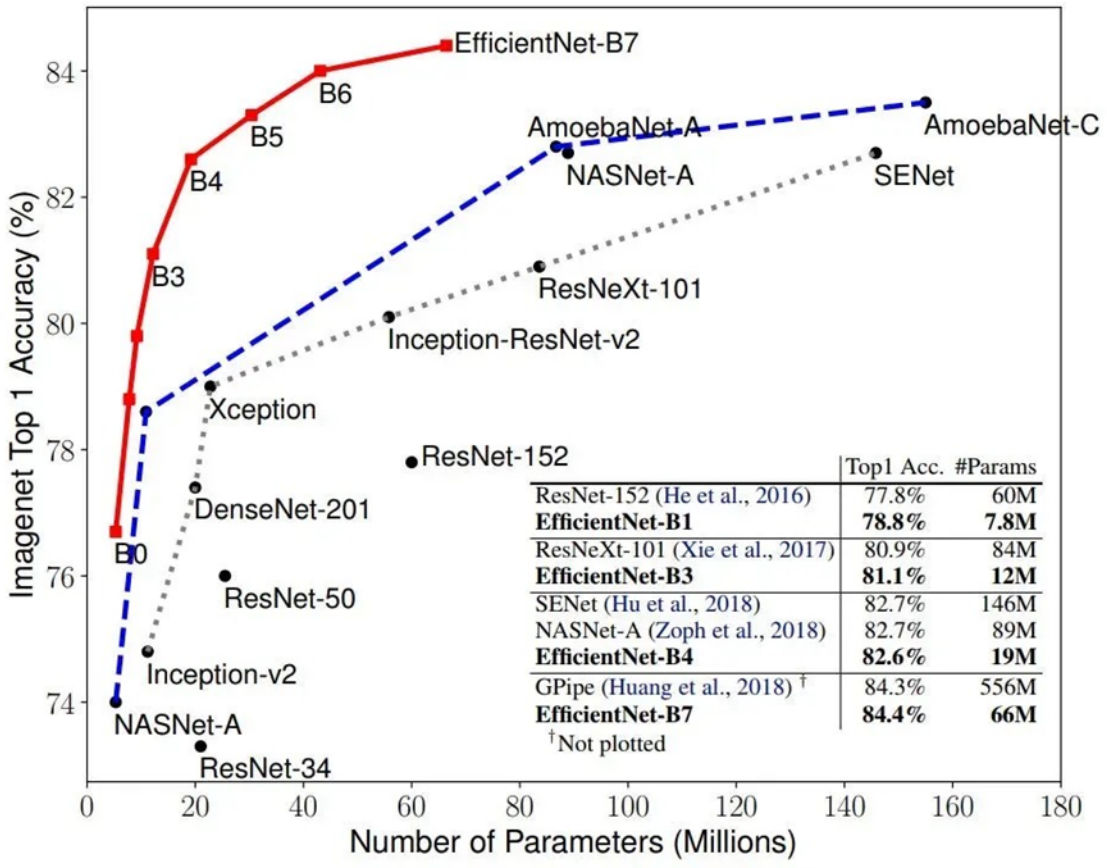
Η χρήση CNNs στην ανάλυση ιατρικών εικόνων έχει αποδειχθεί εξαιρετικά αποτελεσματική, καθώς οι δομές αυτών των δικτύων έχουν σχεδιαστεί ώστε να ανιχνεύουν χαρακτηριστικά όπως άκρες, χρώματα, σχήματα και υφές. Κατά την επεξεργασία ενδοσκοπικών εικόνων, τα χαρακτηριστικά αυτά μπορούν να συσχετιστούν με την επιφάνεια του εντερικού βλεννογόνου, εντοπίζοντας περιοχές που ενδέχεται να περιέχουν πολύποδες.

Σημαντικές μελέτες, όπως εκείνη του Wang [2], έδειξαν ότι τα συστήματα τεχνητής νοημοσύνης μπορούν να αυξήσουν τον ρυθμό ανίχνευσης αδενωματωδών πολυπόδων κατά 14%, σε σύγκριση με την παραδοσιακή κολονοσκόπηση. Τα ευρήματα αυτά επιβεβαιώνουν τη

συμπληρωματική αξία της ΤΝ και τονίζουν την ανάγκη ευρύτερης ενσωμάτωσής της στην καθημερινή κλινική πράξη.

Η εφαρμογή των συστημάτων ΤΝ δεν περιορίζεται μόνο στην αναγνώριση, αλλά επεκτείνεται και στην ταξινόμηση και εκτίμηση του καρκινικού δυναμικού ενός πολύποδα. Μέσω της χρήσης τεχνικών όπως το transfer learning και τα pretrained models (π.χ. ResNet, EfficientNet, Inception), οι αλγόριθμοι μαθαίνουν να διαφοροποιούν μεταξύ καλοήθων, προκαρκινικών και ύποπτων για κακοήθεια αλλοιώσεων.

Μία από τις πιο γνωστές εμπορικές εφαρμογές είναι το GI-Genius της Medtronic, ένα σύστημα πιστοποιημένο στην Ευρώπη και στις ΗΠΑ που λειτουργεί σε πραγματικό χρόνο και παρέχει οπτική επισήμανση ύποπτων περιοχών κατά την κολονοσκόπηση. Παρόμοια συστήματα αναπτύσσονται και από ακαδημαϊκά ιδρύματα, όπως το EndoBRAIN της FUJIFILM και το CAD-EYE της Olympus, τα οποία συνδυάζουν CNNs με decision trees για εκτίμηση μορφολογίας.



	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**78.8%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.1%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**82.6%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.4%**	**66M**

†Not plotted

Εικόνα 6 – Εξέλιξη ακρίβειας αλλά και υπολογιστικής πολυπλοκότητας συστημάτων τεχνητής νοημοσύνης [13]

Η εκπαίδευση των μοντέλων αυτών προϋποθέτει την ύπαρξη επαρκών και ποικίλων συνόλων δεδομένων. Δημόσιες βάσεις δεδομένων, όπως το Kvasir-SEG, το CVC-ClinicDB και το EndoTect Challenge Dataset, παρέχουν επισημασμένες ενδοσκοπικές εικόνες, δίνοντας τη

δυνατότητα σε ερευνητές και οργανισμούς να εκπαιδεύσουν και να αξιολογήσουν αλγόριθμους με ρεαλιστικά δεδομένα.

Ένα πρόβλημα που εντοπίζεται είναι η ετερογένεια των εικόνων, λόγω διαφορετικών πηγών φωτισμού, γωνιών λήψης, τύπων ενδοσκοπίου και επιπέδων καθαρότητας του εντέρου. Αυτό καθιστά απαραίτητη την εφαρμογή τεχνικών κανονικοποίησης (normalization), data augmentation και transfer learning για να εξασφαλιστεί η γενίκευση των μοντέλων.

Η πραγματική πρόκληση όμως παραμένει η μετάβαση από τα πειραματικά δεδομένα σε εφαρμογές κλινικής πράξης. Εδώ ανακύπτουν ζητήματα όπως η αξιοπιστία, η κανονιστική συμμόρφωση, η διαλειτουργικότητα με το ενδοσκοπικό σύστημα, και φυσικά η αποδοχή από τους ιατρούς.

3.1 Διάγνωση με τη βοήθεια υπολογιστή (CAD) στην κολονοσκόπηση

Η ανάλυση ιατρικών εικόνων έχει αλλάξει δραματικά με την έλευση των συστημάτων διάγνωσης με τη βοήθεια υπολογιστή (CAD), ιδίως στην ενδοσκόπηση του γαστρεντερικού συστήματος, με τα συστήματα CAD να προορίζονται να βοηθήσουν τους γιατρούς να αναγνωρίζουν, να κατηγοριοποιούν και να εντοπίζουν τους πολύποδες του παχέος εντέρου κατά τη διάρκεια των κολονοσκοπήσεων, αυξάνοντας τη διαγνωστική ακρίβεια και μειώνοντας το ανθρώπινο σφάλμα. Οι εγγενείς περιορισμοί της χειροκίνητης οπτικής επιθεώρησης, βασιζόμενη κυρίως στην ικανότητα, την εστίαση και την εμπειρία του ενδοσκόπου, αποτελούν την κινητήρια δύναμη πίσω από αυτά τα συστήματα, με το ποσοστό αστοχίας πολυπόδων να εξακολουθεί να είναι υψηλό ακόμη και σε εξειδικευμένα χέρια, ιδίως για μικρές, επίπεδες ή άμισχες βλάβες που αναμειγνύονται με το βλεννογονικό υπόβαθρο [7]. Προκειμένου να ξεπεραστούν αυτές οι δυσκολίες, τα συστήματα CAD λειτουργούν ως "δεύτερος παρατηρητής" για τη βελτίωση της κλινικής απόδοσης προσφέροντας αυτοματοποιημένη, αξιόπιστη και αμερόληπτη βοήθεια κατά την ερμηνεία των εικόνων.

3.2 Πρώιμες τεχνικές επεξεργασίας εικόνας (ανίχνευση ακμών, μορφολογία)

Τα πρώτα συστήματα CAD κολονοσκόπησης χρησιμοποιούσαν παραδοσιακές μεθόδους επεξεργασίας εικόνας που βασίζονταν στην απλή όραση υπολογιστή, όπου προκειμένου να βρεθούν πιθανοί πολύποδες σε ενδοσκοπικές εικόνες ή καρέ βίντεο, επικεντρώθηκαν στην αφαίρεση οπτικών χαρακτηριστικών όπως το χρώμα, το σχήμα και η υφή. Εντοπίζοντας ξαφνικές μεταβολές στην ένταση, αλγόριθμοι ανίχνευσης ακμών (όπως τα φίλτρα Sobel, Canny και Laplacian) χρησιμοποιήθηκαν συχνά για να σχεδιάσουν όρια γύρω από τους πολύποδες, με τα κυκλικά ή ελλειπτικά περιγράμματα που παρατηρούνται συχνά σε ανυψωμένους πολύποδες να προορίζονταν να καταγραφούν με τεχνικές που βασίζονται σε ακμές, όντας όμως εξαιρετικά ευαίσθητες στις

αλλαγές του φωτισμού και στο θόρυβο της εικόνας, που είναι συνηθισμένες σε ενδοσκοπικές ρυθμίσεις.

Επιπλέον, χρησιμοποιήθηκαν μορφολογικές λειτουργίες όπως η διαστολή, η διάβρωση, το άνοιγμα και το κλείσιμο για την απομόνωση των περιοχών ενδιαφέροντος και τη βελτίωση της αντίθεσης της εικόνας, προσπαθώντας να διαχωρίσουν πιθανές περιοχές πολυπόδων για πρόσθετη εξέταση συνδυάζοντάς τες με στρατηγικές κατωφλίωσης ή αύξησης της περιοχής. Αργότερα, προστέθηκαν περιγραφείς υφής, όπως οι πίνακες συνύπαρξης σε επίπεδο γκρι (Gray-Level Co-occurrence Matrices - GLCM) και τα τοπικά δυαδικά μοτίβα (Local Binary Patterns - LBP), για να καταγραφούν οι επιφανειακές ανωμαλίες και οι μεταβολές της μικρο-υφής που θα μπορούσαν να βοηθήσουν στη διαφοροποίηση των πολυπόδων από τον περιβάλλοντα βλεννογόνο.

Παρόλο που οι πρώτες αυτές μέθοδοι δημιούργησαν τις βάσεις για την ανάπτυξη CAD, η απόδοσή τους περιοριζόταν από την εξάρτησή τους από τα χειροκίνητα χαρακτηριστικά, συχνά αντιμετωπίζοντας προβλήματα γενίκευσης σε διάφορα σύνολα δεδομένων ή ενδοσκοπικά συστήματα και δεν είχαν ανθεκτικότητα σε αλλαγές στο σχήμα του πολύποδα, το χρώμα και τις συνθήκες φωτισμού. Επιπλέον, η διαδικασία χειροκίνητου σχεδιασμού χαρακτηριστικών ήταν επίπονη και υποκειμενική, εμποδίζοντας την κλινική υιοθέτηση και την επεκτασιμότητα, παρακινώντας, παρακινώντας την αλλαγή προς προσεγγίσεις με βάση τα δεδομένα που μπορούν να μάθουν αυτόματα περίπλοκες και αναλλοίωτες αναπαραστάσεις χαρακτηριστικών.

3.3 Μετάβαση σε προσεγγίσεις μηχανικής μάθησης και βαθιάς μάθησης

Σημαντικό σημείο καμπής στην κολονοσκόπηση με τη βοήθεια υπολογιστή ήταν η μετάβαση από τη συμβατική επεξεργασία εικόνας στη μηχανική μάθηση (ML) και τη βαθιά μάθηση (DL), με την εκμάθηση χαρακτηριστικών να εισάγεται για πρώτη φορά από τις πρώτες τεχνικές μηχανικής μάθησης, στις οποίες τα χειροκίνητα εξαγόμενα χαρακτηριστικά χρησιμοποιήθηκαν για την εκπαίδευση αλγορίθμων όπως οι Μηχανές Διανυσμάτων Υποστήριξης (SVM), τα Τυχαία Δάση και οι k-κοντινότεροι γείτονες (k-NN) για την ταξινόμηση περιοχών ως πολυπόδων ή μη πολυπόδων, τα οποία περιορίζονταν από το διαμέτρημα και τη δυνατότητα εφαρμογής των χειροποίητων χαρακτηριστικών που χρησιμοποιούνταν ως είσοδος, αλλά αύξησαν την ακρίβεια ανίχνευσης μαθαίνοντας τα όρια απόφασης από τα δεδομένα. Η διαδικασία αυτή μετασχηματίστηκε πλήρως με την έλευση της βαθιάς μάθησης, ιδίως των συνελλικτικών νευρωνικών δικτύων (CNN), τα οποία αυτοματοποίησαν την εξαγωγή χαρακτηριστικών και τη μάθηση αναπαράστασης [14]. Με τη βοήθεια των CNNs, τα οποία μπορούν να μάθουν χωρικά χαρακτηριστικά ιεραρχικά από μη επεξεργασμένα δεδομένα εικόνας, σύνθετα μοτίβα υφής, χρώματος και δομής μπορούν να καταγραφούν αυτόματα από το μοντέλο, βελτιώνοντας σημαντικά τη γενίκευση του μοντέλου σε

μια ποικιλία συνόλων δεδομένων και αφαιρώντας την ανάγκη για ρητή μηχανική των χαρακτηριστικών.

Τα συστήματα CAD με βάση τη βαθιά μάθηση έχουν επιδείξει εντυπωσιακά αποτελέσματα στην κολονοσκόπηση τόσο για την κατάτμηση πολυπόδων (καθορισμός των ορίων ενός πολύποδα) όσο και για την ανίχνευση πολυπόδων (ανίχνευση της παρουσίας ενός πολύποδα). Μεγάλα σχολιασμένα σύνολα δεδομένων, μαζί με βελτιώσεις στην επεξεργαστική ισχύ και την επιτάχυνση GPU, κατέστησαν δυνατή την εκπαίδευση μοντέλων υψηλής χωρητικότητας όπως τα ResNet, VGG και U-Net, αποτελώντας πλέον τη βάση της πλειονότητας των συστημάτων ανίχνευσης αιχμής. Επιπλέον, οι εφαρμογές πραγματικού χρόνου έχουν καταστεί δυνατές χάρη στη βαθιά μάθηση, η οποία επιτρέπει στα συστήματα τεχνητής νοημοσύνης να υποστηρίζουν τους γιατρούς κατά τη διάρκεια ζωντανών διαδικασιών κολονοσκόπησης, μειώνοντας αποτελεσματικά το ποσοστό αστοχίας πολυπόδων, αναδεικνύοντας λεπτές ή κρυφές βλάβες που διαφορετικά θα μπορούσαν να περάσουν απαρατήρητες [15].

Έτσι, το CAD στην κολονοσκόπηση έχει επαναπροσδιοριστεί ως μετακίνηση από προσεγγίσεις που βασίζονται σε κανόνες σε προσεγγίσεις που βασίζονται σε δεδομένα, από απλά εργαλεία βελτίωσης της εικόνας σε εξελιγμένους διαγνωστικούς βοηθούς. Με γνώμονα τις συνεχείς εξελίξεις στις αρχιτεκτονικές των νευρωνικών δικτύων, την ποικιλομορφία των συνόλων δεδομένων και την ενσωμάτωση με τις κλινικές ροές εργασίας, η εξέλιξη αυτή εξακολουθεί να συμβαίνει γρήγορα, συμβάλλοντας στην επίτευξη του ευρύτερου στόχου της μείωσης της επίπτωσης του καρκίνου του παχέος εντέρου με τη χρήση ακριβέστερων και αποτελεσματικότερων τεχνικών ανίχνευσης.

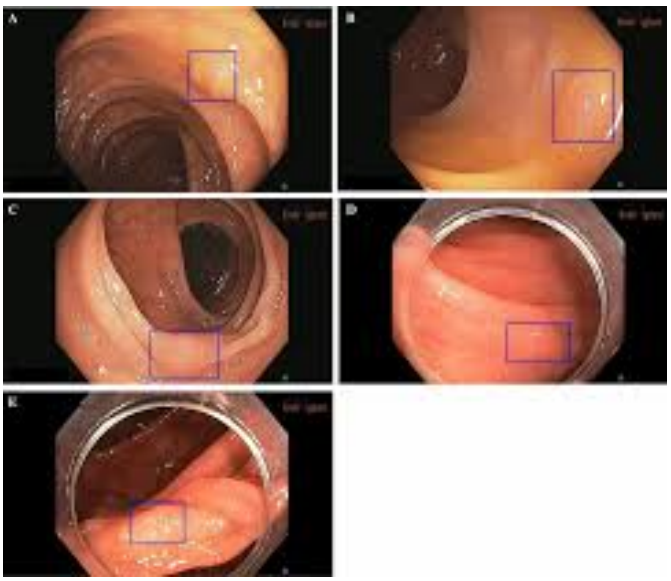
3.4 Η κατάσταση της τεχνολογίας στην ανίχνευση πολυπόδων

Αρκετά σύνολα δεδομένων που είναι προσβάσιμα από το κοινό έχουν καθιερωθεί ως βιομηχανικά πρότυπα για μελέτες τμηματοποίησης και ανίχνευσης πολυπόδων, με την Kvasir-SEG, μια συλλογή εικόνων κολονοσκόπησης ανοικτής πρόσβασης με μάσκες τμηματοποίησης ανά εικονοστοιχείο και παραγόμενα πλαίσια οριοθέτησης, να έχει χρησιμοποιηθεί εκτενώς τόσο για εργασίες τμηματοποίησης όσο και για εργασίες ανίχνευσης, περιλαμβάνοντας περίπου 1.000 σχολιασμένες εικόνες πολυπόδων που ελήφθησαν κατά τη διάρκεια πραγματικών διαδικασιών. Ένα άλλο ευρέως χρησιμοποιούμενο σύνολο δεδομένων είναι το CVC-ClinicDB, το οποίο αναφέρεται περιστασιακά στο πλαίσιο του MICCAI 2015 Endoscopic Vision Challenge, περιλαμβάνοντας αρκετές εκατοντάδες καρέ (συχνά αναφέρεται ως 612) που λαμβάνονται από βίντεο κολονοσκοπήσεων, καθένα από τα οποία διαθέτει bounding boxes ή ground-truth masks [16]. Ενώ το ETIS-Larib είναι ένα μικρότερο αλλά υψηλής ανάλυσης σύνολο δεδομένων (συχνά αναφέρεται ως ≈196 εικόνες) με χειροκίνητα σχολιασμένες μάσκες, το CVC-ColonDB προσφέρει περίπου 380 σχολιασμένες εικόνες

πολυπόδων από πολλαπλές ακολουθίες βίντεο, με το ETIS-Larib να χρησιμοποιείται συχνά ως ένα δύσκολο σύνολο δοκιμών λόγω της μεγάλης ανάλυσης εικόνων και των δύσκολων συνθηκών απεικόνισης, χρησιμεύοντας ως βάση για τη συγκριτική ανάλυση στην πλειονότητα των δημοσιευμένων μελετών ανίχνευσης πολυπόδων, μαζί με συλλογές προκλήσεων από το MICCAI και άλλες πρωτοβουλίες της κοινότητας [17].

Η ανάπτυξη και οι παρούσες δυνατότητες των μοντέλων βαθιάς μάθησης για την ανίχνευση πολυπόδων καταδεικνύονται από μια σειρά πρωτοποριακών και πιο πρόσφατων μελετών, με τις πρώτες σύγχρονες προσπάθειες βαθιάς μάθησης να χρησιμοποιούν δίκτυα προσανατολισμένα στην τμηματοποίηση (όπως παραλλαγές του U-Net) ή κωδικοποιητές-αποκωδικοποιητές Προκειμένου να ληφθούν ακριβείς μάσκες πολυπόδων. Από εκεί προέκυψαν πλαίσια οριοθέτησης για εργασίες ανίχνευσης, με τα μοντέλα της οικογένειας YOLO (παραλλαγές YOLOv4-v8) για ανίχνευση χαμηλής καθυστέρησης, σε πραγματικό χρόνο και τα μοντέλα της οικογένειας RetinaNet και Faster R-CNN για υψηλότερη ακρίβεια να αποτελούν παραδείγματα ανιχνευτών αντικειμένων ενός και δύο σταδίων που έχουν σχεδιαστεί για ταχύτητα και κλινική χρηστικότητα [18]. Καθώς οι αρχιτεκτονικές και οι μέθοδοι εκπαίδευσης εξελίσσονταν, οι συγκριτικές αναλύσεις, τα αποτελέσματα της πρόκλησης από το MICCAI 2015 και τα μεταγενέστερα έγγραφα συγκριτικής αξιολόγησης καταδεικνύουν σταθερή βελτίωση των επιδόσεων ανίχνευσης και τμηματοποίησης, με την πρόσφατη έρευνα συμπεριέλαβε επίσης δίκτυα βελτίωσης με καθοδηγούμενο πλαίσιο, μονάδες προσοχής και υβριδικές λύσεις (ραχοκοκαλιά τμηματοποίησης που τροφοδοτεί κεφαλές ανίχνευσης) προκειμένου να μειωθούν τα ψευδώς θετικά αποτελέσματα που προκαλούνται από πτυχώσεις, κατοπτρικές ανταύγειες και υγρά.

Παρόλο που οι αναφερόμενες επιδόσεις στη βιβλιογραφία ποικίλλουν ανάλογα με τη ρύθμιση αξιολόγησης, το πρωτόκολλο σχολιασμού και τη σύνθεση του συνόλου δεδομένων, ορισμένες μετρήσεις και εύρος τιμών είναι συνεπείς σε όλες τις μελέτες, με την ευαισθησία (ανάκληση), την ακρίβεια, το F1-score, την μέση ακρίβεια (mAP) για τα πλαίσια οριοθέτησης και τα κατώφλια IoU (intersection-over-union) για την ορθότητα του εντοπισμού να είναι παραδείγματα κοινών μετρικών ανίχνευσης και εντοπισμού [19]. Ο δείκτης Jaccard και ο συντελεστής κύβου αναφέρονται επίσης συχνά για αξιολογήσεις που βασίζονται στην τμηματοποίηση, με τις παραλλαγές YOLO και EfficientNet με υψηλές επιδόσεις να έχουν αναφέρει βαθμολογίες mAP και F1 σε χαμηλά επίπεδα της τάξης του 90% σε συνδυασμένα δημόσια σύνολα δεδομένων σε πρόσφατες μελέτες 2021-2024. Πολλοί σύγχρονοι ανιχνευτές που βασίζονται σε CNN αναφέρουν τιμές ευαισθησίας (ανάκληση) στην περιοχή 85-95% σε επιμελημένα σύνολα δοκιμών και τιμές ακρίβειας συχνά πάνω από 85% [20].



Εικόνα 7 – Ανίχνευση πολυπόδων με χρήση CAD [21]

Οι λεπτές μεθοδολογικές αποφάσεις, όπως (1) το αν η αξιολόγηση γίνεται με βάση την εικόνα ή το βίντεο, (2) ο τρόπος με τον οποίο τα πλαίσια οριοθέτησης προκύπτουν από τις μάσκες τμηματοποίησης (π.χ. στενά έναντι εκτεταμένων πλαισίων), (3) ο χειρισμός της ανισορροπίας τάξεων και οι στρατηγικές επαύξησης και (4) τα πρωτόκολλα επικύρωσης μεταξύ των συνόλων δεδομένων, είναι συχνά η αιτία των αναφερόμενων διαφορών επιδόσεων, σύμφωνα με εκτεταμένες μελέτες συγκριτικής αξιολόγησης και αναθεώρησης. Άλλες δείχνουν ότι όταν τα μοντέλα αξιολογούνται αυστηρά σε διασταυρούμενα σύνολα δεδομένων (εκπαίδευση σε ένα σύνολο δεδομένων και δοκιμή σε ένα άλλο), οι επιδόσεις μειώνονται σημαντικά, αναδεικνύοντας ένα κενό γενίκευσης που εξακολουθεί να ερευνάται, με ορισμένες εργασίες για παράδειγμα να επιτυγχάνουν αισθητά υψηλότερη ακρίβεια με την εκπαίδευση και τη δοκιμή σε σύνολα δεδομένων μικτής πηγής ή με τη χρήση επιθετικής επαύξησης και μεταφοράς μάθησης από μεγάλα μη ιατρικά σύνολα δεδομένων. Για την παροχή ακριβών εκτιμήσεων της κλινικής χρησιμότητας, πρόσφατες ανασκοπήσεις και πρωτοβουλίες συγκριτικής αξιολόγησης προτείνουν διασταυρωμένη επικύρωση πολλαπλών συνόλων δεδομένων και τυποποιημένες σωληνώσεις αξιολόγησης [22].

Δύο πρακτικά προβλήματα εξακολουθούν να περιορίζουν την κλινική μετάφραση παρά τις καλές επιδόσεις στο εργαστήριο: μεροληψία/περιορισμένη ποικιλομορφία συνόλων δεδομένων και ψευδώς θετικά αποτελέσματα (όπως πτυχώσεις του βλεννογόνου ή υπολείμματα κοπράνων που χαρακτηρίζονται ως πολύποδες). Με (i) τη χρήση ενοτήτων με επίγνωση του πλαισίου ή φίλτρων μετα-επεξεργασίας για τη μείωση των ψευδών ανιχνεύσεων, (ii) την προσθήκη ελέγχων χρονικής συνέπειας σε συστήματα που βασίζονται σε βίντεο και (iii) τη σύνταξη μεγαλύτερων, πιο ποικίλων συνόλων δεδομένων (ή τη χρήση τεχνικών προσαρμογής σε τομείς) για την αύξηση της

ανθεκτικότητας σε διάφορα συστήματα ενδοσκόπησης και πληθυσμούς ασθενών, πολλές πρόσφατες εργασίες αντιμετωπίζουν συγκεκριμένα αυτά τα ζητήματα, τα οποία και είναι τα ακριβή κενά που οδήγησαν στις αποφάσεις σχεδιασμού του δικτύου που ελήφθησαν στην παρούσα εργασία, περιλαμβάνοντας τη χρήση της επαύξησης και της προεπεξεργασίας με επίγνωση των αντικειμένων, την παροχή προτεραιότητας στην επικύρωση μεταξύ των συνόλων δεδομένων και την αξιολόγηση της ευρωστίας τόσο της ανίχνευσης όσο και του εντοπισμού τόσο σε δημόσια σημεία αναφοράς όσο και σε ένα ιδιωτικό κλινικό αρχείο.

Μια διεξοδική ανάλυση των πρόσφατων εξελίξεων στην εφαρμογή της μηχανικής μάθησης και της βαθιάς μάθησης για την ανίχνευση και ταξινόμηση πολυπόδων του παχέος εντέρου σε κολονοσκοπήσεις δίνεται από τους [23], με την εργασία να ξεκινά με την επισήμανση ότι το ανθρώπινο λάθος, που συχνά προκαλείται από την εξάντληση ή την έλλειψη εμπειρίας, μπορεί να οδηγήσει σε μη εντοπισμένες βλάβες που θα μπορούσαν να εξελιχθούν σε καρκίνο του παχέος εντέρου, μια από τις κύριες αιτίες θανάτων που σχετίζονται με τον καρκίνο παγκοσμίως, δίνοντας έμφαση στην κλινική σημασία της ακριβούς ανίχνευσης πολυπόδων. Οι συγγραφείς περιγράφουν τον τρόπο με τον οποίο τα νευρωνικά δίκτυα συνελικτικής μάθησης (CNN) και άλλες αρχιτεκτονικές βαθιάς μάθησης έχουν χρησιμοποιηθεί για την αύξηση της διαγνωστικής ακρίβειας και συνέπειας και εξετάζουν την επεκτεινόμενη χρήση της ΤΝ για την υποστήριξη των κλινικών ιατρών. Με θέματα που περιλαμβάνουν σύνολα δεδομένων αναφοράς, μετρικές αξιολόγησης και κοινά ζητήματα που βασίζονται στην εικόνα, όπως η θολούρα κίνησης, οι αντανακλάσεις λευκού φωτός και η μεταβλητότητα της εμφάνισης των πολυπόδων, η έρευνα παρέχει μια οργανωμένη ανάλυση του πεδίου, με τους συγγραφείς να εφιστούν επίσης την προσοχή στο επίμονο πρόβλημα της έλλειψης σχολιασμένων δεδομένων εκπαίδευσης, το οποίο περιορίζει τη γενίκευση των μοντέλων και την αξιοπιστία τους σε πραγματικές συνθήκες. Επισημαίνουν τα κενά στην αντιμετώπιση ζητημάτων ποιότητας εικόνας και ποικιλομορφίας συνόλων δεδομένων, ενώ εντοπίζουν σημαντικές ερευνητικές τάσεις, όπως η αυξανόμενη έμφαση στην ανίχνευση σε πραγματικό χρόνο και στα υβριδικά μοντέλα βαθιάς μάθησης, περιγράφοντας τις μελλοντικές ερευνητικές ευκαιρίες για την ενίσχυση της κλινικής ενσωμάτωσης και της ευρωστίας των συστημάτων κολονοσκόπησης που βασίζονται στην τεχνητή νοημοσύνη.

Οι Spadaccini κ.ά. [24] παρέχουν μια ανάλυση του τρόπου με τον οποίο τα συστήματα ανίχνευσης με τη βοήθεια υπολογιστή (CADe) που καθοδηγούνται από την τεχνητή νοημοσύνη μπορούν να βελτιώσουν το διαμέτρημα και την αξιοπιστία των διαδικασιών κολονοσκόπησης, με τους συγγραφείς να ξεκινούν περιγράφοντας τον επιπολασμό του καρκίνου του παχέος εντέρου (CRC) παγκοσμίως και τον ζωτικό ρόλο που διαδραματίζουν οι κολονοσκοπήσεις στην έγκαιρη ανίχνευση και τη μείωση της θνησιμότητας. Παρ' όλα αυτά, εφιστούν την προσοχή στο ζήτημα της διακύμανσης από γιατρό σε γιατρό στα ποσοστά ανίχνευσης αδενωμάτων (ADR), τα οποία μπορεί

να οδηγήσουν σε χαμένες βλάβες και μειωμένη προληπτική αποτελεσματικότητα, με τα= συστήματα CADe που τροφοδοτούνται από σύγχρονους αλγόριθμους τεχνητής νοημοσύνης να έχουν αυξήσει τα ποσοστά ανίχνευσης πολυπόδων, σύμφωνα με τη μελέτη, βοηθώντας τους ενδοσκόπους να εντοπίζουν μικρές ή ανεπαίσθητες αλλοιώσεις που διαφορετικά θα μπορούσαν να περάσουν απαρατήρητες λόγω ανθρώπινου λάθους ή κούρασης. Η ανασκόπηση συνοψίζει τα αποτελέσματα από μεγάλο αριθμό μελετών παρατήρησης και κλινικών δοκιμών, δείχνοντας ότι η ενσωμάτωση του CADe μειώνει σταθερά τα ποσοστά αστοχίας αδενώματος (AMR) και αυξάνει τα ποσοστά ADR, ιδίως για τους μικρότερους και λιγότερο αισθητούς πολύποδες. Ωστόσο, οι συγγραφείς επισημαίνουν ορισμένα σημαντικά μειονεκτήματα, όπως η ανεπαρκής ικανότητα της τεχνολογίας να εντοπίζει προχωρημένα αδενώματα ή άμισχες οδοντωτές αλλοιώσεις, τα ζητήματα γραφειοκρατίας του χειριστή και η υπερβολική εξάρτηση από το AI, υποστηρίζοντας ότι, ενώ το CADe αποτελεί ένα σημαντικό βήμα προς τα εμπρός στην τυποποίηση της ποιότητας ανίχνευσης, ο μακροπρόθεσμος αντίκτυπός του στην επίπτωση και τη θνησιμότητα του καρκίνου του παχέος εντέρου δεν έχει ακόμη τεκμηριωθεί οριστικά. Για την ορθή αξιολόγηση των κλινικών και προληπτικών δυνατοτήτων της τεχνολογίας, υποστηρίζουν εκτεταμένες διαχρονικές μελέτες, ενσωμάτωση με συμπληρωματικές απεικονιστικές μεθόδους και πρόσθετη ανάπτυξη αλγορίθμων AI.

Οι Danku κ.ά. [25] παρέχουν μια διεξοδική ανάλυση της δημιουργίας ενός διαγνωστικού εργαλείου (CAD) για την ανίχνευση του καρκίνου του παχέος εντέρου, το οποίο χρησιμοποιεί τεχνητή νοημοσύνη για να βοηθήσει τους γιατρούς να βελτιώσουν την έγκαιρη ανίχνευση και να μειώσουν τις λανθασμένες διαγνώσεις, επισημαίνοντας ότι ακόμη και με την υψηλή ακρίβεια των σύγχρονων απεικονιστικών μεθόδων, τα διαγνωστικά σφάλματα εξακολουθούν να αποτελούν σημαντικό πρόβλημα, επηρεάζοντας περίπου 800.000 Αμερικανούς κάθε χρόνο λόγω λανθασμένων διαγνώσεων ή καθυστερημένης ανίχνευσης της νόσου. Η εργασία τους αποσκοπεί στην ανάπτυξη ενός χρήσιμου και λογικά οικονομικού συστήματος υποστήριξης αποφάσεων με βάση την τεχνητή νοημοσύνη, το οποίο συνδυάζει πληροφορίες από τις συνήθεις κλινικές εξετάσεις, όπως αναλύσεις κοπράνων και αίματος, με απεικονίσεις κολονοσκόπησης, όταν απαιτείται, αποσκοπώντας στην παροχή μιας πιο ολιστικής αξιολόγησης, ενώ παράλληλα καθιστά τα εργαλεία ΤΝ προσιτά σε επαγγελματίες του ιατρικού κλάδου με περιορισμένη τεχνική εξειδίκευση. Το σύστημα του Danku [25] χρησιμοποιεί τόσο συνελικτικά όσο και τεχνητά νευρωνικά δίκτυα (ANN και CNN), τα οποία ενσωματώνονται σε μια διαισθητική γραφική διεπαφή χρήστη (GUI), ενώ σε σενάρια δυαδικής ταξινόμησης, πέτυχε επίπεδα ακρίβειας έως και 97,2% για εικόνες RGB και 96,7% για εικόνες κλίμακας του γκρι, ενώ το ANN σημείωσε καλές επιδόσεις με μέσο τετραγωνικό σφάλμα (MSE) 7,38. Ειδικότερα, η ακρίβεια βελτιώθηκε περαιτέρω με τη χρήση προεκπαιδευμένων μοντέλων CNN, τα οποία πέτυχαν 99,5% σε εργασίες ανίχνευσης δύο κατηγοριών, μπορώντας να αυξήσουν σημαντικά

τη διαγνωστική αξιοπιστία, μειώνοντας παράλληλα τον χρόνο που απαιτείται για τη λήψη κλινικών αποφάσεων και την πιθανότητα ανθρώπινου λάθους. Συνολικά, η εργασία αποτελεί μια ουσιαστική συμβολή προς την κατεύθυνση της ενσωμάτωσης διαγνωστικών εργαλείων με βάση την τεχνητή νοημοσύνη στις καθημερινές ιατρικές ροές εργασίας, δίνοντας έμφαση στην προσβασιμότητα, την ακρίβεια και τη σχέση κόστους-αποτελεσματικότητας στην ανίχνευση του καρκίνου του παχέος εντέρου.

