



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Μεθοδολογική συγκριτική ανάλυση αρχιτεκτονικών μοντέλων τεχνητής νοημοσύνης στην τμηματοποίηση (segmentation) γαστρεντερικών πολύποδων σε ιατρικές εικόνες του συνόλου δεδομένων Kvasir-SEG.

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ANNA ΔΟΥΣΚΑ**

Επιβλέπων : Γεώργιος Ματσόπουλος
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπουσα : Ουρανία Πετροπούλου
ΕΔΙΠ Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2025

[Αυτή η σελίδα είναι σκοπίμως κενή]



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Μεθοδολογική συγκριτική ανάλυση αρχιτεκτονικών μοντέλων τεχνητής νοημοσύνης στην τμηματοποίηση (segmentation) γαστρεντερικών πολύποδων σε ιατρικές εικόνες του συνόλου δεδομένων Kvasir-SEG.

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ANNA ΔΟΥΣΚΑ**

Επιβλέπων : Γεώργιος Ματσόπουλος
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπουσα : Ουρανία Πετροπούλου
ΕΔΙΠ Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 3^η Φεβρουαρίου 2025.

(Υπογραφή)

.....
Γ. Ματσόπουλος
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....
Α. Παναγόπουλος
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....
Π. Τσανάκας
Καθηγητής Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2025

Άννα Δούσκα

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

Copyright © Άννα Δούσκα, 2024.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη:

Η παρούσα διπλωματική εργασία εστιάζει στη χρήση της τεχνητής νοημοσύνης για την τμηματοποίηση ιατρικών εικόνων κολονοσκόπησης, με σκοπό τον εντοπισμό πολυπόδων και την ενίσχυση της πρόληψης του καρκίνου του παχέος εντέρου. Ο καρκίνος του παχέος εντέρου αποτελεί μία από τις κύριες αιτίες θανάτου παγκοσμίως, και η έγκαιρη διάγνωση και αφαίρεση πολυπόδων μπορεί να μειώσει δραστικά τη θνησιμότητα. Πιο συγκεκριμένα, μελετήθηκε το σύνολο δεδομένων Kvasir-SEG, το οποίο περιέχει 1.000 εικόνες ενδοσκοπικών πολυπόδων, και πραγματοποιήθηκε συγκριτική ανάλυση διαφορετικών αρχιτεκτονικών νευρωνικών δικτύων για την κατάτμηση/τμηματοποίηση (segmentation) των πολυπόδων μετά από προσεκτική προεπεξεργασία των δεδομένων.

Στην έρευνα αξιολογήθηκαν διάφορες αρχιτεκτονικές μοντέλων βαθιάς μάθησης, όπως το UNet, το SegResNet, το PSPNet και το DeepLabv3. Κάθε μοντέλο εκπαιδεύτηκε και αξιολογήθηκε με βάση την απόδοσή του σε μετρικές όπως η ακρίβεια (accuracy), η ευστοχία (precision), η ανάκληση (recall), το f1-score και η διασταύρωση επί ένωση (IoU). Τα αποτελέσματα κατέδειξαν την ανωτερότητα του DeepLabv3 στην ακρίβεια και την ικανότητα γενίκευσης, καθιστώντας το ιδανικό εργαλείο για την τμηματοποίηση πολυπόδων σε ιατρικές εικόνες.

Επιπλέον, η εργασία προτείνει μελλοντικές βελτιώσεις μέσω της ενσωμάτωσης νέων αρχιτεκτονικών και της αξιοποίησης προεκπαιδευμένων δικτύων, όπως τα Transformers, για την περαιτέρω βελτίωση της απόδοσης. Η χρήση της τεχνητής νοημοσύνης αναδεικνύεται ως κρίσιμο εργαλείο για την αυτοματοποίηση και βελτίωση της ιατρικής διάγνωσης, προσφέροντας σημαντικές προοπτικές για την κλινική πρακτική. Συνολικά, η μελέτη ενισχύει την κατανόηση των δυνατοτήτων της τεχνητής νοημοσύνης στην ιατρική απεικόνιση και συμβάλλει στην ανάπτυξη καινοτόμων λύσεων που μπορούν να βελτιώσουν την υγειονομική περίθαλψη.

🔑 Λέξεις κλειδιά

Τεχνητή Νοημοσύνη, Τμηματοποίηση Εικόνας, Πολύποδες, Καρκίνος Παχέος Εντέρου, Καρκίνος Ορθού, Καρκίνος Παχέος Εντέρου Και Ορθού, Βαθιά Μάθηση, Συνελικτικά Νευρωνικά Δίκτυα, Ιατρικές Εικόνες, Κολονοσκόπηση, UNet, SegResNet, PSPNet, DeepLabv3, ResNet

[Αυτή η σελίδα είναι σκοπίμως κενή]

Abstract

This thesis focuses on the application of artificial intelligence for the segmentation of colonoscopy medical images, aimed at polyp detection and enhancing the prevention of colorectal cancer. Colorectal cancer is one of the leading causes of death worldwide, and early diagnosis and removal of polyps can significantly reduce mortality. Specifically, the study utilized the Kvasir-SEG dataset, which contains 1,000 images of endoscopic polyps, and conducted a comparative analysis of different neural network architectures for the segmentation of polyps following meticulous data preprocessing.

The research evaluated various deep learning model architectures, such as UNet, SegResNet, PSPNet, and DeepLabv3. Each model was trained and assessed based on its performance metrics, including accuracy, precision, recall, f1-score, and intersection over union (IoU). The results demonstrated the superiority of DeepLabv3 in terms of accuracy and generalization capabilities, making it an ideal tool for polyp segmentation in medical images.

Furthermore, the study proposes future improvements through the integration of new architectures and the use of pretrained networks, such as Transformers, to further enhance performance. Artificial intelligence emerges as a crucial tool for automating and improving medical diagnosis, offering significant prospects for clinical practice. Overall, the study enhances the understanding of the capabilities of artificial intelligence in medical imaging and contributes to the development of innovative solutions that can improve healthcare.

Keywords

Artificial Intelligence, Image Segmentation, Polyps, Colon Cancer, Rectal Cancer, Colorectal Cancer, Deep Learning, Convolutional Neural Networks, Medical Images, Colonoscopy, UNet, SegResNet, PSPNet, DeepLabv3, ResNet

[Αυτή η σελίδα είναι σκοπίμως κενή]

Ευχαριστίες

Θα ήθελα να ευχαριστήσω ιδιαίτερα τον επιβλέποντα καθηγητή μου Δρ. Γεώργιο Ματσόπουλο, ο οποίος μου έδωσε τη δυνατότητα να ασχοληθώ και να εμβαθύνω στο αντικείμενο της Τεχνητής Νοημοσύνης και την συμβολή της στην Βιοϊατρική Μηχανική μέσω αυτής της διπλωματικής, συνεισφέροντας έτσι και εγώ την ταπεινή μου συμβολή στον εξαιρετικά χρήσιμο για την κοινωνία τομέα. Θα ήθελα να ευχαριστήσω επίσης τον Δρ. Ιωάννη Βεζάκη για τις οδηγίες και την βοήθεια του και τον υποψήφιο διδάκτορα Κωνσταντίνο Γεωργά ο οποίος με καθοδήγησε διεξοδικά με μεγάλη ακρίβεια καθ'όλη τη διάρκεια εκπόνησης της διπλωματικής μου εργασίας, προσφέροντας μου πολύ ουσιαστική βοήθεια. Θα ήθελα ακόμη να ευχαριστήσω ιδιαίτερα την οικογένεια μου για όλα τα εφόδια και τις ευκαιρίες που μου παρείχε μεγαλώνοντας άλλα και την αμέριστη στήριξη και αγάπη σε όλη τη διάρκεια της ζωής μου που με βοήθησαν να διαμορφωθώ στον άνθρωπο που είμαι σήμερα. Θα ήθελα να ευχαριστήσω επίσης, όλους τους φίλους που με βοήθησαν να εξελιχθώ και να φτάσω στο σημείο που βρίσκομαι σήμερα. Τέλος, τον Β.Α. που πιστεύει σε μένα και βρίσκεται συνεχώς δίπλα μου να με στηρίζει και να με βοηθάει για να εξελιχθώ και να γίνω η καλύτερη έκδοση του εαυτού μου.

[Αυτή η σελίδα είναι σκοπίμως κενή]

Πίνακας Περιεχομένων:

Περίληψη:.....	5
Abstract.....	7
Σκοπός εργασίας.....	19
I. ΕΙΣΑΓΩΓΗ.....	19
1) Γενικές Πληροφορίες	19
2) Λειτουργίες της γαστρεντερικής οδού:.....	19
II. Καρκίνος Παχέος Εντέρου-Ορθου και Πολύποδες.....	21
1) Κατηγορίες Πολυπόδων.....	21
i) Αδενωματώδεις πολύποδες (Αδενώματα/Adenoma).....	23
ii) Υπερπλαστικοί πολύποδες (Hyperplastic).....	24
iii) Φλεγμονώδεις πολύποδες (Inflammation).....	24
iv) Πολύποδες οδοντωτού τύπου (Serrated).....	25
v) Αμαρτωματώδεις πολύποδες (Hamartomatous).....	25
2) Πολύποδες και καρκίνος του παχέος εντέρου.....	26
3) Επιδημιολογικά Χαρακτηριστικά.....	27
4) Παράγοντες που επηρεάζουν την ανάπτυξη CRC.....	30
i) αμετάβλητοι παράγοντες:.....	30
ii) μεταβλητοί παράγοντες:.....	31
5) Πρόληψη, διάγνωση, αφαίρεση και θεραπεία.....	32
Καινοτόμες Θεραπευτικές Προσεγγίσεις στον CRC.....	32
Ο Ρόλος της Κολονοσκόπησης.....	32
Θεραπευτικές Μέθοδοι Αφαίρεσης Πολυπόδων.....	33
6) Εξελίξεις στη Διάγνωση και Σταδιοποίηση.....	34
7) Αδύναμα σημεία και αναγκαιότητα τεχνητής νοημοσύνης για Image Segmentation σε εικόνες και βίντεο κολονοσκοπήσεων:.....	35
III. ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ.....	38
1) Η συμβολή της τεχνητής νοημοσύνης.....	38
2) Συνελικτικά Νευρωνικά Δίκτυα.....	39
3) Διαίρεση του συνόλου δεδομένων για τις ανάγκες του νευρωνικού δικτύου.....	40
IV. ΣΤΟΙΧΕΙΑ ΠΕΙΡΑΜΑΤΟΣ.....	43
1) Εργαλεία που χρησιμοποιήθηκαν και Περιβάλλον Εργασίας:.....	43
2) Σύνολο δεδομένων Kvasir-SEG:.....	44
3) Τροποποιήσεις στη συγκεκριμένη μελέτη:.....	46
4) Τεχνικές Προεπεξεργασίας Εικόνων και Μασκών:.....	46
5) Μετρικές Αξιολόγησης:.....	48
6) Αρχιτεκτονικές Μοντέλων Μηχανικής Μάθησης:.....	49
UNet [2015].....	49
PSPNet (Pyramid Scene Parsing Network) [2017] [.....	
SegResNet (2018).....	54

DeepLabv3 (2017).....	56
V. ΣΥΓΚΡΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ.....	59
Training και Validation Loss ανά εποχή:.....	59
Validation Dice Score IoU.....	61
Συμπεράσματα.....	65
VI. ΜΕΛΛΟΝΤΙΚΕΣ ΒΕΛΤΙΩΣΕΙΣ/ΠΡΟΕΚΤΑΣΕΙΣ.....	68
VII. ΑΝΤΙ ΕΠΙΛΟΓΟΥ.....	70
VIII. Εξήγηση ξενόγλωσσων όρων και αρκτικόλεξων:.....	71
IX. Ιστοσελίδες & διαδικτυακά εργαλεία που χρησιμοποιήθηκαν:.....	72
X. ΒΙΒΛΙΟΓΡΑΦΙΑ.....	73

Ευρετήριο Πινάκων

πινακας 1: Χαρακτηριστικά της Tesla T4 NVIDIA GPU.....	43
πινακας 2: Χαρακτηριστικά της A100 NVIDIA GPU.....	43
πινακας 3: Αποτελέσματα μοντέλων πειράματος.....	64