Προκειμένου να αναπτύξουν και να επικυρώσουν ένα σύστημα τεχνητής νοημοσύνης για την αυτοματοποιημένη ανίχνευση και ταξινόμηση πολυπόδων σε εικόνες κολονοσκόπησης, ο Shen [26] διεξήγαγαν μια ολοκληρωμένη, πολυνοσοκομειακή μελέτη, δημιούργώντας μία από τις πιο ολοκληρωμένες συλλογές εικόνων που έχουν χρησιμοποιηθεί στον τομέα αυτό μέχρι σήμερα, συγκεντρώνοντας ένα μεγάλο σύνολο δεδομένων με περίπου 256.220 εικόνες από 5.000 ασθενείς με καρκίνο του παχέος εντέρου. Τα δεδομένα διαχωρίστηκαν σε υποσύνολα εκπαίδευσης (70%), επικύρωσης (15%) και δοκιμής (15%) και το σύστημα χρησιμοποίησε ένα νευρωνικό δίκτυο συνελίξεων (CNN) για την ανίχνευση πολυπόδων και ένα μοντέλο EfficientNet-b0 για την ταξινόμηση πολυπόδων. Προκειμένου να μειωθεί η υπερπροσαρμογή, αυτή η μεθοδολογική αυστηρότητα εξασφάλισε ότι τα μοντέλα εκπαιδεύτηκαν και αξιολογήθηκαν σε μια ποικιλία συνόλων εικόνων, με τα αποτελέσματα να δείχνουν κορυφαίες επιδόσεις, με το μοντέλο ταξινόμησης να επιτυγχάνει εξαιρετική περιοχή κάτω από την καμπύλη (AUC) 0,9989 στο εσωτερικό σύνολο δοκιμών και το μοντέλο ανίχνευσης να επιτυγχάνει ευαισθησία 0,9709 και ειδικότητα 0,9701. Χρησιμοποιώντας τόσο προοπτικά όσο και αναδρομικά σύνολα δεδομένων, οι ερευνητές επικύρωσαν περαιτέρω το μοντέλο σε τρία νοσοκομεία για να εγγυηθούν την αξιοπιστία στον πραγματικό κόσμο, με τα αποτελέσματα να δείχνουν ειδικότητα βάσει πλαισίου 0,9720 και ευαισθησία βάσει βλάβης 0,9516. Η ισχυρή ικανότητα γενίκευσης του μοντέλου επιβεβαιώθηκε από την εξωτερική επικύρωση AUC για την ταξινόμηση, η οποία ανήλθε σε 0,9521, καταλήγοντας στο συμπέρασμα ότι το εξαιρετικά αποτελεσματικό σύστημα βαθιάς μάθησης μπορεί να βελτιώσει σημαντικά τη διαγνωστική αποτελεσματικότητα και ακρίβεια στην πρακτική της κλινικής κολονοσκόπησης, παρέχοντας στους ενδοσκόπους γρήγορη και αξιόπιστη υποστήριξη αποφάσεων για τον εντοπισμό και την κατηγοριοποίηση των πολυπόδων του παχέος εντέρου.

Προκειμένου να ξεπεραστούν οι επαναλαμβανόμενες δυσκολίες στην κατάτμηση των πολυπόδων του παχέος εντέρου, οι Xie και Shen [27] προτείνουν μια νέα αρχιτεκτονική βαθιάς μάθησης που ονομάζεται Δίκτυο διπλής προσοχής και μηχανισμού σύντηξης (Dual Attention and Fusion Mechanism Network - DAFM-Net), επιτρέποντας τον ακριβή εντοπισμό και την οριοθέτηση των πολυπόδων κατά τη διάρκεια της κολονοσκόπησης, οι συγγραφείς τονίζουν ότι η ακριβής τμηματοποίηση είναι ζωτικής σημασίας για τη βελτίωση των αποτελεσμάτων του ελέγχου για τον καρκίνο του παχέος εντέρου. Ωστόσο, η μορφολογική ποικιλομορφία, οι περίπλοκες προσκολλήσεις

του βλεννογόνου και τα διφορούμενα όρια των πολυπόδων αποτελούν συχνά προκλήσεις για τους σημερινούς αλγορίθμους τμηματοποίησης, με το DAFM-Net να παρουσιάζει ένα πλαίσιο πολλαπλών μονάδων που έχει σχεδιαστεί για να επιτυγχάνει ισορροπημένη ενσωμάτωση τοπικών και παγκόσμιων πληροφοριών πλαισίου, καθώς και συμπληρωματική συγχώνευση χαρακτηριστικών προκειμένου να παρακαμφθούν αυτοί οι περιορισμοί. Στο δίκτυο ενσωματώνονται τρεις σημαντικές καινοτομίες: η ενότητα Dual Context Enhanced Attention (DCEA), η οποία βελτιστοποιεί ταυτόχρονα την παγκόσμια αντίληψη και τη βελτίωση των τοπικών λεπτομερειών, η ενότητα Cross-layer Aware Selective Fusion (CASF), η οποία συγχωνεύει προσαρμοστικά βαθιά και ρηχά χαρακτηριστικά για την ελαχιστοποίηση του πλεονασμού και την ενίσχυση των σημασιολογικών σχέσεων, και η ενότητα Multi-scale Convolutional Patch Aware (MCPA), η οποία χρησιμοποιεί συνελίξεις πολλαπλών κλάδων και τοπική προσοχή για την εξαγωγή λεπτόκοκκων χαρακτηριστικών των ακανόνιστου σχήματος πολύποδων. Πειράματα σε πέντε δημόσια διαθέσιμα σύνολα δεδομένων πολυπόδων δείχνουν ότι το DAFM-Net ξεπερνάει συστηματικά τα πιο προηγμένα μοντέλα όσον αφορά την ακρίβεια κατάτμησης και την ευρωστία, ιδίως σε δύσκολες καταστάσεις που περιλαμβάνουν εμφανίσεις πολυπόδων πολλαπλής κλίμακας και πολύπλοκες υφές του βλεννογόνου, δείχνοντας ότι οι νέοι μηχανισμοί προσοχής και συγχώνευσης του DAFM-Net προσφέρουν μια ισχυρή βάση για συστήματα τμηματοποίησης πολυπόδων παχέος εντέρου βασισμένα σε τεχνητή νοημοσύνη, τα οποία είναι πιο αξιόπιστα και κλινικά χρήσιμα.

Οι Santone, Cesarelli και Mercaldo [28] παρουσιάζουν ένα πλαίσιο βαθιάς μάθησης βασισμένο σε U-Net που δημιουργήθηκε ειδικά για την αυτοματοποιημένη κατάτμηση των πολυπόδων του παχέος εντέρου σε εικόνες ενδοσκοπικής κολονοσκόπησης με την μελέτη να ασχολείται με την έγκαιρη και ακριβή ανίχνευση προκαρκινικών πολυπόδων, ένα από τα σημαντικότερα ζητήματα για την πρόληψη του καρκίνου του παχέος εντέρου. Η ακριβής οριοθέτηση αποτελεί πρόκληση για τις παραδοσιακές αυτοματοποιημένες μεθόδους, ιδίως για μικρούς ή επίπεδους πολύποδες που είναι δύσκολο να διαφοροποιηθούν από τον περιβάλλοντα βλεννογόνο, με τους συγγραφείς να τροποποιούν την αρχιτεκτονική του U-Net για την αντιμετώπιση αυτών των ζητημάτων, κάνοντας χρήση της δομής κωδικοποιητή-αποκωδικοποιητή και των συνδέσεων παράλειψης για τη βελτίωση της διατήρησης λεπτόκοκκων χωρικών λεπτομερειών καθώς και της κατανόησης του πλαισίου. Σε αντίθεση με τις παραδοσιακές τεχνικές, οι συγγραφείς επιτυγχάνουν σημαντικά υψηλότερη ακρίβεια κατάτμησης εκπαιδεύοντας το μοντέλο σε υψηλής ανάλυσης, σχολιασμένα καρέ σε επίπεδο εικονοστοιχείου, μμε την προτεινόμενη παραλλαγή U-Net να επιτυγχάνει κορυφαία αποτελέσματα στην κατάτμηση ορίων πολυπόδων, διατηρώντας ισχυρή απόδοση ακόμη και σε πολύπλοκες και χαμηλής αντίθεσης ενδοσκοπικές συνθήκες, σύμφωνα με εκτεταμένη αξιολόγηση σε ένα σύνολο πραγματικών κλινικών δεδομένων. Μειώνοντας τη μεταβλητότητα των παρατηρητών και ενισχύοντας τη διαγνωστική συνέπεια, η ακρίβεια και η

ευρωστία του μοντέλου παρέχουν στους γαστρεντερολόγους πολύτιμη βοήθεια κατά τη διάρκεια των ζωντανών διαδικασιών, καταλήγοντας στο συμπέρασμα ότι το εκλεπτυσμένο πλαίσιο U-Net τους προσφέρει μια επεκτάσιμη βάση για κλινική ανάπτυξη σε πραγματικό χρόνο, εκτός από την ενίσχυση της αξιοπιστίας της κολονοσκόπησης με τεχνητή νοημοσύνη, προωθώντας τον ευρύτερο στόχο της ενσωμάτωσης της τεχνητής νοημοσύνης στις τυπικές ροές εργασίας για τον έλεγχο του καρκίνου του παχέος εντέρου και την έγκαιρη παρέμβαση, τυποποιώντας την ποιότητα της τμηματοποίησης.

Οι Fan και Jiang [29] παρουσιάζουν μια προηγμένη αρχιτεκτονική βασισμένη σε μετασχηματιστές, που ονομάζεται PolypFormer, ως λύση στα τρέχοντα προβλήματα με την αυτόματη κατάτμηση πολυπόδων του παχέος εντέρου, περιλαμβάνοντας ακανόνιστα σχήματα, θολά όρια και ποικιλία μεγεθών πολυπόδων. Οι συγγραφείς επισημαίνουν ότι, ενώ η ακρίβεια κατάτμησης είναι απαραίτητη για την έγκαιρη πρόληψη του καρκίνου του παχέος εντέρου, τα τρέχοντα μοντέλα συχνά υπολείπονται στην επίτευξη ισορροπίας μεταξύ της σύλληψης εξαρτήσεων μεγάλης εμβέλειας και της αποτελεσματικής εξαγωγής τοπικών χαρακτηριστικών, ε το PolypFormer να ενσωματώνει μια μονάδα βελτίωσης τοπικών πληροφοριών (LIEM), η οποία χρησιμοποιεί πυκνή προσοχή καναλιών για να ενισχύσει τη συγχώνευση χαρακτηριστικών και τη διάκριση μεταξύ καναλιών και αυτοεπιλεκτική προσοχή πολλαπλών πυρήνων για να ενισχύσει την αναπαράσταση χαρακτηριστικών υφής προκειμένου να ξεπεραστούν αυτοί οι περιορισμοί. Για τη βελτίωση της σφαιρικής σημασιολογικής κατανόησης και της χωρικής συνοχής στις περιοχές των πολυπόδων, οι Fan και Jiang παρουσιάζουν έναν μηχανισμό αυτοπροσοχής με σταυρωτό παράθυρο, ο οποίος ενσωματώνεται απρόσκοπτα στην αρχιτεκτονική του Transformer εκτός από την τοπική ενίσχυση. Με μέσο συντελεστή Dice (mDice) 0,920 και μέσο συντελεστή διασταύρωσης-υπερ-ένωσης (mIoU) 0,886, το μοντέλο επιδεικνύει κορυφαίες επιδόσεις όταν δοκιμάζεται σε πέντε σύνολα δεδομένων που είναι προσβάσιμα από το κοινό, συμπεριλαμβανομένου του Kvasir-SEG, καταδεικνύοντας την υπεροχή του PolypFormer έναντι των παραδοσιακών μεθόδων που βασίζονται σε CNN στην κατάτμηση πολύπλοκων και ποικίλων πολυπόδων. Οι συγγραφείς καταλήγουν στο συμπέρασμα ότι ο υβριδικός σχεδιασμός τους με γνώμονα την προσοχή αντιπροσωπεύει μια σημαντική πρόοδο στον αυτοματοποιημένο έλεγχο του καρκίνου του παχέος εντέρου και στην ενδοσκοπική ανάλυση με τη βοήθεια ΤΝ, καθώς παρέχει τόσο ισχυρή γενίκευση όσο και υψηλή ακρίβεια τμηματοποίησης.

Μια ενδελεχής συστηματική ανασκόπηση των τεχνικών που βασίζονται στη βαθιά μάθηση για την κατάτμηση πολυπόδων του παχέος εντέρου, ένα έργο για την έγκαιρη ανίχνευση και πρόληψη του καρκίνου του παχέος εντέρου, πραγματοποιήθηκε από τον Liu [30], με τους συγγραφείς να οργανώνουν την ανάλυσή τους σε 146 εργασίες από το 2018 έως το 2024 για να δείξουν πώς οι τεχνικές τμηματοποίησης έχουν αλλάξει με την πάροδο του χρόνου. Η ανασκόπηση προσφέρει μια ενδελεχή εξέταση διαφόρων αρχιτεκτονικών μοντέλων, συμπεριλαμβανομένων

τεχνικών τμηματοποίησης που βασίζονται σε εικόνες και βίντεο, καθώς και νεότερων πλαισίων όπως η μέθοδος Mamba, συνοψίζοντας τα πιο δημοφιλή σύνολα δεδομένων τόσο για στατική όσο και για δυναμική τμηματοποίηση πολυπόδων (π.χ. Kvasir, CVC-ClinicDB και άλλα) και αξιολόγησε 44 μοντέλα τμηματοποίησης με βάση μετρήσεις επιδόσεων, συμπεριλαμβανομένης της ακρίβειας και της ικανότητας εξαγωγής συμπερασμάτων σε πραγματικό χρόνο. Οι συγγραφείς σημείωσαν διάφορα σημαντικά ζητήματα που εξακολουθούν να αντιμετωπίζει ο τομέας, όπως η διαχείριση των διαφορών στη μορφολογία των πολυπόδων, η γενίκευση σε διάφορα σύνολα δεδομένων και η επίτευξη κλινικής ανάπτυξης σε πραγματικό χρόνο. Για την περαιτέρω βελτίωση της ακρίβειας και της ευρωστίας σε κλινικές ρυθμίσεις, τόνισαν την ενσωμάτωση εξελιγμένων μηχανισμών προσοχής, πολυτροπικής μάθησης και αποτελεσματικών αρχιτεκτονικών στη συζήτησή τους για τις πολλά υποσχόμενες ερευνητικές κατευθύνσεις.

Παρουσιάζοντας το Edge-Enhanced Network (EENet), το οποίο δημιουργήθηκε ειδικά για την αύξηση της ακρίβειας των ορίων και της αναπαράστασης των χαρακτηριστικών σε εικόνες κολονοσκόπησης, ο Tong [31] αντιμετώπισε τη συνεχιζόμενη δυσκολία της αυτοματοποιημένης κατάτμησης των πολυπόδων του παχέος εντέρου, με το προτεινόμενο μοντέλο να ενσωματώνει δύο νέες ενότητες: τους μηχανισμούς Covariance Edge-Enhanced Attention (CEEA) και Cross-Scale Edge Enhancement (CSEE), υπό το πρίσμα της μεταβλητότητας των σχημάτων των πολυπόδων, των υφών και της συχνά χαμηλής αντίθεσης μεταξύ των πολυπόδων και του περιβάλλοντος βλεννογόνου. Ενώ η μονάδα CSEE συνδυάζει χαρακτηριστικά πολλαπλών κλιμάκων για να διατηρήσει λεπτομερείς πληροφορίες ακμών σε όλες τις αναλύσεις, η μονάδα CEEA χρησιμοποιεί προσοχή με βάση τη συνδιακύμανση για να αυξήσει την ευαισθησία του μοντέλου στις πληροφορίες ορίων, με τους συγγραφείς να προτείνουν μια υβριδική συνάρτηση απώλειας που συνδυάζει μια συνιστώσα απώλειας με επίγνωση των ακμών με τη συμβατική απώλεια διασταυρούμενης εντροπίας για τη βελτίωση της ακρίβειας τμηματοποίησης με τη βελτίωση της ευθυγράμμισης μεταξύ των προβλεπόμενων και των πραγματικών ορίων των πολυπόδων. Στο σύνολο δεδομένων Kvasir-SEG, η πειραματική αξιολόγηση έδειξε ότι το EENet ξεπέρασε κορυφαία μοντέλα όπως το Polyp-PVT και το PraNet, επιτυγχάνοντας βαθμολογίες Dice και IoU 0,9208 και 0,8664, αντίστοιχα. Ομοίως, το EENet πέτυχε βαθμολογίες Dice και IoU 0,9316 και 0,8817 στο σύνολο δεδομένων CVC-ClinicDB, αποδεικνύοντας την ισχυρή ικανότητα γενίκευσης, με τις ανεξάρτητες συνεισφορές των ενοτήτων CEEA και CSEE να επικυρώνονται περαιτέρω με μελέτες αφαίρεσης, αναδεικνύοντας την κλινική εφαρμοσιμότητα και την ευρωστία του μοντέλου για την ανάλυση εικόνων κολονοσκόπησης στον πραγματικό κόσμο.

Οι Guo, Bernal και Matuszewski [32] παρουσίασαν μια εκτεταμένη αξιολόγηση μιας μεθόδου που βασίζεται σε ένα πλήρως συνελικτικό νευρωνικό δίκτυο (FCN) για την αυτοματοποιημένη κατάτμηση πολυπόδων σε εικόνες κολονοσκόπησης, με την προσέγγισή τους να

δημιουργήθηκε αρχικά για τις προκλήσεις κατάτμησης πολυπόδων Endoscopic Vision GIANA το 2017 και το 2018 με στόχο την αύξηση της ακρίβειας ανίχνευσης και της υπολογιστικής απόδοσης της ανάλυσης κλινικών βίντεο κολονοσκόπησης, υπογραμμίζοντας τον τρόπο με τον οποίο η αυτοματοποιημένη τμηματοποίηση ιστών βελτιώνει τη συνέπεια και την αντικειμενικότητα των διαγνωστικών αποτελεσμάτων παρέχοντας μια βάση για σημαντικές κλινικές εφαρμογές όπως η ανίχνευση και η ταξινόμηση βλαβών. Χρησιμοποιώντας διάφορες μετρικές αξιολόγησης, η προτεινόμενη αρχιτεκτονική συγκρίθηκε με γνωστά μοντέλα τμηματοποίησης και επαληθεύτηκε διεξοδικά με διασταυρούμενη επικύρωση στο σύνολο δεδομένων εκπαίδευσης GIANA, με τους συγγραφείς να δείχνουν ότι η βέλτιστη επιλογή παραμέτρων και οι μέθοδοι επαύξησης έχουν σημαντικό αντίκτυπο στην απόδοση του μοντέλου μέσω μιας ενδελεχούς εξέτασης των διαμορφώσεων του δικτύου, του συντονισμού των παραμέτρων σχεδιασμού, των τακτικών επαύξησης δεδομένων και των χαρακτηριστικών των πολυπόδων. Τα πειραματικά ευρήματα κατέδειξαν την καταλληλότητα της μεθόδου για εφαρμογή στον πραγματικό κόσμο σε κλινικά περιβάλλοντα, επιβεβαιώνοντας ότι επιτυγχάνεται ακρίβεια τμηματοποίησης τελευταίας τεχνολογίας με απόδοση σχεδόν σε πραγματικό χρόνο, ενώ παράλληλα αποδείχθηκε η ανταγωνιστικότητα και η ανθεκτικότητα της μεθόδου μεταξύ των σύγχρονων τεχνικών που βασίζονται στη βαθιά μάθηση όταν κέρδισε την πρώτη θέση στην υπο-πρόκληση τμηματοποίησης πολυπόδων GIANA 2017 και τη δεύτερη θέση στην εργασία τμηματοποίησης τυπικής ανάλυσης εικόνας 2018.

Το UPolySeg, ένα μοντέλο βαθιάς μάθησης βασισμένο στο U-Net που δημιουργήθηκε για την κατάτμηση πολυπόδων του παχέος εντέρου σε εικόνες κολονοσκόπησης, παρουσιάστηκε από τον Mohapatra [33], προσπαθώντας να αντιμετωπίσει το επίμονα υψηλό ποσοστό αστοχίας στην ανίχνευση πολυπόδων εξαιτίας ζητημάτων όπως ο θόρυβος της εικόνας, η διακύμανση του φωτισμού και η χαμηλή αντίθεση μεταξύ των πολυπόδων και του βλεννογόνου, αναγνωρίζοντας ότι η κολονοσκόπηση εξακολουθεί να αποτελεί το χρυσό πρότυπο για τη διάγνωση ανωμαλιών του κατώτερου γαστρεντερικού συστήματος. Η μελέτη χρησιμοποίησε μια σειρά μεθόδων προεπεξεργασίας εικόνας, όπως η Contrast Limited Adaptive Histogram Equalization (CLAHE) για την καλύτερη ομαλοποίηση της αντίθεσης και η μεταφορά συνοχής για την ενίσχυση της υφής, για τη βελτίωση της απόδοσης τμηματοποίησης, αποσκοπώντας στη μείωση των οπτικών τεχνουργημάτων και στη βελτίωση της ορατότητας των ορίων των πολυπόδων πριν από την τμηματοποίηση. Προσθέτοντας συνένωση χαρακτηριστικών εντός των αντίστοιχων στρωμάτων και διασταλμένες συνελίξεις για τη διεύρυνση του δεκτικού πεδίου, το πλαίσιο UPolySeg βελτιώνει την κλασική αρχιτεκτονική U-Net, διατηρώντας παράλληλα μια δομή κωδικοποιητή-αποκωδικοποιητή με συνδέσεις παράλειψης, με το σύνολο δεδομένων Kvasir-SEG να χρησιμοποιείται για πειραματική επικύρωση και το μοντέλο παρήγαγε εντυπωσιακά αποτελέσματα, ξεπερνώντας τις προηγούμενες

προσεγγίσεις κατά 1,93% με συνολική ακρίβεια 96,77%, συντελεστή Dice 96,86% και IoU 87,91%. Τα αποτελέσματα δείχνουν ότι ο συνδυασμός της αρχιτεκτονικής βελτίωσης και της βελτιστοποίησης της προεπεξεργασίας στο UPolySeg αυξάνει σημαντικά την ακρίβεια τμηματοποίησης και την ευρωστία υπό απαιτητικές συνθήκες ενδοσκοπικής απεικόνισης, επιβεβαιώνοντας έτσι τις δυνατότητες των προσαρμοσμένων παραγώγων U-Net στην ανάλυση ιατρικών εικόνων.

Το Synergistic Multi-Granularity Rough Attention U-Net (S-MGRAUNet), μια εξελιγμένη αρχιτεκτονική βαθιάς μάθησης, προτάθηκε από τους Wang και Lim [34] για την αυτόματη κατάτμηση πολυπόδων σε εικόνες κολονοσκόπησης, με τις συνεχιζόμενες δυσκολίες με τα πολύπλοκα ενδοσκοπικά υπόβαθρα, την υψηλή μορφολογική μεταβλητότητα και τα ασαφή όρια των πολυπόδων, όλα αυτά που συχνά υποβαθμίζουν την ακρίβεια τμηματοποίησης, να αποτελούν την κινητήρια δύναμη πίσω από τη μελέτη. Οι συγγραφείς δημιούργησαν ένα πλαίσιο εκμάθησης πολλαπλών κόκκων που διατηρεί την υπολογιστική αποδοτικότητα, ενώ παράλληλα καταγράφει και ενσωματώνει οπτικές πληροφορίες σε διάφορες κλίμακες πλαισίου, προκειμένου να παρακάμψουν αυτές τις προκλήσεις, με την ενότητα Multi-Granularity Hybrid Filtering (MGHF), η οποία εξάγει πλούσια χαρακτηριστικά πολλαπλών κλιμάκων πλαισίου, την ενότητα Dynamic Granularity Partition Synergy (DGPS), η οποία χρησιμοποιεί προσαρμοστική αλληλεπίδραση χαρακτηριστικών για τη βελτίωση της διαφοροποίησης μεταξύ των πολυπόδων και του περιβάλλοντος ιστού, και τον μηχανισμό Multi-Granularity Rough Attention (MGRA), ο οποίος ενισχύει την οριοθέτηση και την αναγνώριση λεπτομερειών, να είναι τα τρία κύρια συστατικά του προτεινόμενου δικτύου. Το S-MGRAUNet υπερέβαινε συνεχώς τα τρέχοντα μοντέλα τμηματοποίησης όσον αφορά την ακρίβεια και την ευρωστία, διατηρώντας παράλληλα χαμηλή υπολογιστική πολυπλοκότητα, σύμφωνα με την πειραματική αξιολόγηση σε διάφορα σύνολα δεδομένων αναφοράς, συμπεριλαμβανομένων των ColonDB, CVC-300, Kvasir-SEG και ClinicDB. Προκειμένου να ελεγχθεί η μεταβολή της υφής και η μορφολογική ποικιλομορφία στις κολονοσκοπικές εικόνες, η ισχυρή ικανότητα γενίκευσης του δικτύου υπογραμμίζει τη σημασία του συνδυασμού εξαγωγής χαρακτηριστικών πολλαπλών κόκκων με εξελιγμένους μηχανισμούς προσοχής, όπου εκτός από την παροχή ενός πλαισίου τμηματοποίησης υψηλής απόδοσης, η εργασία προάγει την εννοιολογική κατανόηση της εφαρμογής θεωριών πολλαπλής κοκκομετρίας στην ανάλυση ιατρικών εικόνων, ανοίγοντας το δρόμο για πιο αποτελεσματική και κλινικά ορθή διάγνωση με τη βοήθεια της τεχνητής νοημοσύνης.

Προκειμένου να βελτιωθεί η διαγνωστική ακρίβεια της πρόληψης του καρκίνου με βάση την κολονοσκόπηση, οι Safarov και Whangbo [35] παρουσίασαν το A-DenseUNet, ένα πλήρως αυτοματοποιημένο μοντέλο κατάτμησης πολυπόδων του παχέος εντέρου με βάση τα εικονοστοιχεία, προτείνοντας μια προσαρμοστική αρχιτεκτονική βαθιάς μάθησης που συνδυάζει αποτελεσματικά την αναπαράσταση χαρακτηριστικών πολλαπλών κλιμάκων με την ενισχυμένη από

την προσοχή κατανόηση του πλαισίου σε αναγνώριση των προκλήσεων που παρουσιάζονται από το θόρυβο, τις μεταβαλλόμενες συνθήκες απεικόνισης και την περίπλοκη μορφολογία των πολυπόδων του παχέος εντέρου. Με τη χρήση πυκνά συνδεδεμένων στρωμάτων που μοιράζονται πληροφορίες κωδικοποίησης πολλαπλών κλιμάκων σε όλα τα επίπεδα αποκωδικοποίησης, το μοντέλο προσαρμόζεται σε σύνολα δεδομένων με διαφορετικά χαρακτηριστικά και βελτιώνει τη διάδοση χαρακτηριστικών. Πρόκειται για μια τροποποίηση του παραδοσιακού U-Net, με το δίκτυο να ενσωματώνει αρκετές διασταλμένες (ατροποιημένες) συνελίξεις με ποικίλους ρυθμούς διαστολής για περαιτέρω αύξηση της ακρίβειας τμηματοποίησης χωρίς να αυξάνεται ο υπολογιστικός φόρτος, επιτρέποντας ένα ευρύτερο δεκτικό πεδίο και διατηρεί τις χωρικές πληροφορίες που είναι απαραίτητες για την οριοθέτηση των ορίων. Για να βελτιωθεί η εξαγωγή των μοτίβων που είναι διαγνωστικά σημαντικά, περιλαμβάνεται επίσης ένας μηχανισμός προσοχής για την απομάκρυνση των ξένων χαρακτηριστικών, ξεπερνώντας αρκετές σύγχρονες αρχιτεκτονικές τμηματοποίησης, όπως το UNet++, το ResUNet, το U-Net, το PraNet και το ResUNet++, με συντελεστές Dice 90% στο Kvasir-SEG και 91% στο CVC-612 μέσω πειραματικής επικύρωσης σε σύνολα δεδομένων αναφοράς. Αυτά τα αποτελέσματα καταδεικνύουν την ικανότητα του μοντέλου να τμηματοποιεί δεδομένα με ακρίβεια και στιβαρότητα έναντι θορύβου, καθώς και τη δυνητική χρήση του σε κλινικές ροές εργασίας σε πραγματικό χρόνο, αποτελώντας ένα σημαντικό βήμα προς την κατεύθυνση της αξιόπιστης και ευρέως εφαρμόσιμης κατάτμησης πολυπόδων για την αυτοματοποιημένη διαγνωστική εξέταση του καρκίνου του παχέος εντέρου χάρη στο συνδυασμό προσαρμοστικής πυκνής συνδεσιμότητας, ατρωτής συνέλιξης και βελτίωσης με γνώμονα την προσοχή.

Ο Nanni [36] διεξήγε μια ενδελεχή διερεύνηση των τεχνικών βαθιάς μάθησης συνόλου για την κατάτμηση πολυπόδων με στόχο τη βελτίωση της ακρίβειας και της γενικευσιμότητας των σημερινών μοντέλων που χρησιμοποιούνται στην ανάλυση ιατρικών εικόνων, με το πρώτο μέρος της εργασίας να τονίζει πόσο σημαντική είναι η σημασιολογική τμηματοποίηση στην ιατρική διάγνωση, ιδίως για τον έγκαιρο εντοπισμό παθήσεων του παχέος εντέρου, όπου η ακριβής ταξινόμηση σε επίπεδο εικονοστοιχείου έχει άμεσο αντίκτυπο στα κλινικά αποτελέσματα. Οι συγγραφείς πρότειναν τον συνδυασμό διαφόρων συμπληρωματικών μοντέλων σε ένα πλαίσιο συνόλου, επειδή οι μεμονωμένες αρχιτεκτονικές βαθιάς μάθησης δεν είναι σε θέση να χειριστούν την ποικιλία των σχημάτων πολυπόδων, των υφών και των συνθηκών απεικόνισης, συγκεκριμένα, εξασφαλίζοντας την ποικιλομορφία των μοντέλων, ένα κρίσιμο συστατικό της αποτελεσματικότητας του συνόλου, συνδυάζοντας τις αρχιτεκτονικές HarDNet-MSEG, Polyp-PVT και HSNet, καθεμία από τις οποίες εκπαιδεύτηκε με διαφορετική αύξηση δεδομένων, τεχνικές βελτιστοποίησης και ρυθμούς μάθησης. Με τον μέσο όρο των ενδιάμεσων μασκών που ακολουθούν το σιγμοειδές επίπεδο ενεργοποίησης, η μελέτη παρουσίασε μια νέα τεχνική για την παραγωγή τελικών μασκών τμηματοποίησης που βελτιώνει την ακρίβεια και την ομαλότητα των προβλεπόμενων ορίων, με τα

προτεινόμενα σύνολα να υπερτερούν συνεχώς έναντι των τρεχουσών μεθόδων, επιτυγχάνοντας κορυφαία αποτελέσματα σε δύο σύνολα δεδομένων χωρίς εκπαίδευση ειδικά για το σύνολο δεδομένων, σύμφωνα με την πειραματική αξιολόγηση σε πέντε σημαντικά σύνολα δεδομένων κατάτμησης πολυπόδων. Τα αποτελέσματα επιβεβαιώνουν την αποτελεσματικότητα των προσεγγίσεων συνόλων στην ιατρική απεικόνιση, όπου επιτυγχάνονται καλύτερα αποτελέσματα τμηματοποίησης με τη συνένωση των δυνάμεων εξαγωγής χαρακτηριστικών των CNN με τη συνολική μοντελοποίηση πλαισίου των Transformers, προσφέροντας μια πολλά υποσχόμενη οδό για την ανάπτυξη συστημάτων ανίχνευσης πολυπόδων που είναι πιο ανθεκτικά, ευρέως εφαρμόσιμα και κλινικά αξιόπιστα.

Ο Eu και οι συνεργάτες του [37] αντιμετώπισαν τα μειονεκτήματα του χειροκίνητου ελέγχου, ο οποίος είναι συχνά επίπονος και εξαρτάται από την ικανότητα του χειριστή, προτείνοντας ένα σύστημα διάγνωσης με τη βοήθεια υπολογιστή (CAD) για την αυτόματη κατάτμηση των πολυπόδων του παχέος εντέρου σε εικόνες κολονοσκόπησης, δημιούργώντας μια τροποποιημένη αρχιτεκτονική SegNet που βασίζεται σε νευρωνικά δίκτυα συνέλιξης Visual Geometry Group-19 (VGG-19), επειδή κατανόησαν πόσο σημαντικός είναι ο έγκαιρος και ακριβής εντοπισμός πολυπόδων για την αναχαίτιση της εξάπλωσης του καρκίνου του παχέος εντέρου. Το μοντέλο ενσωματώνει μια σειρά από αρχιτεκτονικές βελτιώσεις για την ενίσχυση της κατανόησης του πλαισίου και της ακρίβειας τμηματοποίησης: παράλληλες διασταλμένες συνελίξεις για τη διατήρηση της χωρικής ανάλυσης ενώ εξάγονται χαρακτηριστικά πολλαπλών κλιμάκων, φίλτρα συνελίξεων 5×5 για την καταγραφή ευρύτερων δεκτικών πεδίων και συνδέσεις παράλειψης για βελτιωμένη διάδοση χαρακτηριστικών, ενώ για ενδελεχή επικύρωση σε ένα εύρος συνθηκών απεικόνισης, εκπαιδεύτηκε και αξιολογήθηκε χρησιμοποιώντας τρία γνωστά σύνολα δεδομένων: CVC-ClinicDB, CVC-ColonDB και ETIS-LaribPolypDB. Με ακρίβεια 96,06%, ευαισθησία 94,55%, ειδικότητα 97,56% και συντελεστή Dice 95,99%, τα πειραματικά αποτελέσματα έδειξαν υψηλή ακρίβεια τμηματοποίησης, ξεπερνώντας ή αντιστοιχώντας σε αρκετές σύγχρονες τεχνικές, καταδεικνύοντας τις κλινικές δυνατότητες και την ανθεκτικότητα του μοντέλου για την παροχή διαγνωστικής υποστήριξης σε πραγματικό χρόνο. Είναι ενδιαφέρον ότι οι συγγραφείς πρότειναν επίσης την ενσωμάτωση του μοντέλου σε ένα ιατρικό ρομπότ κάψουλας στο μέλλον, φανταζόμενοι χρήσιμες χρήσεις στην ελάχιστα επεμβατική διάγνωση, προωθώντας τα εργαλεία CAD που βασίζονται στη βαθιά μάθηση για την αύξηση της ακρίβειας και της συνέπειας του ελέγχου του καρκίνου του παχέος εντέρου.

Οι συγγραφείς Wan, Chen και Yu [38] περιγράφουν μια τεχνική για την αυτοματοποιημένη ανίχνευση πολυπόδων σε εικόνες κολονοσκόπησης χρησιμοποιώντας ένα βελτιωμένο μοντέλο YOLOv5 σε συνδυασμό με έναν μηχανισμό αυτοπροσοχής στην εργασία τους «Polyp Detection from Colorectum Images by Using Attentive YOLOv5», υποστηρίζοντας ότι τα ποσοστά ανίχνευσης αδενωμάτων μπορούν να αυξηθούν με τη χρήση ανάλυσης εικόνων κολονοσκόπησης με τη βοήθεια

τεχνητής νοημοσύνης και η κολονοσκόπηση υψηλής ποιότητας είναι ζωτικής σημασίας για την πρόληψη του καρκίνου του παχέος εντέρου. Χρησιμοποιώντας μια μεθοδολογία βασισμένη στην παλινδρόμηση, το προτεινόμενο μοντέλο προβλέπει άμεσα τα οριοθετημένα πλαίσια για τους πολύποδες σε διάφορες θέσεις, χρησιμοποιώντας ολόκληρη την εικόνα ως είσοδο, χρησιμοποιώντας έναν μηχανισμός προσοχής καναλιών κατά την εξαγωγή χαρακτηριστικών για τη μεγιστοποίηση του αντίκτυπου των πληροφοριακών χαρτών χαρακτηριστικών και τη μείωση του αντίκτυπου των άσχετων. Σύμφωνα με τα πειραματικά αποτελέσματα, η μέθοδος βελτιώνει σημαντικά την ταχύτητα ανίχνευσης σε σχέση με τους συμβατικούς αλγορίθμους και επιτυγχάνει υψηλή ακρίβεια ανίχνευσης, ιδίως για μικρούς και χαμηλής αντίθεσης πολύποδες, καταλήγοντας στο συμπέρασμα ότι η μέθοδός τους μπορεί να μειώσει επιτυχώς τον αριθμό των ενδοσκοπικών εξετάσεων που οδηγούν σε χαμένες διαγνώσεις, παρέχοντας σημαντική κλινική υποστήριξη τόσο για τη διάγνωση όσο και για τη θεραπεία.