Ευρετήριο Εικόνων

Figure 1: σχεδιάγραμμα οργάνων και αδένων πεπτικού συστήματος.....	20
<i>(http://www.turkupertcentre.net/petanalysis/organ_gi.html)</i>	
Figure 2: δομή εντερικού τοιχώματος.....	21
<i>(https://www.kekos.gr/2014/05/31/polypodes/)</i>	
Figure 3: ανατομία παχέος εντέρου.....	21
<i>(https://www.researchgate.net/figure/Overview-of-Colon-and-Resection-Sites-of-the-105-Tissue-Samples-Samples-cancer-healthy_fig2_330863722)</i>	
Figure 4: κλινική εικόνα κολονοσκόπησης που απεικονίζεται φυσιολογικός ιστός, επίπεδος πολύποδας (sessile) και έμμισχος (pedunculated).....	22
<i>(https://www.uvmhealth.org/healthwise/topic/zm2655)</i>	
Figure 5: κυκλικό διάγραμμα με ποσοστά εμφάνισης ειδών αδενωμάτων (I) και κίνδυνος κακοήθειας (M).....	22
<i>(https://commons.wikimedia.org/wiki/File:Pie_chart_of_colorectal_polyp_etiologies.svg)</i>	
Figure 6: μικροσκοπική δομή κυττάρων αδενώματος.....	23
<i>(https://www.researchgate.net/figure/Figure-8-Histologic-appearance-of-Tubular-adenoma_fig6_340084991)</i>	
Figure 7: μικροσκοπική δομή κυττάρων υπερπλαστικού πολύποδα.....	24
<i>(https://www.mypathologyreport.ca/diagnosis-library/colon-hyperplastic-polyp/)</i>	
Figure 8: μικροσκοπική δομή κυττάρων φλεγμονώδους πολύποδα.....	24
<i>(https://www.mypathologyreport.ca/diagnosis-library/inflammatory-polyp-colon/)</i>	
Figure 9: μικροσκοπική δομή sessile serrated lesion (SSL).....	25
<i>(https://www.eligast.gr/wp-content/uploads/2020/04/01.kyriakos.pdf)</i>	
Figure 10: μικροσκοπική δομή αμαρτώματος.....	26
<i>(https://librepathology.org/wiki/Gastrointestinal_tract_polyps)</i>	
Figure 11: σχεδιάγραμμα εξέλιξης αδενωματούδους πολύποδα παχέος εντέρου σε αδενοκαρκίνωμα.....	26
<i>(https://kalochristianakis.gr/pathiseis/pachy-entero/karkinos-pacheos-enterou/)</i>	
Figure 12: επίσημα επιδημιολογικά στατιστικά από το GCO και τον Παγκόσμιο Οργανισμό Υγείας (2022).....	27
<i>(https://gco.iarc.fr/)</i>	
Figure 13: στατιστικά του 2022 που δείχνουν ότι ο CRC είναι η τρίτη μορφή καρκίνου με τη μεγαλύτερη συχνότητα και ο δεύτερος με μεγαλύτερη θνησιμότητα.....	28
<i>(https://gco.iarc.fr/)</i>	
Figure 14: διαγράμματα για τις χώρες που ο καρκίνος του παχέος εντέρου και του ορθού αποτέλεσε την πιο θανατηφόρα μορφή (2022).....	28

(<https://gco.iarc.fr/>)

Figure 15: ιστόγραμμα για τους 5 καρκίνους με την μεγαλύτερη θνησιμότητα παγκοσμίως και στα δυο φύλα το 2022.....28

(<https://gco.iarc.fr/>)

Figure 16: κυκλικό διάγραμμα που παρουσιάζει την εμφάνιση και την θνησιμότητα του CRC λαμβάνοντας υπόψη τον Δείκτη Ανθρώπινης Ανάπτυξης HDI παγκοσμίως για όλα τα φύλα.....29

(<https://gco.iarc.fr/>)

Figure 17: κολονοσκόπιο της Olympus GIF-1TH190..... 33

(<https://medical.olympusamerica.com/products/evis-exera-iii-pcf-h190dli>)

Figure 18: χρωμοενδοσκόπηση με 0.2 % indigo carmine στην αντλία του νερού ή με καθετήρα..... 34

(<https://www.medarbejder.rm.dk/siteassets/uddannelse/medarbejder/laringsportaler/akademi-for-koloskoporer/endoscopic-diagnosis-of-sessile-serrated-adenoma-polyp-with-and-without-dysplasia-carcinoma.pdf>)

Figure 19: παραδείγματα εικόνων από το σύνολο δεδομένων Kvasir όπου με πράσινο επισημαίνεται το περίγραμμα του ιστού των πολύποδων..... 45

(https://link.springer.com/chapter/10.1007/978-3-030-37734-2_37)

Figure 20: υπόδειγμα πίνακα σύγχυσης..... 49

(https://www.researchgate.net/figure/A-representation-of-Confusion-Matrix_fig3_364159469)

Figure 21: Αρχιτεκτονική U-net (32x32 pixel στη χαμηλότερη ανάλυση). μπλε πλαίσια: πολυκαναλικοί χάρτες χαρακτηριστικών με αριθμό καναλιών (πάνω) | μέγεθος x-y (κάτω αριστερό άκρο κουτιού) | λευκά πλαίσια: αντιγραμμένοι χάρτες χαρακτηριστικών | βέλη: λειτουργίες..... 50

(https://www.researchgate.net/figure/The-original-U-Net-architecture-32-Blue-boxes-multi-channel-feature-maps-White_fig3_359067492)

Figure 22: U-net's segmentation success on PhC-U373 (a-b) and DIC-HeLa (c-d) datasets (a) original image (b) cyan mask: Unet's segmentation | yellow outline: grand truth mask (c) original image (d) colourful masks: Unets segmentation | yellow outlines: grand truth masks..... 52

(https://www.researchgate.net/figure/Results-of-the-method-proposed-in-27-Fig4-a-An-original-image-in-PhC-U373_fig1_358603083)

Figure 23: Επισκόπηση του προτεινόμενου PSPNet (a) εικόνα εισόδου (b) χάρτης χαρακτηριστικών του τελευταίου convolutional layer που προκύπτει από CNN (c) εφαρμογή PPM για τη συλλογή διαφορετικών αναπαραστάσεων υποπεριοχής, ακολουθούμενα από επίπεδα upsampling και συνένωσης για να σχηματιστεί η τελική αναπαράσταση χαρακτηριστικών, η οποία μεταφέρει τόσο τοπικές όσο και καθολικές πληροφορίες περιβάλλοντος (d) τελική πρόβλεψη ανά pixel αφού η αναπαράσταση τροφοδοτείται σε ένα επίπεδο συνέλιξης..... 53

(https://www.researchgate.net/figure/Pyramid-scene-parsing-network-20_fig2_346418029)

Figure 24: παραδείγματα αποτελεσμάτων PSPNet στο Cityscapes dataset.....	54
<i>(https://i.cs.hku.hk/~hszhao/project/pspnet/index.html)</i>	
Figure 25: επισκόπηση της αρχιτεκτονικής δικτύου SegResNet με κωδικοποιητή που κάνει down-sample 4 φορές και οι ποσότητες των residual block κάθε σταδίου είναι 1, 2, 2, 2, 4	55
<i>(https://joser-online.biomedcentral.com/articles/10.1186/s13018-024-04680-5)</i>	
Figure 26: segmentation όγκου σε εικόνα MRI εγκεφάλου με όγκο (αριστερά), πρόβλεψη SegResNet (μέση), μάσκα αναφοράς (δεξιά) κίτρινο: ET (ενισχυόμενος όγκος), πράσινο: WT (ολόκληρος όγκος), μπλε: TC (πυρήνας όγκου).....	55
<i>(https://www.sciencedirect.com/science/article/pii/S1746809424014009)</i>	
Figure 27: αρχιτεκτονική SegResNet.....	55
<i>(αυτή η εικόνα κατασκευάστηκε από την συγγραφέα με βάση την περιγραφή της αρχιτεκτονικής)</i>	
Figure 28: αρχιτεκτονική μοντέλου DeepLabv3 με backbone ResNet50.....	57
<i>(κατασκευάστηκε επίσης από την συγγραφέα με βάση αυτήν https://www.researchgate.net/figure/ResNet50-architecture_fig2_382617634)</i>	
Figure 29: γραφημα απεικόνισης αρχιτεκτονικής DeepLabv3.....	57
<i>(https://www.sciencedirect.com/science/article/pii/S1047320318302931)</i>	
Figure 30: αποτελέσματα οπτικοποίησης στο Cityscapes val set κατά την εκπαίδευση με μόνο training set.....	58
<i>(https://medium.com/towards-data-science/review-deeplabv3-atrous-convolution-semantic-segmentation-6d818bfd1d74)</i>	
Figure 31: αρχιτεκτονική DeepLabv3+.....	58
<i>(https://www.researchgate.net/figure/DeepLabV3-architecture-with-Atrous-Spatial-Pyramid-Pooling-ASPP-module_fig4_356212856)</i>	
Figure 32: αρχιτεκτονική DeepLabv3.....	58
<i>(https://www.researchgate.net/figure/Deeplab-V3-with-core-module-of-atrous-spatial-pyramid-pooling-ASPP-The-feature_fig1_355072835)</i>	
Figure 33: UNet διάγραμμα training και validation loss ανά εποχή.....	59
Figure 34: SegResNet διάγραμμα training και validation loss ανά εποχή.....	59
Figure 35: PSPNet διάγραμμα training και validation loss ανά εποχή.....	60
Figure 36: DeepLabv3 διάγραμμα training και validation loss.....	60
Figure 37: UNet validation dice score (IoU).....	62
Figure 38: SegResNet validation dice score (IoU).....	62
Figure 39: PSPNet validation dice score (IoU).....	63
Figure 40: DeepLabv3 validation dice score (IoU).....	63

Figure 41: απεικόνιση αποτελέσματος DeepLabv3 για εκπαίδευση 200 εποχών. Οι ground truth masks εμφανίζονται blurry στα όρια λόγω του ότι εμφανίζονται σε binary μορφή με μηδενικό threshold..... 67

Σκοπός εργασίας

Ο σκοπός της παρούσας διπλωματικής εργασίας είναι η διερεύνηση και αξιολόγηση της χρήσης τεχνητής νοημοσύνης στην τμηματοποίηση εικόνων κολονοσκόπησης για τον εντοπισμό πολυπόδων, με στόχο τη βελτίωση της πρόληψης του καρκίνου του παχέος εντέρου. Ειδικότερα, η εργασία επικεντρώνεται στην αξιολόγηση και σύγκριση διαφορετικών αρχιτεκτονικών μοντέλων βαθιάς μάθησης, όπως το UNet, το SegResNet, το PSPNet και το DeepLabv3, μέσω του συνόλου δεδομένων Kvasir-SEG. Η ανάλυση των αποτελεσμάτων στοχεύει στην ανάδειξη των πλεονεκτημάτων της τεχνητής νοημοσύνης στην αυτοματοποίηση και βελτίωση της διαγνωστικής ακρίβειας, καθώς και στην πρόταση μελλοντικών βελτιώσεων και προοπτικών για κλινικές εφαρμογές.

I. ΕΙΣΑΓΩΓΗ

1) Γενικές Πληροφορίες

Η γαστρεντερική οδός (GI) αποτελείται από ένα σύνολο οργάνων που συνθέτουν το πεπτικό σύστημα, ένα σύστημα ζωτικής σημασίας για την ανθρώπινη υγεία. Μέσω αυτής πραγματοποιούνται βασικές διεργασίες όπως η πρόσληψη, η επεξεργασία, και η διάσπαση της τροφής, καθώς και η απορρόφηση των απαραίτητων θρεπτικών ουσιών για την επιβίωση και ανάπτυξη του οργανισμού, καθώς και η αποβολή των άχρηστων παραπροϊόντων του μεταβολισμού. Η γαστρεντερική οδός είναι καθοριστική για τη διατήρηση της ζωής και της υγείας, καθώς εξασφαλίζει την παροχή ενέργειας και δομικών συστατικών στα κύτταρα, συμβάλλει στη ρύθμιση της ισορροπίας υγρών και ηλεκτρολυτών, και υποστηρίζει την εύρυθμη λειτουργία του ανοσοποιητικού συστήματος μέσω της δράσης της εντερικής μικροχλωρίδας.