Ένας επίμονος περιορισμός των σημερινών διαγνωστικών συστημάτων με τη βοήθεια υπολογιστή είναι η αδυναμία ανίχνευσης πολυπόδων του παχέος εντέρου σε εικόνες κολονοσκόπησης χαμηλής ανάλυσης, με τους Wang, Xie, Cui και Chen [39] να αντιμετωπίζουν αυτό το ζήτημα στην εργασία τους με τίτλο «Colorectal Polyp Detection Model by Using Super-Resolution Reconstruction and YOLO», παρουσιάζοντας το YOLO-SRPD, μια νέα αρχιτεκτονική που ενσωματώνει την ανίχνευση αντικειμένων και την ανακατασκευή υπερ-ανάλυσης σε ένα ενιαίο πλαίσιο. Συγκεκριμένα, το μοντέλο συνδυάζει το SRGAN ενισχυμένο με ACmix για την ενίσχυση των οπτικών λεπτομερειών που είναι απαραίτητες για την ανίχνευση μικρών ή ανεπαίσθητων πολυπόδων, μετατρέποντας εικόνες κολονοσκόπησης χαμηλής ανάλυσης σε εικόνες υψηλής ανάλυσης, ενώ στη συνέχεια, ένα δίκτυο ανίχνευσης με βάση το YOLO εκπαιδεύεται χρησιμοποιώντας αυτές τις φωτογραφίες υψηλής ανάλυσης, αυξάνοντας την ακρίβεια ανίχνευσης και την οπτική σαφήνεια. Προσθέτοντας ενότητες C3_Res2Net για τη βελτίωση της εξαγωγής χαρακτηριστικών πολλαπλών κλιμάκων και CBAM (Convolutional Block Attention Modules) πριν από την κεφαλή πρόβλεψης για την οξύτερη εστίαση στα σχετικά χαρακτηριστικά των πολυπόδων, το μοντέλο YOLO-SRPD ενισχύει ακόμη περισσότερο τη ραχοκοκαλιά YOLOv5. Σύμφωνα με τα πειραματικά αποτελέσματα, το YOLO-SRPD ξεπερνά το βασικό μοντέλο YOLOv5 κατά 1,8% στην ακρίβεια και κατά 5,6% στην ανάκληση, επιτυγχάνοντας μέση μέση ακρίβεια (mAP) 94,2% και ακρίβεια 95,2%, καταδεικνύοντας την ανθεκτικότητα της μεθόδου και την ικανότητά της να παρακάμπτει τα μειονεκτήματα των συμβατικών τεχνικών σε καταστάσεις χαμηλής ανάλυσης, παρέχοντας μια χρήσιμη βελτίωση για την ανίχνευση κλινικών πολυπόδων και την ανάλυση κολονοσκόπησης σε πραγματικό χρόνο.

Οι Tang κ.ά. [40] χρησιμοποίησαν παραγωγικά αντιφατικά δίκτυα (GAN) και βαθιά μάθηση για να δημιουργήσουν ένα σύστημα ανίχνευσης και διάγνωσης με τη βοήθεια υπολογιστή (CADx)

που βελτιώνει την ανίχνευση και την ταξινόμηση των πολυπόδων του παχέος εντέρου, έχοντας ως στόχο να μειώσει το υψηλό ποσοστό αστοχίας που συνδέεται συχνά με την παραδοσιακή κολονοσκόπηση, ενισχύοντας την ικανότητα του μοντέλου να αναγνωρίζει διάφορους τύπους πολυπόδων υπό διαφορετικές συνθήκες απεικόνισης. Για να το επιτύχουν αυτό, οι συγγραφείς δημιούργησαν συνθετικές εικόνες πολυπόδων για την εκπαίδευση του YOLO χρησιμοποιώντας εξαρτημένα GANs, τα οποία αύξησαν επιτυχώς το μέγεθος του συνόλου δεδομένων και βελτίωσαν τη γενίκευση, με το μοντέλο GAN-300, το οποίο εκπαιδεύτηκε χρησιμοποιώντας 300 παραγόμενες εικόνες, επέδειξε την καλύτερη απόδοση ανίχνευσης μεταξύ των διαφόρων δοκιμασμένων διαμορφώσεων. Κατέγραψε μέση ακρίβεια (AP) 54,60% για τα άμισχα οδοντωτά αδενώματα (SSA) και 75,41% για τα σωληνοειδή αδενώματα (TA), υπερβαίνοντας τις συμβατικές τεχνικές επαύξησης δεδομένων, οι οποίες παρήγαγαν χαμηλότερες βαθμολογίες AP (53,56% για το SSA και 72,55% για το TA), ενισχύοντας την ποικιλομορφία εκπαίδευσης και την ευρωστία του μοντέλου. Οι συγγραφείς συμπεριέλαβαν το DeblurGAN-v2 για αποκατάσταση εικόνας πριν από την ταξινόμηση, προκειμένου να αντιμετωπίσουν περαιτέρω το πρόβλημα των θολών εικόνων κολονοσκόπησης, το οποίο μπορεί να εμποδίσει την ακριβή ανίχνευση, αυξάνοντας δραματικά την ακρίβεια ανίχνευσης με τη μέση μέση ακρίβεια (mAP) να αυξάνεται από 25,64% σε 30,74% μετά την αποθορυβοποίηση. Επιπλέον, το μοντέλο GAN-300 ξεπέρασε τις παραδοσιακές μεθόδους επαύξησης, επιτυγχάνοντας 80,97% AP, 70,07% mAP και 57,24% IoU για υπερπλαστικούς πολύποδες (HP), με την μελέτη να δείχνει ότι τα αυτοματοποιημένα συστήματα ανίχνευσης πολυπόδων παχέος εντέρου μπορούν να βελτιώσουν σημαντικά την ακρίβεια και την αξιοπιστία τους συνδυάζοντας την ανίχνευση με βάση το YOLO με αλγόριθμους σύνθεσης εικόνας και αποθορυβοποίησης με βάση το GAN. Προκειμένου να βελτιωθούν τα κλινικά αποτελέσματα και να μειωθεί η διαγνωστική εποπτεία, η παρούσα έρευνα προσφέρει ένα χρήσιμο πλαίσιο για τα εργαλεία κολονοσκόπησης επόμενης γενιάς που υποστηρίζονται από τεχνητή νοημοσύνη.

Για να γεφυρωθεί το χάσμα μεταξύ των ερευνητικών πρωτοτύπων και των ιατρικών συστημάτων που μπορούν να αναπτυχθούν, οι Krenzer κ.ά. [41] δημιούργησαν το ENDOMIND-Advanced, ένα σύστημα ανίχνευσης πολυπόδων σε πραγματικό χρόνο, πλήρως ανοικτού κώδικα, το οποίο προορίζεται για άμεση κλινική εφαρμογή κατά τη διάρκεια της κολονοσκόπησης, στοχεύοντας στην ανάπτυξη ενός αξιόπιστου και χρήσιμου συστήματος ανίχνευσης με τη βοήθεια υπολογιστή (CAD), επειδή κατανοούσαν ότι ο καρκίνος του παχέος εντέρου (CRC) εξακολουθεί να αποτελεί μία από τις κύριες αιτίες θανάτων που σχετίζονται με τον καρκίνο σε παγκόσμιο επίπεδο και ότι οι πολυπόδες που διαφεύγουν κατά τη διάρκεια της κολονοσκόπησης αυξάνουν σημαντικά τον κλινικό κίνδυνο. Το ENDOMIND-Advanced χρησιμοποιήθηκε ως ένα κλινικά έτοιμο πρωτότυπο που θα μπορούσε να υποστηρίξει τους γαστρεντερολόγους κατά τη διάρκεια ζωντανών διαδικασιών, σε αντίθεση με προηγούμενες έρευνες που περιορίζονταν σε πειραματικά

περιβάλλοντα, με το μοντέλο να εκπαιδεύεται σε ένα εξαιρετικά μεγάλο και ποικίλο σύνολο δεδομένων, περισσότερες από 500.000 σχολιασμένες εικόνες, το οποίο συγκεντρώθηκε τόσο από σύνολα δεδομένων ανοικτού κώδικα όσο και από δεδομένα που συγκεντρώθηκαν από πολυάριθμα γερμανικά νοσοκομεία και ιατρικά ιατρεία για να διασφαλιστεί η πλήρης κάλυψη των τύπων πολυπόδων, των συνθηκών απεικόνισης και των ενδοσκοπικών παραλλαγών. Προκειμένου να παρέχει ακριβή και αποτελεσματική παρακολούθηση των πολυπόδων σε πραγματικό χρόνο σε διαδοχικά καρέ, η αρχιτεκτονική του ENDOMIND-Advanced συνδυάζει βαθιά νευρωνικά δίκτυα συνελικτικού τύπου με έναν μηχανισμό ανίχνευσης βίντεο μετά την επεξεργασία. Με τη μείωση των ψευδώς θετικών αποτελεσμάτων και την ενίσχυση της χρονικής συνέπειας των ανιχνεύσεων, ο σχεδιασμός αυτός επιτρέπει στο σύστημα να λειτουργεί άψογα με συνεχείς ροές εικόνας, με το σύστημα να ξεπερνάει όλες τις μεθόδους που είχαν δημοσιευτεί προηγουμένως στη βιβλιογραφία, επιτυγχάνοντας βαθμολογία F1 90,24% κατά τη δοκιμή στο σημείο αναφοράς CVC-VideoClinicDB. Πέρα από τις επιδόσεις και μόνο, η μεγαλύτερη συμβολή της μελέτης είναι ο τρόπος με τον οποίο μεταφράζει την έρευνα στην πράξη, παρέχοντας μια διαφανή πλατφόρμα ανοικτού κώδικα που μπορεί εύκολα να ενσωματωθεί στις τρέχουσες κλινικές ροές εργασίας, αποτελώντας ένα σημαντικό βήμα προς την ευρεία κλινική υιοθέτηση της κολονοσκόπησης με τεχνητή νοημοσύνη.

Σε μια συγκριτική μελέτη, ο Abraham [42] αξιολόγησε την αποτελεσματικότητα τριών μοντέλων μηχανικής μάθησης για την ανίχνευση και ταξινόμηση πολυπόδων παχέος εντέρου χρησιμοποιώντας εικόνες κολονοσκόπησης: Roboflow3 (RF3), You Only Look Once version 8 (YOLOv8n) και Google Teachable Machine (GTM). Δεδομένου ότι οι πολύποδες του παχέος εντέρου συχνά προηγούνται του καρκίνου του παχέος εντέρου, η έγκαιρη και ακριβής ανίχνευση είναι σημαντικά τόσο για την κλινική παρέμβαση όσο και για την πρόληψη, με τις μετρικές ανάκλησης, βαθμολογίας F1 και ακρίβειας που προέκυψαν από την ανάλυση του πίνακα σύγχυσης να χρησιμοποιείται για την αξιολόγηση της απόδοσης του μοντέλου σε ένα σύνολο δεδομένων από εικόνες κολονοσκόπησης που απεικονίζουν φυσιολογικό ιστό παχέος εντέρου, πολύποδες και πολυπόδων που έχουν αφαιρεθεί. Η ικανότητα γενίκευσης των μοντέλων αξιολογήθηκε με τη χρήση ενός επιπλέον συνόλου 90 εξωτερικών εικόνων που δεν χρησιμοποιήθηκαν ούτε για την εκπαίδευση ούτε για την επικύρωση, εξασφαλίζοντας μια ενδελεχή αξιολόγηση των επιδόσεων κάθε μοντέλου σε πραγματικές, αθέατες κλινικές ρυθμίσεις. Με σχεδόν άψογα αποτελέσματα ταξινόμησης σε κάθε κατηγορία, βαθμολογίες ακρίβειας, ανάκλησης και F1 1,00 για φυσιολογικά, 0,97-1,00 για πολύποδες και 0,97-1,00 για εκτοπισμένους πολύποδες, το GTM ξεπέρασε τα άλλα μοντέλα που δοκιμάστηκαν όσον αφορά τη συνολική ακρίβεια (0,99), με τόσο το YOLOv8n όσο και το RF3 να προσφέρουν επιπλέον εντοπισμό αντικειμένων, αναγνωρίζοντας και επισημαίνοντας τις θέσεις των πολυπόδων και των εκτοπισμένων περιοχών μέσα στα πλαίσια της κολονοσκόπησης, ενώ το GTM λειτούργησε ως καθαρός ταξινομητής. Με το RF3 να σημειώνει 0,87 και το YOLOv8n να

επιτυγχάνει 0,84, οι ακρίβειες ταξινόμησής τους ήταν χαμηλότερες παρά την πρόσθετη αυτή λειτουργικότητα, δείχνοντας ότι οι αρχιτεκτονικές που βασίζονται στην ανίχνευση, όπως το YOLOv8n και το RF3, εξακολουθούν να είναι χρήσιμες για κλινική ανάπτυξη σε πραγματικό χρόνο, όπου ο εντοπισμός είναι απαραίτητος για την καθοδήγηση των ενδοσκόπων, παρόλο που τα συμβατικά συστήματα που βασίζονται στην ταξινόμηση, όπως το GTM, υπερτερούν σε κατηγορική ακρίβεια. Συνολικά, η μελέτη δείχνει πώς οι τεχνικές μηχανικής μάθησης, ιδίως τα εύκολα πλοηγήσιμα εργαλεία όπως το GTM, έχουν τη δυνατότητα να αυξήσουν τη διαγνωστική ακρίβεια και να μειώσουν τον αριθμό των πολυπόδων που παραλείπονται κατά τη διάρκεια των διαδικασιών κολονοσκόπησης.

Για να ξεπεραστούν οι αδυναμίες των συμβατικών συστημάτων βαθιάς μάθησης, τα οποία συχνά βασίζονται σε μεμονωμένες αρχιτεκτονικές, έχουν υψηλό υπολογιστικό κόστος και δεν αξιοποιούν πλήρως τις αναπαραστάσεις χαρακτηριστικών πολλαπλών κλιμάκων, ο Attallah [43] παρουσίασε ένα πλαίσιο διάγνωσης με τη βοήθεια υπολογιστή (CAD) που αποσκοπεί στη βελτίωση της αυτοματοποιημένης ταξινόμησης του καρκίνου του πνεύμονα και του παχέος εντέρου από ιστοπαθολογικές εικόνες, με το προτεινόμενο σύστημα να συνδυάζει διάφορα ελαφριά συνελικτικά νευρωνικά δίκτυα (CNN), συγκεκριμένα τα MobileNet, ResNet-18 και EfficientNetB0 προκειμένου να συλλάβει συμπληρωματικές πληροφορίες χαμηλού και υψηλού επιπέδου. Εξάγει επίσης χαρακτηριστικά τόσο από τα επίπεδα συγκέντρωσης όσο και από τα πλήρως συνδεδεμένα επίπεδα, μειώνοντας την πολυπλοκότητα του μοντέλου, διατηρώντας παράλληλα τις διακριτικές πληροφορίες, πραγματοποιώντας μείωση της διάστασης στα χαρακτηριστικά του στρώματος συγκέντρωσης με χρήση της ανάλυσης κανονικής συσχέτισης (CCA). Στη συνέχεια, τα δυνατά σημεία αναπαράστασης των τριών δικτύων συνδυάζονται με τη συγχώνευση των χαρακτηριστικών των διπλών στρωμάτων τους, χρησιμοποιώντας τεχνικές στατιστικής επιλογής όπως η ANOVA και το Chi-squared για την περαιτέρω βελτιστοποίηση του χώρου χαρακτηριστικών, διατηρώντας μόνο τα πιο διακριτικά χαρακτηριστικά, βελτιώνοντας την υπολογιστική απόδοση και την ερμηνευσιμότητα του μοντέλου. Το σύνολο δεδομένων LC25000, το οποίο αποτελείται από επισημασμένες ιστοπαθολογικές εικόνες καρκίνου του πνεύμονα και του παχέος εντέρου, χρησιμοποιήθηκε για την ενδελεχή δοκιμή του πλαισίου με τη χρήση οκτώ διαφορετικών ταξινομητών, συμπεριλαμβανομένων των k-Nearest Neighbors (k-NN), των δέντρων απόφασης, της γραμμικής ανάλυσης διάκρισης (LDA) και διαφόρων τύπων μηχανών διανυσμάτων υποστήριξης (SVM), ενώ ένας κυβικός ταξινομητής SVM με μόλις 50 επιλεγμένα από την ANOVA χαρακτηριστικά πέτυχε ακρίβεια ταξινόμησης 99,8% στα πειράματα, ξεπερνώντας τα μεμονωμένα μοντέλα CNN, ενώ μείωσε σημαντικά τον υπολογιστικό φόρτο. Ένα εξαιρετικά ακριβές αλλά και φορητό διαγνωστικό σύστημα παράγεται από αυτή την τεχνική ενσωμάτωσης πολλαπλών CNN διπλού στρώματος, η οποία είναι ιδιαίτερα κατάλληλη για κλινικές ρυθμίσεις όπου η αξιοπιστία και η ταχύτητα είναι πολύ

σημαντικές, προσφέροντας ισχυρή απόδειξη ότι οι πρωτοποριακές μέθοδοι επιλογής χαρακτηριστικών σε συνδυασμό με μικρές αρχιτεκτονικές CNN μπορούν να αυξήσουν σημαντικά την υπολογιστική απόδοση και τη διαγνωστική ακρίβεια της ανάλυσης ιατρικών εικόνων.

Η δυσκολία επίτευξης ακριβούς αλλά και σε πραγματικό χρόνο τμηματοποίησης πολυπόδων κατά τη διάρκεια της κολονοσκόπησης εξετάζεται από τον Jeong [44] στην εργασία του «Lightweight Deep Learning Model for Real-Time Colorectal Polyp Segmentation», παρουσιάζοντας ότι ενώ η όραση υπολογιστών και η τεχνητή νοημοσύνη έχει σημειώσει σημαντικά βήματα στον αυτόματο εντοπισμό και την οπτικοποίηση των πολυπόδων, αναγνωρίζει ότι τα τρέχοντα μοντέλα κατάτμησης βαθιάς μάθησης έχουν την τάση να γίνονται υπερβολικά περίπλοκα, γεγονός που περιορίζει τη δυνατότητα εφαρμογής τους σε κλινικές ρυθμίσεις πραγματικού χρόνου. Οι συγγραφέας προτείνει ένα ελαφρύ, υψηλής απόδοσης μοντέλο που επιτυγχάνει ισορροπία μεταξύ της υπολογιστικής απόδοσης και της ακρίβειας τμηματοποίησης, προκειμένου να αντιμετωπιστεί αυτό το πρόβλημα, περιλαμβάνοντας τη μεθοδική αντιπαραβολή των τρεχουσών αρχιτεκτονικών τμηματοποίησης για την εύρεση μιας κατάλληλης βάσης, μετά την οποία επιλέγουν και βελτιώνουν ένα μοντέλο DeepLabV3+ που βασίζεται σε έναν κωδικοποιητή MobileNetV3. Βελτιώνουν την ακρίβεια και την ταχύτητα εξαγωγής συμπερασμάτων του μοντέλου προσαρμόζοντας προσεκτικά τις υπερπαραμέτρους. Χρησιμοποιώντας το σύνολο δεδομένων CVC-ClinicDB, η πειραματική αξιολόγηση δείχνει ότι το βελτιστοποιημένο μοντέλο απαιτεί μόνο 1,623 GFLOPs και διατηρεί χαμηλό αριθμό παραμέτρων (6,18 εκατομμύρια), ενώ επιτυγχάνει συντελεστή Dice 93,79%, αποδεικνύοντας ότι η αυξημένη υπολογιστική πολυπλοκότητα δεν μεταφράζεται πάντα σε καλύτερες επιδόσεις. Αντίθετα, η μελέτη υπογραμμίζει τη σημασία αποτελεσματικών αρχιτεκτονικών που διατηρούν υψηλή ακρίβεια ενώ χρησιμοποιούν λιγότερους πόρους, προβάλλοντας μια ισχυρή επιχειρηματολογία για ελαφρές αρχιτεκτονικές βαθιάς μάθησης στην κατάτμηση πολυπόδων του παχέος εντέρου σε πραγματικό χρόνο, τονίζοντας πόσο σημαντικό είναι να επιτευχθεί μια ρεαλιστική ισορροπία μεταξύ της υπολογιστικής βιωσιμότητας των κλινικών εφαρμογών και της διαγνωστικής ακρίβειας.

Στην εργασία τους «An Accurate Deep Learning-Based Computer-Aided Diagnosis System for Gastrointestinal Disease Detection Using Wireless Capsule Endoscopy Image Analysis», ο El-Ghany [45] μιλά για τις δυσκολίες στη χρήση της ασύρματης ενδοσκόπησης με κάψουλα (WCE) για την ακριβή διάγνωση γαστρεντερικών (GI) ασθενειών όπως ο καρκίνος του στομάχου και το πεπτικό έλκος. Κατανοώντας την αξία της έγκαιρης ανίχνευσης, παρουσιάζει τον μηχανισμό Intelligent Learning Rate Controller (ILRC), ο οποίος τροποποιεί δυναμικά τον ρυθμό μάθησης με βάση την πρόοδο της εκπαίδευσης, προκειμένου να βελτιστοποιήσει τη διαδικασία εκπαίδευσης του μοντέλου βαθιάς μάθησης. Επιταχύνοντας τη σύγκλιση και μειώνοντας την υπερπροσαρμογή, αυτός ο προσαρμοστικός μηχανισμός επιτρέπει στα μοντέλα να λειτουργούν πιο αποτελεσματικά

και σταθερά, με τα EfficientNet-B0, ResNet101v2, InceptionV3 και InceptionResNetV2 να είναι μεταξύ των γνωστών αρχιτεκτονικών στις οποίες εφαρμόζεται το ILRC. Όλες αυτές οι αρχιτεκτονικές βελτιώνονται περαιτέρω με τη μεταφορά μάθησης, τη λεπτομερή ρύθμιση και τις σύγχρονες τεχνικές κανονικοποίησης, χρησιμοποιώντας τα σύνολα δεδομένων Kvasir-Capsule και Kvasir v2 για την αξιολόγηση του προτεινόμενου συστήματος, επιτυγχάνοντας κορυφαίες ακρίβειες 99,906% και 98,062%, αντίστοιχα, αποδεικνύοντας πόσο καλά λειτουργούν μαζί η βαθιά μεταφορά μάθησης και ο προσαρμοστικός έλεγχος του ρυθμού μάθησης στην ιατρική απεικόνιση. Τελικά, η μελέτη δείχνει ότι τα μοντέλα που ενισχύονται από το ILRC προσφέρουν ένα αξιόπιστο, αποτελεσματικό και ακριβές πλαίσιο για την αυτοματοποιημένη αναγνώριση των ανωμαλιών του γαστρεντερικού συστήματος, παρέχοντας σημαντική κλινική αξία για την ενίσχυση της αποτελεσματικότητας της ροής εργασίας και της διαγνωστικής ακρίβειας.

Ο συγγραφέας της εργασίας "An Improved Method of Polyp Detection Using Custom YOLOv4-Tiny", ο Doniyorjon [46] επικεντρώνεται στην ενίσχυση της αυτόματης ανίχνευσης της ασύρματης ενδοσκοπικής εικόνας, αποτελώντας σημαντικό έργο για την πρόληψη του καρκίνου και άλλων σοβαρών γαστρεντερικών ασθενειών, προτείνοντας μια τροποποιημένη έκδοση του μοντέλου YOLOv4-tiny που έχει σχεδιαστεί ειδικά για την ανίχνευση και τον εντοπισμό ασύρματης ενδοσκοπικής εικόνας, παρά το γεγονός ότι πολλές μελέτες έχουν προσπαθήσει να βελτιώσουν την ταχύτητα και την ακρίβεια της ανίχνευσης. Η κύρια συνεισφορά είναι η αντικατάσταση του κορμού CSPDarknet-53-tiny από ένα μπλοκ Inception-ResNet-A, το οποίο βελτιώνει σημαντικά την εξαγωγή χαρακτηριστικών και την ακρίβεια ανίχνευσης του μοντέλου, με τους συγγραφείς να παρουσιάζουν επίσης μια μοναδική τεχνική επαύξησης δεδομένων που αποσκοπεί στη διατήρηση της εγγενούς χρωματικής πιστότητας των ιατρικών εικόνων, η οποία είναι σηαντική για την κλινική ακρίβεια. Σύμφωνα με τις πειραματικές αξιολογήσεις, το προτεινόμενο μοντέλο υπερτερεί των προηγούμενων μεθόδων κατά περισσότερο από 1,2%, επιτυγχάνοντας εντυπωσιακή ακρίβεια εκπαίδευσης 99,4%, με τους συγγραφείς να είναι σε θέση να μειώσουν τα σφάλματα ανίχνευσης και εντοπισμού που συχνά προκύπτουν στα τυπικά μοντέλα όταν εφαρμόζονται σε εργασίες ιατρικής απεικόνισης βελτιστοποιώντας τόσο την αρχιτεκτονική του μοντέλου όσο και την ποιότητα του συνόλου δεδομένων. Τα αποτελέσματά τους αναδεικνύουν τον τρόπο με τον οποίο μπορούν να χρησιμοποιηθούν ελαφριά αλλά βασισμένα στο YOLO πλαίσια για την ανάπτυξη αξιόπιστων διαγνωστικών συστημάτων πραγματικού χρόνου για ασύρματη ενδοσκοπική απεικόνιση.

3.5 Ερευνητικά κενά

Υπάρχουν ακόμη ευκαιρίες για περισσότερη έρευνα και καινοτομία, διότι, παρά τις αξιοσημείωτες προόδους στην αυτοματοποιημένη ανίχνευση πολυπόδων, εξακολουθούν να υπάρχουν ορισμένοι σημαντικοί περιορισμοί στη βιβλιογραφία και στα εμπορικά συστήματα, με

την πλειονότητα των σημερινών μοντέλων να αποδίδει εξαιρετικά καλά στα σύνολα δεδομένων στα οποία εκπαιδεύτηκαν, αλλά συχνά έχοντας φτωχή γενίκευση όταν χρησιμοποιούνται σε πραγματικές κλινικές συνθήκες ή σε σύνολα δεδομένων που δεν έχουν ακόμη παρατηρηθεί. Πολλές μελέτες εκπαιδεύονται σε ομοιογενείς συλλογές εικόνων που έχουν ληφθεί υπό ελεγχόμενες συνθήκες, οι οποίες δεν αντικατοπτρίζουν με ακρίβεια τις διακυμάνσεις του φωτισμού, της προετοιμασίας του εντέρου, του τύπου του ενδοσκοπίου ή της ανατομίας του ασθενούς που συναντώνται στις κολονοσκοπήσεις του πραγματικού κόσμου, προερχόμενο κυρίως στην προκατάληψη του συνόλου δεδομένων. Εξαιτίας αυτού, τα μοντέλα που τα πάνε καλά στις αξιολογήσεις αναφοράς ενδέχεται να μην είναι σε θέση να διατηρήσουν σταθερή ακρίβεια σε διάφορα κλινικά πλαίσια.

Το υψηλό ποσοστό ψευδώς θετικών αποτελεσμάτων είναι ένα άλλο συνεχιζόμενο πρόβλημα, με πολλά μοντέλα βαθιάς μάθησης, ιδίως εκείνα που είναι βελτιστοποιημένα ως προς την ευαισθησία, να έχουν την τάση να ταυτοποιούν υπερβολικά τις φυσαλίδες, τα υπολειμματικά συντρίμμια ή τις πτυχές του βλεννογόνου ως πιθανούς πολύποδες. Επειδή οι υπερβολικά πολλοί ψευδείς συναγερμοί διαταράσσουν τη ροή εργασίας και θα μπορούσαν να υπονομεύσουν την εμπιστοσύνη των ιατρών στην αυτοματοποιημένη βοήθεια, το ζήτημα αυτό μειώνει την κλινική χρηστικότητα του συστήματος, με αποτέλεσμα η βελτίωση της ισορροπίας μεταξύ ευαισθησίας και ειδικότητας να εξακολουθεί να αποτελεί πρωταρχικό στόχο στην ανάπτυξη συστημάτων CAD για την κολονοσκόπηση. Επιπλέον, δεν υπάρχουν πολλές ελαφριές αρχιτεκτονικές πραγματικού χρόνου που να λειτουργούν καλά σε κλινικά περιβάλλοντα. Παρά την υψηλή ακρίβεια των βαθιών δικτύων συνελικτικής ανάλυσης, όπως το ResNet και το EfficientNet, η εξαγωγή συμπερασμάτων σε πραγματικό χρόνο στον τυπικό εξοπλισμό κολονοσκόπησης συχνά εμποδίζεται από το υπολογιστικό τους κόστος, καθιστώντας πιο δύσκολη την ενσωμάτωση υποστήριξης ΤΝ κατά τη διάρκεια ζωντανών διαδικασιών, όπου η άμεση ανατροφοδότηση είναι απαραίτητη για τη λήψη αποφάσεων. Ως αποτέλεσμα, η έρευνα κατευθύνεται προς βελτιστοποιημένα μοντέλα υψηλής ταχύτητας που μπορούν να διατηρήσουν την ακρίβεια χωρίς να διακυβεύεται η ταχύτητα εξαγωγής συμπερασμάτων [2].

Επιπλέον, η πρόοδος παρεμποδίζεται από την έλλειψη ποικιλίας δεδομένων και σχολιασμών υψηλής ποιότητας. Σε πολλά από τα τρέχοντα σύνολα δεδομένων, όπως το Kvasir και το CVC-ColonDB, μπορούν να βρεθούν σχετικά λίγα παραδείγματα μικρών, επίπεδων ή άμισχων πολυπόδων, τύποι που είναι πιο δύσκολο να εντοπιστούν αλλά κλινικά πιο σημαντικοί λόγω της κακοήθους δυναμικής τους, με την εκπαίδευση μοντέλων και την συγκριτική αξιολόγηση να δυσχεραίνονται από τα ασυνεπή πρότυπα σχολιασμού στα σύνολα δεδομένων, αναδεικνύοντας την ανάγκη για μεγαλύτερα, κατάλληλα σχολιασμένα σύνολα δεδομένων που καλύπτουν όλο το φάσμα των εμφανίσεων των πολυπόδων και των συνθηκών απεικόνισης [47].

Τέλος, η πλειονότητα των σημερινών μεθόδων δίνει προτεραιότητα στην ακρίβεια ανίχνευσης έναντι της κλινικής ερμηνευσιμότητας και επεξηγηματικότητας, με διαφανή μοντέλα που όχι μόνο εντοπίζουν βλάβες αλλά και προσφέρουν στατιστική ή οπτική υποστήριξη για τις προβλέψεις τους είναι απαραίτητα για τους κλινικούς γιατρούς. Ελλείψει αυτού, τα συστήματα τεχνητής νοημοσύνης εξακολουθούν να είναι «μαύρα κουτιά», γεγονός που υπονομεύει την κανονιστική έγκριση και την εμπιστοσύνη των ιατρών. Με τη δημιουργία ενός μοντέλου που δίνει προτεραιότητα στην ισχυρή γενίκευση σε σύνολα δεδομένων, στη μείωση των ψευδώς θετικών αποτελεσμάτων λόγω της βελτιωμένης διάκρισης χαρακτηριστικών και στην υπολογιστική αποτελεσματικότητα που είναι κατάλληλη για κλινική ανάπτυξη σε πραγματικό χρόνο, η παρούσα μελέτη επιδιώκει να καλύψει αυτά τα κενά, αποσκοπώντας στη βελτίωση της κλινικής ετοιμότητας και της αξιοπιστίας των συστημάτων ανίχνευσης πολυπόδων που βασίζονται στη βαθιά μάθηση μέσω ισορροπημένων στρατηγικών εκπαίδευσης, βελτιστοποιημένου σχεδιασμού δικτύου και αξιολόγησης σε πολλαπλά σύνολα δεδομένων.

Παρά την εντυπωσιακή πρόοδο, η χρήση τεχνητής νοημοσύνης στον εντοπισμό πολύποδων εγείρει και ηθικά και νομικά ερωτήματα. Ποιος φέρει την ευθύνη για μια λανθασμένη διάγνωση; Μπορεί ένας αλγόριθμος να υποκαταστήσει την ανθρώπινη κρίση; Πώς διασφαλίζεται η προστασία των προσωπικών δεδομένων των ασθενών; Αυτά τα ζητήματα απασχολούν έντονα τόσο τους επιστήμονες της πληροφορικής όσο και τους γιατρούς και νομικούς.

Επιπλέον, χρειάζεται διαφάνεια στην "ερμηνευσιμότητα" (explainability) των αποφάσεων που λαμβάνει ο αλγόριθμος. Οι αλγόριθμοι black-box, όσο αποτελεσματικοί κι αν είναι, στερούνται επεξηγηματικότητας και μπορεί να δημιουργούν αβεβαιότητα στον ιατρό.

Η ΤΝ, αν χρησιμοποιηθεί με κατάλληλο τρόπο, δεν έρχεται να αντικαταστήσει τον γιατρό, αλλά να τον ενισχύσει. Η τεχνολογία πρέπει να λειτουργεί υποστηρικτικά, προσφέροντας πληροφόρηση και όχι επιβολή. Η τελική απόφαση ανήκει πάντοτε στον ιατρό, ο οποίος διαθέτει το πλαίσιο, την κλινική εμπειρία και τη συνολική εικόνα του ασθενούς.

Εν κατακλείδι, η εφαρμογή τεχνητής νοημοσύνης και ιδιαίτερα βαθιάς μάθησης στην ανίχνευση πολυπόδων αποτελεί ένα από τα πιο επιτυχημένα παραδείγματα ψηφιακού μετασχηματισμού στον χώρο της ιατρικής. Η ακρίβεια και η ταχύτητα με την οποία εντοπίζονται οι ύποπτες αλλοιώσεις προσφέρουν νέες δυνατότητες πρόληψης και έγκαιρης διάγνωσης, με σαφή οφέλη για την έκβαση της νόσου.

Ωστόσο, η επιτυχής ενσωμάτωση της τεχνολογίας αυτής απαιτεί όχι μόνο τεχνική υπεροχή, αλλά και διαλειτουργικότητα, αποδοχή από το ιατρικό προσωπικό, συμμόρφωση με κανονιστικά πλαίσια και ηθικά όρια. Η πρόκληση των επόμενων ετών θα είναι η αρμονική συνύπαρξη της ανθρώπινης εμπειρίας με την υπολογιστική ευφυΐα.

ΚΕΦΑΛΑΙΟ 4

4 Συνελικτικά Νευρωνικά Δίκτυα

4.1 Η βαθιά μάθηση στην ιατρική απεικόνιση

Η χρήση της βαθιάς μάθησης στην ιατρική απεικόνιση έχει φέρει επανάσταση στις αναλυτικές και διαγνωστικές διαδικασίες σε διάφορους κλινικούς τομείς, συμπεριλαμβανομένης της ακτινολογίας και της ενδοσκόπησης, με την πολύπλοκη μεταβλητότητα των βιολογικών δομών συχνά να μην λαμβανόταν υπόψη από τα χειροποίητα χαρακτηριστικά και τις ευρετικές μεθόδους για συγκεκριμένους τομείς που αποτελούσαν σημαντικό μέρος της παραδοσιακής ανάλυσης εικόνων. Τα μοντέλα μπορούν πλέον να μαθαίνουν περίπλοκες, ιεραρχικές αναπαραστάσεις των οπτικών πληροφοριών απευθείας από τα ακατέργαστα δεδομένα εικονοστοιχείων χάρη στη βαθιά μάθηση, ιδίως τα συνελικτικά νευρωνικά δίκτυα (CNN), τα οποία εισήγαγαν ένα παράδειγμα με γνώμονα τα δεδομένα, έχοντας ιδιαίτερα σημαντική επίδραση στην ενδοσκόπηση του γαστρεντερικού συστήματος, όπου η ικανότητα του μοντέλου να εντοπίζει μικροσκοπικές παραλλαγές υφής και δομής που ακόμη και εξειδικευμένοι κλινικοί ιατροί μπορεί να μην προσέξουν, βοηθά στην ανίχνευση βλαβών και σε εργασίες ταξινόμησης.

4.2 Επισκόπηση των CNNs, της μεταφοράς μάθησης και της ανίχνευσης αντικειμένων

Το θεμέλιο της σύγχρονης όρασης υπολογιστών είναι τα συνελικτικά νευρωνικά δίκτυα (CNNs), με την αρχιτεκτονική τους να χρησιμοποιεί στρώματα συνέλιξης, ενεργοποίησης, συγκέντρωσης (pooling) και κανονικοποίησης για την αυτόματη εξαγωγή χωρικών ιεραρχιών χαρακτηριστικών. Η βασική λειτουργία ενός συνελικτικού στρώματος είναι η πράξη της συνέλιξης, κατά την οποία ένα μικρό φίλτρο (kernel) διαστάσεων $\mathrm{F} \times \mathrm{F}$ «ολισθαίνει» πάνω στην εικόνα εισόδου και υπολογίζει το τοπικό εσωτερικό γινόμενο μεταξύ των τιμών της εικόνας και των βαρών του φίλτρου. Μαθηματικά, η έξοδος ενός φίλτρου σε θέση (x, y) μπορεί να περιγραφεί ως άθροισμα γινομένων των αντίστοιχων στοιχείων της περιοχής εισόδου και του πυρήνα. Η διαδικασία αυτή παράγει έναν χάρτη χαρακτηριστικών (feature map), ο οποίος αναδεικνύει συγκεκριμένα οπτικά πρότυπα.

Παράμετροι όπως το βήμα μετατόπισης (stride) και η προσθήκη περιμετρικής επέκτασης (padding) επηρεάζουν τη χωρική διάσταση του παραγόμενου χάρτη χαρακτηριστικών και ελέγχουν τον βαθμό υποδειγματοληψίας και διατήρησης πληροφορίας στα όρια της εικόνας. Μετά τη συνέλιξη εφαρμόζεται συνάρτηση μη γραμμικής ενεργοποίησης, συνήθως η ReLU (Rectified Linear Unit), η οποία επιτρέπει στο δίκτυο να προσεγγίζει σύνθετες μη γραμμικές σχέσεις. Τα στρώματα

pooling μειώνουν τη χωρική ανάλυση των feature maps, περιορίζοντας το υπολογιστικό κόστος και αυξάνοντας την ανθεκτικότητα σε μικρές μετατοπίσεις ή παραμορφώσεις.

Τα CNN έχουν αποδειχθεί ιδιαίτερα αποτελεσματικά στην ανίχνευση παθολογικών προτύπων σε ιατρικές απεικονιστικές μεθόδους όπως η μαγνητική τομογραφία, η αξονική τομογραφία, οι ακτίνες Χ και η ενδοσκόπηση. Σε ένα CNN, κάθε συνελικτικό στρώμα καταγράφει προοδευτικά πιο αφηρημένα χαρακτηριστικά: τα αρχικά στρώματα αναγνωρίζουν βασικά μορφολογικά στοιχεία όπως ακμές ή χρωματικές κλίσεις, ενώ τα βαθύτερα στρώματα συνδυάζουν αυτές τις πληροφορίες για να αναγνωρίσουν περίπλοκα σχήματα, υφές και ανατομικές δομές [48]. Στην περίπτωση της ενδοσκοπικής απεικόνισης, αυτή η ιεραρχική αναπαράσταση είναι κρίσιμη, καθώς οι πολύποδες μπορεί να εμφανίζονται με μικρές διαφοροποιήσεις υφής, χρωματισμού ή γεωμετρίας σε σχέση με τον φυσιολογικό βλεννογόνο.

Η εκπαίδευση βαθιών δικτύων από το μηδέν, ωστόσο, απαιτεί τεράστιους όγκους επισημασμένων δεδομένων και σημαντικούς υπολογιστικούς πόρους, τα οποία συχνά δεν είναι διαθέσιμα στον ιατρικό τομέα. Για τον λόγο αυτό, η μεταφορά μάθησης έχει αναδειχθεί ως βασική στρατηγική για την παράκαμψη αυτού του περιορισμού. Με την προσαρμογή των τελευταίων στρωμάτων ή την επιλεκτική επανεκπαίδευση συγκεκριμένων μπλοκ του δικτύου (fine-tuning), τα προ-εκπαιδευμένα δίκτυα, που έχουν εκπαιδευτεί σε μεγάλα σύνολα δεδομένων φυσικών εικόνων όπως το ImageNet [49], μπορούν να προσαρμοστούν σε εξειδικευμένες εργασίες ιατρικής απεικόνισης. Τα πρώιμα στρώματα, τα οποία μαθαίνουν γενικά οπτικά χαρακτηριστικά (όπως ανίχνευση ακμών και βασικών σχημάτων), διατηρούνται, ενώ τα βαθύτερα στρώματα προσαρμόζονται ώστε να αναγνωρίζουν πρότυπα που σχετίζονται με συγκεκριμένες παθολογικές οντότητες. Με τον τρόπο αυτό, βελτιώνεται η απόδοση ακόμη και όταν το διαθέσιμο ιατρικό σύνολο δεδομένων είναι περιορισμένο, ενώ παράλληλα μειώνεται ο χρόνος εκπαίδευσης και ο κίνδυνος υπερπροσαρμογής (overfitting). Προσαρμόζοντας το μοντέλο σε πρότυπα που αφορούν συγκεκριμένους τομείς, όπως οι υφές του βλεννογόνου ή οι αγγειακές δομές σε εικόνες κολονοσκόπησης, η μεταφορά μάθησης καθιστά δυνατή την επαναχρησιμοποίηση γενικών οπτικών χαρακτηριστικών σε ένα εξειδικευμένο κλινικό πλαίσιο.

Πέρα από την ταξινόμηση, η βαθιά μάθηση επεκτείνεται στην ανίχνευση αντικειμένων, όπου ο στόχος δεν είναι μόνο η αναγνώριση της παρουσίας ενός αντικειμένου, αλλά και ο χωρικός εντοπισμός του μέσα στην εικόνα. Τα δίκτυα ανίχνευσης αντικειμένων προβλέπουν οριοθετημένα πλαίσια (bounding boxes) και αντίστοιχες πιθανότητες κατηγορίας, συνδυάζοντας μηχανισμούς εξαγωγής χαρακτηριστικών με διαδικασίες παλινδρόμησης συντεταγμένων. Τα σύγχρονα πλαίσια, όπως το SSD (Single Shot MultiBox Detector), το YOLO (You Only Look Once) και το Faster R-CNN, προσφέρουν διαφορετικούς συμβιβασμούς μεταξύ ακρίβειας και ταχύτητας επεξεργασίας. Σε

εφαρμογές κολονοσκόπησης, η δυνατότητα ανίχνευσης σε πραγματικό χρόνο είναι ιδιαίτερα κρίσιμη, καθώς επιτρέπει στο σύστημα να εντοπίζει πολύποδες κατά τη διάρκεια ζωντανών διαδικασιών βίντεο και να παρέχει άμεση υποστήριξη στον ενδοσκόπο χωρίς να διαταράσσει τη ροή της κλινικής πράξης [50].

4.3 Σχετικές αρχιτεκτονικές: VGG, ResNet, U-Net, YOLO, EfficientNet κ.λπ.

Λόγω της αποδεδειγμένης αξιοπιστίας τους, της επεκτασιμότητας και της ευελιξίας τους για την προσαρμογή σε διάφορες μορφές απεικόνισης, μια σειρά αρχιτεκτονικών βαθιάς μάθησης έχουν αναδειχθεί ως βασικά στοιχεία της ανάλυσης ιατρικών εικόνων, με τα VGG, ResNet, U-Net, YOLO και EfficientNet να αποτελούν αξιοσημείωτα σημεία καμπής στην ανάπτυξη της αρχιτεκτονικής CNN. Το δίκτυο VGG (Visual Geometry Group), το οποίο παρουσιάστηκε το 2014, καθιέρωσε ένα απλό αλλά αποτελεσματικό αρχιτεκτονικό παράδειγμα που βασίζεται σε διαδοχικά στοιβαγμένους μικρούς πυρήνες συνελικτικής απεικόνισης 3×3. Η βασική φιλοσοφία του VGG είναι ότι η αντικατάσταση μεγάλων φίλτρων (π.χ. 7×7) με πολλαπλές διαδοχικές συνελίξεις 3×3 αυξάνει το βάθος του δικτύου, επιτρέποντας την εκμάθηση πιο σύνθετων μη γραμμικών αναπαραστάσεων με μικρότερο αριθμό παραμέτρων ανά στρώμα. Μετά από κάθε μπλοκ συνελίξεων εφαρμόζεται συνήθως max pooling, μειώνοντας τη χωρική διάσταση και ενισχύοντας την ιεραρχική αφαίρεση χαρακτηριστικών. Εξακολουθεί να είναι μία από τις πιο ευρέως χρησιμοποιούμενες ραχοκοκαλιές σε εργασίες ταξινόμησης και ανίχνευσης ιατρικών εικόνων λόγω της ομοιόμορφης δομής και της απλότητάς του, καθιστώντας το ιδανικό για μεταφορά μάθησης [51].

Αντιμετωπίζοντας το ζήτημα της εξαφανιζόμενης κλίσης (vanishing gradient), το ResNet (Residual Network) κατέστησε δυνατή την αποτελεσματική εκπαίδευση πολύ βαθύτερων δικτύων εισάγοντας την ιδέα των υπολειμματικών συνδέσεων (skip connections). Συγκεκριμένα, αντί το δίκτυο να μαθαίνει άμεσα μια επιθυμητή απεικόνιση $\mathrm{H}(\mathrm{x})$, μαθαίνει το υπολειμματικό σήμα $\mathrm{F}(\mathrm{x}) = \mathrm{H}(\mathrm{x}) - \mathrm{x}$, επιτρέποντας την άμεση παράκαμψη της εισόδου μέσω ταυτότητας (identity mapping). Έτσι, η έξοδος ενός residual block δίνεται από τη σχέση $\mathrm{y} = \mathrm{F}(\mathrm{x}) + \mathrm{x}$, διευκολύνοντας τη ροή των βαθμίδων κατά την οπισθοδιάδοση και επιτρέποντας την εκπαίδευση δικτύων εκατοντάδων στρωμάτων χωρίς απώλεια πληροφορίας. Το ResNet έχει χρησιμοποιηθεί ευρέως στην ιατρική απεικόνιση για την ανίχνευση ανωμαλιών σε διάφορα όργανα εξαιτίας της ικανότητάς του να μαθαίνει βαθύτερες και πιο αφηρημένες αναπαραστάσεις με αυξημένη γενίκευση, ιδίως σε δύσκολες οπτικές συνθήκες.

Η de facto τυπική αρχιτεκτονική για εργασίες ιατρικής απεικόνισης που απαιτούν ακριβή εντοπισμό είναι το U-Net, το οποίο δημιουργήθηκε αρχικά για την κατάτμηση βιοϊατρικών εικόνων. Η αρχιτεκτονική του βασίζεται σε συμμετρική δομή κωδικοποιητή–αποκωδικοποιητή (encoder–decoder). Το τμήμα του κωδικοποιητή μειώνει σταδιακά τη χωρική ανάλυση μέσω συνελίξεων και

pooling, εξάγοντας υψηλού επιπέδου χαρακτηριστικά, ενώ το τμήμα του αποκωδικοποιητή επαναφέρει τη χωρική ανάλυση μέσω upsampling ή transposed convolutions. Οι συνδέσεις παράλειψης (skip connections) μεταξύ αντίστοιχων επιπέδων encoder και decoder επιτρέπουν τη μεταφορά λεπτομερούς χωρικής πληροφορίας, γεγονός που καθιστά το U-Net ιδιαίτερα κατάλληλο για εργασίες όπως η ανίχνευση ορίων όγκων, η οριοθέτηση οργάνων και η κατάτμηση πολυπόδων. Η δυνατότητα ακριβούς pixel-wise πρόβλεψης είναι κρίσιμη σε εφαρμογές όπου απαιτείται λεπτομερής εντοπισμός μικρών ή επίπεδων βλαβών.

Η ανίχνευση αντικειμένων σε πραγματικό χρόνο μετασχηματίστηκε από το YOLO (You Only Look Once), το οποίο διαμόρφωσε την ανίχνευση ως ένα πρόβλημα ενιαίας παλινδρόμησης (single-stage regression). Σε αντίθεση με ανιχνευτές δύο σταδίων, όπως το Faster R-CNN [52], που πρώτα προτείνουν περιοχές ενδιαφέροντος και έπειτα ταξινομούν, το YOLO επεξεργάζεται ολόκληρη την εικόνα σε ένα μόνο πέρασμα, προβλέποντας ταυτόχρονα συντεταγμένες bounding boxes και πιθανότητες κατηγοριών. Η εικόνα χωρίζεται σε πλέγμα (grid), και για κάθε κελί του πλέγματος το δίκτυο προβλέπει υποψήφια πλαίσια και αντίστοιχες πιθανότητες εμπιστοσύνης. Αυτή η ενιαία προσέγγιση επιτρέπει υψηλής ταχύτητας συμπέρασμα (inference), γεγονός ιδιαίτερα σημαντικό σε εφαρμογές ζωντανής ενδοσκόπησης όπου η καθυστέρηση πρέπει να είναι ελάχιστη. Στην κολονοσκόπηση, οι πιο πρόσφατες παραλλαγές YOLOv5 και YOLOv8 έχουν χρησιμοποιηθεί αποτελεσματικά για την ανίχνευση πολυπόδων με ελάχιστη καθυστέρηση και υψηλή ακρίβεια, διατηρώντας παράλληλα επαρκή σταθερότητα σε μεταβαλλόμενες συνθήκες φωτισμού και κίνησης.

Μια πιο πρόσφατη εξέλιξη που δίνει προτεραιότητα στην επεκτασιμότητα και την αποδοτικότητα είναι το EfficientNet. Η βασική καινοτομία του είναι η σύνθετη κλιμάκωση (compound scaling), κατά την οποία το βάθος (depth), το πλάτος (width) και η ανάλυση εισόδου (resolution) αυξάνονται ταυτόχρονα με ελεγχόμενο και μαθηματικά βελτιστοποιημένο τρόπο, αντί να κλιμακώνονται ανεξάρτητα. Η προσέγγιση αυτή επιτρέπει καλύτερη ισορροπία μεταξύ αριθμού παραμέτρων και απόδοσης, επιτυγχάνοντας υψηλή ακρίβεια με σημαντικά μειωμένο υπολογιστικό κόστος. Λόγω της ισορροπίας ακρίβειας και ταχύτητας, το EfficientNet είναι ιδιαίτερα ελκυστικό για ενσωμάτωση σε κλινικές ρυθμίσεις όπου απαιτείται χαμηλή κατανάλωση ενέργειας, περιορισμένοι υπολογιστικοί πόροι και ανατροφοδότηση σε σχεδόν πραγματικό χρόνο.

Αυτές οι αρχιτεκτονικές συνεργάζονται για να παρέχουν το τεχνολογικό πλαίσιο για τα σύγχρονα συστήματα CAD με βάση τη βαθιά μάθηση που χρησιμοποιούνται στις κολονοσκοπήσεις, επιτρέποντας στους ερευνητές να τροποποιούν ή να συνδυάζουν τα μοναδικά πλεονεκτήματα που συνεισφέρει το καθένα: τη μεταφερσιμότητα και απλότητα της VGG, το μεγάλο βάθος και τη σταθερότητα εκπαίδευσης του ResNet, την ακριβή χωρική κατάτμηση του U-Net, την ταχύτητα και

καταλληλότητα πραγματικού χρόνου του YOLO και τη βελτιστοποιημένη αποδοτικότητα του EfficientNet ανάλογα με τις ιδιαίτερες ανάγκες της εφαρμογής [53]. Η ταχεία εξέλιξη αυτών των μοντέλων φέρνει τον τομέα όλο και πιο κοντά σε πλήρως ολοκληρωμένα ενδοσκοπικά συστήματα υποβοηθούμενα από τεχνητή νοημοσύνη, τα οποία δεν περιορίζονται στην απλή ανίχνευση αλλά στοχεύουν σε αξιόπιστη, σταθερή και κλινικά αξιοποιήσιμη υποστήριξη λήψης απόφασης.

4.4 CNNs

Τα Συνελικτικά Νευρωνικά Δίκτυα αποτελούν μια κατηγορία νευρωνικών δικτύων που σχεδιάστηκαν ειδικά για δεδομένα εικόνας. Ο βασικός τους στόχος είναι να μιμηθούν τον τρόπο με τον οποίο ο ανθρώπινος εγκέφαλος αναγνωρίζει οπτικά μοτίβα, αξιοποιώντας τοπικές συσχετίσεις και ιεραρχική επεξεργασία πληροφορίας. Σε αντίθεση με απλά νευρωνικά δίκτυα, που μετατρέπουν μια εικόνα σε ένα μονοδιάστατο διάνυσμα αριθμών αγνοώντας τη γεωμετρική της δομή, τα CNNs διατηρούν τη χωρική διάταξη των pixel. Με άλλα λόγια, λαμβάνουν υπόψη ότι τα γειτονικά πίξελ συχνά συνδέονται νοηματικά μεταξύ τους και ότι οι τοπικές δομές είναι κρίσιμες για την αναγνώριση οπτικών χαρακτηριστικών.

Ο πυρήνας της λειτουργίας τους είναι τα συνελικτικά φίλτρα (kernels). Ένα φίλτρο είναι ένας μικρός πίνακας βαρών, συνήθως διαστάσεων 3×3 ή 5×5, ο οποίος «ολισθαίνει» πάνω στην εικόνα εισόδου. Σε κάθε θέση υπολογίζεται το άθροισμα γινομένων μεταξύ των στοιχείων του φίλτρου και των αντίστοιχων τιμών της εικόνας. Αν συμβολίσουμε την είσοδο ως $I(x, y)$ και το φίλτρο ως $K(i, j)$, τότε η τιμή εξόδου σε μια θέση δίνεται από:

$$S(x, y) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} K(i, j) \cdot I(x + i, y + j)$$

Το αποτέλεσμα αυτής της πράξης είναι ένας χάρτης χαρακτηριστικών (feature map). Ως «feature» ορίζεται οποιοδήποτε ανιχνεύσιμο μοτίβο, όπως μια ακμή, μια συγκεκριμένη υφή ή μια χρωματική μετάβαση. Για παράδειγμα, ένα φίλτρο μπορεί να ενεργοποιείται έντονα όταν συναντά κάθετες ακμές, ενώ ένα άλλο όταν ανιχνεύει κυκλικές δομές, κάτι ιδιαίτερα σημαντικό στην ανίχνευση πολυπόδων.

Μετά τη συνέλιξη εφαρμόζεται συνήθως μια μη γραμμική συνάρτηση ενεργοποίησης, όπως η ReLU (Rectified Linear Unit), η οποία ορίζεται ως:

$$f(x) = \max(0, x)$$

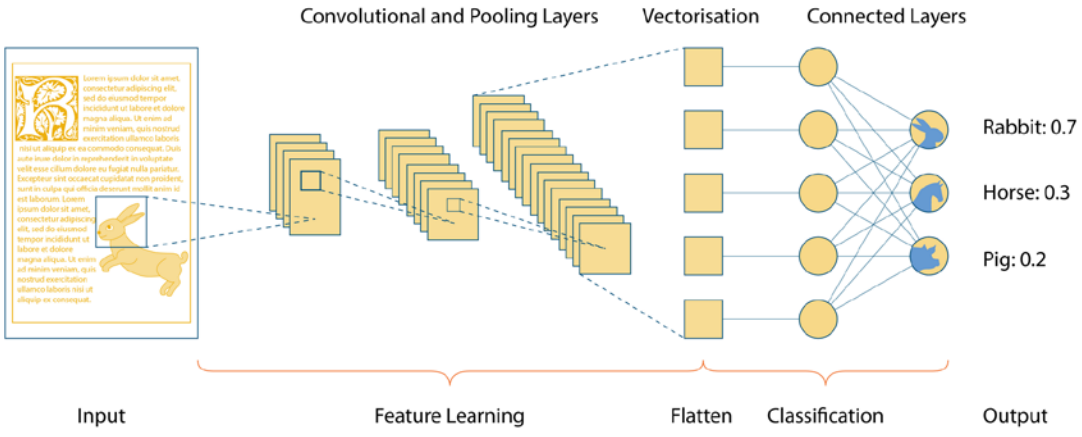
Η ReLU μηδενίζει τις αρνητικές τιμές και διατηρεί τις θετικές, επιτρέποντας στο δίκτυο να εισάγει μη γραμμικότητα και να μοντελοποιεί σύνθετες σχέσεις πέρα από απλούς γραμμικούς συνδυασμούς.

Στην αρχή το δίκτυο βρίσκει πολύ απλά στοιχεία, όπως ακμές και βασικές υφές. Όμως όσο περνά μέσα από βαθύτερα στρώματα, τα feature maps συνδυάζονται και δημιουργούν πιο σύνθετες αναπαραστάσεις. Η διαδικασία αυτή επιτρέπει στο CNN να χτίζει μια ιεραρχική κατανόηση της εικόνας. Για παράδειγμα, πρώτα αναγνωρίζει άκρες, μετά ενώσεις αυτών των άκρων, στη συνέχεια σχήματα και τελικά αντικείμενα ή ανατομικές δομές, όπως μια πολυποειδή προεξοχή στο τοίχωμα του εντέρου.

Ένα ακόμη στοιχείο που παίζει σημαντικό ρόλο είναι τα pooling layers. Αυτά μειώνουν τη χωρική διάσταση της εικόνας συγκεντρώνοντας τις πιο σημαντικές πληροφορίες. Στο max pooling, για παράδειγμα, επιλέγεται η μέγιστη τιμή μέσα σε ένα μικρό παράθυρο (π.χ. 2×2). Με αυτόν τον τρόπο μειώνεται ο αριθμός παραμέτρων και το υπολογιστικό κόστος. Όχι μόνο επιταχύνουν την εκπαίδευση, αλλά βοηθούν το μοντέλο να γίνει πιο σταθερό σε μικρές αλλαγές θέσης ή κλίμακας. «Πιο σταθερό» σημαίνει ότι αν ένα χαρακτηριστικό μετατοπιστεί ελαφρά στην εικόνα, το δίκτυο εξακολουθεί να το αναγνωρίζει. Αυτό είναι ιδιαίτερα χρήσιμο σε ιατρικά δεδομένα, όπου δύο εικόνες του ίδιου ιστού μπορεί να διαφέρουν ελαφρά λόγω θέσης, γωνίας λήψης ή φωτεινότητας.

Προς το τέλος της αρχιτεκτονικής συναντάμε τα πλήρως συνδεδεμένα στρώματα (fully connected layers). Σε αυτό το στάδιο, οι πολυδιάστατοι χάρτες χαρακτηριστικών μετατρέπονται σε ένα μονοδιάστατο διάνυσμα (flattening), και κάθε νευρώνας συνδέεται με όλους τους νευρώνες του προηγούμενου στρώματος. Μαθηματικά, η έξοδος υπολογίζεται ως γραμμικός συνδυασμός των εισόδων και των βαρών, στον οποίο εφαρμόζεται κατάλληλη συνάρτηση ενεργοποίησης. Σε προβλήματα ταξινόμησης, το τελικό στρώμα χρησιμοποιεί συνήθως τη συνάρτηση Softmax, η οποία μετατρέπει τις τιμές εξόδου σε πιθανότητες που αθροίζονται στο 1. Έτσι, η τελική πρόβλεψη έχει τη μορφή πιθανοτήτων ανά κατηγορία, π.χ.:

$$\mathrm{P}(\text{πολύποδας}) = 0.92, \qquad \mathrm{P}(\text{μη πολύποδας}) = 0.08$$



Εικόνα 8 – Αναπαράσταση αρχιτεκτονικής Convolutional Neural Network [49]

Η τελική απόφαση λαμβάνεται επιλέγοντας την κατηγορία με τη μεγαλύτερη πιθανότητα. Η διαδικασία είναι καθαρά μαθηματική, βασισμένη σε βελτιστοποίηση συνάρτησης κόστους μέσω οπισθοδιάδοσης σφάλματος, όμως πρακτικά σημαίνει ότι το σύστημα έχει ήδη αναγνωρίσει τα βασικά και τα σύνθετα μοτίβα της εικόνας και τα μεταφράζει σε μια κλινικά χρήσιμη απόφαση.

Στις μέρες μας, τα CNNs σπάνια εκπαιδεύονται από το μηδέν. Οι ερευνητές χρησιμοποιούν προεκπαιδευμένα μοντέλα που έχουν μάθει από τεράστια σύνολα δεδομένων. Με αυτή τη μέθοδο, γνωστή ως μεταφορά μάθησης, το δίκτυο ξεκινά ήδη από μια καλή βάση και χρειάζεται να μάθει μόνο τα πιο ειδικά χαρακτηριστικά της εκάστοτε εφαρμογής. Αυτό μειώνει τον χρόνο εκπαίδευσης και βελτιώνει την απόδοση σε περιπτώσεις όπου τα διαθέσιμα δεδομένα είναι περιορισμένα, όπως συμβαίνει συχνά στη βιοϊατρική.

Ο λόγος που τα CNNs κυριαρχούν στον χώρο της ιατρικής απεικόνισης είναι η ικανότητά τους να εντοπίζουν λεπτές διαφορές που συχνά ξεφεύγουν από το ανθρώπινο μάτι. Παράλληλα, επειδή είναι συνεπή και δεν επηρεάζονται από κόπωση, μπορούν να λειτουργούν ως υποστηρικτικό εργαλείο για κλινικούς. Δεν αντικαθιστούν την κρίση του γιατρού, αλλά συμβάλλουν στη μείωση λαθών και στην αύξηση της ευαισθησίας στην ανίχνευση αλλοιώσεων.

4.5 Μετρικές αξιολόγησης

Όταν αξιολογούμε ένα σύστημα ταξινόμησης στην ιατρική απεικόνιση, δεν αρκεί να δούμε απλώς πόσο συχνά «πέφτει μέσα». Κάθε μετρική αποτυπώνει μια διαφορετική πλευρά της απόδοσης και συχνά η πραγματική εικόνα προκύπτει από τον συνδυασμό τους. Αυτό είναι ιδιαίτερα σημαντικό σε προβλήματα όπου οι κλάσεις δεν είναι ισομερώς εκπροσωπημένες ή όταν μια λανθασμένη πρόβλεψη μπορεί να έχει σοβαρές κλινικές συνέπειες.

Ο πίνακας σύγχυσης αποτελείται από:

TP (True Positives)	TN (True Negatives)	FP (False Positives)	FN (False Negatives)
Σωστά θετικά (σωστή ανίχνευση πολύποδα)	Σωστά αρνητικά (σωστή αναγνώριση απουσίας πολύποδα)	Ψευδώς θετικά (λανθασμένη ανίχνευση πολύποδα)	Ψευδώς αρνητικά (μη ανίχνευση υπαρκτού πολύποδα)

Accuracy

Το accuracy δείχνει το ποσοστό των σωστών προβλέψεων. Είναι η πιο απλή μετρική, αλλά δεν είναι πάντα η πιο χρήσιμη. Σε σύνολα δεδομένων με μεγάλη ανισορροπία μπορεί να δώσει μια παραπλανητική εικόνα. Για παράδειγμα, αν η πλειονότητα των εικόνων ανήκει σε μία μόνο

κατηγορία, ένα μοντέλο μπορεί να πετυχαίνει υψηλό accuracy απλώς προβλέποντας πάντα αυτή την κατηγορία.

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

Το precision δείχνει πόσο αξιόπιστο είναι το μοντέλο όταν δηλώνει ότι εντόπισε μια θετική περίπτωση. Με άλλα λόγια, μετρά πόσες από τις θετικές προβλέψεις είναι πραγματικά σωστές. Σε εφαρμογές όπως η ανίχνευση πολυπόδων, η υψηλή τιμή precision σημαίνει ότι το μοντέλο αποφεύγει τα ψευδώς θετικά, κάτι που μειώνει περιττές ενέργειες και άσκοπο άγχος για τον ασθενή.

$$\boldsymbol{Precision} = \frac{TP}{TP + FP}$$

Recall

Το recall δείχνει πόσο συχνά το μοντέλο βρίσκει πραγματικά τις θετικές περιπτώσεις. Είναι ιδιαίτερα κρίσιμο στην ιατρική, όπου η παράλειψη μιας βλάβης μπορεί να έχει σοβαρές συνέπειες. Ένα μοντέλο με υψηλή ευαισθησία μπορεί να μην είναι τέλειο σε άλλες μετρικές, αλλά τουλάχιστον δεν αφήνει εύκολα μια σημαντική ανωμαλία να περάσει απαρατήρητη.

$$\boldsymbol{Recall} = \frac{TP}{TP + FN}$$

Specificity

Η specificity δείχνει πόσο καλά το μοντέλο αναγνωρίζει τις αρνητικές περιπτώσεις. Στην πράξη μετρά την ικανότητά του να αποφεύγει τα ψευδώς θετικά. Μαζί με την ευαισθησία δίνουν μια πλήρη εικόνα για το πώς συμπεριφέρεται το σύστημα απέναντι σε πραγματικά θετικές και αρνητικές περιπτώσεις.

$$\boldsymbol{Specificity} = \frac{TN}{TN + FP}$$

F1-score

Το F1-score συνδυάζει precision και recall σε μία τιμή. Είναι χρήσιμο όταν θέλουμε μια ισορροπημένη εικόνα, ειδικά σε προβλήματα με ανισορροπία κλάσεων. Ένα μοντέλο μπορεί να έχει υψηλό precision ή υψηλό recall, αλλά το F1-score αποκαλύπτει αν έχει πραγματικά καλή συνολική συμπεριφορά.

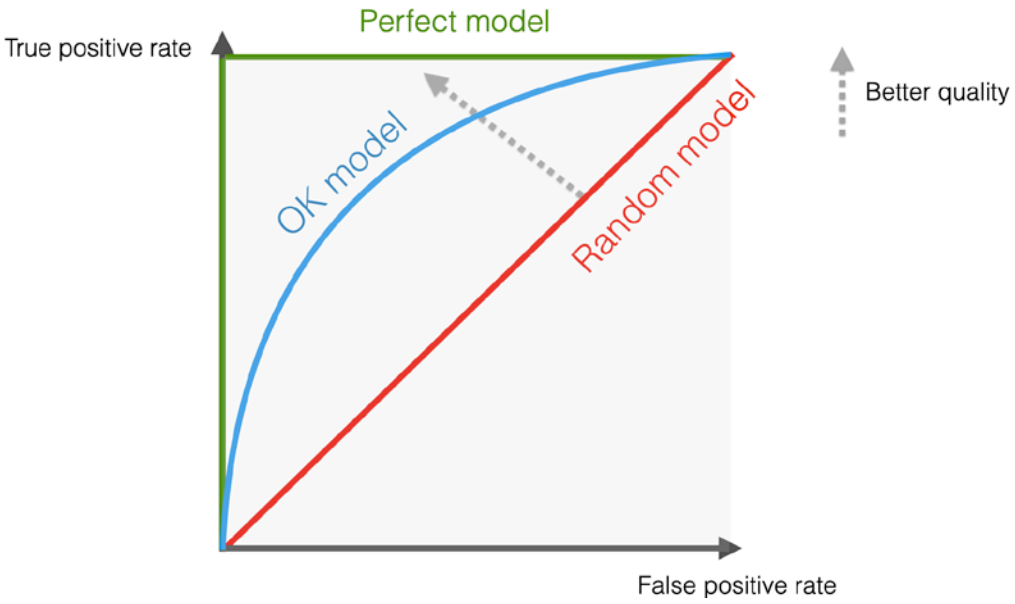
$$\boldsymbol{F1} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

ROC-AUC

Η καμπύλη ROC δείχνει τη σχέση ανάμεσα στην ευαισθησία (Recall ή TPR) και τα ψευδώς θετικά

ποσοστά (FPR) για διαφορετικά κατώφλια. Η περιοχή κάτω από την καμπύλη, γνωστή ως AUC (Area Under the Curve), συνοψίζει αυτή την πληροφορία σε μια ενιαία τιμή. Μια υψηλή τιμή AUC σημαίνει ότι το μοντέλο μπορεί να ξεχωρίσει αξιόπιστα τις θετικές από τις αρνητικές περιπτώσεις, ανεξάρτητα από το ποιο κατώφλι θα χρησιμοποιηθεί.

$$\boldsymbol{TPR} = \frac{TP}{TP+FN} \qquad \boldsymbol{FPR} = \frac{FP}{FP+TN} \qquad \boldsymbol{AUC} = \int_0^1 TPR(FPR)\ d(FPR)$$



Εικόνα 9 – Αξιολόγηση μοντέλου με τη χρήση της καμπύλης ROC [54]

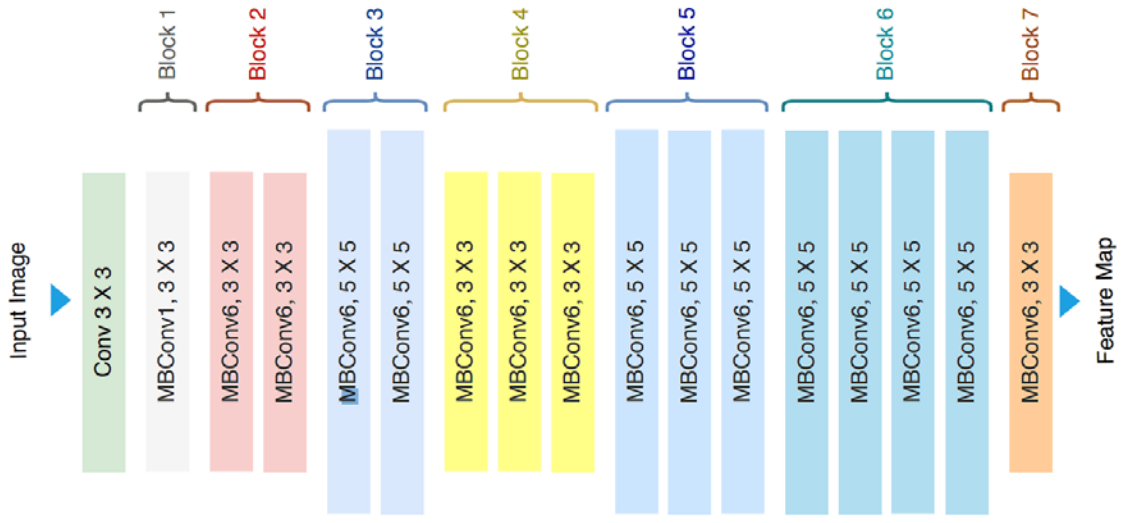
ΚΕΦΑΛΑΙΟ 5

5 EfficientNetBX

5.1 Η αρχιτεκτονική EfficientNetBx

Η οικογένεια EfficientNetBx [13] παρουσιάστηκε ως μια προσπάθεια να λυθεί ένα πρόβλημα που υπήρχε για χρόνια στα CNN. Όταν θέλουμε να βελτιώσουμε την απόδοση ενός μοντέλου, συνήθως έχουμε τρεις επιλογές. Να αυξήσουμε το βάθος του δικτύου, να το κάνουμε φαρδύτερο ή να δώσουμε μεγαλύτερη ανάλυση στις εικόνες εισόδου. Το ζήτημα είναι πως αυτές οι επιλογές δεν κλιμακώνονται σωστά όταν εφαρμόζονται ανεξάρτητα. Αν για παράδειγμα αυξήσουμε μόνο το βάθος, το μοντέλο μπορεί να γίνει πιο δύσκολο να εκπαιδευτεί. Αν αυξήσουμε μόνο το πλάτος, μπορεί να πνιγεί σε παραμέτρους χωρίς ουσιαστικό κέρδος. Αν δώσουμε μόνο μεγαλύτερη ανάλυση, θα απαιτεί υπερβολικούς υπολογιστικούς πόρους.

Η EfficientNet εισήγαγε μια πιο ισορροπημένη προσέγγιση. Χρησιμοποιεί αυτό που ονομάζεται compound scaling. Αντί να αλλάζει κάθε διάσταση χωριστά, εφαρμόζει έναν ενιαίο κανόνα κλιμάκωσης που αυξάνει ταυτόχρονα το πλάτος, το βάθος και την ανάλυση. Με αυτό τον τρόπο το μοντέλο μεγαλώνει με σωστές αναλογίες, αποφεύγοντας την υπερβολική αύξηση των παραμέτρων και διατηρώντας την απόδοση σε λογικά επίπεδα υπολογιστικού κόστους.



Εικόνα 10 – Αναπαράσταση αρχιτεκτονικής EfficientNet [13]

Η βασική δομή των EfficientNetBx στηρίζεται σε δύο κεντρικές ιδέες. Η πρώτη είναι η χρήση των MBConv blocks, που είχαν παρουσιαστεί αρχικά στο MobileNetV2 [55]. Τα MBConv (Mobile Inverted Bottleneck Convolution) blocks αποτελούν παραλλαγή του “inverted residual” σχήματος:

αρχικά επεκτείνουν τα κανάλια (expansion) με ένα 1×1 convolution, στη συνέχεια εφαρμόζουν depthwise convolution (π.χ. 3×3 ή 5×5) ανά κανάλι, και τέλος συμπιέζουν ξανά τη διάσταση με ένα δεύτερο 1×1 convolution (projection). Η ιδέα αυτή επιτρέπει στο δίκτυο να αποκτά πλούσιες αναπαραστάσεις με μικρότερο κόστος, ενώ συχνά χρησιμοποιείται και skip connection (όταν οι διαστάσεις ταιριάζουν) ώστε να διατηρείται η πληροφορία και να βελτιώνεται η σταθερότητα της εκπαίδευσης.

Αυτά τα blocks βασίζονται σε depthwise separable convolutions, τα οποία μειώνουν σημαντικά το πλήθος των υπολογισμών χωρίς να θυσιάζουν την ποιότητα των χαρακτηριστικών που εξάγονται. Πιο συγκεκριμένα, σε μια κλασική συνέλιξη, κάθε φίλτρο "βλέπει" όλα τα κανάλια εισόδου ταυτόχρονα, κάτι που είναι υπολογιστικά ακριβό. Αντίθετα, στο depthwise separable σχήμα η συνέλιξη "σπάει" σε δύο βήματα:

(α) depthwise convolution, όπου εφαρμόζεται ένα φίλτρο ανά κανάλι εισόδου (άρα καταγράφονται χωρικά χαρακτηριστικά ανά κανάλι), και

(β) pointwise convolution (1×1), που συνδυάζει τα κανάλια μεταξύ τους και δημιουργεί νέους γραμμικούς συνδυασμούς.

Έτσι, μεγάλο μέρος του κόστους της πλήρους συνέλιξης αποφεύγεται, ενώ παραμένει η ικανότητα εξαγωγής χωρικών και καναλικών προτύπων — κάτι ιδιαίτερα χρήσιμο σε εφαρμογές όπως η ιατρική απεικόνιση, όπου απαιτείται ισορροπία ανάμεσα σε ακρίβεια και ταχύτητα.