2) Λειτουργίες της γαστρεντερικής οδού:

Το πεπτικό σύστημα αποτελείται από τη γαστρεντερική οδό και τους προσαρτημένους σε αυτήν αδένες, όπως οι σιελογόνοι αδένες, το πάγκρεας και το ήπαρ (συκώτι). Η γαστρεντερική οδός περιλαμβάνει τη στοματική κοιλότητα, τον φάρυγγα, τον οισοφάγο, το στομάχι, το λεπτό και το παχύ έντερο [1]. Στη στοματική κοιλότητα πραγματοποιείται η μάσηση και ο σχηματισμός του βλωμού με τη βοήθεια του σάλιου, που διασπά το άμυλο. Στη συνέχεια μετά την κατάποση, ο βλωμός περνά διαμέσου του φάρυγγα και του οισοφάγου με περισταλτικές κινήσεις στο στομάχι, όπου αρχίζει η πέψη των πρωτεϊνών με την βοήθεια των γαστρικών υγρών που αποτελούνται από πεψινογόνο και υδροχλωρικό οξύ. Το λεπτό έντερο ολοκληρώνει την πέψη και απορροφά τα θρεπτικά συστατικά. Το λεπτό έντερο έχει διάμετρο 2,5cm και μήκος 6-7 m, στο πρώτο τμήμα του (δωδεκαδάκτυλο, τα πρώτα 25cm) μέσω ενός κοινού πόρου εκκρίνονται η χολή και παγκρεατικές εκκρίσεις. Η χολή παράγεται στο ήπαρ, περιέχει νερό, ανόργανα και χολικά άλατα, χοληστερόλη, λεκιθίνη και χολοχρωστικές (πχ. χολερυθρίνη) με αποτέλεσμα να γαλακτωματοποιεί τα λίπη [2]. Το παγκρεατικό υγρό περιέχει ένζυμα για τη διάσπαση υδατανθράκων, πρωτεϊνών, λιπών και νουκλεϊνικών οξέων. (όπως

είναι η ινσουλίνη). Ο βλεννογόνος ιστός του λεπτού εντέρου παρουσιάζει πολυάριθμες πτυχώσεις με προεκβολές, τις λάχνες, τα κύτταρα της επιφάνειας των οποίων παρουσιάζουν μικρολάχνες με ένζυμα που διασπούν τα συστατικά του χυλού σε μικρά απορροφήσιμα μόρια. Η απορροφητική επιφάνεια του εντέρου λόγω της μορφολογίας του φτάνει τα 250m^2 στους ενήλικες. Στο λεπτό έντερο ολοκληρώνεται η πέψη των πρωτεϊνών, των υδατανθράκων και των λιπών και γίνεται η απορρόφηση των θρεπτικών συστατικών της τροφής. Το παχύ έντερο έχει διάμετρο περίπου 6,5 cm και μήκος 1,5m, εκεί αποθηκεύεται προσωρινά το υλικό που δεν έχει υποστεί διασπαστεί μέχρι να αποβληθεί. Όσο το υλικό κινείται στο παχύ έντερο γίνεται απορρόφηση νερού, αλάτων και ορισμένων βιταμινών. Η απορρόφηση νερού έχει ως αποτέλεσμα την δημιουργία κοπράνων, από τα άπεπτα υπολείμματα των τροφών, χρωστικές της χολής και βακτήρια που μεταβολίζουν τις άπεπτες ουσίες και παράγουν οξέα και άλλες ενώσεις με χαρακτηριστική οσμή. Ορισμένα από τα βακτήρια που υπάρχουν στο παχύ έντερο παράγουν βιταμίνες εξαιρετικά χρήσιμες στον οργανισμό όπως η βιταμίνη K που συμμετέχει στην πήξη του αίματος. Το παχύ έντερο δηλαδή αποθηκεύει και αποβάλλει τα άπεπτα υπολείμματα, απορροφώντας ταυτόχρονα νερό και βιταμίνες.

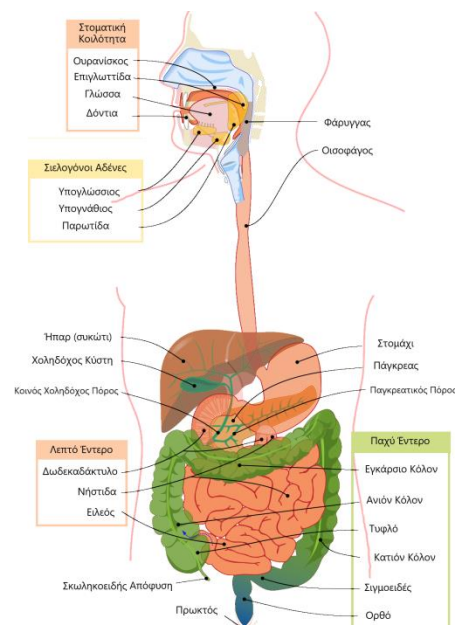


Figure 1: σχεδιάγραμμα οργάνων και αδένων πεπτικού συστήματος

Το παχύ έντερο, που εκτείνεται από τη μετάβαση του λεπτού εντέρου (τυφλό) έως τον πρωκτό, περιλαμβάνει το ανιόν, εγκάρσιο, κατión και σιγμοειδές κόλον και το ορθό [3], παίζει καίριο ρόλο στη διαδικασία της πέψης καθώς είναι υπεύθυνο για την απορρόφηση νερού και ηλεκτρολυτών, τη ρύθμιση της εντερικής μικροχλωρίδας και τον σχηματισμό των κοπράνων. Ωστόσο, το παχύ έντερο δεν αποτελεί απλώς ένα όργανο επεξεργασίας, είναι συχνά το σημείο εμφάνισης σημαντικών παθολογικών καταστάσεων, όπως είναι οι πολύποδες.

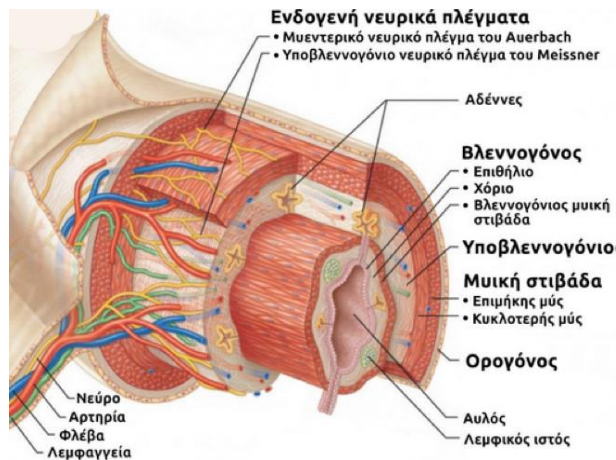


Figure 2: δομή εντερικού τοιχώματος

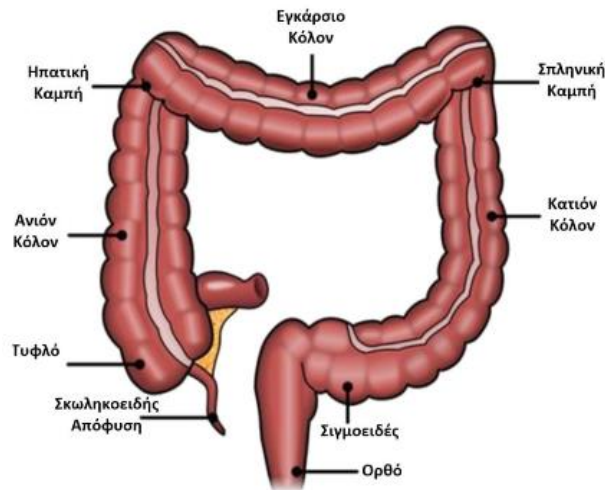


Figure 3: ανατομία παχέος εντέρου

II. Καρκίνος Παχέος Εντέρου-Ορθου και Πολύποδες

1) Κατηγορίες Πολυπόδων

Οι πολύποδες είναι ανώμαλες εξογκώσεις, συνήθως καλοήθεις, που εμφανίζονται στην εσωτερική επιφάνεια του παχέος εντέρου και είναι βλάβες που ανιχνεύονται όταν ο βλεννογόνος ιστός υπερφύεται. Συνήθως οι πολύποδες είναι ασυμπτωματικοί και σπανιότερα, ανάλογα με τη θέση και το μέγεθός τους, μπορεί να προκαλέσουν αιμορραγία, τεινισμό¹, εντερική απόφραξη, διάρροια ή πόνο στην κοιλιακή χώρα [4]. Οι πολύποδες μπορεί να είναι επίπεδοι (sessile), ανυψωμένοι ή έμμοιχοι (pedunculated) και διακρίνονται από τον φυσιολογικό βλεννογόνο ως προς το χρώμα και το ανάγλυφο της επιφάνειας. Μπορεί να

¹τεινισμός = (ιατρικός όρος) σφίξιμο, δυσκολία, πόνος κατά την ούρηση ή αφόδευση

έχουν ποικίλα σχήματα, μεγέθη και χαρακτηριστικά, και ορισμένες φορές αποτελούν προκαρκινικές αλλοιώσεις, αυξάνοντας τον κίνδυνο ανάπτυξης καρκίνου του παχέος εντέρου. Γενικά όσον αφορά το σχήμα οι πολύποδες μπορεί να είναι:

i. Πεταλοειδείς (Pedunculated):

- έμμισχοι, με εμφανές στέλεχος που συνδέει τον πολύποδα με το τοίχωμα του εντέρου και αφαιρούνται εύκολα με ενδοσκοπικές μεθόδους

ii. Επίπεδοι (Sessile):

- προεξέχουν ελάχιστα από τον βλεννογόνο καθιστώντας δύσκολο τον εντοπισμό και την αφαίρεσή τους και έχουν αυξημένο κίνδυνο καρκινικής εξέλιξης

iii. Υπερυψωμένοι (Elevated):

- παρουσιάζουν μικρή προεξοχή από την επιφάνεια του εντέρου και είναι ενδιάμεσου κινδύνου

iv. Λαχνωτοί (Villous):

- Έχουν υφή και εμφάνιση που μοιάζει με λαχνές (μικρές προβολές) και συνδέονται με αυξημένο κίνδυνο κακοήθειας

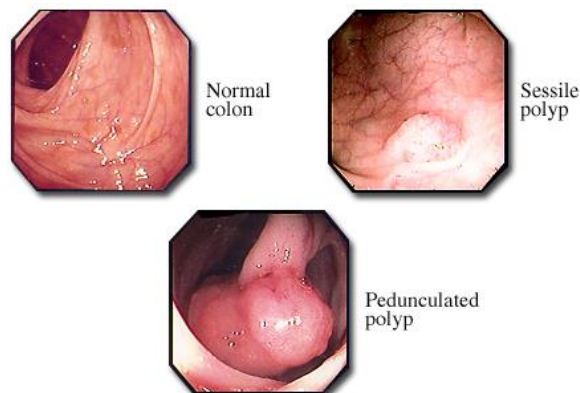


Figure 4: κλινική εικόνα κολονοσκόπησης που απεικονίζεται φυσιολογικός ιστός, επίπεδος πολύποδας (sessile) και έμμισχος (pedunculated)

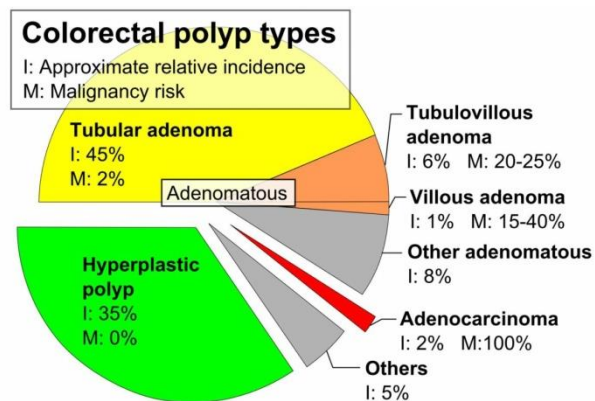


Figure 5: κυκλικό διάγραμμα με ποσοστά εμφάνισης ειδών αδενωμάτων (I) και κίνδυνος κακοήθειας (M)

Οι πολύποδες του παχέος εντέρου, ανάλογα με τη μορφολογία, τη δομή και τον κίνδυνο κακοήθειας, διακρίνονται σε διάφορους κύριους τύπους:

Ι. Αδενωματώδεις πολύποδες (Αδενώματα/Adenoma) [5]

- Είναι οι πιο συνηθισμένοι τύποι πολυπόδων. Σαν μορφολογία μπορεί να είναι:
 - επίπεδοι (sessile): καλύπτουν μεγαλύτερη επιφάνεια του βλεννογόνου και είναι πιο δύσκολη η αφαίρεσή τους
 - πεταλοειδείς (pedunculated) δηλαδή έμμισχοι, με σχήμα σαν μανιτάρι, λόγω στελέχους που τους συνδέει με το τοίχωμα του εντέρου και είναι εύκολο να αφαιρεθούν ενδοσκοπικά.

Ως υφή η επιφάνεια τους μακροσκοπικά μπορεί να είναι λεία, τραχιά ή λαχνωτή, ανάλογα με τη μορφολογία τους. Σε μικροσκοπικό επίπεδο χαρακτηρίζονται από μια ιστολογική δομή που περιλαμβάνει ανώμαλη ανάπτυξη των αδενικών κυττάρων του βλεννογόνου.

- Κίνδυνος: Οι επίπεδοι πολύποδες και εκείνοι με μέγεθος μεγαλύτερο από 1 cm έχουν αυξημένο κίνδυνο κακοήθειας. Γενικά, τα αδενώματα θεωρούνται προκαρκινικοί πολύποδες και αφαιρούνται προληπτικά.

Υποκατηγορίες αδενωμάτων:

- σωληνώδη (tubular): το πιο κοινό είδος, με τον χαμηλότερο κίνδυνο κακοήθειας, συνήθως πρόκειται για καλοήθης όγκους με δομή σωληνοειδών αδενίων.
- λαχνωτά (villous): δομή με λαχνωτές προεκβολές που συνδέεται με υψηλότερο κίνδυνο κακοήθειας.
- σωληνολαχνωτά (tubulovillous): συνδυάζουν χαρακτηριστικά των παραπάνω τύπων και έχουν μέτριο κίνδυνο κακοήθειας.