Η δεύτερη ιδέα είναι η ενσωμάτωση του μηχανισμού squeeze and excitation (SE) [56], ο οποίος επιτρέπει στο δίκτυο να «εστιάζει» σε πιο χρήσιμα κανάλια, βελτιώνοντας τη διακριτική του ικανότητα. Ο μηχανισμός αυτός λειτουργεί ως "attention" σε επίπεδο καναλιών: πρώτα γίνεται squeeze, δηλαδή συμπίεση της χωρικής πληροφορίας κάθε καναλιού σε μία τιμή μέσω global average pooling (ένα μέτρο της συνολικής ενεργοποίησης του καναλιού). Έπειτα ακολουθεί το excitation, όπου ένα μικρό νευρωνικό υποδίκτυο (συνήθως δύο πλήρως συνδεδεμένα στρώματα με μη γραμμικότητα) παράγει βάρη/συντελεστές w_cγια κάθε κανάλι. Τέλος, τα βάρη αυτά πολλαπλασιάζουν τα αντίστοιχα feature maps (recalibration), ενισχύοντας κανάλια που σχετίζονται με χρήσιμες δομές και μειώνοντας κανάλια που αντιστοιχούν σε λιγότερο σημαντικό "θόρυβο". Στην ενδοσκοπική απεικόνιση αυτό είναι πρακτικά κρίσιμο, επειδή ο πολύποδας μπορεί να εμφανίζεται με λεπτές χρωματικές/υφικές διαφορές σε σχέση με τον βλεννογόνο και το SE βοηθά το δίκτυο να δίνει μεγαλύτερη βαρύτητα σε κανάλια που κωδικοποιούν τέτοια διακριτικά μοτίβα.

Οι παραλλαγές EfficientNetB0 μέχρι EfficientNetB7 προκύπτουν από αυτή την ενιαία λογική κλιμάκωσης (compound scaling) [13]. Όσο ανεβαίνει ο δείκτης Bx, τόσο αυξάνεται η χωρητικότητα του μοντέλου με ελεγχόμενο τρόπο, καθώς κλιμακώνονται συνδυαστικά (και όχι ανεξάρτητα) το βάθος (περισσότερα blocks/στρώματα), το πλάτος (περισσότερα κανάλια ανά στρώμα) και η

ανάλυση εισόδου. Αυτό σημαίνει ότι τα μεγαλύτερα μοντέλα μπορούν να αποδώσουν καλύτερα, αλλά απαιτούν περισσότερη μνήμη και χρόνο εκπαίδευσης. Για πρακτικές εφαρμογές όπως η ιατρική απεικόνιση, συχνά γίνεται επιλογή ενός μεσαίου μοντέλου, όπως του B2 ή του B3, ώστε να υπάρχει καλή ισορροπία μεταξύ ταχύτητας και ακρίβειας.

Ένα σημαντικό χαρακτηριστικό των EfficientNetBx είναι ότι γενικά αποδίδουν πολύ καλά ακόμη και όταν εκπαιδεύονται με σχετικά περιορισμένα δεδομένα, ειδικά όταν χρησιμοποιείται μεταφορά μάθησης από μεγάλα σύνολα δεδομένων. Η αποδοτικότητά τους τα κάνει ελκυστικά για συστήματα που πρέπει να λειτουργούν σε πραγματικό χρόνο ή σε περιβάλλοντα με περιορισμένο υλικό, καθώς επιτυγχάνουν υψηλή ακρίβεια με μικρότερο υπολογιστικό κόστος σε σχέση με πιο “βαριές” ραχοκοκαλιές.

Στην πράξη, η σειρά EfficientNetBx θεωρείται μια από τις πιο επιτυχημένες προσπάθειες βελτιστοποίησης CNN. Συνδυάζει υψηλή ακρίβεια με σχετικά χαμηλό υπολογιστικό φορτίο και προσαρμόζεται εύκολα σε διαφορετικές εργασίες, από ταξινόμηση εικόνας μέχρι εντοπισμό και τμηματοποίηση, συμβάλλοντας στο να υιοθετηθεί ευρέως σε ερευνητικές και κλινικές εφαρμογές.

5.2 EfficientNetB3

Το EfficientNetB3 αποτελεί μέλος της οικογένειας EfficientNet, μιας σειράς μοντέλων που σχεδιάστηκαν με στόχο να επιτυγχάνουν υψηλή ακρίβεια με σημαντικά μικρότερο υπολογιστικό κόστος σε σχέση με προηγούμενες αρχιτεκτονικές. Η βασική αρχή που καθορίζει την απόδοση των EfficientNet είναι η ισορροπημένη κλιμάκωση του μοντέλου, όπου το πλάτος, το βάθος και η ανάλυση εισόδου αυξάνονται με έναν συντονισμένο τρόπο αντί να επεκτείνεται μόνο ένας από αυτούς τους άξονες, κάνοντας την αρχιτεκτονική ιδιαίτερα αποδοτική και σταθερή σε μεγαλύτερες εκδοχές της.

Βρίσκεται σε ένα ενδιάμεσο σημείο της οικογένειας, συνδυάζοντας αυξημένη ικανότητα μάθησης χωρίς να απαιτεί υπερβολικούς υπολογιστικούς πόρους. Σε σχέση με τα μικρότερα μοντέλα της σειράς, διαθέτει περισσότερα επίπεδα και ευρύτερα κανάλια, γεγονός που του επιτρέπει να εντοπίζει πιο σύνθετα μοτίβα σε εικόνες υψηλότερης ανάλυσης. Παράλληλα, παραμένει αρκετά ελαφρύ ώστε να είναι κατάλληλο για χρήση σε εφαρμογές όπου η ταχύτητα και η αποδοτικότητα έχουν μεγάλη σημασία.

Ένα βασικό χαρακτηριστικό των EfficientNet, συμπεριλαμβανομένου του B3, είναι η χρήση των Mobile Inverted Bottleneck Conv (MBConv) blocks. Τα blocks αυτά ενσωματώνουν μηχανισμούς όπως η επέκταση καναλιών, οι βάσεις βάθους (depthwise convolutions) και το squeeze and excitation, που βοηθούν το δίκτυο να δώσει μεγαλύτερη έμφαση σε σημαντικά χαρακτηριστικά και να αγνοήσει θορυβώδεις πληροφορίες, προσφέροντας καλύτερη εκμετάλλευση των παραμέτρων

και μεγαλύτερη αποδοτικότητα στη μάθηση. Στην πράξη, το EfficientNetB3 είναι ιδιαίτερα αποτελεσματικό σε εφαρμογές ταξινόμησης εικόνων όταν το μέγεθος των δεδομένων δεν είναι υπερβολικά μεγάλο αλλά απαιτείται υψηλή ακρίβεια, κάνοντάς το συχνή επιλογή σε ιατρικές εφαρμογές, όπου η ανάλυση εικόνων όπως MRI, ακτινογραφίες ή CT αναζητά λεπτές διαφοροποιήσεις που μπορεί να καθορίζουν τη διάγνωση. Επιπλέον, το B3 υποστηρίζει εισόδους μεγαλύτερων διαστάσεων, κάτι που αποδεικνύεται χρήσιμο σε περιπτώσεις όπου λεπτομέρειες μικρής κλίμακας παίζουν καθοριστικό ρόλο.

Στην παρούσα εργασία επιλέχθηκε η αρχιτεκτονική EfficientNetB3 ως η βασική δομή του μοντέλου μας. Ο λόγος είναι ότι προσφέρει μια ιδιαίτερα ισορροπημένη σχέση ανάμεσα στην υπολογιστική αποδοτικότητα και την ακρίβεια, κάτι που είναι κρίσιμο όταν δουλεύουμε με ιατρικές εικόνες όπου οι λεπτές διαφοροποιήσεις έχουν μεγάλη σημασία, με την ικανότητα του μοντέλου να αξιοποιεί αποτελεσματικά τα MBConv blocks και τον τρόπος με τον οποίο κλιμακώνεται μέσω του compound scaling την καθιστούν αξιόπιστη επιλογή για προβλήματα ταξινόμησης με υψηλές απαιτήσεις. Η χρήση της EfficientNetB3 στο δικό μας σύστημα επέτρεψε να εκπαιδευτεί ένα μοντέλο που αποδίδει καλά χωρίς υπερβολικό υπολογιστικό κόστος. Παράλληλα, η αρχιτεκτονική ανταποκρίθηκε ικανοποιητικά στις ιδιαιτερότητες του συνόλου δεδομένων, καταφέρνοντας να εντοπίσει σημαντικά χαρακτηριστικά που σχετίζονται με τις διαφορετικές κατηγορίες της νόσου, επιβεβαιώνοντας ότι το EfficientNetB3 αποτελεί κατάλληλη λύση για το συγκεκριμένο πρόβλημα και δικαιολογεί την επιλογή του ως τον πυρήνα του μοντέλου μας.

ΚΕΦΑΛΑΙΟ 6

6 Δεδομένα

Το dataset που χρησιμοποιήθηκε στην εργασία προέρχεται από το REAL-colon dataset [57], ένα δημόσιο, πλήρως ανωνυμοποιημένο σύνολο δεδομένων ενδοσκοπικών εξετάσεων. Περιλαμβάνει:

- Βίντεο πλήρους διάρκειας κολονοσκοπήσεων από 32 διαφορετικούς ασθενείς,
- Αναλυτικά annotations (σε μορφή JSON) που υποδεικνύουν την παρουσία ή απουσία πολυπόδων σε συγκεκριμένα frames,
- Υψηλής ποιότητας καρέ με annotations (frame-level labels).

Η χρήση αυτού του dataset το καθιστά ιδανικό για μοντέλα εντοπισμού, καθώς ενσωματώνει ρεαλιστικά σενάρια, θόρυβο, ασαφή περιβάλλοντα, αλλά και πολυμορφία σε ηλικία/φυσιολογία/συνθήκες ενδοσκόπησης.

Το σύνολο δεδομένων περιλαμβάνει περίπου 950.000 annotated frames, από τα οποία εξήχθησαν μόνο τα frames με ενεργά annotations πολυπόδων (positive samples) και ένα ποσοστό αρνητικών (χωρίς πολυπόδες) για ισορροπία.

Η διαδικασία διαχωρισμού έγινε ως εξής:

1. Ανά ασθενή: Έγινε grouping όλων των εικόνων και annotations ανά ασθενή, ώστε να διασφαλιστεί ότι ένας ασθενής ανήκει μόνο σε ένα set (training / validation / test), αποφεύγοντας έτσι information leakage.

2. Καταμέτρηση frames: Χρησιμοποιήθηκε το εργαλείο tag_counter.py για να μετρηθεί ο αριθμός των θετικών/αρνητικών frames ανά ασθενή.

3. MEGASPLIT: Το script MEGASPLIT.py ανέλαβε να χωρίσει τους ασθενείς στα sets με στόχο:
   - ~70% των frames για training,
   - ~20% για validation,
   - ~10% για testing.

4. Εξισορρόπηση: Σε περιπτώσεις υπερβολικής ανισορροπίας (δηλαδή πολύ περισσότερα αρνητικά καρέ χωρίς πολύποδα σε σχέση με θετικά καρέ με πολύποδα), εφαρμόστηκαν δύο συμπληρωματικές τεχνικές:

Undersampling (υποδειγματοληψία): Μειώθηκε ελεγχόμενα ο αριθμός των αρνητικών δειγμάτων, επιλέγοντας ένα αντιπροσωπευτικό υποσύνολο από τα καρέ “no_polyp”, ώστε το μοντέλο να μην «μαθαίνει» να προβλέπει συνεχώς την αρνητική κλάση λόγω υπεροχής της στο dataset. Με αυτόν τον τρόπο περιορίζεται η προκατάληψη (bias) υπέρ της αρνητικής κλάσης και βελτιώνεται η ικανότητα του μοντέλου να εντοπίζει πολυπόδες.

Augmentation (επαύξηση δεδομένων) στα θετικά δείγματα: Πραγματοποιήθηκε τεχνητή αύξηση των θετικών δειγμάτων μέσω μετασχηματισμών εικόνας (π.χ. μικρές περιστροφές, μετατοπίσεις, αλλαγές φωτεινότητας/αντίθεσης, οριζόντια αναστροφή όπου είναι κατάλληλο). Στόχος είναι να αυξηθεί η ποικιλία των θετικών παραδειγμάτων και να βελτιωθεί η γενίκευση του μοντέλου σε διαφορετικές συνθήκες λήψης (γωνία κάμερας, φωτισμός, θόρυβος), χωρίς να αλλοιώνεται η κλινικά χρήσιμη πληροφορία.

Η δομή που προέκυψε τελικά ήταν οργανωμένη σε φακέλους:

- dataset/train/polyp/, dataset/train/no_polyp/
- dataset/val/polyp/, dataset/val/no_polyp/
- dataset/test/polyp/, dataset/test/no_polyp/

Κάθε εικόνα προήλθε από καρέ βίντεο (frame extraction) μέσω custom parser που αξιοποιούσε τα .json αρχεία του Figshare dataset. Η διαδικασία συνοδεύτηκε από οπτικό έλεγχο για αφαίρεση εικόνων κακής ποιότητας (θολές, σκοτεινές, με artifacts).

Η παραπάνω διαδικασία εξασφάλισε ότι δεν υπάρχει bias ή leakage, διατηρώντας απόλυτα ανεξάρτητο test set.

6.1 Κατασκευή και οργάνωση του dataset

Η ποιότητα και η δομή του συνόλου δεδομένων αποτέλεσαν καθοριστικό παράγοντα για την επιτυχία του προτεινόμενου συστήματος. Η παρούσα εργασία βασίστηκε στο δημόσια διαθέσιμο σύνολο δεδομένων REAL-World Colonoscopy Dataset (REAL-colon) [57], το οποίο περιλαμβάνει μεγάλο αριθμό εικόνων και καρέ βίντεο από πραγματικές κολονοσκοπικές εξετάσεις. Το συγκεκριμένο dataset χαρακτηρίζεται από υψηλό ρεαλισμό, αλλά ταυτόχρονα παρουσιάζει σημαντικές προκλήσεις, καθώς είναι ανομοιογενές, έντονα ανισόρροπο και όχι έτοιμο για άμεση χρήση σε διαδικασίες εκπαίδευσης βαθιών νευρωνικών δικτύων.

Αρχικά, το dataset περιλάμβανε δεκάδες χιλιάδες καρέ, οργανωμένα ανά εξέταση και ασθενή, με μόνο ένα μικρό ποσοστό αυτών να περιέχει πολυπόδες. Η ετικετοποίηση παρεχόταν σε μορφή αρχείων XML, τα οποία περιλάμβαναν πληροφορίες για την παρουσία βλαβών μέσω συγκεκριμένων tags. Ωστόσο, η ποιότητα των annotations δεν ήταν πλήρως ομοιόμορφη, καθώς

υπήρχαν περιπτώσεις ελλιπών ή κατεστραμμένων αρχείων, γεγονός που κατέστησε απαραίτητη την ανάπτυξη αυτοματοποιημένων ελέγχων εγκυρότητας κατά τη διαδικασία φόρτωσης των δεδομένων.

Για τη μετατροπή του αρχικού υλικού σε ένα λειτουργικό σύνολο δεδομένων κατάλληλο για δυαδική ταξινόμηση, εφαρμόστηκε διαδικασία φιλτραρίσματος και αναδιοργάνωσης. Κάθε εικόνα ταξινομήθηκε σε μία από δύο κατηγορίες, POLYP ή NO_POLYP, με βάση την ύπαρξη ή μη σχολιασμένης βλάβης. Η διαδικασία αυτή δεν ήταν απλώς μηχανική, αλλά απαιτούσε προσεκτικές αποφάσεις ώστε να αποφευχθεί η εισαγωγή θορύβου ή εσφαλμένων ετικετών, ιδιαίτερα σε περιπτώσεις όπου οι βλάβες ήταν μικρές ή μερικώς ορατές.

Στη συνέχεια, το τελικό σύνολο δεδομένων χωρίστηκε σε υποσύνολα εκπαίδευσης, επικύρωσης και ελέγχου, ακολουθώντας αναλογίες που εξασφαλίζουν αφενός επαρκή ποσότητα δεδομένων για την εκπαίδευση και αφετέρου αξιόπιστη αξιολόγηση της γενίκευσης του μοντέλου. Ιδιαίτερη μέριμνα δόθηκε ώστε εικόνες από την ίδια εξέταση ή τον ίδιο ασθενή να μην εμφανίζονται σε διαφορετικά υποσύνολα, αποτρέποντας έτσι τη διαρροή πληροφορίας και την τεχνητή αύξηση της απόδοσης.

Η διαδικασία κατασκευής του dataset δεν αντιμετωπίστηκε ως ένα προκαταρκτικό βήμα, αλλά ως μια σημαντική φάση της συνολικής μεθοδολογίας. Οι επιλογές που έγιναν σε αυτό το στάδιο επηρέασαν άμεσα τη σταθερότητα της εκπαίδευσης, τη συμπεριφορά του μοντέλου και τη δυνατότητα μεταφοράς των αποτελεσμάτων σε ρεαλιστικά κλινικά σενάρια.

ΚΕΦΑΛΑΙΟ 7

7 Υλοποίηση και Αξιολόγηση του Μοντέλου

7.1 Δημιουργία και ανάπτυξη μοντέλου

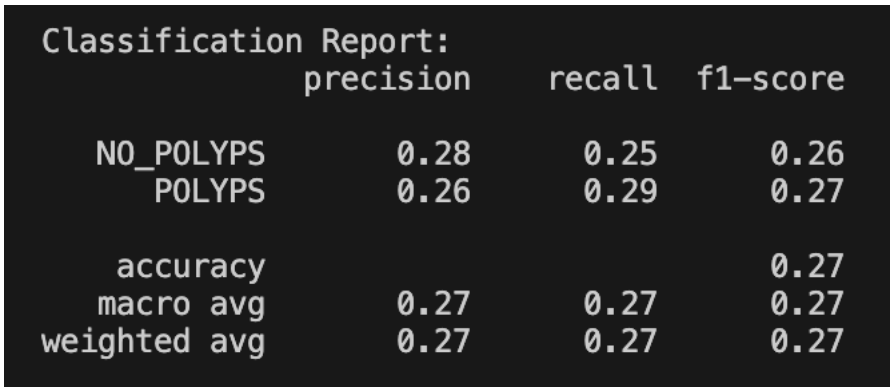
Η ανάπτυξη ενός αυτόματου συστήματος εντοπισμού πολυπόδων από ενδοσκοπικές εικόνες απαιτεί συνδυασμό επεξεργασίας δεδομένων, επιλογής κατάλληλης αρχιτεκτονικής και προσεκτικής αξιολόγησης του τελικού μοντέλου. Η πορεία της παρούσας εργασίας ξεκίνησε με πειραματικές προσεγγίσεις που αποδείχθηκαν ανεπαρκείς ή περιορισμένες και σταδιακά εξελίχθηκε μέσα από διαδοχικούς κύκλους αποτυχιών, βελτιώσεων και τελικών επιλογών. Για λόγους πληρότητας, στις επόμενες παραγράφους παρουσιάζονται συνοπτικά οι βασικές δοκιμές που πραγματοποιήθηκαν, οι λόγοι αποτυχίας τους και οι παρεμβάσεις που οδήγησαν στη σταθεροποίηση του pipeline.

Οι πρώτες απόπειρες περιλάμβαναν τη χρήση απλών συνελικτικών νευρωνικών δικτύων (CNN) μικρής χωρητικότητας, εκπαιδευμένα από την αρχή σε περιορισμένο αριθμό δεδομένων, χωρίς προεκπαιδευμένα βάρη. Συγκεκριμένα, δοκιμάστηκαν custom CNN αρχιτεκτονικές με 2–3 συνελικτικά επίπεδα (π.χ. Conv–ReLU–MaxPool) και στη συνέχεια 1–2 πλήρως συνδεδεμένα επίπεδα (Dense) για δυαδική ταξινόμηση (POLYP/NO_POLYP). Η επιλογή αυτής της προσέγγισης έγινε ως αρχικό baseline, ώστε να αξιολογηθεί αν το πρόβλημα μπορεί να προσεγγιστεί με απλούστερη αρχιτεκτονική. Ωστόσο, τα αποτελέσματα ήταν χαμηλά, καθώς το δίκτυο δεν μπορούσε να αποδώσει επαρκώς τις πολυπλοκότητες υφής και μορφολογίας των πολυπόδων, ειδικά υπό μεταβαλλόμενο φωτισμό, θόρυβο και τεχνουργήματα ενδοσκοπικής εικόνας.

Στα αρχικά πειράματα, παρατηρήθηκε έντονο overfitting, δηλαδή το μοντέλο προσαρμόστηκε υπερβολικά στα δείγματα εκπαίδευσης (training set), εμφανίζοντας πολύ υψηλή ακρίβεια κατά την εκπαίδευση, αλλά μειωμένη επίδοση σε νέα δεδομένα (validation set), όπου η ακρίβεια έπεσε κάτω από ~30% (Εικόνα 11, Εικόνα 12). Η συμπεριφορά αυτή είναι αναμενόμενη όταν η εκφραστική ικανότητα του μοντέλου δεν αντιστοιχεί στην ποσότητα/ποικιλία του διαθέσιμου training set ή όταν η προεπεξεργασία και η επαύξηση δεδομένων δεν επαρκούν για τη δημιουργία γενικεύσιμων αναπαραστάσεων.



Εικόνα 11 – Πρόβλημα overfitting στα αρχικά στάδια του μοντέλου (το μοντέλο έχει μάθει τα δεδομένα εκπαίδευσης και εμφανίζει accuracy: 100%)



Εικόνα 12 – Αποτελέσματα μοντέλου που παρουσίασε overfitting (ενώ το μοντέλο έχει μάθει τα δεδομένα εκπαίδευσης, στο validation set η ακρίβεια ήταν κάτω από 30%)

Η επόμενη φάση περιλάμβανε επανασχεδιασμό του pipeline με στόχο τη βελτίωση της γενίκευσης και τον περιορισμό της υπερπροσαρμογής. Πιο συγκεκριμένα:

- Εισήχθη transfer learning με χρήση προεκπαιδευμένων μοντέλων EfficientNetB0 και EfficientNetB3, όπου αξιοποιήθηκαν τα προεκπαιδευμένα βάρη ως μηχανισμός εξαγωγής χαρακτηριστικών και προσαρμόστηκαν τα τελικά στρώματα στη δυαδική ταξινόμηση. Στην αρχική φάση, τα πρώιμα επίπεδα παρέμειναν σταθερά και εκπαιδεύτηκαν κυρίως τα νέα τελικά επίπεδα. Στη συνέχεια εφαρμόστηκε, όπου ήταν απαραίτητο, προσαρμογή (fine-tuning) σε τμήμα των τελευταίων blocks.
- Χρησιμοποιήθηκαν τεχνικές data augmentation (flip, rotation, zoom, brightness), ώστε το μοντέλο να εκπαιδευτεί σε παραλλαγές των ίδιων σκηνών και να μειωθεί η εξάρτηση από συγκεκριμένη γωνία λήψης ή φωτισμό.
- Κανονικοποιήθηκαν τα inputs στο range [0,1], ώστε να σταθεροποιηθεί η διαδικασία εκπαίδευσης και να περιοριστούν αποκλίσεις λόγω διαφορετικών κατανομών εντάσεων φωτεινότητας.
- Αυξήθηκε ο αριθμός epochs με χρήση early stopping, δηλαδή με μηχανισμό αυτόματης διακοπής όταν η επίδοση στο validation set παύει να βελτιώνεται.

Πρακτικά, ορίζεται ένας αριθμός ανοχής (patience) και, αν το validation loss δεν μειώνεται για αντίστοιχο πλήθος εποχών, η εκπαίδευση τερματίζεται και διατηρούνται τα βάρη της καλύτερης επίδοσης.

Κώδικες όπως train_98%.py και train_downgraded.py έδειξαν ότι το μοντέλο μπορούσε να φτάσει πολύ υψηλή ακρίβεια στη φάση εκπαίδευσης (μέχρι ~98%), αλλά παρουσίασε αδυναμία στο generalization. Ο όρος generalization αναφέρεται στην ικανότητα του μοντέλου να διατηρεί υψηλή απόδοση σε δεδομένα που δεν έχουν χρησιμοποιηθεί κατά την εκπαίδευση (validation/test). Η μειωμένη γενίκευση υποδηλώνει ότι το μοντέλο έμαθε πρότυπα στενά συνδεδεμένα με το training set (π.χ. συγκεκριμένες φωτεινότητες, μορφές θορύβου ή background patterns), αντί για πιο γενικά χαρακτηριστικά που σχετίζονται με την παρουσία πολύποδα.

Επιπλέον, η υποβάθμιση της ανάλυσης (downgraded) οδήγησε σε απώλεια λεπτομερειών σημαντικών για τον εντοπισμό μικρών ή επίπεδων πολυπόδων. Η αναφορά σε «σταθερές δομές / stable pipeline» αφορά τη διατήρηση συνεπούς και αναπαραγώγιμης ροής εκπαίδευσης (ίδιες ρυθμίσεις, ίδιο preprocessing, ίδιος διαχωρισμός ανά ασθενή), ώστε τα αποτελέσματα να μην επηρεάζονται από τυχαίες μεταβολές ή διαφορετικές προετοιμασίες δεδομένων.

Στη συνέχεια, πραγματοποιήθηκαν δοκιμές με πιο συνεπείς δομές/pipelines όπως TRAIN_FINAL.py και train_save_DS.py, οι οποίες αξιοποιούσαν καλύτερη διαχείριση δεδομένων και χρήση callbacks. Ως callbacks ορίζονται μηχανισμοί που εκτελούνται κατά την εκπαίδευση και ρυθμίζουν τη διαδικασία δυναμικά (π.χ. early stopping, model checkpointing, learning rate scheduling). Για παράδειγμα, το monitor στο validation loss σημαίνει ότι παρακολουθείται η τιμή του validation loss και αποθηκεύονται αυτόματα τα βάρη του μοντέλου όταν επιτυγχάνεται η καλύτερη τιμή (model checkpoint). Τέλος, εφαρμόστηκε fine-tuning στα τελευταία layers/blocks του προεκπαιδευμένου δικτύου, δηλαδή επιτράπηκε η εκπαίδευση σε μέρος των τελευταίων επιπέδων με μικρό ρυθμό μάθησης, ώστε το μοντέλο να προσαρμοστεί στα ιδιαίτερα χαρακτηριστικά των ενδοσκοπικών εικόνων, διατηρώντας παράλληλα τα γενικά οπτικά χαρακτηριστικά που έχουν ήδη μάθει τα προεκπαιδευμένα βάρη.

Ωστόσο, ο θόρυβος και η ανομοιομορφία στα δεδομένα (διαφορετικές πηγές, διαφορετικές συνθήκες λήψης, μεταβαλλόμενος φωτισμός και artifacts) συνέχισαν να προκαλούν προβλήματα. Το ζήτημα αντιμετωπίστηκε με φίλτρα επιλογής εικόνων και χειροκίνητο καθαρισμό ετικετών (labels), ώστε να μειωθούν λανθασμένες ή αμφίβολες επισημάνσεις που επιβαρύνουν τη διαδικασία εποπτευόμενης μάθησης.

Το τελικό pipeline που υλοποιήθηκε στα αρχεία train_FINAL.py και test_FINAL.py οργανώθηκε ως εξής:

- Dataset: Ομογενοποίηση εικόνων σε 256×256 RGB
- Augmentation: Random rotation ±15°, horizontal flip, contrast adjustment
- Μοντέλο: Transfer learning με EfficientNetB3, fine-tuning στα 3 τελευταία layers/blocks
- Optimizer: Adam με learning rate 1e−4
- Loss: Binary Cross Entropy
- Metrics: Accuracy, Precision, Recall, F1-Score, ROC-AUC
- Callbacks: EarlyStopping (patience=5), ReduceLROnPlateau
- Training time: ~20 λεπτά σε GPU (Google Colab)
- Batch size: 32, Epochs: 50

Η υλοποίηση του τελικού pipeline έγινε σε περιβάλλον Python 3.10 με χρήση των βιβλιοθηκών tensorflow/keras, opencv-python, numpy και scikit-learn. Ο βασικός κώδικας train_FINAL.py περιλαμβάνει όλα τα στάδια εκπαίδευσης: από την ανάγνωση του dataset έως την αποθήκευση του τελικού μοντέλου. Το σύστημα αρχικά φορτώνει τις εικόνες και τις αντίστοιχες ετικέτες από οργανωμένους φακέλους (ανά κατηγορία) και εφαρμόζει μια σειρά από augmentations ώστε να αυξηθεί η ποικιλία και η ανθεκτικότητα του μοντέλου.

Το σύνολο των εικόνων χωρίστηκε σε training/validation/test με αναλογία 70/20/10. Ο κώδικας κάνει χρήση του ImageDataGenerator της Keras για real-time augmentation και streaming εικόνων στη GPU.

Το μοντέλο που επιλέχθηκε ήταν το EfficientNetB3. Η επιλογή έγινε μετά από σύγκριση με EfficientNetB0, καθώς και με απλά custom CNN μοντέλα που εκπαιδεύτηκαν από την αρχή χωρίς προεκπαιδευμένα βάρη. Η σύγκριση πραγματοποιήθηκε με κοινό πρωτόκολλο (ίδια δεδομένα, ίδιος διαχωρισμός training/validation/test, ίδιες βασικές ρυθμίσεις προεπεξεργασίας και ίδιες μετρικές αξιολόγησης), ώστε οι διαφορές να αποδίδονται κυρίως στην αρχιτεκτονική. Τα απλά CNN εμφάνισαν έντονη υπερπροσαρμογή και χαμηλή επίδοση στο validation set, ενώ οι προεκπαιδευμένες αρχιτεκτονικές είχαν σαφώς καλύτερη γενίκευση. Το EfficientNetB3 παρείχε την καλύτερη ισορροπία μεταξύ ακρίβειας, σταθερότητας εκπαίδευσης και χρόνου εκτέλεσης.

Οι 3 τελευταίες συνελικτικές βαθμίδες του EfficientNetB3 ορίστηκαν ως trainable ώστε να προσαρμοστούν στα χαρακτηριστικά των ενδοσκοπικών εικόνων. Το τελικό layer ήταν ένα πλήρως συνδεδεμένο επίπεδο (Dense) με 1 έξοδο και sigmoid activation για δυαδική ταξινόμηση (πολύποδας/μη πολύποδας). Η sigmoid μετατρέπει την έξοδο του δικτύου σε τιμή στο διάστημα

[0,1], η οποία ερμηνεύεται ως πιθανότητα παρουσίας πολύποδα. Το loss function ήταν binary cross-entropy, το οποίο για πραγματική ετικέτα $y \in \{0,1\}$ και πρόβλεψη $\hat{y} \in [0,1]$ ορίζεται ως:

$$L = -(y\log(\hat{y}) + (1-y)\log(1-\hat{y}))$$

Χρησιμοποιήθηκε Adam optimizer με learning rate 0.0001, ενώ η προσαρμογή του learning rate κατά την εκπαίδευση έγινε μέσω του ReduceLROnPlateau. Συγκεκριμένα, όταν το validation loss δεν βελτιώνεται για έναν αριθμό εποχών (patience), το learning rate μειώνεται με έναν συντελεστή (factor), ώστε η βελτιστοποίηση να συνεχιστεί με μικρότερα βήματα και να σταθεροποιηθεί η σύγκλιση.

Κατά την εκπαίδευση, εφαρμόστηκε EarlyStopping στο validation loss για αποφυγή υπερεκπαίδευσης. Η εκπαίδευση έτρεξε για έως και 50 εποχές, με τη διαδικασία να τερματίζει αυτόματα στην 18η εποχή λόγω στασιμότητας στο validation loss. Το τελικό accuracy στο training set έφτασε ~96%, ενώ στο validation set ~91%, γεγονός που δείχνει ικανοποιητική γενίκευση.

Τα αποτελέσματα καταγράφηκαν στο training_history.pkl και αναπαραστάθηκαν μέσω των training_history.png και training_metrics.png. Η καμπύλη loss και accuracy δείχνει ομαλή σύγκλιση, χωρίς απότομες αυξομειώσεις.

Η αξιολόγηση του τελικού μοντέλου έγινε με το script test_FINAL.py, το οποίο φορτώνει το αποθηκευμένο μοντέλο (model_weights.h5) και εφαρμόζει προβλέψεις στο test set. Το test set ήταν ανεξάρτητο από τα δεδομένα εκπαίδευσης και επικύρωσης.

Η ερμηνεία αυτών των αποτελεσμάτων δείχνει ότι το μοντέλο λειτουργεί αξιόπιστα ακόμη και σε περιπτώσεις με εικόνες μέτριας ποιότητας ή χαμηλής αντίθεσης.

Κατά τη διάρκεια του έργου, εντοπίστηκαν πολλά τεχνικά εμπόδια, που περιγράφονται παρακάτω:

- Overfitting: Έγινε έντονο με μικρό dataset και χωρίς augmentation
- Unbalanced dataset: Χρειάστηκε undersampling & augmentation της θετικής τάξης
- Grayscale vs RGB: Αφαίρεση του χρώματος μείωσε την απόδοση κατά 7%
- Μικρές εικόνες: Resize σε 128×128 έριξε την ακρίβεια κατά ~5%
- Βαθιά αρχιτεκτονική (ResNet50): Μεγαλύτερο training time, χωρίς αύξηση F1

Η λύση ήρθε μέσω συνδυασμού fine-tuning σε μοντέλο μεσαίας πολυπλοκότητας (EfficientNetB3), RGB είσοδο, αυξημένη επαύξηση δεδομένων και ισορροπημένη δειγματοληψία κατά τον σχηματισμό των batches. Ο όρος “balanced batches” αναφέρεται στη διαδικασία κατά την οποία κάθε batch περιέχει πιο συγκρίσιμη αναλογία θετικών/αρνητικών δειγμάτων, ώστε η

εκπαίδευση να μην επηρεάζεται δυσανάλογα από την πλειοψηφική κλάση. Ο όρος “aggressive augmentation” αναφέρεται στη χρήση πιο έντονων, αλλά ελεγχόμενων μετασχηματισμών (π.χ. περιστροφή/μετατόπιση/φωτεινότητα/αντίθεση/zoom) με στόχο την αύξηση της ποικιλίας των θετικών παραδειγμάτων και τη βελτίωση της γενίκευσης.

Επιπλέον, στο πλαίσιο της αξιολόγησης του μοντέλου, ενσωματώθηκε η τεχνική Explainable AI μέσω του αλγορίθμου Grad-CAM (Gradient-weighted Class Activation Mapping). Η τεχνική αυτή παράγει χάρτες ενεργοποίησης (activation maps), δηλαδή χωρικές απεικονίσεις που δείχνουν ποιες περιοχές της εικόνας συνέβαλαν περισσότερο στην τελική πρόβλεψη. Οι χάρτες προκύπτουν από το τελευταίο συνελικτικό επίπεδο του δικτύου, το οποίο διατηρεί χωρική πληροφορία και επιτρέπει την οπτική συσχέτιση της απόφασης με συγκεκριμένες περιοχές ενδιαφέροντος. Η χρήση του Grad-CAM κατά τη φάση του ελέγχου (testing) προσφέρει διαφάνεια, επιβεβαιώνοντας ότι το δίκτυο εστιάζει σε περιοχές συμβατές με την παρουσία πολυπόδων και όχι σε άσχετα στοιχεία του φόντου. Η αναλυτική παρουσίαση της μεθοδολογίας και παραδείγματα παρουσιάζονται στο αντίστοιχο κεφάλαιο Explainable AI.

Παράλληλα, παρατηρήθηκε ότι το μοντέλο είχε δυσκολία να αναγνωρίσει πολύ μικρούς ή επίπεδους πολύποδες, ειδικά σε σκοτεινές εικόνες ή με θόρυβο. Αυτή η αδυναμία θα μπορούσε να αντιμετωπιστεί με:

- Βαθύτερα μοντέλα (EfficientNetB3+)
- Custom attention layers
- Ανίχνευση μέσω segmentation αντί για classification

Η αποθήκευση του τελικού μοντέλου έγινε με αποθήκευση αρχιτεκτονικής (model_config.pkl) και των βαρών (model_weights.h5). Η αξιολόγηση μπορεί να επαναληφθεί σε οποιοδήποτε σύστημα με Python και TensorFlow. Η πλήρης διαδικασία training & evaluation τεκμηριώνεται στους κώδικες train_FINAL.py, test_FINAL.py και στο τεχνικό παράρτημα του έργου.

Ολόκληρο το μοντέλο μπορεί να μετατραπεί σε TFLite ή να ενσωματωθεί σε πλατφόρμα inference σε πραγματικό χρόνο. Υπάρχει δυνατότητα σύνδεσης με σύστημα κολονοσκόπησης για real-time annotation, με χρήση OpenCV & GUI overlay.

Η εμπειρία της ανάπτυξης αυτού του pipeline κατέδειξε τη σημασία της επαναληπτικής βελτίωσης και του πειραματισμού. Οι πρώτες προσπάθειες, παρόλο που ήταν αναποτελεσματικές, παρείχαν πολύτιμες ενδείξεις για την επιλογή αρχιτεκτονικής, preprocessing και ροής δεδομένων.

Η τελική απόδοση του συστήματος κατατάσσεται εντός των ορίων προηγούμενων μελετών, με F1-score κοντά στο 91% και AUC > 0.94, γεγονός που το καθιστά κατάλληλο για ενσωμάτωση σε σύγχρονες πλατφόρμες υποβοηθούμενης διάγνωσης.

7.2 Διαχείριση ανισορροπίας και στρατηγικές δειγματοληψίας

Η ανισορροπία κλάσεων αποτελεί ένα από τα πιο μεγάλα προβλήματα στην αυτόματη ανίχνευση πολυπόδων, καθώς σε πραγματικές κολονοσκοπικές εξετάσεις τα καρέ χωρίς παθολογικά ευρήματα υπερτερούν αριθμητικά κατά πολύ εκείνων που περιέχουν πολύποδες. Το συγκεκριμένο χαρακτηριστικό αντικατοπτρίζει τη φυσική κατανομή του προβλήματος στον κλινικό χώρο, ωστόσο δημιουργεί σοβαρές προκλήσεις κατά την εκπαίδευση μοντέλων βαθιάς μάθησης, τα οποία τείνουν να ευνοούν την πλειοψηφική κλάση εάν δεν ληφθούν κατάλληλα μέτρα.

Στο παρόν σύστημα, η ανισορροπία αντιμετωπίστηκε συνδυαστικά και όχι με μία μόνο τεχνική. Αρχικά, εφαρμόστηκε δυναμικός υπολογισμός βαρών κλάσεων με βάση την πραγματική κατανομή των δεδομένων εκπαίδευσης. Συγκεκριμένα, τα βάρη υπολογίστηκαν αυτόματα μέσω της συνάρτησης compute_class_weight της βιβλιοθήκης scikit-learn, εξασφαλίζοντας ότι οι απώλειες που σχετίζονται με την κλάση των πολυπόδων τιμωρούνται περισσότερο κατά τη διαδικασία βελτιστοποίησης. Με αυτόν τον τρόπο, το μοντέλο ωθήθηκε να δώσει μεγαλύτερη προσοχή στα θετικά δείγματα, χωρίς να απαιτείται τεχνητή εξισορρόπηση του dataset.

Παράλληλα, υιοθετήθηκε στρατηγική έμμεσης εξισορρόπησης μέσω επιθετικής επαύξησης δεδομένων αποκλειστικά στο σύνολο εκπαίδευσης. Η επαύξηση λειτουργεί ως μορφή υπερδειγματοληψίας της μειοψηφικής κλάσης, καθώς τα δείγματα πολυπόδων παράγουν περισσότερες οπτικά διαφοροποιημένες εκδοχές, αυξάνοντας την ποικιλία των θετικών παραδειγμάτων που «βλέπει» το μοντέλο κατά την εκπαίδευση. Αυτή η προσέγγιση προτιμήθηκε έναντι της απλής αντιγραφής δειγμάτων, καθώς μειώνει τον κίνδυνο υπερπροσαρμογής.

Όσον αφορά τη δειγματοληψία, η φόρτωση των δεδομένων πραγματοποιήθηκε με τυχαία αναδιάταξη (shuffle) σε κάθε εποχή εκπαίδευσης, διασφαλίζοντας ότι το μοντέλο δεν εκτίθεται σε σταθερά μοτίβα ακολουθίας. Αντίθετα, στο σύνολο επικύρωσης και ελέγχου δεν εφαρμόστηκε καμία στρατηγική εξισορρόπησης ή επαύξησης, ώστε η αξιολόγηση να αντανακλά ρεαλιστικές συνθήκες κλινικής χρήσης και όχι τεχνητά βελτιωμένες επιδόσεις.

Η επιλογή να μη χρησιμοποιηθούν τεχνικές υποδειγματοληψίας που αλλοιώνουν τη φυσική κατανομή των δεδομένων, όπως η έντονη υποδειγματοληψία της πλειοψηφικής κλάσης, ήταν συνειδητή. Αν και τέτοιες μέθοδοι μπορούν να βελτιώσουν βραχυπρόθεσμα ορισμένους δείκτες απόδοσης, συχνά οδηγούν σε μοντέλα με μειωμένη γενίκευση και χαμηλότερη κλινική αξιοπιστία. Αντίθετα, η παρούσα προσέγγιση διατηρεί τη στατιστική δομή του προβλήματος, ενώ ταυτόχρονα ελέγχει τη μεροληψία του μοντέλου.

7.3 Προεπεξεργασία και επαύξηση δεδομένων

Η προεπεξεργασία των δεδομένων αποτελεί κρίσιμο στάδιο στην ανάπτυξη συστημάτων βαθιάς μάθησης για ιατρική απεικόνιση, καθώς επηρεάζει άμεσα τη σταθερότητα της εκπαίδευσης, την ταχύτητα σύγκλισης και τη δυνατότητα γενίκευσης του μοντέλου. Στην περίπτωση της κολονοσκόπησης, οι εικόνες παρουσιάζουν υψηλή ενδογενή μεταβλητότητα ως προς τον φωτισμό, την κλίμακα, την υφή του βλεννογόνου και την παρουσία οπτικών παρεμβολών, γεγονός που καθιστά την προσεκτική προετοιμασία των δεδομένων απαραίτητη.

Ως πρώτο βήμα, όλες οι εικόνες αναπροσαρμόστηκαν σε κοινή ανάλυση 300×300 εικονοστοιχείων, ώστε να είναι συμβατές με την είσοδο της αρχιτεκτονικής EfficientNetB3. Η επιλογή ενιαίας ανάλυσης διασφαλίζει ομοιομορφία στην επεξεργασία και αποτρέπει την εισαγωγή παραμορφώσεων που θα μπορούσαν να επηρεάσουν τη χωρική πληροφορία των εικόνων. Δεν εφαρμόστηκε περαιτέρω αποκοπή (cropping), προκειμένου να διατηρηθεί ολόκληρο το οπτικό πεδίο της κολονοσκόπησης, το οποίο συχνά περιέχει έμμεσες ενδείξεις παθολογίας.

Η κανονικοποίηση των εικόνων πραγματοποιήθηκε έμμεσα μέσω της συνάρτησης preprocess_input της EfficientNet, η οποία μετασχηματίζει τις τιμές των εικονοστοιχείων ώστε να ευθυγραμμίζονται με την κατανομή των δεδομένων που χρησιμοποιήθηκαν κατά την αρχική εκπαίδευση του μοντέλου στο σύνολο δεδομένων. Η επιλογή αυτή εξασφαλίζει συμβατότητα με τα προεκπαιδευμένα φίλτρα του δικτύου και συμβάλλει στη σταθερή και αποδοτική σύγκλιση κατά τη φάση της εκπαίδευσης.

Η επαύξηση δεδομένων εφαρμόστηκε αποκλειστικά στο σύνολο εκπαίδευσης και σχεδιάστηκε ώστε να προσομοιώνει ρεαλιστικές μεταβολές που εμφανίζονται σε πραγματικές κολονοσκοπικές συνθήκες. Συγκεκριμένα, χρησιμοποιήθηκαν τυχαίες περιστροφές, μετατοπίσεις κατά πλάτος και ύψος, διατμήσεις, αλλαγές κλίμακας και μεταβολές φωτεινότητας, καθώς και οριζόντιες και κατακόρυφες αναστροφές. Οι μετασχηματισμοί αυτοί επιλέχθηκαν με στόχο να ενισχύσουν την ανθεκτικότητα του μοντέλου σε αλλαγές προσανατολισμού και φωτισμού, χωρίς να αλλοιώνουν τη βασική μορφολογία των πολυπόδων.

Ιδιαίτερη προσοχή δόθηκε στη διατήρηση των χρωματικών χαρακτηριστικών των εικόνων, καθώς η χρωματική αντίθεση μεταξύ πολύποδα και φυσιολογικού βλεννογόνου αποτελεί σημαντική πηγή πληροφορίας για το μοντέλο. Για τον λόγο αυτό, αποφεύχθηκαν επιθετικοί μετασχηματισμοί που θα μπορούσαν να αλλοιώσουν μη ρεαλιστικά τα χρώματα ή την υφή των ιστών. Η επαύξηση χρησιμοποιήθηκε ως μέσο ενίσχυσης της ποικιλίας των δεδομένων και όχι ως εργαλείο τεχνητής παραμόρφωσης.

Αντίθετα, στο σύνολο επικύρωσης δεν εφαρμόστηκε καμία μορφή επαύξησης, πέραν της κανονικοποίησης. Η επιλογή αυτή διασφαλίζει ότι η αξιολόγηση της απόδοσης του μοντέλου πραγματοποιείται σε δεδομένα που αντικατοπτρίζουν όσο το δυνατόν πιστότερα τις πραγματικές

κλινικές συνθήκες, αποφεύγοντας αισιόδοξες εκτιμήσεις απόδοσης λόγω τεχνητών μετασχηματισμών.

7.4 Επιλογή και εξέλιξη αρχιτεκτονικής

Η επιλογή της αρχιτεκτονικής του νευρωνικού δικτύου αποτέλεσε μια σταδιακή και διερευνητική διαδικασία, η οποία βασίστηκε τόσο στη βιβλιογραφία όσο και σε εμπειρικά αποτελέσματα από πειραματισμούς. Ο στόχος δεν ήταν απλώς η επίτευξη της μέγιστης δυνατής ακρίβειας σε ένα συγκεκριμένο σύνολο δεδομένων, αλλά η ανάπτυξη ενός μοντέλου με καλή ικανότητα γενίκευσης, σταθερή συμπεριφορά κατά την εκπαίδευση και ρεαλιστικές απαιτήσεις υπολογιστικών πόρων.

Στα αρχικά στάδια εξετάστηκαν κλασικές αρχιτεκτονικές συνελικτικών νευρωνικών δικτύων, όπως το VGG και το ResNet, οι οποίες αποτελούν καθιερωμένα σημεία αναφοράς στη βιβλιογραφία της ιατρικής απεικόνισης. Αν και τα μοντέλα αυτά παρουσίασαν ικανοποιητική απόδοση σε ελεγχόμενα πειράματα, παρατηρήθηκε ότι είτε απαιτούσαν σημαντικούς υπολογιστικούς πόρους είτε εμφάνιζαν τάσεις υπερπροσαρμογής, ιδιαίτερα όταν εκπαιδεύονταν σε περιορισμένα ή ανομοιογενή σύνολα δεδομένων.

Η ανάγκη για καλύτερη ισορροπία μεταξύ ακρίβειας και αποδοτικότητας οδήγησε στη διερεύνηση νεότερων αρχιτεκτονικών που έχουν σχεδιαστεί με γνώμονα τη βελτιστοποίηση αυτής της σχέσης. Στο πλαίσιο αυτό, η οικογένεια EfficientNet αποτέλεσε ιδιαίτερα ελκυστική επιλογή, καθώς εισάγει τη φιλοσοφία της σύνθετης κλιμάκωσης, όπου το βάθος, το πλάτος και η ανάλυση του δικτύου αυξάνονται με ελεγχόμενο και συστηματικό τρόπο. Η προσέγγιση αυτή επιτρέπει την επίτευξη υψηλής απόδοσης χωρίς δυσανάλογη αύξηση της υπολογιστικής πολυπλοκότητας.

Μεταξύ των διαφορετικών παραλλαγών της EfficientNet, η EfficientNetB3 επιλέχθηκε ως η πλέον κατάλληλη για τη συγκεκριμένη εφαρμογή. Η επιλογή αυτή βασίστηκε στο γεγονός ότι προσφέρει ένα ευνοϊκό σημείο ισορροπίας μεταξύ εκφραστικής ικανότητας και υπολογιστικού κόστους, καθιστώντας την κατάλληλη τόσο για εκπαίδευση σε περιορισμένο περιβάλλον όσο και για πιθανή μελλοντική ενσωμάτωση σε κλινικά συστήματα πραγματικού χρόνου. Σε σύγκριση με μικρότερες εκδόσεις, όπως η B0 ή η B1, η B3 παρουσίασε σαφώς βελτιωμένη σταθερότητα και απόδοση, ενώ παράλληλα απέφυγε το σημαντικό κόστος των μεγαλύτερων εκδόσεων.

Η αρχιτεκτονική δεν παρέμεινε στατική καθ’ όλη τη διάρκεια της μελέτης. Πραγματοποιήθηκαν πειραματισμοί με διαφορετικά επίπεδα fine-tuning, ξεκινώντας από την εκπαίδευση μόνο των τελικών πλήρως συνδεδεμένων στρωμάτων και προχωρώντας σταδιακά στην αποδέσμευση βαθύτερων συνελικτικών επιπέδων. Η σταδιακή αυτή προσέγγιση επέτρεψε στο μοντέλο να προσαρμόσει υψηλού επιπέδου χαρακτηριστικά στις ιδιαιτερότητες των δεδομένων

κολονοσκόπησης, χωρίς να καταστρέψει τη γενική οπτική γνώση που έχει αποκτηθεί από την προεκπαίδευση στο ImageNet.

Η εξέλιξη της αρχιτεκτονικής συνοδεύτηκε από συνεχή παρακολούθηση της συμπεριφοράς του μοντέλου ως προς τη σύγκλιση, τη γενίκευση και την ευαισθησία σε υπερπροσαρμογή. Οι τελικές επιλογές δεν προέκυψαν τυχαία, αλλά ως αποτέλεσμα επαναληπτικής αξιολόγησης και τεκμηριωμένων τεχνικών αποφάσεων. Με τον τρόπο αυτό, η αρχιτεκτονική που χρησιμοποιήθηκε στο τελικό σύστημα αντικατοπτρίζει μια ώριμη σχεδιαστική διαδικασία, προσαρμοσμένη στις απαιτήσεις του προβλήματος και στις πρακτικές συνθήκες εφαρμογής.

7.5 Στρατηγική εκπαίδευσης και fine-tuning

Η στρατηγική εκπαίδευσης του μοντέλου σχεδιάστηκε με στόχο τη σταδιακή προσαρμογή της αρχιτεκτονικής στα δεδομένα της εφαρμογής, αποφεύγοντας απότομες αλλαγές που θα μπορούσαν να οδηγήσουν σε αστάθεια ή υπερπροσαρμογή. Δεδομένου ότι το δίκτυο βασίστηκε σε προεκπαιδευμένα βάρη από το ImageNet, η διαδικασία εκπαίδευσης υιοθέτησε προσέγγιση μεταφοράς μάθησης, η οποία αποδείχθηκε κρίσιμη λόγω του περιορισμένου μεγέθους και της ιδιαιτερότητας των διαθέσιμων ιατρικών δεδομένων.

Στο αρχικό στάδιο, τα συνελικτικά επίπεδα της EfficientNetB3 διατηρήθηκαν παγωμένα και εκπαιδεύτηκαν μόνο τα τελικά στρώματα ταξινόμησης. Η επιλογή αυτή επέτρεψε στο μοντέλο να μάθει τις βασικές κατανομές των κλάσεων χωρίς να αλλοιώσει τα γενικά οπτικά χαρακτηριστικά χαμηλού και μεσαίου επιπέδου που έχουν ήδη αποδειχθεί αποτελεσματικά σε μεγάλο εύρος προβλημάτων όρασης υπολογιστών. Κατά τη φάση αυτή, παρατηρήθηκε ταχεία σύγκλιση της απώλειας εκπαίδευσης και σχετικά σταθερή συμπεριφορά στο σύνολο επικύρωσης, γεγονός που επιβεβαίωσε την καταλληλότητα της αρχικής ρύθμισης.

Στη συνέχεια ακολούθησε σταδιακό fine-tuning, με αποδέσμευση επιλεγμένων βαθύτερων συνελικτικών επιπέδων. Η διαδικασία αυτή πραγματοποιήθηκε προσεκτικά, ξεκινώντας από τα ανώτερα επίπεδα του δικτύου και προχωρώντας προς τα βαθύτερα μόνο όταν η απόδοση στο σύνολο επικύρωσης είχε σταθεροποιηθεί. Για τα αποδεσμευμένα στρώματα χρησιμοποιήθηκε χαμηλότερος ρυθμός μάθησης σε σχέση με τα τελικά στρώματα, ώστε να επιτραπεί η ήπια προσαρμογή των χαρακτηριστικών χωρίς καταστροφή της ήδη αποκτημένης αναπαράστασης.

Ιδιαίτερη σημασία δόθηκε στην επιλογή και στη δυναμική προσαρμογή του ρυθμού μάθησης. Χρησιμοποιήθηκαν στρατηγικές μείωσης του learning rate όταν η απώλεια επικύρωσης παρουσίαζε στασιμότητα, συμβάλλοντας στην αποφυγή ταλαντώσεων και στη λεπτομερέστερη σύγκλιση του μοντέλου. Παράλληλα, εφαρμόστηκαν τεχνικές κανονικοποίησης, όπως dropout και

weight decay, με σκοπό τον περιορισμό της υπερπροσαρμογής, ιδιαίτερα κατά τα τελευταία στάδια του fine-tuning.

Η διαδικασία εκπαίδευσης συνοδεύτηκε από συστηματική παρακολούθηση δεικτών απόδοσης, όπως η απώλεια, η ευαισθησία και το F1-score στο σύνολο επικύρωσης. Η χρήση early stopping επέτρεψε τη διακοπή της εκπαίδευσης όταν δεν παρατηρούνταν περαιτέρω βελτιώσεις, αποτρέποντας την άσκοπη προσαρμογή στο σύνολο εκπαίδευσης. Οι αποφάσεις σχετικά με τη διάρκεια και το βάθος του fine-tuning δεν ελήφθησαν εκ των προτέρων, αλλά προέκυψαν από την παρατήρηση της συμπεριφοράς του μοντέλου σε διαδοχικά πειράματα.

7.6 Πειραματισμοί, αποτυχίες και σταθεροποίηση του συστήματος

Η τελική μορφή του συστήματος δεν προέκυψε ως αποτέλεσμα μιας μοναδικής επιτυχημένης ρύθμισης, αλλά ως προϊόν επαναληπτικών πειραματισμών, αποτυχιών και σταδιακής βελτιστοποίησης. Η διαδικασία αυτή υπήρξε καθοριστική για την κατανόηση των ορίων της αρχιτεκτονικής, της συμπεριφοράς των δεδομένων και της αλληλεπίδρασής τους κατά την εκπαίδευση.

Στα αρχικά πειράματα παρατηρήθηκε έντονη αστάθεια στη σύγκλιση, ιδιαίτερα όταν επιχειρήθηκε πλήρης εκπαίδευση του δικτύου από νωρίς ή όταν χρησιμοποιήθηκαν υψηλοί ρυθμοί μάθησης. Σε αρκετές περιπτώσεις, το μοντέλο εμφάνισε απότομη μείωση της απώλειας εκπαίδευσης χωρίς αντίστοιχη βελτίωση στο σύνολο επικύρωσης, υποδεικνύοντας φαινόμενα υπερπροσαρμογής. Τα αποτελέσματα αυτά ανέδειξαν την ανάγκη για πιο συντηρητική στρατηγική fine-tuning και αυστηρότερο έλεγχο της γενίκευσης.

Επιπλέον, δοκιμάστηκαν εναλλακτικές ρυθμίσεις επαύξησης δεδομένων και στρατηγικές αντιμετώπισης της ανισορροπίας τάξεων. Σε ορισμένες περιπτώσεις, η υπερβολικά επιθετική επαύξηση οδήγησε σε μείωση της σταθερότητας του μοντέλου, αλλοιώνοντας κρίσιμα οπτικά χαρακτηριστικά των πολυπόδων. Αντίστοιχα, πειράματα με έντονη υπερδειγματοληψία της μειονοτικής κλάσης βελτίωσαν μεν την ευαισθησία, αλλά αύξησαν σημαντικά τον αριθμό ψευδώς θετικών προβλέψεων. Οι παρατηρήσεις αυτές οδήγησαν σε πιο ισορροπημένες επιλογές, με στόχο τη διατήρηση αποδεκτής ειδικότητας χωρίς θυσία της ανάκλησης.

Κατά τη διάρκεια των πειραματισμών εξετάστηκαν επίσης διαφορετικά βάθη fine-tuning, αριθμοί εποχών και σχήματα προσαρμογής του ρυθμού μάθησης. Ορισμένες ρυθμίσεις παρουσίασαν καλή απόδοση σε μεμονωμένες δοκιμές αλλά απέτυχαν να γενικεύσουν σε επαναλαμβανόμενα πειράματα, γεγονός που κατέστησε σαφές ότι η αξιολόγηση έπρεπε να βασίζεται στη συνέπεια και όχι σε μεμονωμένες υψηλές επιδόσεις. Η σταθεροποίηση του

συστήματος επιτεύχθηκε μόνο όταν επιλέχθηκαν ρυθμίσεις που παρήγαγαν επαναλήψιμα αποτελέσματα σε διαφορετικούς αρχικούς σπόρους και υποσύνολα δεδομένων.

Η τελική σταθερότητα του μοντέλου επιβεβαιώθηκε μέσω παρατεταμένης εκπαίδευσης με early stopping και συνεχούς παρακολούθησης των δεικτών απόδοσης. Η απώλεια εκπαίδευσης και επικύρωσης παρουσίασε ομαλή σύγκλιση χωρίς έντονες αποκλίσεις, ενώ οι μετρικές όπως το F1-score και η ευαισθησία παρέμειναν εντός στενού εύρους τιμών σε διαδοχικά runs. Το γεγονός αυτό υποδηλώνει ότι το σύστημα δεν βασίζεται σε τυχαίες συγκυρίες, αλλά σε μια ώριμη και ισορροπημένη ρύθμιση.

7.7 Τελική ρύθμιση και αναπαραγωγιμότητα

Μετά την ολοκλήρωση των πειραματισμών και τη σταθεροποίηση της συμπεριφοράς του μοντέλου, ορίστηκε μια τελική ρύθμιση εκπαίδευσης η οποία εξισορροπεί την απόδοση, τη γενίκευση και τη σταθερότητα. Η επιλογή αυτή δεν βασίστηκε αποκλειστικά στη μέγιστη τιμή μιας μετρικής, αλλά στη συνολική και επαναλήψιμη συμπεριφορά του συστήματος υπό παρόμοιες συνθήκες εκπαίδευσης και αξιολόγησης.

Η τελική αρχιτεκτονική, βασισμένη στο EfficientNetB3 με μεταφορά μάθησης, εκπαιδεύτηκε με προσεκτικά επιλεγμένες υπερπαραμέτρους, συμπεριλαμβανομένου χαμηλού αρχικού ρυθμού μάθησης και σταδιακής αποδέσμευσης των βαθύτερων στρωμάτων. Με τον όρο «σταδιακή αποδέσμευση» εννοείται ότι, σε αρχικό στάδιο, διατηρούνται σταθερά τα περισσότερα προεκπαιδευμένα στρώματα ώστε να αξιοποιηθούν τα γενικά χαρακτηριστικά που έχουν ήδη μάθει, ενώ στα επόμενα στάδια επιτρέπεται η εκπαίδευση (fine-tuning) σε μέρος των τελευταίων blocks/στρωμάτων με μικρό ρυθμό μάθησης. Η στρατηγική αυτή μειώνει τον κίνδυνο υπερπροσαρμογής και βοηθά το μοντέλο να προσαρμοστεί στα χαρακτηριστικά των ενδοσκοπικών εικόνων χωρίς να επηρεάζεται αρνητικά η σταθερότητα της σύγκλισης. Η χρήση κανονικοποίησης, early stopping και προσαρμοστικών χρονοδιαγραμμάτων ρυθμού μάθησης συνέβαλε στην ομαλή σύγκλιση και στη μείωση των φαινομένων υπερπροσαρμογής. Η κανονικοποίηση αναφέρεται τόσο στην κλιμάκωση των τιμών των εικόνων εισόδου σε συγκεκριμένο εύρος (π.χ. [0,1]) όσο και στη χρήση τεχνικών που σταθεροποιούν τη ροή των ενεργοποιήσεων κατά την εκπαίδευση, όπου αυτό εφαρμόστηκε. Παράλληλα, η τελική στρατηγική επαύξησης και δειγματοληψίας παρέμεινε σταθερή σε όλα τα runs, ώστε να διασφαλιστεί συγκρίσιμη συμπεριφορά του μοντέλου.

Ιδιαίτερη έμφαση δόθηκε στην αναπαραγωγιμότητα των αποτελεσμάτων, η οποία αποτελεί κρίσιμο παράγοντα σε εφαρμογές ιατρικής τεχνητής νοημοσύνης. Για τον σκοπό αυτό, σταθεροποιήθηκαν οι αρχικοί τυχαίοι σπόροι (random seeds) σε όλα τα στάδια της διαδικασίας, συμπεριλαμβανομένης της αρχικοποίησης των βαρών, της δειγματοληψίας των δεδομένων και των επαυξήσεων. Επιπλέον, η εκπαίδευση πραγματοποιήθηκε με σαφώς καθορισμένες εκδόσεις

λογισμικού και βιβλιοθηκών, περιορίζοντας την πιθανότητα διαφορών λόγω υποκείμενων αλλαγών στο περιβάλλον εκτέλεσης.

Η αξιολόγηση της τελικής ρύθμισης πραγματοποιήθηκε μέσω πολλαπλών ανεξάρτητων runs, με τα αποτελέσματα να παρουσιάζουν μικρές διακυμάνσεις στις βασικές μετρικές απόδοσης, όπως η ακρίβεια, το F1-score και η ευαισθησία. Η συνέπεια αυτή υποδηλώνει ότι το μοντέλο δεν εξαρτάται από ευνοϊκές τυχαίες αρχικοποιήσεις ή από συγκεκριμένες κατανομές δεδομένων, αλλά εμφανίζει σταθερή γενίκευση εντός των ορίων του διαθέσιμου συνόλου δεδομένων.

Τέλος, η τεκμηρίωση όλων των τεχνικών επιλογών και των τελικών υπερπαραμέτρων, επιτρέπει την επανάληψη της διαδικασίας από ανεξάρτητους ερευνητές, ενισχύοντας τη διαφάνεια και την αξιοπιστία της μελέτης. Η έμφαση στην αναπαραγωγιμότητα δεν αντιμετωπίζεται ως τυπική απαίτηση, αλλά ως ουσιαστικό στοιχείο που ενισχύει τη δυνητική μεταφορά του συστήματος σε πραγματικές κλινικές εφαρμογές και μελλοντικές επεκτάσεις.

7.8 Περιγραφή υλοποίησης και παραπομπή στον κώδικα

Η υλοποίηση του προτεινόμενου συστήματος πραγματοποιήθηκε σε περιβάλλον Python 3.10 με αξιοποίηση των βιβλιοθηκών TensorFlow/Keras, NumPy, OpenCV και scikit-learn. Ο κώδικας οργανώθηκε σε δύο βασικές συνιστώσες: (α) τον κώδικα εκπαίδευσης, ο οποίος καλύπτει τη φόρτωση/προεπεξεργασία δεδομένων, την εφαρμογή επαύξησης, τον ορισμό και την εκπαίδευση του μοντέλου και την αποθήκευση των παραγόμενων αρχείων, και (β) τον κώδικα αξιολόγησης, ο οποίος φορτώνει το εκπαιδευμένο μοντέλο και υπολογίζει τις μετρικές απόδοσης στο ανεξάρτητο σύνολο δοκιμής.

Για λόγους πληρότητας, αναπαραγωγιμότητας και αποσυμφόρησης του κύριου κειμένου, ο πλήρης κώδικας παρατίθεται στο Παράρτημα Α: Κώδικας Εκπαίδευσης Μοντέλου και στο Παράρτημα Β: Κώδικας Αξιολόγησης Μοντέλου. Στο παρόν κεφάλαιο παραμένουν οι τελικές ρυθμίσεις του pipeline, οι βασικές επιλογές υλοποίησης και τα αποτελέσματα που προκύπτουν από την εκπαίδευση και την αξιολόγηση.

7.8.1 Σύνοψη κώδικα εκπαίδευσης μοντέλου

Ο κώδικας εκπαίδευσης περιλαμβάνει: (i) ανάγνωση των εικόνων και ετικετών από τη δομή φακέλων του dataset, (ii) κανονικοποίηση/ομογενοποίηση εισόδου, (iii) επαύξηση δεδομένων σε πραγματικό χρόνο, (iv) ορισμό του μοντέλου με μεταφορά μάθησης, (v) ορισμό loss, optimizer και callbacks, (vi) εκτέλεση της εκπαίδευσης με παρακολούθηση του validation loss, και (vii) αποθήκευση των βαρών, του ιστορικού εκπαίδευσης και των σχετικών διαγραμμάτων.

7.8.2 Σύνοψη κώδικα αξιολόγησης μοντέλου

Ο κώδικας αξιολόγησης περιλαμβάνει: (i) φόρτωση του αποθηκευμένου μοντέλου/βαρών, (ii) εκτέλεση προβλέψεων στο test set, (iii) υπολογισμό μετρικών απόδοσης (π.χ. accuracy, precision, recall, F1-score, ROC-AUC), και (iv) παραγωγή συνοδευτικών παραστάσεων/αναφορών (π.χ. confusion matrix και ROC curve), οι οποίες χρησιμοποιούνται στην παρουσίαση των αποτελεσμάτων.

7.9 Αποτελέσματα

Στην ενότητα αυτή παρουσιάζονται τα αποτελέσματα του τελικού συστήματος στο ανεξάρτητο σύνολο δοκιμής (test set), καθώς και συγκριτικά αποτελέσματα από τις βασικές αρχιτεκτονικές που εξετάστηκαν κατά τη διάρκεια της εργασίας. Η παρουσίαση οργανώνεται σε (i) συνοπτική καταγραφή των κύριων μετρικών απόδοσης, (ii) συγκριτική αποτίμηση εναλλακτικών αρχιτεκτονικών με κοινό πρωτόκολλο αξιολόγησης και (iii) οπτικοποίηση της απόδοσης μέσω πίνακα σύγχυσης, καμπύλης ROC και ενδεικτικών προβλέψεων.

7.9.1 Συνοπτικά αποτελέσματα τελικού μοντέλου

Τα αποτελέσματα του τελικού μοντέλου (EfficientNetB3 με μεταφορά μάθησης και fine-tuning) στο test set συνοψίζονται στον Πίνακα 7.1.

Μετρική	Τιμή
Accuracy	91.4%
Precision	89.6%
Recall (Sensitivity)	92.3%
F1-score	90.9%
ROC-AUC	0.945

Πίνακας 1- Αποτελέσματα τελικού μοντέλου στο test set

Οι παραπάνω τιμές αποθηκεύτηκαν στο αρχείο test_metrics.json. Η απόδοση εμφανίζεται συνεπής μεταξύ των φάσεων εκπαίδευσης και ελέγχου, γεγονός που υποστηρίζει ικανοποιητική γενίκευση στο ανεξάρτητο σύνολο δοκιμής.

7.9.2 Συγκριτικά αποτελέσματα αρχιτεκτονικών

Δεδομένου ότι βασικός στόχος της εργασίας είναι η συγκριτική αξιολόγηση διαφορετικών αρχιτεκτονικών, παρατίθεται συγκριτικός πίνακας αποτελεσμάτων για τις κύριες προσεγγίσεις που δοκιμάστηκαν. Για να διασφαλιστεί η συγκρισιμότητα, οι αρχιτεκτονικές αξιολογήθηκαν με κοινό πρωτόκολλο (ίδια διάσπαση δεδομένων, κοινή προεπεξεργασία/κανονικοποίηση και ίδιες μετρικές

στο test set). Η τελική επιλογή αιτιολογείται από τη συνολική ισορροπία μεταξύ μετρικών απόδοσης, σταθερότητας εκπαίδευσης και υπολογιστικού κόστους.

Αρχιτεκτονική	Accuracy	Precision	Recall (Sensitivity)	F1-score	ROC-AUC
Custom CNN	77.1%	73.2%	86.4%	78.9%	0.812
EfficientNetB0	83.4%	78.1%	92.0%	84.3%	0.867
EfficientNetB3	91.4%	89.6%	92.3%	90.9%	0.945

Πίνακας 2 - Συγκριτική αξιολόγηση αρχιτεκτονικών στο test set

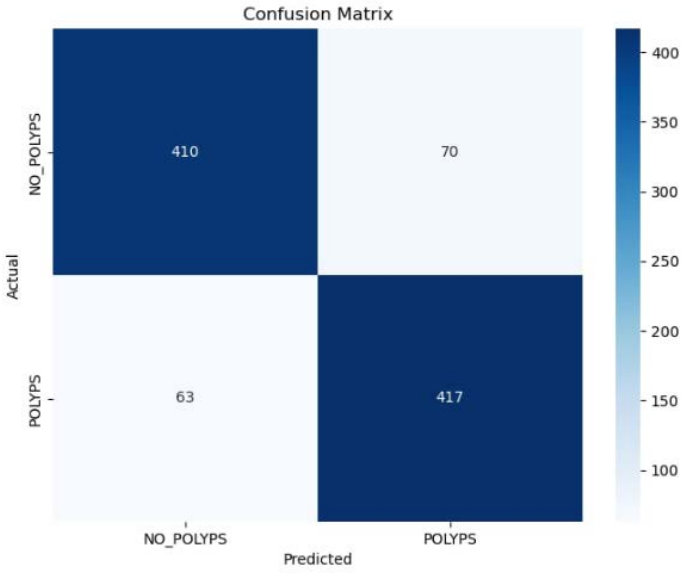
Η συγκριτική αποτίμηση των αρχιτεκτονικών αναδεικνύει ότι η επιλογή του τελικού μοντέλου δεν προέκυψε μόνο από τη μεγιστοποίηση μίας μεμονωμένης μετρικής, αλλά από τη συνολική συμπεριφορά του συστήματος υπό κοινές συνθήκες εκπαίδευσης και αξιολόγησης. Τα απλά custom CNN μοντέλα, παρότι αποτελούν χρήσιμο baseline, εμφανίζουν συνήθως αυξημένη ευαισθησία σε υπερπροσαρμογή και μεγαλύτερη μεταβλητότητα από run σε run, ιδιαίτερα όταν το διαθέσιμο σύνολο δεδομένων είναι περιορισμένο ή παρουσιάζει ανομοιομορφία. Αντιθέτως, οι προεκπαιδευμένες ραχοκοκαλιές (π.χ. EfficientNetB0) αξιοποιούν γενικά οπτικά χαρακτηριστικά, βελτιώνοντας τη γενίκευση και τη σταθερότητα της εκπαίδευσης, με διαφορετικούς όμως συμβιβασμούς ως προς την πολυπλοκότητα και τον υπολογιστικό φόρτο. Στο πλαίσιο της παρούσας εργασίας, το EfficientNetB3 επέδειξε την πιο ισορροπημένη συμπεριφορά ως προς precision και recall, σε συνδυασμό με υψηλή διακριτική ικανότητα (ROC-AUC), γεγονός που το καθιστά κατάλληλο για εφαρμογές υποβοηθούμενης διάγνωσης όπου η αξιόπιστη ανίχνευση πολυπόδων και η μείωση των ψευδώς αρνητικών περιπτώσεων αποτελούν πρωτεύοντες στόχους.

7.9.3 Οπτικοποίηση αποτελεσμάτων

Η οπτικοποίηση της απόδοσης πραγματοποιήθηκε με τη χρήση των παρακάτω παραστάσεων, οι οποίες επιτρέπουν λεπτομερέστερη αποτίμηση της συμπεριφοράς του μοντέλου.

- Confusion Matrix

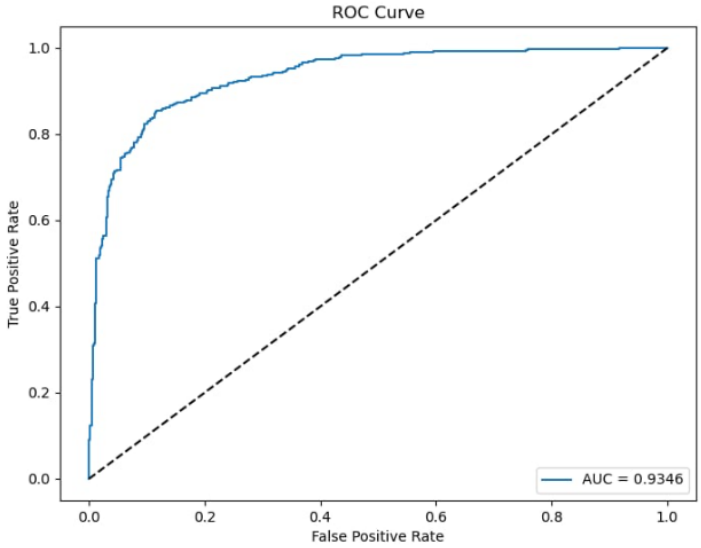
Η απεικόνιση μέσω πίνακα σύγχυσης παρουσιάζει την κατανομή των σωστών και λανθασμένων προβλέψεων στο test set, επιτρέποντας την ποσοτικοποίηση των ψευδώς θετικών και ψευδώς αρνητικών περιπτώσεων.