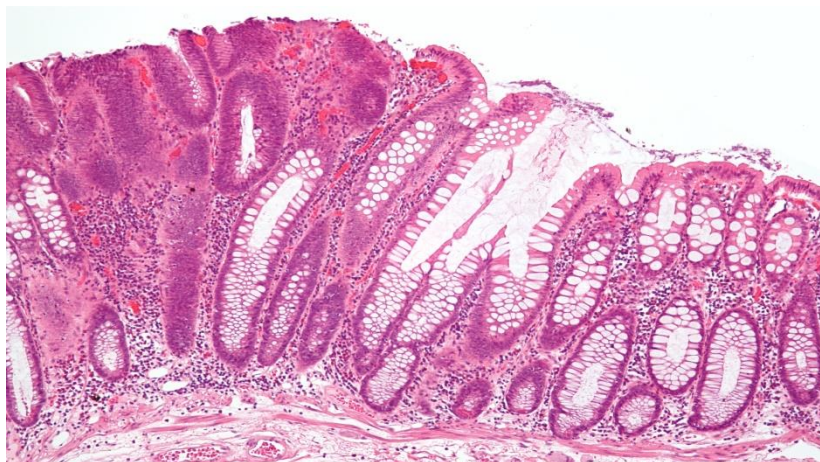


Figure 6: μικροσκοπική δομή κυττάρων αδενώματος

ii. Υπερπλαστικοί πολύποδες (Hyperplastic)

- Συνήθως είναι μικρού μεγέθους (<5 mm) με στρογγυλό ή θολωτό σχήμα και είναι επίπεδοι ή ελαφρώς υπερυψωμένοι. Το χρώμα τους μπορεί να είναι πιο ανοιχτό από τον περιβάλλοντα ιστό με ομαλή και συμπαγή υφή. Εμφανίζονται κυρίως στο ορθό και το κατώτερο τμήμα του παχέος εντέρου (σιγμοειδές) [6].
- Κίνδυνος: Έχουν χαμηλή πιθανότητα καρκινικής εξέλιξης εκτός από τους μεγαλύτερους (>1 cm) που ενδέχεται να σχετίζονται με υψηλότερο κίνδυνο κακοήθειας.

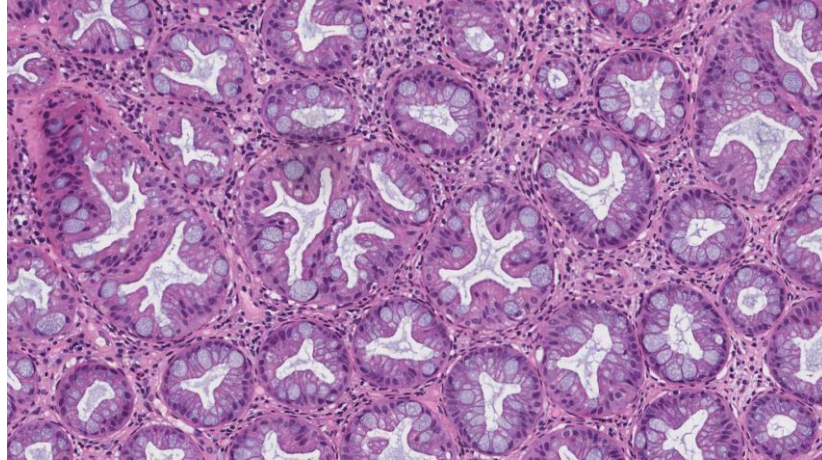


Figure 7: μικροσκοπική δομή κυττάρων υπερπλαστικού πολύποδα

iii. Φλεγμονώδεις πολύποδες (Inflammation)

- Εμφανίζονται ως ελκώδεις ή οζώδεις αλλοιώσεις και σχετίζονται με φλεγμονώδεις νόσους του εντέρου όπως η ελκώδης κολίτιδα και η νόσος του Crohn. Είναι συνήθως μικροί, ασύμμετροι, ανώμαλοι και ασχημάτιστοι με μαλακή και ερεθισμένη επιφάνεια, λόγω της φλεγμονής στον βλεννογόνο.
- Κίνδυνος: Κατά κανόνα δεν είναι προκαρκινικοί, αλλά η παρατεταμένη φλεγμονή αυξάνει τον συνολικό κίνδυνο ανάπτυξης καρκίνου.

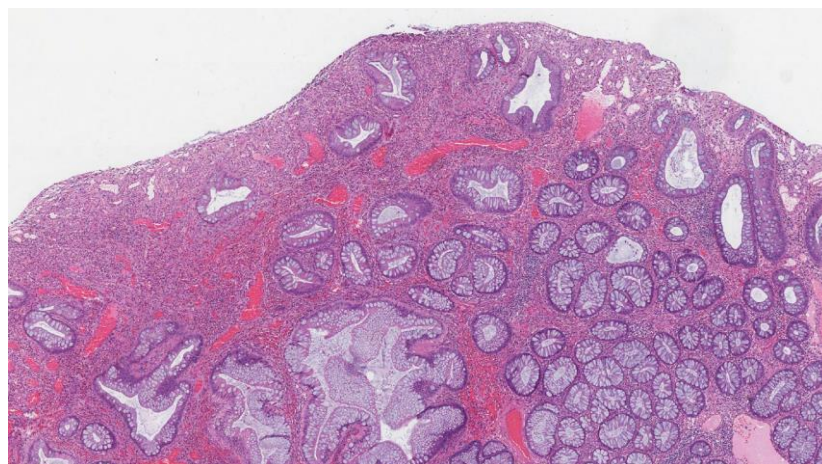


Figure 8: μικροσκοπική δομή κυττάρων φλεγμονώδους πολύποδα

iv. Πολύποδες οδοντωτού τύπου (Serrated) [7]

- Συνδυάζουν ιδιότητες των υπερπλαστικών και αδενωματοδών πολυπόδων. Εντοπίζονται κυρίως ως επίπεδοι που απλώνονται πλατιά πάνω στο τοίχωμα του βλενογόνου και είναι δύσκολο να εντοπιστούν. Πιο σπάνια μπορεί να είναι και πεταλοειδής που όμως εμφανίζονται σε μεγάλα μεγέθη. Η εξωτερική τους όψη είναι “οδοντωτή” και ακανόνιστη και το μέγεθος τους κυμαίνεται από <5mm έως και >1cm.
- Κίνδυνος: Οι παραδοσιακοί οδοντωτοί πολύποδες και οι μεγαλύτεροι σε μέγεθος έχουν αυξημένο κίνδυνο καρκινικής μετατροπής. Οι οδοντωτοί χωρίς δυσπλασία έχουν χαμηλό κίνδυνο κακοήθειας, ενώ οι οδοντωτοί με δυσπλασία έχουν υψηλότερο όπως το sessile serrated adenoma (SSA).

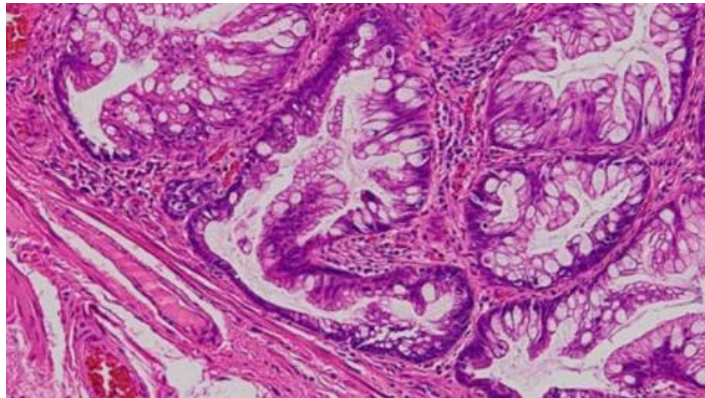


Figure 9: μικροσκοπική δομή sessile serrated lesion (SSL)

v. Αμαρτωματώδεις πολύποδες (Hamartomatous)

- Σχηματίζονται λόγω γενετικών διαταραχών, συχνά σε σύνδρομο πολυποδίασης. Είναι καλά οριοθετημένοι και συχνά έχουν πεταλοειδή μορφή με κοντό στέλεχος. Περιέχουν μίγμα διαφορετικών ιστών (π.χ. βλεννογόνους αδένες, αγγεία) και για αυτό έχουν ασυνήθιστη μορφή. Λόγω αυξημένης αγγείωσης μπορεί να είναι πολύχρωμοι (ροζ ή κόκκινοι) και μεγάλοι σε μέγεθος, ιδιαίτερα σε άτομα με γενετικά σύνδρομα. Παραδείγματα:
 - Σύνδρομο Peutz-Jeghers (PJS): Πολύποδες σε συνδυασμό με μελαγχρωματικές βλάβες στο δέρμα. Ενώ οι πολύποδες είναι κατά κανόνα καλοήθεις, οι ασθενείς έχουν αυξημένο κίνδυνο ανάπτυξης διαφόρων τύπων καρκίνου.
 - Νεανικοί πολύποδες (JPS): Εμφανίζονται κυρίως σε παιδιά και είναι συνήθως καλοήθεις.

Επιπλέον η οικογενής αδενωματώδης πολυποδίαση (FAP) είναι σπάνια γενετική διαταραχή που χαρακτηρίζεται από την ανάπτυξη πολυάριθμων αδενωμάτων σε όλο το παχύ έντερο και στο ορθό. Εάν δεν αφαιρεθούν οι πολύποδες, οι ασθενείς με οικογενή αδενωματώδη πολυποδίαση έχουν εξαιρετικά υψηλό κίνδυνο ανάπτυξης καρκίνου παχέος εντέρου.

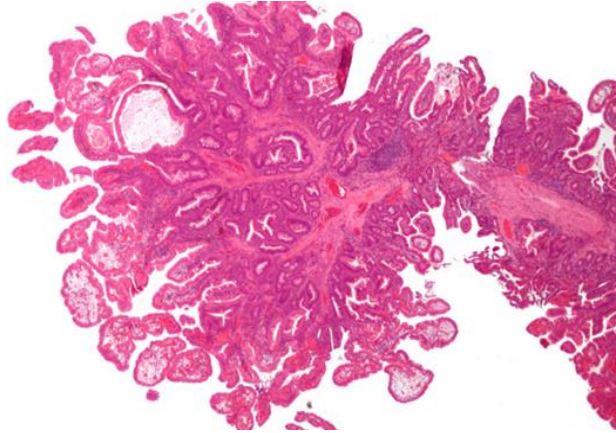


Figure 10: μικροσκοπική δομή αμαρτώματος

2) Πολύποδες και καρκίνος του παχέος εντέρου

Οι περισσότεροι τύποι πολυπόδων, όπως οι υπερπλαστικοί, είναι αβλαβείς. Ωστόσο, άλλοι, όπως τα αδενώματα, τα οποία αποτελούν το 75% των περιπτώσεων, μπορεί να εξελιχθούν σε κακοήθειες, προκαλώντας καρκίνο του παχέος εντέρου και ορθού (ΚΠΕ-Ο/CRC). Αυτός ο καρκίνος είναι ο δεύτερος πιο συχνός στις γυναίκες και ο τρίτος στους άνδρες [8]. Οι πολύποδες που συχνά αποτελούν πρόδρομους του καρκίνου απαντώνται στο 50% των ατόμων άνω των 50 που υποβάλλονται σε προληπτική κωλονοσκόπηση, με τη συχνότητα να αυξάνεται με την ηλικία [9].

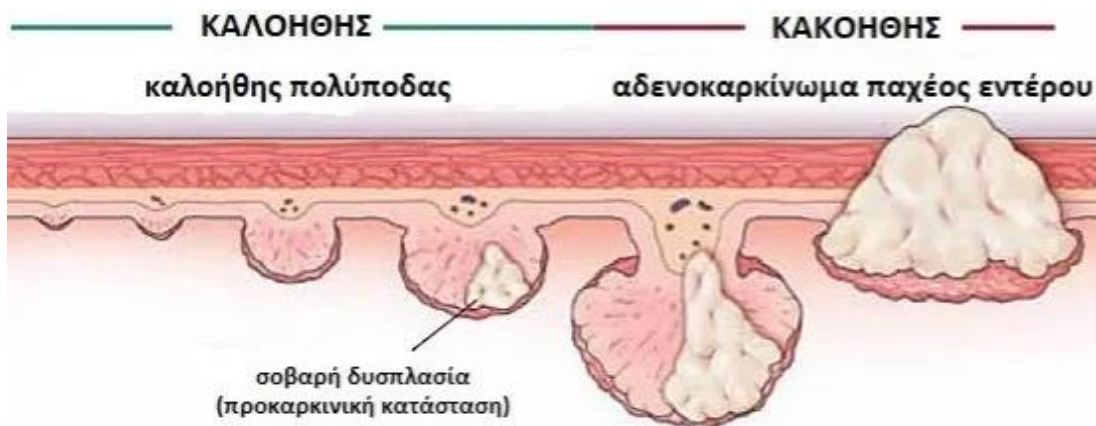


Figure 11: σχεδιάγραμμα εξέλιξης αδενωματώδους πολύποδα παχέος εντέρου σε αδενοκαρκίνωμα

Ο καρκίνος του παχέος εντέρου και ορθού αποτελεί σταθερά την τρίτη κύρια αιτία θανάτου από καρκίνο παγκοσμίως και η επίπτωση του αυξάνεται σταθερά στις αναπτυσσόμενες χώρες. Το 96% των περιπτώσεων καρκίνου του παχέος εντέρου και του ορθού είναι διηθητικά αδενοκαρκίνωμα. Το αδενοκαρκίνωμα προκύπτει από γενετικές και επιγενετικές μεταλλάξεις στα αδενικά, επιθηλιακά κύτταρα που τους παρέχουν επιλεκτικό πλεονέκτημα οδηγεί στην δημιουργία δομών οι οποίες είναι γνωστές ως πολύποδες [10]. Με ασυνήθιστα αυξημένη αναπαραγωγή και επιβίωση οδηγώντας σε υπερπολλαπλασιασμό δημιουργούνται αρχικά καλοήγη αδενώματα, τα οποία μπορεί να εξελιχθούν σε καρκινώματα και να κάνουν μεταστάσεις με την πάροδο των ετών [11]. Συχνά, ο όγκος αναπτύσσεται χωρίς εμφανή συμπτώματα έως ότου επεκταθεί σε βαθύτερα στρώματα του εντέρου ή κάνει μετάσταση μέσω της κυκλοφορίας του αίματος. Ο καρκίνος διηθεί στην συνέχεια τον ιστό, την περίμετρο του εντέρου, τον υποβλεννογόνο και τις υπόλοιπες στοιβάδες [12]. Επίσης, μέσω της κυκλοφορίας του αίματος μπορεί να μεταφέρονται καρκινικά κύτταρα δημιουργώντας μεταστάσεις σε γειτονικές δομές [13]. Μόνο το 10% περίπου όλων των αδενωμάτων εξελίσσεται σε διηθητικό καρκίνο, αν και ο κίνδυνος καρκίνου αυξάνεται όσο ο πολύποδας μεγαλώνει.