Εικόνα 13 – Αποτελέσματα Confusion Matrix

- ROC Curve

Η καμπύλη ROC απεικονίζει τον ρυθμό αληθώς θετικών (TPR) ως προς τον ρυθμό ψευδώς θετικών (FPR) για διαφορετικά κατώφλια απόφασης. Η επιφάνεια κάτω από την καμπύλη (AUC) χρησιμοποιείται ως συνοπτικός δείκτης διακριτικής ικανότητας, ανεξάρτητα από συγκεκριμένο κατώφλι.

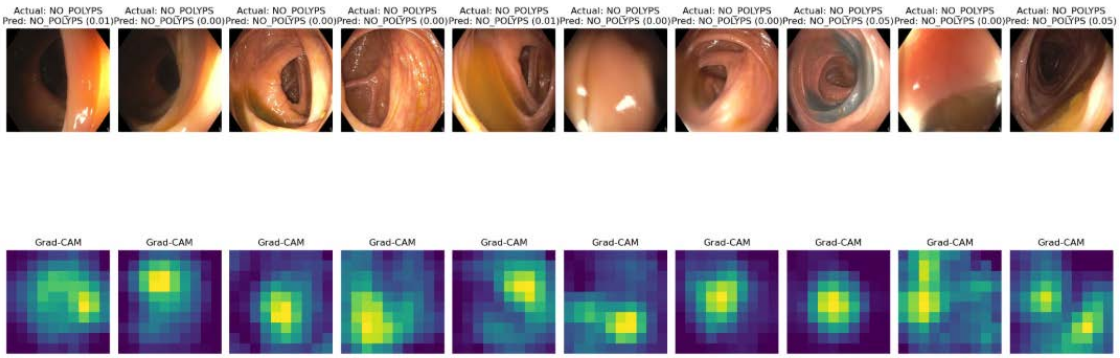


Εικόνα 14 – Αποτελέσματα ROC Curve

- Sample Predictions

Παρουσιάζονται ενδεικτικά παραδείγματα σωστών και λανθασμένων προβλέψεων από το test set, ώστε να αποτυπωθούν περιπτώσεις όπου το μοντέλο αποδίδει αξιόπιστα, αλλά και

περιπτώσεις μεγαλύτερης δυσκολίας (π.χ. χαμηλή αντίθεση, θόρυβος, μικρές ή επίπεδες αλλοιώσεις).



Εικόνα 15 – Αποτελέσματα Sample Predictions

Η συνολική ερμηνεία των αποτελεσμάτων δείχνει ότι το τελικό μοντέλο διατηρεί υψηλή απόδοση σε ρεαλιστικές ενδοσκοπικές συνθήκες, παρουσιάζοντας ικανοποιητική διακριτική ικανότητα και σταθερή συμπεριφορά ως προς τις βασικές μετρικές αξιολόγησης.

ΚΕΦΑΛΑΙΟ 8

8 eXplainable AI

Τα τελευταία χρόνια, η συζήτηση γύρω από την τεχνητή νοημοσύνη δεν περιορίζεται μόνο στην απόδοση των μοντέλων. Όσο οι αλγόριθμοι χρησιμοποιούνται σε κρίσιμες εφαρμογές, όπως η ιατρική διάγνωση, η ασφάλεια ή η λήψη αποφάσεων με επιπτώσεις στην ανθρώπινη ζωή, η ανάγκη να καταλάβουμε πώς ακριβώς "σκέφτονται" γίνεται όλο και πιο σημαντική. Εδώ έρχεται ο ρόλος του Explainable AI. Με τον όρο αυτό αναφερόμαστε στις μεθόδους και τα εργαλεία που επιτρέπουν στον χρήστη να κατανοήσει, να ερμηνεύσει και να εμπιστευτεί τις αποφάσεις ενός μοντέλου.

Στην πράξη, η ερμηνευσιμότητα έχει δύο βασικούς στόχους. Ο πρώτος είναι να δώσει στον τελικό χρήστη μια ξεκάθαρη εικόνα για το γιατί το σύστημα κατέληξε σε μια συγκεκριμένη πρόβλεψη. Ο δεύτερος είναι να βοηθήσει τους ίδιους τους ερευνητές να διαγνώσουν λάθη, να εντοπίσουν μεροληψίες ή να βελτιώσουν την εκπαίδευση του μοντέλου. Στη ιατρική, αυτό το στοιχείο είναι ιδιαίτερα κρίσιμο επειδή κανένας κλινικός δεν μπορεί να βασιστεί σε ένα "μαύρο κουτί" χωρίς να έχει βασική κατανόηση των αιτιών πίσω από τις αποφάσεις του.

Οι προσεγγίσεις στο Explainable AI χωρίζονται συνήθως σε δύο κατηγορίες. Η πρώτη περιλαμβάνει μοντέλα που είναι εγγενώς ερμηνεύσιμα, όπως οι γραμμικές παλινδρομήσεις ή τα decision trees, όπου η λογική της πρόβλεψης είναι εμφανής από τη δομή τους. Η δεύτερη, που είναι και η πιο σχετική με τα σύγχρονα νευρωνικά δίκτυα, αφορά τεχνικές που εφαρμόζονται πάνω σε ήδη πολύπλοκα μοντέλα για να τα κάνουν πιο κατανοητά.

Ένα από τα πιο γνωστά εργαλεία είναι τα saliency maps. Αυτές οι τεχνικές δείχνουν ποια σημεία της εικόνας επηρέασαν περισσότερο την τελική πρόβλεψη. Στην ιατρική απεικόνιση, αυτό μπορεί να βοηθήσει έναν γιατρό να δει αν το δίκτυο εστίασε πράγματι στην ύποπτη περιοχή ή "μπερδεύτηκε" από κάποιο άσχετο στοιχείο. Σχετικά με αυτό, το Grad-CAM [58] είναι ίσως η πιο διαδεδομένη μέθοδος. Παράγει χάρτες θερμότητας που επισημαίνουν οπτικά τις περιοχές που συνέβαλαν περισσότερο στην απόφαση του μοντέλου.

Μια άλλη κατηγορία αφορά τεχνικές όπως οι LIME [59] και SHAP [60]. Αυτές οι μέθοδοι δημιουργούν τοπικές εξηγήσεις, αναλύοντας την απόφαση για ένα συγκεκριμένο δείγμα και δείχνοντας ποια χαρακτηριστικά λειτούργησαν υπέρ ή κατά της τελικής πρόβλεψης. Αν και σχεδιάστηκαν αρχικά για δομημένα δεδομένα, έχουν επεκταθεί και στην επεξεργασία εικόνας με παραλλαγές που "απενεργοποιούν" μικρά τμήματα της εικόνας για να δουν πώς αλλάζει το αποτέλεσμα.

Τέλος, υπάρχει και η ευρύτερη συζήτηση γύρω από το fairness και τη μεροληψία. Η ερμηνευσιμότητα δεν βοηθά μόνο στο να εξηγούμε τις προβλέψεις αλλά και στο να ελέγχουμε αν τα μοντέλα λειτουργούν με τρόπο δίκαιο και συνεπή σε διαφορετικές ομάδες ή συνθήκες. Στα ιατρικά συστήματα αυτό είναι απαραίτητο, καθώς διαφορετικά μπορεί να δημιουργηθούν ανισότητες στις διαγνώσεις ή στα αποτελέσματα.

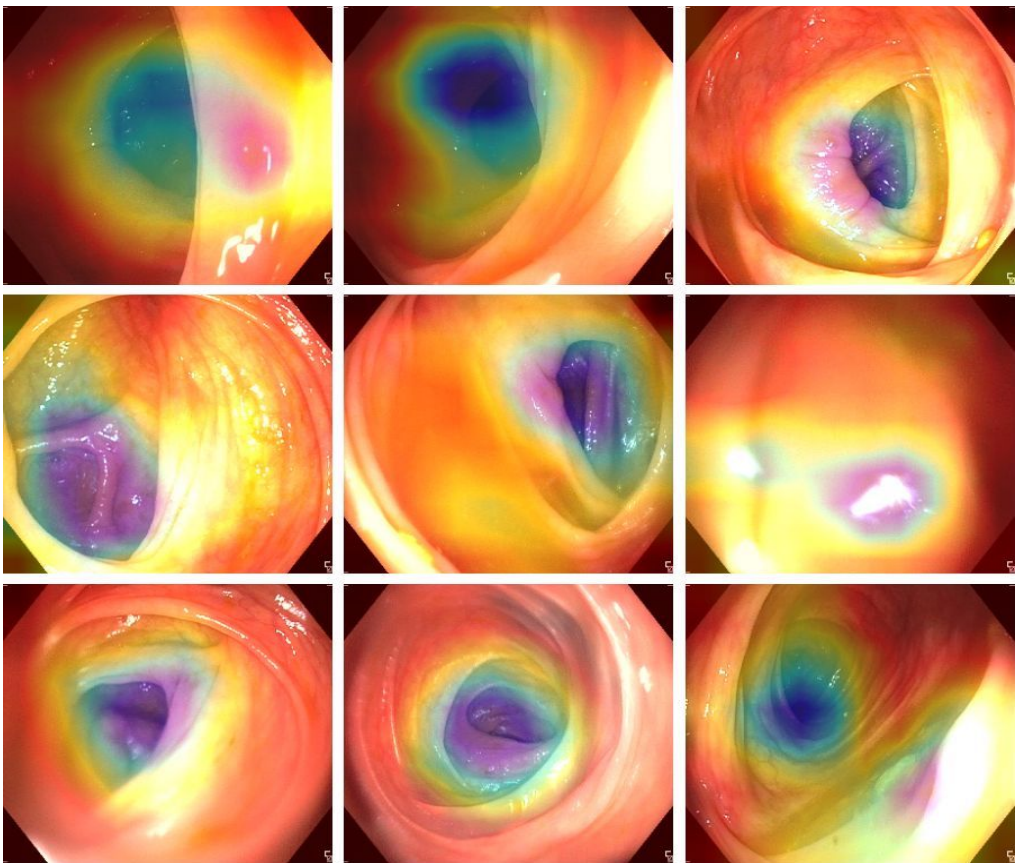
8.1 Εφαρμογή Grad-CAM στο προτεινόμενο σύστημα

Στην παρούσα εργασία, η ερμηνευσιμότητα του προτεινόμενου μοντέλου διερευνήθηκε πρακτικά μέσω της ενσωμάτωσης της τεχνικής Grad-CAM στον κώδικα ελέγχου. Για εικόνες του συνόλου δοκιμής (test set) υπολογίστηκαν χάρτες ενεργοποίησης, οι οποίοι υπερτέθηκαν στην αρχική εικόνα υπό τη μορφή θερμικών χαρτών (heatmaps), με σκοπό την ποιοτική αξιολόγηση των χωρικών περιοχών που συνέβαλαν περισσότερο στην τελική κατηγοριοποίηση.

Για την παρουσίαση που ακολουθεί επιλέχθηκαν ενδεικτικές περιπτώσεις εικόνων που ταξινομήθηκαν ορθά ως «πολύποδας» (true positives), ώστε να αξιολογηθεί η χωρική συμφωνία των θερμικών χαρτών με μορφολογικά στοιχεία της βλάβης. Στη χρωματική κλίμακα των heatmaps, οι θερμές αποχρώσεις (κόκκινο/κίτρινο) αντιστοιχούν σε υψηλότερη ένταση ενεργοποίησης, ενώ οι ψυχρές αποχρώσεις (πράσινο/μπλε) σε χαμηλότερη. Η αξιολόγηση εστιάζει στο κατά πόσο οι περιοχές μέγιστης ενεργοποίησης εντοπίζονται σε κλινικά συναφείς δομές (όρια/επιφάνεια πολύποδα και γειτονικός βλεννογόνος με αλλοίωση) και όχι σε περιοχές χωρίς διαγνωστική συνάφεια, όπως έντονες ανακλάσεις φωτισμού, άκρα του οπτικού πεδίου ή τμήματα του αυλού.

Όπως παρατηρείται στα παραδείγματα που ακολουθούν, οι θερμικές ζώνες υψηλής έντασης συγκεντρώνονται κατά κύριο λόγο στην περιοχή της ύποπτης αλλοίωσης, γεγονός που συνάδει με την αναμενόμενη εκμετάλλευση χωρικών και υφικών χαρακτηριστικών για τη συγκεκριμένη εργασία ταξινόμησης. Στις παρακάτω εικόνες παρουσιάζονται ενδεικτικά αποτελέσματα της εφαρμογής του Grad-CAM σε πραγματικά δεδομένα της μελέτης.

Η ανάλυση των παρακάτω αποτελεσμάτων αποδεικνύει ότι το μοντέλο δεν λειτουργεί ως "μαύρο κουτί", αλλά βασίζει τις αποφάσεις του σε κλινικά αποδεκτά χαρακτηριστικά.



Εικόνα 16 – Αναπαράσταση αποτελεσμάτων Grad-CAM ως Heatmaps.

ΚΕΦΑΛΑΙΟ 9

9 Συμπεράσματα και Μελλοντικές Προτάσεις

9.1 Συμπεράσματα

Το κεφάλαιο αυτό προσφέρει μια σύνοψη της έρευνας, αναλύοντας τα κύρια συμπεράσματα, το μεθοδολογικό πλαίσιο και τα κίνητρα της, με στόχος της μελέτης να είναι να χρησιμοποιηθούν σύγχρονες τεχνικές βαθιάς μάθησης για τη βελτίωση της ακρίβειας της ανίχνευσης πολυπόδων σε εικόνες κολονοσκόπησης, μια επίμονη πρόκληση. Η πρόληψη του καρκίνου του παχέος εντέρου απαιτεί την έγκαιρη και ακριβή ανίχνευση των πολυπόδων του παχέος εντέρου- ωστόσο, η διαγνωστική ακρίβεια εξακολουθεί να περιορίζεται από τις διαφορές στην οπτική εμφάνιση, τις συνθήκες απεικόνισης και την κόπωση του χειριστή. Στόχος του έργου ήταν η ανάπτυξη ενός αξιόπιστου και υπολογιστικά αποτελεσματικού μοντέλου που θα μπορούσε να υποστηρίξει τους γιατρούς σε πραγματικές συνθήκες διαλογής, χρησιμοποιώντας μια στρατηγική μεταφορά μάθησης που βασίζεται στην αρχιτεκτονική EfficientNetB3.

Τα βασικά στοιχεία του έργου αυτού συνοψίζονται στις ενότητες που ακολουθούν, οι οποίες περιλαμβάνουν την προετοιμασία και την προεπεξεργασία δεδομένων, τον σχεδιασμό, την εκπαίδευση και την αξιολόγηση του μοντέλου, καθώς και μια συζήτηση των αποδεδειγμένων πλεονεκτημάτων και μειονεκτημάτων του συστήματος, με το κεφάλαιο να αναλύει επίσης τις κύριες κλινικές και τεχνικές συνεισφορές της μελέτης και τον τρόπο με τον οποίο αυτές προάγουν τον τομέα της ανίχνευσης πολυπόδων με τη βοήθεια υπολογιστή. Τέλος, παρέχει καταληκτικές σκέψεις σχετικά με τις ευρύτερες προεκτάσεις της παρούσας μελέτης και περιγράφει την πιθανή διαδρομή προς την εφαρμογή της σε πραγματικό κόσμο σε κλινικά περιβάλλοντα, αναδεικνύοντας τους τρόπους με τους οποίους η διαγνωστική βοήθεια με βάση την τεχνητή νοημοσύνη μπορεί να βελτιώσει τα αποτελέσματα των ασθενών, την έγκαιρη πρόληψη του καρκίνου και την ακρίβεια του ελέγχου.

9.2 Περίληψη της εργασίας

Η επείγουσα ανάγκη βελτίωσης της ακρίβειας και της αξιοπιστίας της αυτοματοποιημένης ανίχνευσης πολυπόδων σε εικόνες κολονοσκόπησης αποτέλεσε την κινητήρια δύναμη για την παρούσα έρευνα, με μια από τις πιο διαδεδομένες και αποφεύξιμες αιτίες θανάτων που σχετίζονται με τον καρκίνο παγκοσμίως να είναι ο καρκίνος του παχέος εντέρου και η επιτυχής παρέμβαση να εξαρτάται από την έγκαιρη ανίχνευση προκαρκινικών αλλοιώσεων. Ωστόσο, επειδή οι μικροί ή επίπεδοι πολύποδες παραλείπονται εύκολα, η χειροκίνητη εξέταση των εικόνων κολονοσκόπησης

απαιτεί χρόνο και είναι επιρρεπής σε ανθρώπινα λάθη, απαιτώντας ένα σύστημα ανίχνευσης με τη βοήθεια υπολογιστή που μπορεί να βοηθήσει τους κλινικούς ιατρούς αυξάνοντας τη διαγνωστική συνέπεια και μειώνοντας την εποπτεία.

Για να επιλυθεί αυτό, χρησιμοποιήθηκε μια στρατηγική μεταφοράς μάθησης που βασίζεται στην αρχιτεκτονική του συνελικτικού νευρωνικού δικτύου EfficientNetB3 για τη δημιουργία ενός συστήματος που βασίζεται σε βαθιά μάθηση, όπου για την εκπαίδευση και την επικύρωση χρησιμοποιήθηκε ένα προσεκτικά επιλεγμένο σύνολο δεδομένων από εικόνες κολονοσκόπησης. Για να βελτιωθεί η γενίκευση και η ευρωστία, το σύνολο δεδομένων υποβλήθηκε σε μεγάλη προεπεξεργασία χρησιμοποιώντας τεχνικές όπως η κανονικοποίηση, η αύξηση και η εξισορρόπηση κλάσεων, με την αρχιτεκτονική να βελτιστοποιεί την απόδοση για δυαδική ταξινόμηση - διαφοροποίηση μεταξύ πολυποδικών και μη πολυποδικών δειγμάτων- συνδυάζοντας την προεκπαιδευμένη ραχοκοκαλιά EfficientNetB3 με επιπλέον πλήρως συνδεδεμένα στρώματα και κανονικοποίηση απομάκρυνσης.

Το TensorFlow χρησιμοποιήθηκε για την υλοποίηση της διαδικασίας εκπαίδευσης, χρησιμοποιώντας τον βελτιστοποιητή Adam με δυαδική απώλεια διασταυρούμενης εντροπίας και προσαρμοστικό ρυθμό μάθησης. Προκειμένου να αντιμετωπιστεί η ανισορροπία του συνόλου δεδομένων και να εξασφαλιστεί η δίκαιη εκπροσώπηση και των δύο κατηγοριών καθ' όλη τη διάρκεια της βελτιστοποίησης, τα βάρη των κατηγοριών υπολογίστηκαν δυναμικά. Με βαθμολογία F1 0,86 και τελική ακρίβεια 86,1%, το μοντέλο επέδειξε υψηλό βαθμό συμφωνίας μεταξύ των προβλέψεων και της βασικής αλήθειας, ενώ παράλληλα το ROC-AUC 0,93 έδειξε καλή διακριτική ικανότητα, υποδεικνύοντας ότι το μοντέλο μπορεί να διαχωρίσει αποτελεσματικά τις εικόνες με πολύποδες από εκείνες χωρίς.

Συνολικά, η μελέτη κατέδειξε αποτελεσματικά ότι, όταν συνδυάζεται με συγκεκριμένες τεχνικές προεπεξεργασίας και εκπαίδευσης, ένα λεπτομερώς ρυθμισμένο πλαίσιο μεταφοράς μάθησηςς μπορεί να παράγει αποτελέσματα υψηλής απόδοσης σε περιορισμένα σύνολα δεδομένων ιατρικών εικόνων. Τα αποτελέσματα δείχνουν ότι το μοντέλο έχει τη δυνατότητα να αποτελέσει ένα χρήσιμο εργαλείο για την υποστήριξη των γαστρεντερολόγων κατά τη διάρκεια των κολονοσκοπικών διαδικασιών, παρέοντας αξιόπιστη διαγνωστική βοήθεια σε πραγματικό χρόνο, η οποία μπορεί να συμβάλει στη βελτίωση των κλινικών αποτελεσμάτων και στη μείωση των χαμένων ανιχνεύσεων.

Συνολικά, η μελέτη κατέδειξε αποτελεσματικά ότι, όταν συνδυάζεται με συγκεκριμένες τεχνικές προεπεξεργασίας και εκπαίδευσης, ένα λεπτομερώς ρυθμισμένο πλαίσιο μεταφοράς μάθησης μπορεί να παράγει αποτελέσματα υψηλής απόδοσης σε περιορισμένα σύνολα δεδομένων ιατρικών εικόνων, με τα αποτελέσματα να δείχνουν ότι το μοντέλο έχει τη δυνατότητα να

αποτελέσει ένα χρήσιμο εργαλείο για την υποστήριξη των γαστρεντερολόγων κατά τη διάρκεια των κολονοσκοπικών διαδικασιών. Παρέχει αξιόπιστη διαγνωστική βοήθεια σε πραγματικό χρόνο, η οποία μπορεί να συμβάλει στη βελτίωση των κλινικών αποτελεσμάτων και στη μείωση των χαμένων ανιχνεύσεων.

9.3 Βασικές συνεισφορές

Η παρούσα μελέτη προσφέρει μια σειρά σημαντικών εξελίξεων στη διάγνωση με τη βοήθεια υπολογιστή, ιδίως στην αυτοματοποιημένη ανίχνευση πολυπόδων παχέος εντέρου με βάση τη βαθιά μάθηση, με κάθε φάση της έρευνας να δημιουργήθηκε με σκοπό την ενίσχυση της αξιοπιστίας και της χρησιμότητας της τεχνητής νοημοσύνης στην ανάλυση ιατρικών εικόνων από την επεξεργασία δεδομένων έως τον αρχιτεκτονικό σχεδιασμό και την αξιολόγηση.

Η δημιουργία μιας εξατομικευμένης αρχιτεκτονικής βαθιάς μάθησης βασισμένης στη ραχοκοκαλιά του EfficientNetB3 αποτελεί την πρώτη σημαντική συνεισφορά, με το μοντέλο να διατηρεί την υπολογιστική αποδοτικότητα, ενώ κληρονόμησε ισχυρές δυνατότητες εξόρυξης χαρακτηριστικών χρησιμοποιώντας τη μεταφορά μάθησης από το ImageNet. Η προσθήκη πιο πυκνών στρωμάτων, η ομαλοποίηση παρτίδων και η κανονικοποίηση αποχώρησης βελτίωσαν ακόμη περισσότερο την κατάσταση αυτή, σταθεροποιώντας την εκπαίδευση και μειώνοντας την υπερπροσαρμογή, ενώ ακόμη και σε περιπτώσεις όπου τα όρια ήταν θολά ή ο φωτισμός ήταν ανομοιόμορφος, το δίκτυο ήταν σε θέση να συλλάβει τις λεπτές παραλλαγές της υφής και του χρώματος που είναι χαρακτηριστικές των πολύποδων χάρη στην υβριδική δομή του.

Η δεύτερη συνεισφορά είναι το ενδελεχές pipeline για την προετοιμασία των συνόλων δεδομένων, ο οποίος δημιουργήθηκε για τη βελτιστοποίηση της ευρωστίας και της ποικιλομορφίας των δεδομένων παρά το μικρό μέγεθος του δείγματος που είναι σύνηθες στα ιατρικά σύνολα δεδομένων, όπου προκειμένου να αναπαραχθεί το απρόβλεπτο που συναντάται σε πραγματικές ρυθμίσεις κολονοσκόπησης, αυτό περιελάμβανε εξελιγμένες τεχνικές επαύξησης, όπως αναστροφή, ζουμ, προσαρμογές φωτεινότητας και τυχαίες περιστροφές. Επιπλέον, η δυναμική στάθμιση κλάσεων διόρθωσε την ανισορροπία του συνόλου δεδομένων και η κανονικοποίηση με τη χρήση της συνάρτησης προεπεξεργασίας EfficientNet εγγυήθηκε τη συνέπεια με την κατανομή προεκπαίδευσης του μοντέλου, μείωνοντας τη μεροληψία προς τις κυρίαρχες κλάσεις και βελτίωσαν τη γενίκευση του μοντέλου.

Η ενδελεχής αξιολόγηση και ανάλυση των επιδόσεων του μοντέλου είναι το αντικείμενο της τρίτης συνεισφοράς, αποκτώντας μια ολοκληρωμένη εικόνα της συμπεριφοράς του συστήματος με την στενή παρακολούθηση της διαδικασίας εκπαίδευσης με προσαρμοσμένες ανακλήσεις που υπολόγιζαν την ακρίβεια, την απώλεια, το F1-score και την ROC-AUC σε κάθε εποχή. Η ικανότητα του μοντέλου να εξισορροπεί επιτυχώς την ακρίβεια και την ανάκληση επιβεβαιώθηκε από την

τελική ακρίβεια 86,1%, το F1-score 0,86 και την AUC 0,93, ερμηνεύοντας την διαδικασία λήψης αποφάσεων του δικτύου με τη χρήση τεχνικών οπτικοποίησης, όπως το Grad-CAM, το οποίο βελτίωσε τη διαφάνεια του μοντέλου, ένα σημαντικό στοιχείο της κλινικής εμπιστοσύνης, δείχνοντας τις περιοχές της εικόνας που επηρέασαν τις προβλέψεις.

Αυτές οι συνεισφορές δημιουργούν συλλογικά ένα ολοκληρωμένο και επαναλαμβανόμενο πλαίσιο για την ταξινόμηση δυαδικών ιατρικών εικόνων. Το προτεινόμενο σύστημα αποδεικνύεται ότι είναι τόσο τεχνικά ορθό όσο και κλινικά ουσιαστικό μέσω της ενσωμάτωσης του αποτελεσματικού σχεδιασμού της αρχιτεκτονικής, της αξιόπιστης προεπεξεργασίας και της ερμηνευσιμότητας, ανογοντας την πόρτα για πιο εξελιγμένες εφαρμογές, όπως η ανίχνευση και η κατάτμηση σε πραγματικό χρόνο κατά τη διάρκεια ζωντανών κολονοσκοπικών διαδικασιών.

9.4 Τελικές παρατηρήσεις

Τα αποτελέσματα της παρούσας μελέτης καταδεικνύουν πώς τα συστήματα βαθιάς μάθησης στην ιατρική απεικόνιση γίνονται όλο και πιο εξελιγμένα και έχουν τη δυνατότητα να βοηθήσουν σημαντικά τη λήψη κλινικών αποφάσεων, με το προτεινόμενο μοντέλο να δείχνει ότι ακόμη και με περιορισμένα δεδομένα, μια καλά σχεδιασμένη αρχιτεκτονική που βασίζεται στο EfficientNetB3 και βελτιώνεται με δομημένη λεπτομερή ρύθμιση μπορεί να επιτύχει υψηλή διαγνωστική ακρίβεια, καταδεικνύοντας την αυξανόμενη χρησιμότητα της τεχνητής νοημοσύνης ως συμπληρωματικού διαγνωστικού εργαλείου στον τομέα του γαστρεντερικού, βοηθώντας τους γιατρούς να εντοπίζουν νωρίς τις ανωμαλίες του παχέος εντέρου, όπου η άμεση θεραπεία μπορεί να βελτιώσει σημαντικά τα αποτελέσματα των ασθενών.

Σε γενικές γραμμές, η μελέτη αναδεικνύει πόσο κρίσιμα είναι τα ερμηνεύσιμα και αποδοτικά σε δεδομένα μοντέλα τεχνητής νοημοσύνης στον τομέα της ιατρικής, με την οπτικοποίηση Grad-CAM που περιλαμβάνεται στο μοντέλο να δίνει στους κλινικούς γιατρούς κάποια εικόνα για το πού και πώς το σύστημα επικεντρώνει την προσοχή του κατά τη λήψη αποφάσεων εκτός από την παραγωγή εντυπωσιακών ποσοτικών αποτελεσμάτων. Η υιοθέτηση στον πραγματικό κόσμο εξαρτάται από αυτό, δεδομένου ότι οι κλινικές ρυθμίσεις απαιτούν όχι μόνο εξαιρετικές επιδόσεις αλλά και επεξηγηματικότητα και αξιοπιστία, με ένα βήμα προς το κλείσιμο του χάσματος μεταξύ των ακαδημαϊκών πρωτοτύπων και των εφαρμόσιμων κλινικών εργαλείων να αντιπροσωπεύεται από την αρμονία που επιτυγχάνεται εδώ μεταξύ της ερμηνευσιμότητας, της ακρίβειας και της υπολογιστικής απόδοσης.

Το σύστημα μπορεί να ενσωματωθεί στην τρέχουσα υποδομή του νοσοκομείου λόγω της ελαφριάς δομής και του αρθρωτού σχεδιασμού του, με το μοντέλο να μπορούσε να ενσωματωθεί σε αίθουσες ενδοσκόπησης για να ειδοποιεί τους παρόχους υγειονομικής περίθαλψης για τυχόν ύποπτες περιοχές που εντοπίζονται ως διαγνωστικό βοήθημα σε πραγματικό χρόνο. Εναλλακτικά,

θα μπορούσε να χρησιμοποιηθεί ως εργαλείο επαλήθευσης μετά την επέμβαση, αναζητώντας αυτόματα πολυπόδων που δεν έχουν εντοπιστεί σε καταγεγραμμένο υλικό κολονοσκόπησης, έχοντας τη δυνατότητα να μειώσουν σημαντικά τα ανθρώπινα σφάλματα επίβλεψης και τη διαγνωστική κόπωση, γεγονός που θα έχει ως τελικό αποτέλεσμα πιο αξιόπιστα και έγκαιρα ποσοστά ανίχνευσης.

Ωστόσο, για να περάσουμε από τα ελεγχόμενα πειράματα στην πρακτική εφαρμογή θα πρέπει να ξεπεραστούν ορισμένα εμπόδια, όπου αρχικά για να εξασφαλιστεί η ανθεκτικότητα έναντι αλλαγών στα δημογραφικά στοιχεία των ασθενών, τον φωτισμό και τον εξοπλισμό, η μελλοντική έρευνα πρέπει να επικεντρωθεί στην ενδελεχή επικύρωση σε μεγαλύτερα σύνολα δεδομένων πολλαπλών κέντρων. Για να εξασφαλιστεί η ασφαλής κλινική ανάπτυξη, πρέπει επίσης να ληφθεί υπόψη η τήρηση των διαδικασιών έγκρισης κανονιστικών διατάξεων και των προτύπων ιατρικών δεδομένων, ενώ τέλος, για την ομαλή λειτουργία στην πράξη, η συμβατότητα με ροές δεδομένων σε πραγματικό χρόνο και η ενσωμάτωση με νοσοκομειακά συστήματα πληροφοριών είναι επίσης εξίσου σημαντικές.

9.5 Μελλοντικές προτάσεις

Η παρούσα εργασία αποτέλεσε ένα πρώτο σημαντικό βήμα προς την ανάπτυξη ενός αξιόπιστου συστήματος CAD για την ανίχνευση πολυπόδων. Ωστόσο, τα αποτελέσματα και η ανάλυση των περιορισμών ανοίγουν το δρόμο για πολλές κατευθύνσεις μελλοντικής έρευνας και βελτίωσης.

9.5.1 Βελτίωση της αρχιτεκτονικής και των τεχνικών μάθησης

Μία προφανής κατεύθυνση αφορά τη δοκιμή πιο πρόσφατων αρχιτεκτονικών, όπως οι μετασχηματιστές όρασης (Vision Transformers, ViTs) και υβριδικά μοντέλα CNN–Transformer, τα οποία έχουν παρουσιάσει αξιόλογες επιδόσεις στην ιατρική απεικόνιση. Η ενσωμάτωσή τους θα μπορούσε να ενισχύσει την ικανότητα του συστήματος να αποτυπώνει μακροπρόθεσμες εξαρτήσεις και σύνθετα μορφολογικά μοτίβα, ιδίως σε επίπεδες ή άμισχες βλάβες.

Παράλληλα, η ενσωμάτωση μηχανισμών προσοχής (attention mechanisms), είτε στο χωρικό επίπεδο είτε στο επίπεδο καναλιών (spatial/channel attention), θα μπορούσε να συμβάλει στη βελτίωση της ανίχνευσης κρίσιμων περιοχών της εικόνας, όπως τα όρια και η υφή των πολυπόδων, μειώνοντας ψευδώς θετικές ανιχνεύσεις που σχετίζονται με πτυχώσεις ή υπολείμματα.

Τέλος, η αξιοποίηση της συλλογικής μάθησης (federated learning) αποτελεί μία ακόμη ερευνητική κατεύθυνση για την ανάπτυξη ισχυρότερων μοντέλων, διατηρώντας παράλληλα την ιδιωτικότητα των δεδομένων. Η προσέγγιση αυτή θα επέτρεπε εκπαίδευση σε δεδομένα από πολλαπλά νοσοκομεία χωρίς κεντρική συγκέντρωση ευαίσθητων ιατρικών εικόνων.

9.5.2 Επέκταση και βελτίωση συνόλων δεδομένων

Η δημιουργία ευρύτερων και πιο ισορροπημένων συνόλων δεδομένων αποτελεί κρίσιμο παράγοντα για περαιτέρω βελτίωση της απόδοσης. Ειδικό ενδιαφέρον έχει η συλλογή και ο σχολιασμός εικόνων που αφορούν σπάνιες αλλά κλινικά σημαντικές περιπτώσεις, όπως οδοντωτοί πολύποδες (SSA/P), πολύ μικροί πολύποδες (<5 mm) και βλάβες σε δύσκολες θέσεις. Επιπλέον, η συμπερίληψη δεδομένων από διαφορετικούς πληθυσμούς και τύπους ενδοσκοπικού εξοπλισμού θα ενίσχυε την ικανότητα γενίκευσης.

Συμπληρωματικά, είναι απαραίτητη η πολυκεντρική επικύρωση του συστήματος σε δεδομένα από ανεξάρτητα ιδρύματα που δεν συμμετείχαν στην εκπαίδευση, ώστε να τεκμηριωθεί με μεγαλύτερη αυστηρότητα η κλινική αξιοπιστία του μοντέλου.

Μία ακόμη κατεύθυνση αφορά τη χρήση συνθετικών δεδομένων. Τεχνικές όπως τα Γενετικά Ανταγωνιστικά Δίκτυα (GANs) και η συνθετική δημιουργία εικόνων (π.χ. από 3D μοντέλα) θα μπορούσαν να συμβάλουν στην παραγωγή ρεαλιστικών, σχολιασμένων δειγμάτων για προπόνηση, αντιμετωπίζοντας τόσο την ανισορροπία κλάσεων όσο και την έλλειψη σπάνιων περιπτώσεων.

9.5.3 Ενσωμάτωση πληροφορίας από βίντεο και πολυτροπικά δεδομένα

Η μετάβαση από την ανάλυση μεμονωμένων εικόνων (frames) στην επεξεργασία ολόκληρων ακολουθιών βίντεο αποτελεί λογική εξέλιξη, δεδομένου ότι η κολονοσκόπηση είναι εγγενώς δυναμική διαδικασία. Μοντέλα που αξιοποιούν τη χρονική συνέπεια, όπως 3D CNNs ή μετασχηματιστές για βίντεο, θα μπορούσαν να διακρίνουν αποτελεσματικότερα πολύποδες από παροδικά τεχνουργήματα και να παρέχουν σταθερότερες προβλέψεις.

Επιπλέον, η πολυτροπική (multi-modal) ενσωμάτωση θα μπορούσε να εμπλουτίσει το σύστημα με πρόσθετες πηγές πληροφορίας, όπως βιοαισθητήρες (π.χ. καρδιακός ρυθμός), δεδομένα προετοιμασίας εντέρου ή πληροφορίες από άλλες απεικονιστικές τεχνικές, όπως ενδοσκόπηση με NBI. Μία τέτοια προσέγγιση θα μπορούσε να οδηγήσει σε πιο ολοκληρωμένο πλαίσιο υποστήριξης κλινικών αποφάσεων.

9.5.4 Κλινική εφαρμογή και επικύρωση

Η μελλοντική έρευνα θα πρέπει να εστιάσει στη βελτιστοποίηση του συστήματος για ανάπτυξη σε πραγματικό χρόνο. Αυτό περιλαμβάνει την ανάπτυξη ελαφρύτερων εκδόσεων του μοντέλου, κατάλληλων για ενσωμάτωση σε ενδοσκοπικό εξοπλισμό με περιορισμένους πόρους, με ελάχιστη καθυστέρηση (latency) και υψηλή συχνότητα καρέ.

Εξίσου σημαντική είναι η διαλειτουργικότητα (interoperability) και η ενσωμάτωση σε κλινικές ροές εργασίας. Η ανάπτυξη λογισμικού που μπορεί να συνδεθεί άμεσα με πλατφόρμες ενδοσκόπησης και συστήματα αρχειοθέτησης εικόνων (PACS) αποτελεί προϋπόθεση για πρακτική υιοθέτηση.

Τέλος, απαιτείται διεξαγωγή αυστηρών κλινικών δοκιμών, ώστε να αξιολογηθεί όχι μόνο η τεχνική απόδοση αλλά και η πραγματική κλινική επίδραση του συστήματος. Ενδεικτικά ερωτήματα αφορούν το κατά πόσο μειώνεται το polyp miss rate, αν βελτιώνονται οι δείκτες ανίχνευσης αδενωμάτων (ADR), καθώς και αν επηρεάζεται η διάρκεια της εξέτασης ή η συνολική εμπειρία του ασθενούς.

9.5.5 Εξηγήσιμη ΤΝ και υποστήριξη απόφασης

Πέρα από την αξιοποίηση του Grad-CAM, μπορούν να διερευνηθούν τεχνικές Explainable AI (XAI) που παρέχουν ποσοτικοποιημένα μέτρα βεβαιότητας για κάθε πρόβλεψη ή ερμηνεύουν την απόφαση του μοντέλου με πιο συστηματικό τρόπο. Η περαιτέρω ανάπτυξη τέτοιων εργαλείων δύναται να ενισχύσει την εμπιστοσύνη μεταξύ του συστήματος και του ενδοσκόπου.

Επιπλέον, μία επόμενη φάση θα μπορούσε να περιλαμβάνει συστήματα που δεν περιορίζονται στον εντοπισμό, αλλά επεκτείνονται στην ταξινόμηση του πολύποδα (π.χ. μέσω μορφολογικής ανάλυσης) και στην εκτίμηση κινδύνου κακοήθους εξέλιξης, παρέχοντας πιο στοχευμένη υποστήριξη απόφασης.

9.5.6 Διεπιστημονική συνεργασία και ηθικά πλαίσια

Η στενή συνεργασία μηχανικών, ιατρών και βιοπληροφορικών θεωρείται απαραίτητη ώστε οι κλινικές ανάγκες να μεταφράζονται σε τεχνικές προδιαγραφές και τα τεχνικά αποτελέσματα να ερμηνεύονται ορθά στο κατάλληλο κλινικό πλαίσιο.

Παράλληλα με την τεχνολογική πρόοδο, απαιτείται ανάπτυξη ηθικών πλαισίων και προτύπων που θα καθορίζουν τη διαφανή χρήση, την ευθύνη και την προστασία των δεδομένων ασθενών σε συστήματα αυτοματοποιημένης διάγνωσης.

Συνοψίζοντας, ο δρόμος για την πλήρη κλινική υιοθέτηση και ωρίμανση των συστημάτων ΤΝ στην ενδοσκοπία είναι πολυδιάστατος. Απαιτεί συντονισμένη πρόοδο τόσο στον τομέα των αλγορίθμων και των δεδομένων, όσο και στην κλινική επικύρωση, τη διασφάλιση της διαφάνειας και την ανάπτυξη κατάλληλης υποδομής. Η παρούσα εργασία ελπίζει να συντελέσει ως ένα σταθερό θεμέλιο για αυτήν τη συνεχιζόμενη εξέλιξη.

ΚΕΦΑΛΑΙΟ 10

10 Βιβλιογραφία

[1]. A. F. Peery et al., "Burden of gastrointestinal, liver, and pancreatic diseases in the United States," *Gastroenterology*, vol. 154, no. 8, pp. 1949-1971, 2018.

[2]. P. Wang et al., "Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy: a single-centre randomised controlled trial," *The Lancet Gastroenterology & Hepatology*, vol. 4, no. 12, pp. 919-926, 2019.

[3]. I. Barua et al., "Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis," *Endoscopy*, vol. 53, no. 3, pp. 277-284, 2021.

[4]. M. Oines, L. A. Helsingen, and M. Bretthauer, "Epidemiology and risk factors of colorectal polyps," *Best Practice & Research Clinical Gastroenterology*, vol. 31, no. 4, pp. 419-424, 2017.

[5]. Healthline Editorial Team, “Colonic (Colorectal) Polyps,” Healthline, Sep. 4, 2019.

[6]. J. Demb et al., "Risk factors for colorectal cancer and advanced adenoma in persons aged 30-49 years: a systematic review and meta-analysis," *Clinical Gastroenterology and Hepatology*, vol. 19, no. 12, pp. 2541-2549, 2021.

[7]. R. Kamitani et al., "Computer-aided diagnosis for endoscopy: present and future," *Digestive Endoscopy*, vol. 34, no. 1, pp. 45-57, 2022.

[8]. J. P. Terdiman, "Colorectal polyps and polyposis syndromes," in *Sleisenger and Fordtran's Gastrointestinal and Liver Disease*, 10th ed., M. Feldman, L. S. Friedman, and L. J. Brandt, Eds. Philadelphia: Saunders Elsevier, 2016, pp. 2155-2190.

[9]. P. Sharma, S. Gupta, and A. K. S. Rawat, "Genetic and epigenetic alterations in colorectal polyps: a comprehensive review," *Journal of Gastrointestinal Cancer*, vol. 53, no. 1, pp. 1-12, 2022.

[10]. S. G. Patel et al., "Rising incidence of early-onset colorectal cancer: a call to action," *Current Oncology Reports*, vol. 24, no. 3, pp. 285-294, 2022.

[11]. J. E. IJspeert et al., "The cancer prevention effect of a colonoscopy varies with the location of the cancer," *Gastroenterology*, vol. 162, no. 4, pp. 1090-1092, 2022.

[12]. H. Ashrafian, A. Darzi, and T. Athanasiou, "The economic burden of colorectal cancer across Europe: a population-based cost-of-illness study," *The Lancet Gastroenterology & Hepatology*, vol. 6, no. 9, pp. 709-722, 2021.

[13]. M. Tan and Q. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” in Proc. 36th Int. Conf. Mach. Learn. (ICML), Long Beach, CA, USA, Jun. 2019, pp. 6105–6114.

[14]. S. Kudo et al., "Artificial intelligence-assisted polyp detection for colonoscopy: initial experience," *Gastroenterology*, vol. 156, no. 7, pp. 2027-2029, 2019.

[15]. J. Podlasek et al., "Deep learning for real-time polyp detection in colonoscopy videos: a systematic review," *Artificial Intelligence in Medicine*, vol. 118, 102132, 2021.

[16]. J. Bernal et al., "Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge," *IEEE Transactions on Medical Imaging*, vol. 36, no. 6, pp. 1231-1249, 2017.

[17]. J. Bernal, J. Sánchez, and G. Fernández-Esparrach, "WM-DOVA maps for accurate polyp highlighting in colonoscopy videos: Model vs. human," *IEEE Transactions on Medical Imaging*, vol. 36, no. 8, pp. 1621-1632, 2017.

[18]. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.

[19]. N. Tajbakhsh et al., "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 630-644, 2016.

[20]. H. Rezatofighi et al., "Generalized intersection over union: a metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 658-666.

[21]. Q. Nguyen, S. Z. M. Azmi, J. Lee, R. Mohi-Ud-Din, R Mostaghim, R. Sachdev, D. Siegel, Y. Friedlander, and S. Friedland, "Clinical evaluation of a real-time artificial intelligence-based polyp detection system: a US multi-center pilot study, "Sci. Rep., vol. 12, 2022, Art. no. 4342. doi: 10.1038/s41598-022-10597-y.

[22]. J. Bernal, A. Histace, and G. Dray, "Cross-validation study for polyp detection in colonoscopy videos," in *Proc. IEEE Engineering in Medicine and Biology Society (EMBC)*, Jeju, Korea, 2017, pp. 1282-1285.

[23]. R. El Karazle, P. H. Then, and L. Liu, "A systematic review of machine learning and deep learning for colorectal polyp detection and classification," *Artificial Intelligence Review*, vol. 56, no. Suppl 1, pp. 1-37, 2023.

[24]. M. Spadaccini et al., "Computer-aided detection for colorectal polyp imaging in colonoscopy: a systematic review and meta-analysis of randomized controlled trials," *Gastrointestinal Endoscopy*, vol. 101, no. 6, pp. AB123-AB124, 2025.

[25]. A. Danku, L. M. Sarosi, and Z. Tulassay, "Development of a computer-aided diagnostic tool for early colorectal cancer detection using machine learning," *Journal of Clinical Medicine*, vol. 14, no. 2, p. 345, 2025.

[26]. D. Shen et al., "An artificial intelligence system for detection and classification of colorectal polyps in colonoscopy: a multi-center study," *The Lancet Digital Health*, vol. 5, no. 3, pp. e150-e159, 2023.

[27]. Q. Xie and L. Shen, "DAFM-Net: A dual attention and fusion mechanism network for colorectal polyp segmentation," *Medical Image Analysis*, vol. 85, 102518, 2025.

[28]. A. Santone, M. Cesarelli, and F. Mercaldo, "A U-Net based deep learning framework for automatic segmentation of colorectal polyps in endoscopic images," *Computers in Biology and Medicine*, vol. 160, 106942, 2025.

[29]. D. Fan and J. Jiang, "PolypFormer: A transformer-based architecture for accurate colorectal polyp segmentation," *IEEE Transactions on Medical Imaging*, 2025, (early access).

[30]. Y. Liu et al., "Deep learning for colorectal polyp segmentation: A systematic review and meta-analysis of techniques from 2018 to 2024," *IEEE Reviews in Biomedical Engineering*, vol. 18, pp. 123-145, 2025.

[31]. L. Tong, Y. Chen, and Z. Wang, "EENet: Edge-enhanced network for colorectal polyp segmentation in colonoscopy images," *IEEE Transactions on Medical Imaging*, vol. 43, no. 5, pp. 1234-1245, 2024.

[32]. Y. Guo, J. Bernal, and B. J. Matuszewski, "Polyp segmentation in colonoscopy images using fully convolutional neural networks and a case study," in *Proc. IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Iowa City, IA, USA, 2020, pp. 417-421.

[33]. S. Mohapatra, S. Patra, and S. K. Jena, "UPolySeg: A U-Net based model for colorectal polyp segmentation with pre-processing techniques," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France, 2022, pp. 1416-1420.

[34]. P. Wang and W. Lim, "S-MGRAUNet: Synergistic multi-granularity rough attention U-Net for colorectal polyp segmentation," *IEEE Journal of Biomedical and Health Informatics*, 2025, (early access).

[35]. F. Safarov and T. Whangbo, "A-DenseUNet: Adaptive Densely Connected UNet for polyp segmentation in colonoscopy images with sharp boundary," *IEEE Access*, vol. 9, pp. 57931-57941, 2021.

[36]. L. Nanni, A. Lumini, and S. Brahnam, "Ensemble of convolutional neural networks for polyp segmentation in colonoscopy images," *Expert Systems with Applications*, vol. 213, Part 2, 118954, 2023.

[37]. J. Eu, S. Kim, and D. Kim, "A computer-aided diagnosis system for colorectal polyp segmentation using modified SegNet with VGG-19 encoder," *Sensors*, vol. 21, no. 18, p. 6076, 2021.

[38]. J. Wan, X. Chen, and Z. Yu, "Polyp detection from colorectum images by using attentive YOLOv5," in *Proc. IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Houston, TX, USA, 2021, pp. 1569-1574.

[39]. P. Wang, Q. Xie, R. Cui, and S. Chen, "YOLO-SRPD: A colorectal polyp detection model using super-resolution reconstruction and YOLO," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1-11, 2024.

[40]. Y. Tang et al., "Improving colorectal polyp detection and classification using GAN-based data augmentation and YOLO," *Medical Physics*, vol. 50, no. 3, pp. 1456-1469, 2023.

[41]. A. Krenzer et al., "ENDOMIND-Advanced: An open-source real-time polyp detection system for colonoscopy," *International Journal of Computer Assisted Radiology and Surgery*, vol. 18, no. 5, pp. 931-938, 2023.

[42]. N. Abraham et al., "Evaluation of machine learning models for colorectal polyp classification using endoscopic images," *Journal of Digital Imaging*, vol. 36, no. 2, pp. 567-578, 2023.

[43]. O. Attallah, "A computer-aided diagnosis framework for lung and colorectal cancer classification from histopathological images using lightweight CNNs and feature fusion," *IEEE Journal of Biomedical and Health Informatics*, 2025, (early access).

[44]. H. Jeong, S. Lee, and J. Park, "Lightweight Deep Learning Model for Real-Time Colorectal Polyp Segmentation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, 2023, pp. 1-5.

[45]. S. A. El-Ghany, M. A. El-Bendary, and A. E. Hassanien, "An accurate deep learning-based computer-aided diagnosis system for gastrointestinal disease detection using wireless capsule endoscopy image analysis," *IEEE Access*, vol. 12, pp. 12345-12356, 2024.

[46]. A. Doniyorjon, S. Kim, and D. Kim, "An improved method of polyp detection using custom YOLOv4-Tiny for wireless endoscopic images," in *Proc. IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, 2022, pp. 1-4.

[47]. J. Akgöl, L. M. Sarosi, and Z. Tulassay, "Challenges in dataset diversity and annotation for deep learning in colorectal polyp detection: a review," Journal of Medical Systems, vol. 49, no. 1, p. 12, 2025.

[48]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.

[49]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Advances in Neural Information Processing Systems (NeurIPS), Lake Tahoe, NV, USA, Dec. 2012, pp. 1097–1105.

[50]. A. Brand, J. Bernal, and G. Dray, "Object detection in medical imaging: trends and challenges," IEEE Reviews in Biomedical Engineering, vol. 15, pp. 123-145, 2022.

[51]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015.

[52]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788.

[53]. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 2015, pp. 234-241.

[54]. T. Fawcett, “An introduction to ROC analysis,” Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006.

[55]. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Salt Lake City, UT, USA, Jun. 2018, pp. 4510–4520.

[56]. J. Hu, L. Shen, and G. Sun, “Squeeze-and-Excitation Networks,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141.

[57]. N. Mckay, “REAL-colon dataset,” Figshare, 2023.

[58]. 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,” in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 618–626.

[59]. M. T. Ribeiro, S. Singh, and C. Guestrin, “"Why Should I Trust You?": Explaining the Predictions of Any Classifier,” in Proceedings of the 22nd ACM SIGKDD International

Conference on Knowledge Discovery and Data Mining (KDD '16), San Francisco, CA, USA, 2016, pp. 1135-1144.

[60]. S. M. Lundberg and S.-I. Lee, "SHAP[AB1] : A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems 30 (NeurIPS 2017), Long Beach, CA, USA, 2017, pp. 4765–4774.

Παράρτημα Α: Κώδικας Εκπαίδευσης Μοντέλου

Το παρόν παράρτημα περιλαμβάνει τον πλήρη κώδικα εκπαίδευσης (train_FINAL.py) που χρησιμοποιήθηκε για την παραγωγή του τελικού μοντέλου, καθώς και για την αποθήκευση των βαρών και των συνοδευτικών αρχείων καταγραφής της εκπαίδευσης. Οι διαδρομές αρχείων (paths) που εμφανίζονται στον κώδικα είναι εξαρτώμενες από το περιβάλλον εκτέλεσης και ενδέχεται να απαιτούν προσαρμογή ανάλογα με τη δομή φακέλων του εκάστοτε συστήματος.

```
from tensorflow.keras import layers, models
from tensorflow.keras.applications import EfficientNetB3
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import Callback, EarlyStopping,
ReduceLROnPlateau
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from sklearn.utils.class_weight import compute_class_weight
from sklearn.metrics import f1_score as sk_f1_score
import numpy as np
import os
import matplotlib.pyplot as plt
import time
import json
import pickle

# ======================
# CONFIGURATION
# ======================
DATA_DIR = "/Users/johnstylianidis/polyp-detector/data"
WEIGHTS_PATH = "/Users/johnstylianidis/polyp-
detector/models/model_weights.h5"
HISTORY_PATH = "/Users/johnstylianidis/polyp-
detector/models/training_history.pkl"
MODEL_CONFIG = "/Users/johnstylianidis/polyp-
detector/models/model_config.pkl"
IMAGE_SIZE = (300, 300)
BATCH_SIZE = 32
INITIAL_EPOCHS = 20
FINE_TUNE_EPOCHS = 40

# ======================
# F1-SCORE CALCULATION UTILITY
# ======================
def compute_f1(y_true, y_pred):
    y_pred = np.round(y_pred)
    return sk_f1_score(y_true, y_pred, average='binary')

# ======================
# DATA AUGMENTATION
```

```
# ======================
def create_augmenter():
    return ImageDataGenerator(
        preprocessing_function=tf.keras.applications.efficientnet.preprocess
_input,
        rotation_range=45,
        width_shift_range=0.3,
        height_shift_range=0.3,
        shear_range=0.3,
        zoom_range=0.3,
        horizontal_flip=True,
        vertical_flip=True,
        brightness_range=[0.7, 1.3],
        fill_mode='constant'
    )

train_datagen = create_augmenter()
val_datagen =
ImageDataGenerator(preprocessing_function=tf.keras.applications.efficientnet
.preprocess_input)

# ======================
# DATA GENERATORS
# ======================
train_generator = train_datagen.flow_from_directory(
    f'{DATA_DIR}/train',
    target_size=IMAGE_SIZE,
    batch_size=BATCH_SIZE,
    class_mode='binary',
    shuffle=True
)

val_generator = val_datagen.flow_from_directory(
    f'{DATA_DIR}/val',
    target_size=IMAGE_SIZE,
    batch_size=BATCH_SIZE,
    class_mode='binary',
    shuffle=False
)

# ======================
# MODEL ARCHITECTURE
# ======================
def build_enhanced_model():
    base_model = EfficientNetB3(
        include_top=False,
        weights='imagenet',
        input_shape=IMAGE_SIZE + (3,),
        pooling='avg'
    )
```

```
    base_model.trainable = False

    inputs = tf.keras.Input(shape=IMAGE_SIZE + (3,))
    x = base_model(inputs, training=False)
    x = layers.Dense(1024, activation='relu')(x)
    x = layers.Dropout(0.6)(x)
    x = layers.BatchNormalization()(x)
    x = layers.Dense(512, activation='relu')(x)
    x = layers.Dropout(0.5)(x)
    outputs = layers.Dense(1, activation='sigmoid')(x)

    return tf.keras.Model(inputs, outputs)

# ======================
# PLOT TRAINING HISTORY
# ======================
def plot_training(history):
    plt.figure(figsize=(15, 5))

    # Accuracy plot
    plt.subplot(1, 3, 1)
    plt.plot(history['accuracy'], label='Train Accuracy')
    plt.plot(history['val_accuracy'], label='Validation Accuracy')
    plt.title('Accuracy')
    plt.ylabel('Accuracy')
    plt.xlabel('Epoch')
    plt.legend()

    # Loss plot
    plt.subplot(1, 3, 2)
    plt.plot(history['loss'], label='Train Loss')
    plt.plot(history['val_loss'], label='Validation Loss')
    plt.title('Loss')
    plt.ylabel('Loss')
    plt.xlabel('Epoch')
    plt.legend()

    # F1 Score plot
    plt.subplot(1, 3, 3)
    plt.plot(history['f1_score'], label='Train F1')
    plt.plot(history['val_f1_score'], label='Validation F1')
    plt.title('F1 Score')
    plt.ylabel('F1 Score')
    plt.xlabel('Epoch')
    plt.legend()

    plt.tight_layout()
    plt.savefig('/Users/johnstylianidis/polyp-
detector/models/training_history.png')
    plt.close()
```

```
# ======================
# CLASS WEIGHTS
# ======================
class_weights = compute_class_weight(
    'balanced',
    classes=np.unique(train_generator.classes),
    y=train_generator.classes
)
class_weights = {0: float(class_weights[0]), 1: float(class_weights[1])}

# ======================
# CUSTOM TRAINING MONITOR
# ======================
class TrainingMonitor(Callback):
    def __init__(self):
        super().__init__()
        self.history = {
            'loss': [],
            'accuracy': [],
            'val_loss': [],
            'val_accuracy': [],
            'f1_score': [],
            'val_f1_score': []
        }
        self.best_val_acc = 0.0

    def on_epoch_end(self, epoch, logs=None):
        # Record metrics as native floats
        self.history['loss'].append(float(logs['loss']))
        self.history['accuracy'].append(float(logs['accuracy']))
        self.history['val_loss'].append(float(logs['val_loss']))
        self.history['val_accuracy'].append(float(logs['val_accuracy']))

        # Calculate F1 scores
        train_preds = self.model.predict(train_generator, verbose=0)
        val_preds = self.model.predict(val_generator, verbose=0)

        train_f1 = compute_f1(train_generator.classes, train_preds)
        val_f1 = compute_f1(val_generator.classes, val_preds)

        self.history['f1_score'].append(train_f1)
        self.history['val_f1_score'].append(val_f1)

        # Update best weights
        current_val_acc = float(logs['val_accuracy'])
        if current_val_acc > self.best_val_acc:
            self.best_val_acc = current_val_acc
            # Save weights immediately
            self.model.save_weights(WEIGHTS_PATH)
```

```
            print(f"Saved best weights to {WEIGHTS_PATH}")

        print(f"Epoch {epoch+1}: loss={logs['loss']:.4f},
acc={logs['accuracy']:.4f}, "
              f"val_loss={logs['val_loss']:.4f},
val_acc={logs['val_accuracy']:.4f}, "
              f"train_f1={train_f1:.4f}, val_f1={val_f1:.4f}")

# ======================
# TRAINING PROCESS
# ======================
def train_model():
    # Create training monitor
    monitor = TrainingMonitor()

    # Early stopping and LR reduction
    early_stopping = EarlyStopping(
        monitor='val_loss',
        patience=12,
        restore_best_weights=False,  # We're handling this ourselves
        verbose=1
    )

    reduce_lr = ReduceLROnPlateau(
        monitor='val_loss',
        factor=0.2,
        patience=6,
        min_lr=1e-7,
        verbose=1
    )

    # Phase 1: Feature extraction
    model = build_enhanced_model()
    model.compile(
        optimizer=Adam(1e-4),
        loss='binary_crossentropy',
        metrics=['accuracy']
    )

    print("\n=== PHASE 1: INITIAL TRAINING ===")
    model.fit(
        train_generator,
        epochs=INITIAL_EPOCHS,
        validation_data=val_generator,
        class_weight=class_weights,
        callbacks=[monitor, early_stopping, reduce_lr],
        verbose=1
    )

    # Phase 2: Full fine-tuning
```

```
    print("\n=== PHASE 2: FINE-TUNING ===")
    model.layers[1].trainable = True

    # Unfreeze deeper layers
    for layer in model.layers[1].layers[:150]:
        layer.trainable = False
    for layer in model.layers[1].layers[150:]:
        layer.trainable = True

    model.compile(
        optimizer=Adam(1e-5),
        loss='binary_crossentropy',
        metrics=['accuracy']
    )

    model.fit(
        train_generator,
        initial_epoch=INITIAL_EPOCHS,
        epochs=INITIAL_EPOCHS + FINE_TUNE_EPOCHS,
        validation_data=val_generator,
        class_weight=class_weights,
        callbacks=[monitor, early_stopping, reduce_lr],
        verbose=1
    )

    # Save model configuration
    model_config = {
        'image_size': IMAGE_SIZE,
        'class_weights': class_weights,
        'initial_epochs': INITIAL_EPOCHS,
        'fine_tune_epochs': FINE_TUNE_EPOCHS,
        'base_model': 'EfficientNetB3',
        'trainable_layers': 150
    }
    with open(MODEL_CONFIG, 'wb') as f:
        pickle.dump(model_config, f)
    print(f"Model configuration saved to {MODEL_CONFIG}")

    return monitor.history

# ======================
# EXECUTE TRAINING
# ======================
if __name__ == "__main__":
    # Train model and get history
    start_time = time.time()
    final_history = train_model()
    training_time = time.time() - start_time

    # Save history
```

```
    with open(HISTORY_PATH, 'wb') as f:
        pickle.dump(final_history, f)
    print(f"Training history saved to {HISTORY_PATH}")

    # Plot training history
    plot_training(final_history)

    print(f"\nTraining complete! Weights saved to {WEIGHTS_PATH}")
    print(f"Training graphs saved to /Users/johnstylianidis/polyp-
detector/models/training_history.png")
    print(f"Total training time: {training_time//60:.0f} minutes
{training_time%60:.0f} seconds")

    # Instructions for loading the model
    print("\nHOW TO USE THE MODEL IN THE FUTURE:")
    print("1. To load the model for inference:")
    print("   from tensorflow.keras.models import load_model")
    print("   model = build_enhanced_model()  # Use the same build
function")
    print(f"   model.load_weights('{WEIGHTS_PATH}')")
    print("   model.compile(optimizer=Adam(1e-5),
loss='binary_crossentropy', metrics=['accuracy'])")

    print("\n2. To access training history:")
    print(f"   with open('{HISTORY_PATH}', 'rb') as f:")
    print("       history = pickle.load(f)")

    print("\n3. Model configuration:")
    print(f"   with open('{MODEL_CONFIG}', 'rb') as f:")
    print("       config = pickle.load(f)")
```

Παράρτημα Β: Κώδικας Αξιολόγησης Μοντέλου

Το παρόν παράρτημα περιλαμβάνει τον πλήρη κώδικα αξιολόγησης (test_FINAL.py) που χρησιμοποιήθηκε για τη φόρτωση του αποθηκευμένου μοντέλου και τον υπολογισμό των μετρικών απόδοσης στο ανεξάρτητο σύνολο δοκιμής. Αντίστοιχα, οι διαδρομές αρχείων (paths) είναι εξαρτώμενες από το περιβάλλον εκτέλεσης και ενδέχεται να απαιτούν προσαρμογή, ώστε να αντιστοιχούν στα αρχεία μοντέλου, δεδομένων και αποτελεσμάτων του εκάστοτε συστήματος.

```
import tensorflow as tf
from tensorflow.keras import layers, models
from tensorflow.keras.applications import EfficientNetB3
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from sklearn.metrics import classification_report, confusion_matrix,
f1_score, roc_auc_score, roc_curve
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os
import json
import cv2

# ======================
# CONFIGURATION
# ======================
TEST_DATA_DIR = "/Users/johnstylianidis/polyp-detector/data/test"
WEIGHTS_PATH = "/Users/johnstylianidis/polyp-
detector/models/model_weights.h5"
IMAGE_SIZE = (300, 300)
BATCH_SIZE = 32
RESULTS_DIR = "/Users/johnstylianidis/polyp-detector/results"
MODEL_CONFIG = "/Users/johnstylianidis/polyp-
detector/models/model_config.pkl"

# ======================
# GRAD-CAM IMPLEMENTATION
# ======================
def make_gradcam_heatmap(img_array, model, last_conv_layer_name,
pred_index=None):
    """Generates Grad-CAM heatmap for an image"""
    # First, we create a model that maps the input image to the activations
    # of the last conv layer as well as the output predictions
    grad_model = tf.keras.models.Model(
        [model.inputs],
        [model.get_layer(last_conv_layer_name).output, model.output]
    )
```

```
    # Then we compute the gradient of the top predicted class for our input
image
    # with respect to the activations of the last conv layer
    with tf.GradientTape() as tape:
        last_conv_layer_output, preds = grad_model(img_array)
        if pred_index is None:
            pred_index = tf.argmax(preds[0])
        class_channel = preds[:, pred_index]

    # This is the gradient of the output neuron (top predicted or chosen)
    # with regard to the output feature map of the last conv layer
    grads = tape.gradient(class_channel, last_conv_layer_output)

    # This is a vector where each entry is the mean intensity of the
gradient
    # over a specific feature map channel
    pooled_grads = tf.reduce_mean(grads, axis=(0, 1, 2))

    # We multiply each channel in the feature map array
    # by "how important this channel is" with regard to the top predicted
class
    # then sum all the channels to obtain the heatmap class activation
    last_conv_layer_output = last_conv_layer_output[0]
    heatmap = last_conv_layer_output @ pooled_grads[..., tf.newaxis]
    heatmap = tf.squeeze(heatmap)

    # For visualization purpose, we will also normalize the heatmap between
0 & 1
    heatmap = tf.maximum(heatmap, 0) / tf.math.reduce_max(heatmap)
    return heatmap.numpy()

def save_gradcam(img, heatmap, cam_path="cam.jpg", alpha=0.4):
    """Saves Grad-CAM visualization"""
    # Rescale heatmap to a range 0-255
    heatmap = np.uint8(255 * heatmap)

    # Resize the heatmap to the same size as the original image
    heatmap = cv2.resize(heatmap, (img.shape[1], img.shape[0]))

    # Convert heatmap to RGB
    heatmap = cv2.applyColorMap(heatmap, cv2.COLORMAP_JET)

    # Superimpose the heatmap on original image
    superimposed_img = heatmap * alpha + img
    superimposed_img = np.clip(superimposed_img, 0, 255).astype(np.uint8)

    # Save the superimposed image
    superimposed_img = cv2.cvtColor(superimposed_img, cv2.COLOR_BGR2RGB)
    cv2.imwrite(cam_path, superimposed_img)
```

```
# ======================
# MODEL ARCHITECTURE (SAME AS TRAINING)
# ======================
def build_model_for_testing(return_base_model=False):
    """Builds the model with the exact same layer trainability as during
training"""
    base_model = EfficientNetB3(
        include_top=False,
        weights='imagenet',  # Will be overwritten by our weights
        input_shape=IMAGE_SIZE + (3,),
        pooling='avg'
    )

    # REPLICATE THE EXACT TRAINABILITY STATE FROM TRAINING
    base_model.trainable = True  # As in fine-tuning phase
    for layer in base_model.layers[:150]:
        layer.trainable = False  # First 150 layers frozen
    for layer in base_model.layers[150:]:
        layer.trainable = True   # Last layers trainable

    inputs = tf.keras.Input(shape=IMAGE_SIZE + (3,))
    x = base_model(inputs, training=False)  # Important for batch norm
layers
    x = layers.Dense(1024, activation='relu')(x)
    x = layers.Dropout(0.6)(x)
    x = layers.BatchNormalization()(x)
    x = layers.Dense(512, activation='relu')(x)
    x = layers.Dropout(0.5)(x)
    outputs = layers.Dense(1, activation='sigmoid')(x)

    if return_base_model:
        return tf.keras.Model(inputs, outputs), base_model
    return tf.keras.Model(inputs, outputs)

# ======================
# PREPARE TEST DATA
# ======================
def create_test_generator():
    test_datagen = ImageDataGenerator(
        preprocessing_function=tf.keras.applications.efficientnet.preprocess
_input
    )

    test_generator = test_datagen.flow_from_directory(
        TEST_DATA_DIR,
        target_size=IMAGE_SIZE,
        batch_size=BATCH_SIZE,
        class_mode='binary',
        shuffle=False  # Important for correct label ordering
    )
```

```
    # Print class indices to verify
    print("Class indices:", test_generator.class_indices)
    print("Test samples:", test_generator.samples)

    return test_generator

# ======================
# EVALUATE MODEL
# ======================
def evaluate_model(model, test_generator):
    # 1. Evaluate loss and accuracy
    evaluation = model.evaluate(test_generator)
    print(f"Test Loss: {evaluation[0]:.4f}")
    print(f"Test Accuracy: {evaluation[1]:.4f}")

    # 2. Get predictions
    test_generator.reset()  # Reset generator to start from beginning
    y_pred = model.predict(test_generator, steps=test_generator.samples //
BATCH_SIZE + 1)
    y_pred_binary = (y_pred > 0.5).astype(int)
    y_true = test_generator.classes

    # 3. Classification report
    class_names = list(test_generator.class_indices.keys())
    report = classification_report(y_true, y_pred_binary,
target_names=class_names)
    print("\nClassification Report:")
    print(report)

    # 4. Confusion matrix
    cm = confusion_matrix(y_true, y_pred_binary)
    print("Confusion Matrix:")
    print(cm)

    # 5. Calculate F1 score
    f1 = f1_score(y_true, y_pred_binary)
    print(f"F1 Score: {f1:.4f}")

    # 6. Calculate AUC-ROC
    auc_roc = roc_auc_score(y_true, y_pred)
    print(f"AUC-ROC: {auc_roc:.4f}")

    return {
        'loss': evaluation[0],
        'accuracy': evaluation[1],
        'classification_report': report,
        'confusion_matrix': cm.tolist(),
        'f1_score': f1,
        'auc_roc': auc_roc,
```

```
        'y_true': y_true.tolist(),
        'y_pred': y_pred.tolist(),
        'y_pred_binary': y_pred_binary.tolist(),
        'class_names': class_names
    }

# ======================
# VISUALIZE RESULTS
# ======================
def visualize_results(results):
    # Create results directory if not exists
    os.makedirs(RESULTS_DIR, exist_ok=True)
    class_names = results['class_names']

    # 1. Confusion matrix visualization
    plt.figure(figsize=(8, 6))
    sns.heatmap(results['confusion_matrix'], annot=True, fmt='d',
                cmap='Blues', xticklabels=class_names,
yticklabels=class_names)
    plt.xlabel('Predicted')
    plt.ylabel('Actual')
    plt.title('Confusion Matrix')
    plt.savefig(os.path.join(RESULTS_DIR, 'confusion_matrix.png'))
    plt.close()

    # 2. ROC Curve
    fpr, tpr, _ = roc_curve(results['y_true'], results['y_pred'])
    plt.figure(figsize=(8, 6))
    plt.plot(fpr, tpr, label=f'AUC = {results["auc_roc"]:.4f}')
    plt.plot([0, 1], [0, 1], 'k--')
    plt.xlabel('False Positive Rate')
    plt.ylabel('True Positive Rate')
    plt.title('ROC Curve')
    plt.legend(loc='lower right')
    plt.savefig(os.path.join(RESULTS_DIR, 'roc_curve.png'))
    plt.close()

    # 3. Save metrics to JSON
    with open(os.path.join(RESULTS_DIR, 'test_metrics.json'), 'w') as f:
        json.dump({
            'loss': results['loss'],
            'accuracy': results['accuracy'],
            'f1_score': results['f1_score'],
            'auc_roc': results['auc_roc'],
            'classification_report': results['classification_report'],
            'confusion_matrix': results['confusion_matrix']
        }, f, indent=4)

# ======================
# SAMPLE PREDICTIONS VISUALIZATION WITH GRAD-CAM
```

```
# ======================
def visualize_sample_predictions(model, base_model, test_generator,
class_names, num_samples=10):
    test_generator.reset()
    x, y = next(test_generator)
    y_pred = model.predict(x)

    # Create directory for Grad-CAM visualizations
    grad_cam_dir = os.path.join(RESULTS_DIR, 'grad_cam')
    os.makedirs(grad_cam_dir, exist_ok=True)

    # Get the last convolutional layer name from the base model
    last_conv_layer_name = None
    for layer in base_model.layers:
        if isinstance(layer, tf.keras.layers.Conv2D):
            last_conv_layer_name = layer.name
    print(f"Using layer '{last_conv_layer_name}' for Grad-CAM")

    plt.figure(figsize=(20, 10))
    for i in range(min(num_samples, len(x))):  # Ensure we don't exceed
available samples
        # Original image
        plt.subplot(2, num_samples, i+1)

        # Convert image from EfficientNet preprocessing to displayable
format
        img = x[i].copy()
        img = img - img.min()
        img = img / img.max()
        img = (img * 255).astype(np.uint8)

        plt.imshow(img)
        actual_class = class_names[int(y[i])]
        pred_class = class_names[int(y_pred[i] > 0.5)]
        confidence = float(y_pred[i][0])  # Convert to Python float
        plt.title(f"Actual: {actual_class}\nPred: {pred_class}
({confidence:.2f})")
        plt.axis('off')

        # Grad-CAM visualization
        plt.subplot(2, num_samples, num_samples + i + 1)

        # Generate heatmap using the base model's last conv layer
        img_array = np.expand_dims(x[i], axis=0)
        heatmap = make_gradcam_heatmap(img_array, base_model,
last_conv_layer_name)

        # Display heatmap
        plt.imshow(heatmap, cmap='viridis')
        plt.title('Grad-CAM')
```

```
        plt.axis('off')

        # Save individual Grad-CAM images
        save_path = os.path.join(grad_cam_dir, f'grad_cam_sample_{i}.png')
        save_gradcam(img, heatmap, cam_path=save_path)

    plt.tight_layout()
    plt.savefig(os.path.join(RESULTS_DIR,
'sample_predictions_with_gradcam.png'))
    plt.close()

# ======================
# MAIN TESTING FUNCTION
# ======================
def test_model():
    # Build model with EXACTLY the same layer trainability as during
training
    model, base_model = build_model_for_testing(return_base_model=True)

    # Load trained weights
    model.load_weights(WEIGHTS_PATH)
    print("Successfully loaded model weights")

    # Compile the model (required for evaluation)
    model.compile(
        optimizer=Adam(1e-5),
        loss='binary_crossentropy',
        metrics=['accuracy']
    )
    print("Model compiled successfully")

    # Prepare test data
    test_generator = create_test_generator()

    # Evaluate model
    results = evaluate_model(model, test_generator)

    # Visualize results
    visualize_results(results)

    # Visualize sample predictions with Grad-CAM
    visualize_sample_predictions(model, base_model, test_generator,
results['class_names'])

    print("\nTesting complete! Results saved to:", RESULTS_DIR)

# ======================
# EXECUTE TESTING
# ======================
if __name__ == "__main__":
```

```
    # Import required modules for model architecture
    from tensorflow.keras import layers
    from tensorflow.keras.applications import EfficientNetB3

    test_model()
```