3) Επιδημιολογικά Χαρακτηριστικά

Σύμφωνα με τα στοιχεία του Global Cancer Organization (GCO) του Παγκόσμιου Οργανισμού Υγείας για το 2020 ο καρκίνος του παχέος εντέρου ήταν ο τέταρτος συχνότερα εμφανιζόμενος καρκίνος στον κόσμο, ενώ ο καρκίνος του ορθού ο όγδοος [14]. Συνδυαστικά, ο CRC (Colorectal Cancer) αποτελεί την τρίτη πιο συχνά διαγνωσμένη μορφή καρκίνου παγκοσμίως, αποτελώντας το 11% των συνολικών διαγνώσεων [15].

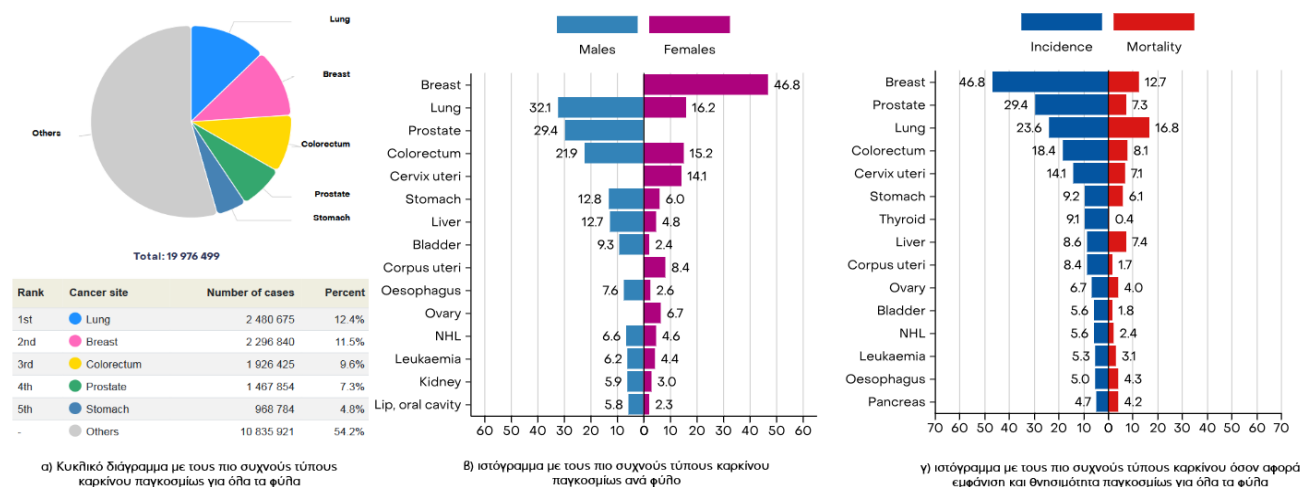


Figure 12: επίσημα επιδημιολογικά στατιστικά από το GCO και τον Παγκόσμιο Οργανισμό Υγείας (2022)

Το 2022 από τα περίπου 19,976,499 νέα περιστατικά καρκίνου που διαγνώστηκαν συνολικά, τα 1,926,425 αφορούσαν CRC (9.6%), γεγονός που τον καθιστά την τρίτη πιο κοινή μορφή καρκίνου που πλήττει και τα δυο φύλα. Αποτελεί και για κάθε φύλο ξεχωριστά την τρίτη πιο

συχνή μορφή καρκίνου με ποσοστό 15.2% στις γυναίκες και 21.9% στους άνδρες επί των

	Males	Females	Both sexes
Population	3 972 735 747	3 912 335 034	7 885 070 781
Mortality			
Number of cancer deaths	5 430 284	4 313 548	9 743 832
Age-standardized mortality rate	109.8	76.9	91.7
Risk of dying from cancer before the age of 75 years (cum. risk %)	11.4	8.0	9.6
Top 3 leading cancers (ranked by deaths)**	Lung Liver Colorectum	Breast Lung Colorectum	Lung Colorectum Liver

Cancer	New cases				Deaths			
	Number	Rank	(%)	Cum.risk	Number	Rank	(%)	Cum.risk
Lung	2 480 675	1	12.4	2.9	1 817 469	1	18.7	2.0
Breast	2 296 840	2	11.5	5.1	666 103	4	6.8	1.4
Colorectum	1 926 425	3	9.6	2.1	904 019	2	9.3	0.84
Prostate	1 467 854	4	7.3	3.7	397 430	8	4.1	0.61
Stomach	968 784	5	4.9	1.1	660 175	5	6.8	0.69
All cancers	19 976 499	-	-	20.0	9 743 832	-	-	9.6
All cancers excl. NMSC	18 741 966	-	-	19.2	9 674 416	-	-	9.6

Figure 13: στατιστικά του 2022 που δείχνουν ότι ο CRC είναι η τρίτη μορφή καρκίνου με τη μεγαλύτερη συχνότητα και ο δεύτερος με μεγαλύτερη θνησιμότητα

συνολικών διαγνώσεων μορφών καρκίνου το συγκεκριμένο έτος παγκοσμίως.

Με βάση τα διαθέσιμα στατιστικά είναι ο καρκίνος που έχει την μεγαλύτερη θνησιμότητα στα δύο φύλα μετά από τον καρκίνο του πνεύμονα. Ο CRC αποτέλεσε την πιο επιθετική μορφή καρκίνου στο Μεξικό, το Ομάν, το Πουέρτο Ρίκο, την γαλλική Γουαδαλούπη και την Σαουδική Αραβία για το 2022.

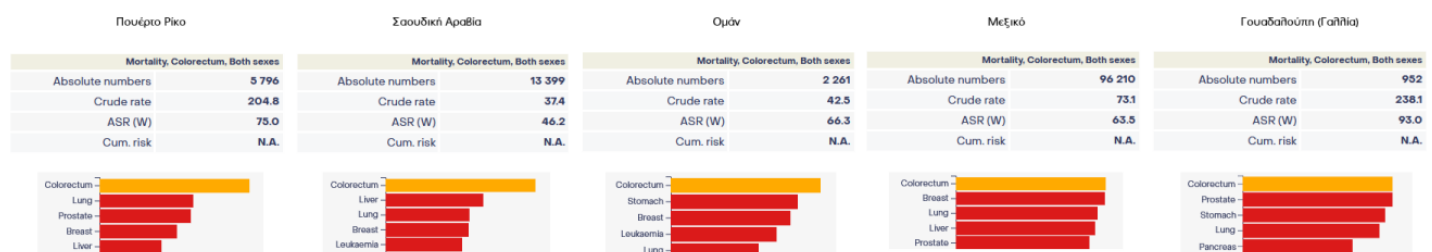


Figure 14: διάγραμμα για τις χώρες που ο καρκίνος του παχέος εντέρου και του ορθού αποτέλεσε την πιο θανατηφόρα μορφή (2022)

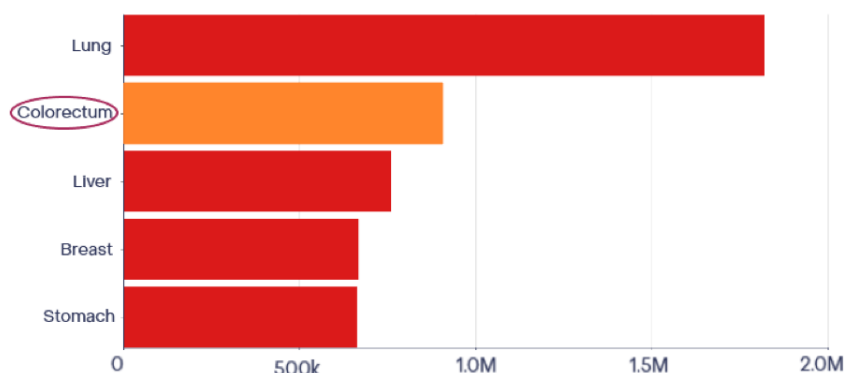


Figure 15: ιστόγραμμα για τους 5 καρκίνους με την μεγαλύτερη θνησιμότητα παγκοσμίως και στα δυο φύλα το 2022

Επιπλέον ο CRC αποτέλεσε το 2022 την δεύτερη μορφή καρκίνου που σε απόλυτα μεγέθη προκάλεσε περισσότερους θανάτους πασχόντων με περίπου 904,019 θανάτους, με την μεγαλύτερη δηλαδή θνησιμότητα στον συνολικό πληθυσμό μετά τον καρκίνο του πνεύμονα με 1,817,469 θανάτους.

Ο CRC είναι η πιο συχνή μορφή καρκίνου που διαγιγνώσκεται στους άνδρες σε 8 χώρες (Ηνωμένα Αραβικά Εμιράτα, Σουλτανάτο του Μπρουνέι, Σιγκαπούρη, Μπαχρέιν, Κατάρ, Υεμένη, Σαουδική Αραβία, Αιθιοπία) από τις 191 χώρες που συμπεριλαμβάνονται στα στατιστικά, ενώ σε καμία χώρα δεν αποτελεί τον πιο διαγνωσμένο καρκίνο μεταξύ των γυναικών. Η επίπτωση της συγκεκριμένης μορφής καρκίνου φαίνεται να είναι μεγαλύτερη στους άνδρες σε σχέση με της γυναίκες και 3-4 φορές μεγαλύτερη στις ανεπτυγμένες από ό,τι στις αναπτυσσόμενες χώρες.

Τα κανονικοποιημένα ηλικιακά (παγκόσμια) ποσοστά εμφάνισης του CRC ανά 100,000 άτομα είναι 18.4 και για τα δύο φύλα, με την θνησιμότητα να ανέρχεται στο 8.1. Στους άνδρες τα ποσοστά εμφάνισης είναι 22.0 και η θνησιμότητα 9.9, ενώ στις γυναίκες είναι 15.2 και 6.5 αντίστοιχα [16].

Επίσης το κανονικοποιημένο ηλικιακό ποσοστό επίπτωσης για όλα τα φύλα είναι 51.1% στα κράτη με πολύ υψηλό δείκτη ανθρώπινης ανάπτυξης (High Human Development Index – HDI), 39.1% για τα κράτη με υψηλό HDI, 7.4% για μέτριο HDI και 2.3% για χαμηλό. Αντίστοιχα τα ποσοστά θνησιμότητας ακολουθούν παρόμοια κατανομή σε σχέση με τον HDI, όπως φαίνεται και στο παρακάτω σχήμα, εικόνα που ενισχύεται με την πάροδο των ετών αν ανατρέξουμε και σε παλαιότερα στατιστικά όπως του 2018 και παλαιότερα.

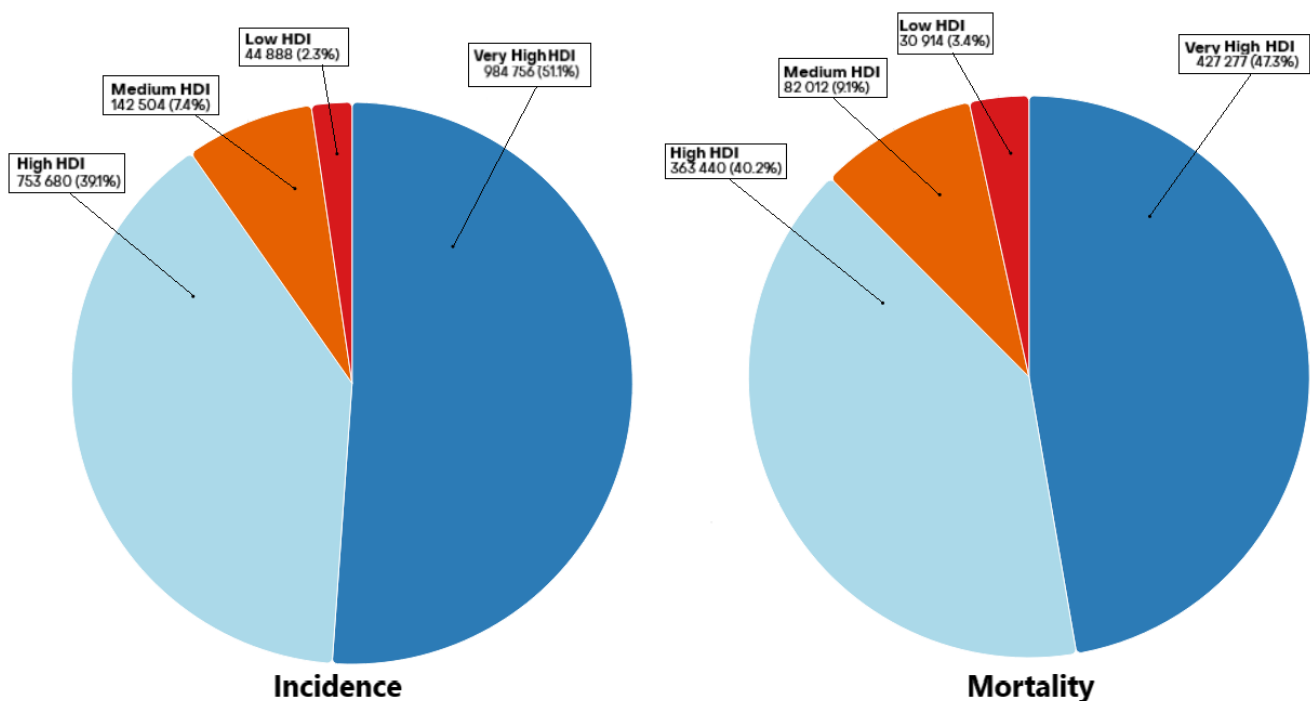


Figure 16: κυκλικό διάγραμμα που παρουσιάζει την εμφάνιση και την θνησιμότητα του CRC λαμβάνοντας υπόψη τον Δείκτη Ανθρώπινης Ανάπτυξης HDI παγκοσμίως για όλα τα φύλα

Η επίπτωση του CRC είναι υψηλότερη σε χώρες με πολύ υψηλό δείκτη ανθρωπίνης ανάπτυξης (HDI), όπως η Αυστραλία και οι ΗΠΑ, και χαμηλότερη σε χώρες με χαμηλό HDI, όπως η Αφρική και η νότια Ασία. Συνολικά, η συχνότητα εμφάνισης του CRC είναι εξαιρετικά μεταβλητή ανά περιοχή με μεγάλες διακυμάνσεις ανάμεσα στις χώρες. Σε χώρες που βρίσκονται σε μείζονα αναπτυξιακή μετάβαση, τα ποσοστά επίπτωσης τείνουν να αυξάνονται ομοιόμορφα με την αύξηση του δείκτη ανθρωπίνης ανάπτυξης (HDI), γεγονός που υποδηλώνει αιτιολογική σχέση μεταξύ των δύο παραμέτρων.

Η Δανία παρουσιάζει τα υψηλότερα ποσοστά εμφάνισης, ενώ η Σιέρα Λεόνε τα χαμηλότερα. Το 2018, η Ουγγαρία είχε τα υψηλότερα ποσοστά θνησιμότητας από καρκίνο του παχέος εντέρου στους άνδρες, και η Σιγκαπούρη στις γυναίκες το 2022.

Η παγκόσμια επιβάρυνση από τον CRC αναμένεται να αυξηθεί κατά 60% μέχρι το 2030, με περισσότερα από 2,2 εκατομμύρια νέα περιστατικά και 1,1 εκατομμύρια θανάτους ετησίως. Αυτή η αύξηση συνδέεται με την οικονομική ανάπτυξη, τις περιβαλλοντικές αλλαγές και τις συνήθειες στις χώρες με χαμηλό έως μεσαίο HDI, καθώς και τις γενεαλογικές αλλαγές στις ήδη ανεπτυγμένες χώρες [17].

4) Παράγοντες που επηρεάζουν την ανάπτυξη CRC

Μετά από πολλές έρευνες που έχουν γίνει ανά τα χρόνια σε πολλούς ασθενείς και διεξοδική μελέτη και σύγκριση του ιστορικού και των συνηθειών τους παγκοσμίως έχει φανεί ότι τα αίτια που φαίνεται να σχετίζονται με την πιθανότητα εξέλιξης πολυπόδων σε καρκίνο του παχέος εντέρου και ορθού χωρίζονται σε δυο κατηγορίες:

i. αμετάβλητοι παράγοντες:

- φύλο: οι άνδρες παγκοσμίως έχουν περίπου 1,5 φορές μεγαλύτερη πιθανότητα να αναπτύξουν CRC σε σχέση με τις γυναίκες [18]. Ωστόσο, οι γυναίκες είναι πιο πιθανό να αναπτύξουν CRC στη δεξιά πλευρά του παχέος εντέρου, που συνδέεται με πιο επιθετικές μορφές νεοπλασίας [19].
- ηλικία: στις ΗΠΑ οι άνω των 65 ετών έχουν περίπου τριπλάσια πιθανότητα εμφάνισης CRC σε σχέση με την ηλικιακή ομάδα 50-64 ετών και σχεδόν τριάντα φορές μεγαλύτερη από την ομάδα 25-49 ετών.
- φυλή/εθνικότητα: πιθανώς σχετίζεται με ανισότητες στην πρόσβαση σε ποιοτική υγειονομική περίθαλψη, προληπτικές εξετάσεις, και υγιεινή διατροφή, παρά με γενετικούς παράγοντες [20].
- κληρονομικότητα: 7-10% των περιπτώσεων σχετίζονται με κληρονομικούς παράγοντες όπως το σύνδρομο Lynch [21]. Άτομα με σύνδρομο αδενωματώδους πολυποδίασης (FAP, MAP) ή αμαρτωματώδους πολυποδίασης (PJS, JPS, PHTS) παρουσιάζουν αυξημένο

κίνδυνο ανάπτυξης [22]. Επίσης συγγενείς πρώτου βαθμού πασχόντων CRC έχουν από 2-4 φορές υψηλότερο κίνδυνο [23].

- φλεγμονώδης νόσος εντέρου (IBD): οι χρόνια ασθενείς έχουν διπλάσιο κίνδυνο να αναπτύξουν CRC, λόγω της παρατεταμένης φλεγμονής που προκαλεί ανώμαλη κυτταρική ανάπτυξη. Οι κύριες αιτίες πίσω από την IBD είναι η ελκώδης κολίτιδα και η νόσος του Crohn [24].

- ακτινοβολία στην κοιλιακή χώρα: οι επιζώντες παιδικού καρκίνου που έχουν υποστεί ακτινοβολία στην κοιλιακή χώρα έχουν αυξημένο κίνδυνο ανάπτυξης CRC ανάλογο με την δόση της ακτινοβολίας [25]. Το ίδιο συμβαίνει και με άτομα που έχουν δεχτεί ακτινοθεραπεία για θεραπεία του καρκίνου του προστάτη και παρουσιάζουν αυξημένο κίνδυνο για καρκίνο του ορθού [26].

- κυστική ίνωση: ασθενείς με κυστική ίνωση έχουν κίνδυνο εμφάνισης CRC 10 φορές υψηλότερο [27].

- χολοκυστεκτομή: σχετίζεται με αυξημένο κίνδυνο CRC στο δεξιό κόλον, πιθανώς λόγω εντερικής έκθεσης στη χολή [28].

- θεραπεία στέρησης ανδρογόνων: ασθενείς με καρκίνο του προστάτη που υποβάλλονται στην θεραπεία αυτή με αγωνιστές εκλυτικής ορμόνης γοναδοτροπινών (GnRH) έχουν αύξηση του κινδύνου να αναπτύξουν CRC από 12-53% ενώ ασθενείς που τους έχει διενεργηθεί ορχεκτομή έχουν αύξηση κινδύνου από 14-66% [29].

i. μεταβλητοί παράγοντες:

- Η παχυσαρκία και η σωματική αδράνεια είναι αθροιστικοί παράγοντες κινδύνου για την ανάπτυξη CRC. Οι σωματικά δραστήριοι άνθρωποι έχουν 25% μικρότερο κίνδυνο, ενώ όσοι ζουν καθιστική ζωή έχουν έως και 50% αυξημένο κίνδυνο. Η αδράνεια μπορεί να οδηγήσει σε παχυσαρκία και να μεταβάλει τη μικροχλωρίδα του εντέρου, προκαλώντας φλεγμονή που μπορεί να συμβάλλει στην καρκινογένεση [30]. Το αυξημένο σωματικό βάρος μπορεί επίσης να διαταράξει τις μεταβολικές διεργασίες, αυξάνοντας την απελευθέρωση μεταλλαξιογόνων ελεύθερων ριζών οξυγόνου.

- διατροφή: οι διατροφικές συνήθειες μπορούν να επηρεάσουν θετικά ή αρνητικά την ανάπτυξη CRC μέσω του μικροβιώματος του παχέος εντέρου. Τροφές όπως το κόκκινο και επεξεργασμένο κρέας αυξάνουν τον κίνδυνο, ενώ τροφές πλούσιες σε ασβέστιο, φυτικές ίνες, βιταμίνη D, φρούτα, και λαχανικά μειώνουν τον κίνδυνο [31].

- κάπνισμα: η χρήση καπνού συνδέεται με αυξημένο κίνδυνο CRC λόγω μεταλλαξιογόνων ουσιών στον καπνό και η διακοπή του καπνίσματος συσχετίζεται με βελτιωμένη επιβίωση [32].

- κατανάλωση αλκοολ: μέτρια έως υπερβολική κατανάλωση αλκοόλ αυξάνει τον κίνδυνο CRC, με 2-3 ποτά ημερησίως να αυξάνουν τον κίνδυνο κατά 20% και 4+ ποτά έως 40% [33].
- φάρμακα: Η μακροχρόνια χρήση μη-στεροειδών αντιφλεγμονωδών φαρμάκων (ΜΣΑΦ) μπορεί να μειώσει τον κίνδυνο CRC, αλλά λόγω κινδύνου γαστρορραγίας δεν συνιστώνται για γενική πρόληψη [34]. Άλλες θεραπείες, όπως τα διφωσφονικά για την οστεοπόρωση [35], σουλινδάκη και οι αναστολείς του ACE για την υπέρταση, έχουν δείξει προστατευτικά αποτελέσματα [36].
- διαβήτης και αντίσταση στην ινσουλίνη: ο σακχαρώδης διαβήτης και κυρίως ο τύπου 2 αυξάνει τον κίνδυνο CRC, πιθανώς λόγω κοινών παραγόντων κινδύνου όπως η παχυσαρκία και η καθιστική ζωή. Η μετφορμίνη έχει αναφερθεί ότι μειώνει τον κίνδυνο, αλλά τα ευρήματα δεν είναι πάντοτε συνεπή [37].
- πρόληψη: η πιο σημαντική παράμετρος που επηρεάζει την ανάπτυξη CRC και είναι καθοριστικής σημασίας για τη μείωση της θνησιμότητας είναι ο έλεγχος, η πρόληψη και η θεραπεία που θα αναλυθούν στην επόμενη ενότητα.

5) Πρόληψη, διάγνωση, αφαίρεση και θεραπεία

Καινοτόμες Θεραπευτικές Προσεγγίσεις στον CRC

Οι πολύποδες του παχέος εντέρου αποτελούν σημαντικό αντικείμενο μελέτης λόγω της άμεσης σχέσης τους με τον καρκίνο του παχέος εντέρου, έναν από τους πιο συχνούς και θανατηφόρους καρκίνους παγκοσμίως. Η ενημέρωση, η πρόληψη και οι τακτικοί διαγνωστικοί έλεγχοι είναι ζωτικής σημασίας για τη μείωση της θνησιμότητας από τη νόσο.

Η μείωση της θνησιμότητας του καρκίνου του παχέος εντέρου έχει επιτευχθεί μέσω καινοτόμων θεραπευτικών προσεγγίσεων, παρά την αυξημένη συχνότητα. Σημαντική πρόοδος σημειώθηκε με τη χρήση διαγνωστικών τεχνικών έγκαιρης ανίχνευσης, όπως η κολονοσκόπηση, η υπολογιστική κολονοσκόπηση (CTC), η σιγμοειδοσκόπηση, η ανάλυση κοπράνων για αίμα, και άλλες μέθοδοι. Αυτές οι τεχνικές, αν και αρχικά οδήγησαν σε αύξηση των ποσοστών διάγνωσης λόγω εντοπισμού προηγούμενων αδιάγνωστων περιπτώσεων, συνέβαλαν στη μείωση της θνησιμότητας μέσω της αφαίρεσης προκαρκινικών και μη μεταστατικών πολυπόδων και της έγκαιρης θεραπείας.

Η πρώιμη ανίχνευση του καρκίνου του παχέος εντέρου έχει σημαντικό αντίκτυπο στην επιβίωση των ασθενών, καθώς επιτρέπει την άμεση αφαίρεση του ιστού, αποτρέποντας έτσι την εξέλιξη και την εξάπλωση του καρκίνου σε άλλα σημεία του σώματος μέσω της κυκλοφορίας του αίματος [38], [39].

Ο Ρόλος της Κολονοσκόπησης

Η κολονοσκόπηση είναι η πιο αξιόπιστη διαγνωστική μέθοδος για την ανίχνευση και αξιολόγηση πολυπόδων του παχέος εντέρου. Χρησιμοποιείται επίσης για την ανεύρεση ελκών,

όγκων, περιοχών φλεγμονής ή αιμορραγίας. Κατά τη διαδικασία, εξετάζεται ολόκληρο το κόλον καθώς και το κατώτερο τμήμα του λεπτού εντέρου. Προτού γίνει η εξέταση, απαιτείται καθαρισμός του παχέος εντέρου, διαδικασία που διαρκεί μία έως δύο ημέρες, ανάλογα με τη μέθοδο που θα επιλεγεί, για να είναι το έντερο καθαρό.

Κατά τη διάρκεια της εξέτασης, ο γαστρεντερολόγος χρησιμοποιεί ένα κολονοσκόπιο, έναν λεπτό, εύκαμπτο σωλήνα με κάμερα που συνδέεται σε οθόνη, επιτρέποντας την εξέταση του εσωτερικού του εντέρου. Το κολονοσκόπιο διαθέτει επίσης κανάλι εργασίας μέσω του οποίου μπορούν να περάσουν εργαλεία για τη λήψη δειγμάτων ιστών (βιοψία) και την αφαίρεση ανώμαλων αναπτύξεων, όπως οι πολύποδες.

Άλλες διαγνωστικές μέθοδοι περιλαμβάνουν την ανάλυση κοπράνων για την ανίχνευση αίματος, η οποία έχει χαμηλή ευαισθησία, τη σιγμοειδοσκόπηση που εξετάζει μόνο το ορθό και το κατώτερο τμήμα του παχέος εντέρου, τον ανοσοχημικό έλεγχο κοπράνων (FIT), και την αξονική κολονογραφία (VC / CTC) με ακτίνες X που δεν μπορεί να εντοπίσει μικροσκοπικούς πολύποδες και δεν επιτρέπει τη λήψη βιοψιών ή πολυπεκτομής [40]. Συνεπώς η κολονοσκόπηση υπερτερεί εναντι των άλλων προληπτικών εξετάσεων για τον καρκίνο του παχέος εντέρου λόγω της ακρίβειας, τους εύρους της περιοχής που εξετάζεται αλλά κυρίως της δυνατότητας άμεσης αφαίρεσης πολυπόδων και λήψης βιοψιών.



Figure 17: κολονοσκόπιο της Olympus GIF-1TH190

Θεραπευτικές Μέθοδοι Αφαίρεσης Πολυπόδων

Ο όρος "καρκίνος του παχέος εντέρου και ορθού" στην πραγματικότητα αναφέρεται σε μια ομάδα νόσων με σημαντική ετερογένεια, προκαλούμενες από ένα ευρύ φάσμα γενετικών μεταλλάξεων και μεταλλαξιογόνων παραγόντων. Λόγω αυτής της ποικιλομορφίας, είναι δύσκολο να σχεδιαστεί μια γενική μοριακή θεραπεία που να καλύπτει όλες τις περιπτώσεις [41]. Η αφαίρεση πολυπόδων μπορεί να πραγματοποιηθεί κατά τη διάρκεια της κολονοσκόπησης όταν οι πολύποδες είναι μικροί, ενώ για μεγαλύτερους πολύποδες μπορεί να απαιτείται χειρουργική επέμβαση. Η χειρουργική παραμένει η κύρια προσέγγιση στην έγκαιρη διάγνωση, αλλά δεν είναι αποτελεσματική σε προχωρημένα στάδια όπου ο καρκίνος έχει μετασταθεί, όπως συμβαίνει στο 25% των περιπτώσεων [42]. Σε τέτοιες περιπτώσεις, η αποτελεσματικότητα των νεοεπιχειρουργικών και κυτταροτοξικών θεραπειών μειώνεται λόγω της ανθεκτικότητας στα φάρμακα και της υποτροπής του καρκίνου [43].

Η επιλογή της μεθόδου αφαίρεσης εξαρτάται από το μέγεθος, τη θέση και τον τύπο του πολύποδα. Οι πιο συνηθισμένες μέθοδοι περιλαμβάνουν:

- 1. Πολυτεκτομή με Βρόγχο (Snare Polypectomy):** χρησιμοποιείται για την αφαίρεση μεσαίων και μεγάλων πολυπόδων
- 2. Ψυχρή Πολυτεκτομή (Cold Snare Polypectomy):** χρησιμοποιείται για την αφαίρεση μικρών πολυπόδων χωρίς ηλεκτροκαυτηρίαση
- 3. Ενδοσκοπική Εκτομή Βλεννογόνου (EMR):** περιλαμβάνει την ανύψωση και αφαίρεση μεγαλύτερων ή επιπέδων πολυπόδων
- 4. Ενδοσκοπική Υποβλεννογόνια Ανατομή (ESD):** χρησιμοποιείται για την αφαίρεση μεγάλων ή διηθητικών πολυπόδων σε ένα κομμάτι
- 5. Χειρουργική Αφαίρεση:** απαιτείται για πολύ μεγάλους ή ύποπτους για κακοήθεια πολύποδες
- 6. Πολυτεκτομή με Λαβίδα (Forceps Polypectomy):** κατάλληλη για μικρούς πολύποδες

Οι πολύποδες αφαιρούνται χωρίς πόνο, ενώ οι κίνδυνοι επιπλοκών, όπως αιμορραγία και διάτρηση, είναι σπάνιοι και ελέγχονται κατάλληλα

6) Εξελίξεις στη Διάγνωση και Σταδιοποίηση

Η εισαγωγή προηγμένων τεχνικών, όπως η εικονική χρωμοενδοσκόπηση, συμβάλλει στη βελτίωση της ακρίβειας διάγνωσης. Αυτή η τεχνική χρησιμοποιεί μηχανικά ή ηλεκτρονικά φίλτρα για να περιορίσει διαφορετικά μήκη κύματος φωτός, ενισχύοντας την απεικόνιση συγκεκριμένων βλαβών. Επιπλέον, η ταξινόμηση NICE [44], η κατηγοριοποίηση Paris [45] για την τυποποίηση της μορφολογίας των πολυπόδων, η ταξινόμηση πλάγιας εξάπλωσης όγκων [46], καθώς και τα κριτήρια WASP [47] για τους οδοντωτούς πολύποδες, παίζουν σημαντικό ρόλο στην βελτίωση της ακρίβειας διάγνωσης.

Μια επιπλέον χρήσιμη τεχνική είναι η έγχυση υγρού κάτω από τον πολύποδα για την ανύψωσή του από τον υποκείμενο ιστό. Αυτή η μέθοδος μειώνει τον κίνδυνο μηχανικής ή ηλεκτροκαυτηριαστικής βλάβης και με την προσθήκη αραιωμένης καρμίνης indigo, διευκολύνει την ακριβή αναγνώριση των ορίων του πολύποδα.

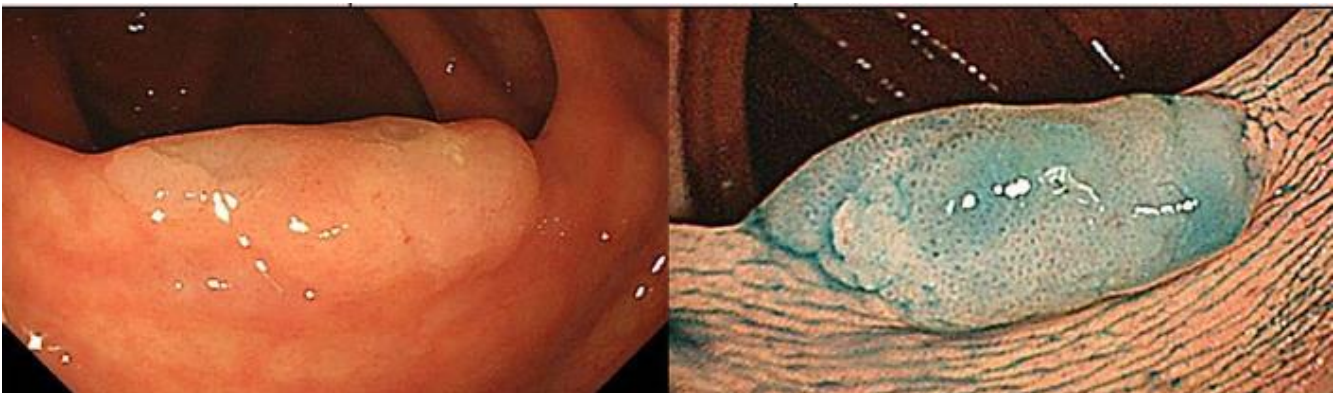


Figure 18: χρωμοενδοσκόπηση με 0.2 % indigo carmine στην αντλία του νερού ή με καθετήρα

Η σταδιοποίηση μέσω του συστήματος TNM της American Joint Committee on Cancer (AJCC) και του Union for International Cancer Control (UICC) είναι κρίσιμη για την πρόγνωση και τη θεραπευτική καθοδήγηση. Το σύστημα TNM βασίζεται σε τρία κλινικά δεδομένα: την έκταση του όγκου (T), την εξάπλωση σε κοντινούς λεμφαδένες (N), και τη μετάσταση σε απομακρυσμένες περιοχές (M) και κατηγοριοποιεί τους όγκους σε στάδια:

- Στάδιο 0: Εντοπισμένο καρκίνωμα (insitu) στο βλεννογόνο
- Στάδιο I: ο όγκος διηθεί το υποβλεννογόνιο ή τον μυϊκό χιτώνα
- Στάδιο IIa: ο όγκος διηθεί τον μυϊκό χιτώνα και επεκτείνεται έως τον υπο-ορογόνο ή εισβάλλει σε γειτονικούς ιστούς στον ενδοπεριτοναϊκό χώρο
- Στάδιο IIb: ο όγκος διηθεί το σπλαχνικό περιτόναιο και/ή απευθείας διηθεί τα όργανα ή τις δομές στον ενδοπεριτοναϊκό χώρο
- Στάδιο IIIa: ο όγκος διηθεί το υποβλεννογόνιο ή τον μυϊκό χιτώνα και έχει εξαπλωθεί σε 1-3 τοπικούς λεμφαδένες
- Στάδιο IIIb: ο όγκος διηθεί τον ορογόνο, το σπλαχνικό περιτόναιο ή γειτονικά όργανα, και έχει εξαπλωθεί σε 1-3 τοπικούς λεμφαδένες
- Στάδιο IIIc: ο όγκος έχει εξαπλωθεί σε 4 ή περισσότερους τοπικούς λεμφαδένες
- Στάδιο IV: ο όγκος έχει εξαπλωθεί σε απομακρυσμένα όργανα και/ή έχει εξαπλωθεί σε τοπικούς λεμφαδένες

Τα συστήματα σταδιοποίησης έχουν εξελιχθεί με την πάροδο των ετών [48] και συνεχίζουν να βελτιώνονται, ενσωματώνοντας νέες μορφολογικές και μοριακές παραμέτρους για την καλύτερη ταξινόμηση των ασθενών [49].

7) Αδύναμα σημεία και αναγκαιότητα τεχνητής νοημοσύνης για Image Segmentation σε εικόνες και βίντεο κολονοσκοπήσεων:

Η έγκαιρη ανίχνευση και αφαίρεση πολυπόδων μέσω προληπτικών ελέγχων αποτελεί θεμέλιο για τη μείωση της θνησιμότητας από καρκίνο του παχέος εντέρου. Η κολονοσκόπηση, ως η πλέον αποτελεσματική μέθοδος, διαδραματίζει καθοριστικό ρόλο τόσο στη διάγνωση όσο και στην άμεση αντιμετώπιση της νόσου.

Παρόλα αυτά δυστυχώς αρκετές μελέτες έχουν δείξει ότι οι πολύποδες συχνά παραβλέπονται κατά τις κολονοσκοπήσεις, με ποσοστά απώλειας πολύποδων 14%-30% ανάλογα με τον τύπο και το μέγεθος τους [50]. Έτσι, η αυτόματη ανίχνευση περισσότερων πολυπόδων σε πρώιμο στάδιο μπορεί να διαδραματίσει κρίσιμο ρόλο στη βελτίωση τόσο της πρόληψης της ασθένειας όσο και της επιβίωσης από τον καρκίνο του παχέος εντέρου [51] και αυτό είναι και το κύριο κίνητρο πίσω από την ανάπτυξη συνόλων ιατρικών δεδομένων όπως το Kvasir-SEG,

αλλά και όλων των μελετών γύρω από το segmentation και το classification ιατρικών εικόνων κολονοσκοπήσεων με την βοήθεια μοντέλων τεχνητής νοημοσύνης.

Η ανίχνευση και η ακριβής οριοθέτηση (segmentation) των πολυπόδων του παχέος εντέρου σε ιατρικές εικόνες αποτελούν κρίσιμα βήματα για τη διάγνωση, τη θεραπεία και την παρακολούθηση της νόσου. Οι συμβατικές μέθοδοι ανάλυσης εικόνας, όπως η οπτική παρατήρηση από ειδικούς γαστρεντερολόγους, είναι χρονοβόρες, υποκειμενικές και μπορούν να οδηγήσουν σε παραλείψεις, ειδικά σε περιπτώσεις μικρών ή δύσκολα αναγνωρίσιμων πολυπόδων. Σε αυτό το πλαίσιο, η τεχνητή νοημοσύνη (AI) και οι τεχνικές βαθιάς μάθησης (Deep Learning) αναδεικνύονται ως ισχυρά εργαλεία για την αυτοματοποιημένη ανάλυση ιατρικών εικόνων με σκοπό:

- ✓ Να αντιμετωπιστεί η υποκειμενικότητα: έχει παρατηρηθεί ότι οι διαγνώσεις που βασίζονται αποκλειστικά στην ανθρώπινη παρατήρηση υπόκεινται σε διαφορές στην εμπειρία και την κρίση του κάθε ιατρού. Ένα μοντέλο τεχνητής νοημοσύνης όταν εκπαιδευτεί κατάλληλα μπορεί να προσφέρει αντικειμενικές και επαναληπτικές αξιολογήσεις, εξαλείφοντας την υποκειμενικότητα και διασφαλίζοντας μεγαλύτερη ακρίβεια στη διάγνωση.
- ✓ Να μπορούν να αναλυθούν σε εύλογο χρόνο μεγάλοι όγκοι δεδομένων: η ευρεία εφαρμογή προληπτικών κολονοσκοπήσεων έχει οδηγήσει σε τεράστιο όγκο εικόνων που απαιτούν ανάλυση και μπορεί να είναι χρήσιμες τόσο για διάγνωση όσο και για επιστημονική μελέτη και εξαγωγή συμπερασμάτων. Τα μοντέλα τεχνητής νοημοσύνης μπορούν να επεξεργαστούν σε πολύ σύντομο χρόνο και με μεγάλη απόδοση μεγάλο αριθμό εικόνων, μειώνοντας έτσι και τον φόρτο εργασίας των γαστρεντερολόγων και τον χρόνο που απαιτείται για τη διάγνωση και επιτρέπει την άμεση λήψη αποφάσεων..
- ✓ Να ανιχνευτούν δύσκολοι και μικροί σε μέγεθος πολύποδες: οι μικροί ή επίπεδοι πολύποδες είναι συχνά δύσκολο να εντοπιστούν λόγω της χαμηλής αντίθεσης και της ομοιότητας τους με τον υποκείμενο ιστό. Τα μοντέλα τεχνητής νοημοσύνης μπορούν μέσω λεπτομερούς ανάλυσης της εισόδου με φίλτρα και προεπεξεργασία, να την βελτιώσουν, να κάνουν εμφανείς τις λεπτομέρειες και να ανιχνεύσουν τους πολύποδες αυτούς με εξαιρετική ακρίβεια.

Η τεχνητή νοημοσύνη δεν έρχεται με σκοπό να αντικαταστήσει τους ειδικούς, αλλά ως ένα εργαλείο που αν το συμβουλευτούν και το χρησιμοποιήσουν σωστά είναι δυνατόν να εμπλουτίσουν την εμπειρία τους, να διευρύνουν την κρίση αλλά και να αποκτήσουν μια “επέκταση” των (βιολογικών) δυνατοτήτων τους, πάντα με σκοπό την πρόοδο της επιστήμης, την προσωπική βελτίωση και το καλό του ασθενούς. Πιο συγκεκριμένα με την τεχνητή νοημοσύνη οι γαστρεντερολόγοι μπορούν να πετύχουν:

- ✓ Αυτόματη και Ακριβής Οριοθέτηση Πολυπόδων: διαχωρισμός πολυπόδων από περιβάλλοντα ιστό που βοηθά στην ακριβή μέτρηση έκτασης του πολύποδα και εκτίμηση της πιθανότητας κακοήθειας

- ✓ Αναγνώριση Προκαρκινικών Αλλοιώσεων: ανίχνευση χαρακτηριστικών που δεν είναι ορατά με γυμνό μάτι ή που δεν αναγνωρίζονται εύκολα όπως μορφολογικές αλλαγές που υποδηλώνουν δυσπλασία
- ✓ Μείωση Ψευδών Θετικών και Ψευδών Αρνητικών: εκπαίδευση σε πολύ μεγάλες βάσεις δεδομένων εικόνων, επιτυγχάνοντας βελτίωση στα ποσοστά ευαισθησίας και ειδικότητας και άρα μειώνουν τις λάθος προβλέψεις (false positive και false negative)
- ✓ Μεγαλύτερη ακρίβεια και ασφάλεια στον εντοπισμό και την αφαίρεση πολυπόδων: με ενίσχυση των ενδοσκοπικών διαδικασιών σε πραγματικό χρόνο με ενσωμάτωση αλγορίθμων τεχνητής νοημοσύνης σε συστήματα ενδοσκόπησης, παρέχοντας έτσι υποστήριξη σε πραγματικό χρόνο κατά τη διάρκεια της εξέτασης.
- ✓ Εκπαίδευση και Υποστήριξη Ιατρών: οι αλγόριθμοι τεχνητής νοημοσύνης μπορούν να λειτουργήσουν ως εκπαιδευτικά εργαλεία για νέους ιατρούς, προσφέροντας συνεχείς και στοχευμένες συστάσεις και καθοδήγηση κατά την ενδοσκοπική διαδικασία.

Η τεχνητή νοημοσύνη έχει τη δυνατότητα να αναδιαμορφώσει τη διάγνωση και θεραπεία των πολυπόδων του παχέος εντέρου, ωστόσο υπάρχουν και περιορισμοί, όπως η ανάγκη για μεγάλες και καλά ανεπτυγμένες βάσεις δεδομένων, ζητήματα ηθικής και απορρήτου σχετικά με τη χρήση ευαίσθητων ιατρικών δεδομένων και η αδυναμία πλήρους αντικατάστασης της ανθρώπινης κρίσης. Συμπερασματικά, οι μέθοδοι image segmentation με τεχνητή νοημοσύνη μπορούν να προσφέρουν σημαντική προστιθέμενη αξία στην κλινική πρακτική, βελτιώνοντας την ακρίβεια, την ταχύτητα και την αποτελεσματικότητα της διάγνωσης και θεραπείας πολυπόδων, και συμβάλλουν καθοριστικά στη μείωση της θνησιμότητας από καρκίνο του παχέος εντέρου.

III. ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

1) Η συμβολή της τεχνητής νοημοσύνης

Η ραγδαία ανάπτυξη της τεχνητής νοημοσύνης (TN) έχει επιτρέψει στους υπολογιστές να εκτελούν διαδικασίες με ακρίβεια και λεπτομέρεια που ξεπερνούν τις ανθρώπινες δυνατότητες παρατήρησης. Στον τομέα της κολονοσκόπησης, οι αλγόριθμοι TN, και ιδιαίτερα τα συνελκτικά νευρωνικά δίκτυα (CNN) που εκπαιδεύονται σε παραδείγματα εικόνων και βίντεο, διαδραματίζουν κρίσιμο ρόλο στην τμηματοποίηση εικόνων για τον εντοπισμό και την ανίχνευση μεγαλύτερου ποσοστού πολυπόδων, συμβάλλοντας έτσι στην πρόληψη του καρκίνου του παχέος εντέρου.

Η τμηματοποίηση εικόνας (image segmentation) είναι η διαδικασία διαίρεσης των εικόνων σε περιοχές ενδιαφέροντος (ROI) που είναι ευκολότερες στην ανάλυση και ερμηνεία. Η τεχνητή νοημοσύνη επιτρέπει την αυτοματοποίηση αυτής της διαδικασίας, βελτιώνοντας την παρακολούθηση της παθολογίας και την ακρίβεια της διάγνωσης, μειώνοντας την ανάγκη για χειρωνακτική παρέμβαση [52]. Παρά την πρόοδο, η χειροκίνητη τμηματοποίηση από ειδικούς παραμένει σημείο αναφοράς για εφαρμογές όπως η μαγνητική τομογραφία (MRI), διαδικασία χρονοβόρα και υποκειμενική. Υπάρχει ανάγκη για αυτοματοποιημένες και αποτελεσματικές μεθόδους τμηματοποίησης εικόνας λόγω της δυσκολίας συλλογής και σχολιασμού μεγάλων όγκων δεδομένων υψηλής ποιότητας.

Οι αλγόριθμοι μηχανικής μάθησης (ML) έχουν αποδείξει την αποτελεσματικότητά τους στην τμηματοποίηση ιατρικών εικόνων. Αν και οι αλγόριθμοι βαθιάς μάθησης (DL) προσφέρουν υψηλή ακρίβεια, απαιτούν μεγάλες ποσότητες επισημασμένων δεδομένων, τα οποία είναι δαπανηρά και χρονοβόρα για συλλογή. Οι ερευνητές έχουν αναπτύξει διάφορες αρχιτεκτονικές DL, όπως τα U-Net και FCN, για τη βελτίωση της τμηματοποίησης πολυπόδων. Αν και αυτές οι μέθοδοι επιτυγχάνουν ακριβή αποτελέσματα, μπορεί να δυσκολεύονται να διαχειριστούν την ποικιλία των χαρακτηριστικών των πολυπόδων λόγω των περιορισμένων δεδομένων που έχουν να εκπαιδευτούν. Συνολικά, η TN προσφέρει σημαντικά οφέλη στην ιατρική απεικόνιση, αν και προκλήσεις όπως η ανάγκη για δεδομένα και ηθικά ζητήματα παραμένουν.

Η ανάπτυξη τέτοιων αυτοματοποιημένων μεθόδων αποτελεί πρόκληση, καθώς οι σύγχρονες τεχνικές μηχανικής εκμάθησης απαιτούν μεγάλους όγκους σχολιασμένων και ετικετοποιημένων δεδομένων υψηλής ποιότητας, τα οποία είναι εξαιρετικά δύσκολα. Ο σχολιασμός των ιατρικών εικόνων, όπως αυτές που περιλαμβάνουν πολύποδες, είναι μια επίπονη διαδικασία που απαιτεί χρόνο, πόρους και την απασχόληση εξειδικευμένων επαγγελματιών, όπως γαστρεντερολόγοι, οι οποίοι συχνά δεν είναι εύκολα διαθέσιμοι. Παράλληλα λόγω της φύσης των δεδομένων υπάρχουν προκλήσεις σχετικά με τη συλλογή ιατρικών εικόνων όσον αφορά την διασφάλιση της ιδιωτικότητας και τη προστασία των δεδομένων των ασθενών που καθιστούν ακόμα πιο δύσκολη τη συλλογή και διαχείριση αυτών των δεδομένων. Πολλοί επιστήμονες υπογραμμίζουν την ανάγκη δημιουργίας συνόλων

δεδομένων με σχολιασμούς και ground truth masks συμβατές με τα σύγχρονα ιατρικά πρότυπα [53].

Η τεχνητή νοημοσύνη (TN) ως επιστημονικό πεδίο εστιάζει στον σχεδιασμό και την ανάπτυξη συστημάτων που εκτελούν εργασίες που απαιτούν ανθρώπινη νοημοσύνη, όπως η αυτόματη μάθηση, η αναγνώριση προτύπων, η κατανόηση της φυσικής γλώσσας, η αναγνώριση φωνής και η ανάλυση εικόνων. Χρησιμοποιώντας συνδυαστικά τεχνικές όπως τα νευρωνικά δίκτυα, η μηχανική όραση και η μηχανική μάθηση, η τεχνητή νοημοσύνη επιτρέπει στα υπολογιστικά συστήματα να αναλύουν δεδομένα, να εξάγουν συμπεράσματα και να λαμβάνουν αποφάσεις σε πολύπλοκα περιβάλλοντα. Η διαδικασία εργασίας της τεχνητής νοημοσύνης συχνά περιλαμβάνει τις εξής βασικές φάσεις:

- Συλλογή Δεδομένων: Τα μοντέλα χρειάζονται μεγάλο όγκο δεδομένων για να εκπαιδευτούν. Αυτά τα δεδομένα μπορεί να είναι κείμενα, εικόνες, ήχοι κλπ ανάλογα με τις ανάγκες της εφαρμογής
- Επεξεργασία Δεδομένων: Τα δεδομένα αυτά υπόκεινται σε διάφορες διαδικασίες επεξεργασίας, όπως καθαριότητα, προσπέλαση και μετατροπή τους σε μορφή κατάλληλη για τη χρήση από μοντέλα
- Εκπαίδευση Μοντέλου: ένα μοντέλο εκπαιδεύεται χρησιμοποιώντας τα επεξεργασμένα δεδομένα, με σκοπό να μάθει τα πρότυπα και τις σχέσεις μεταξύ των δεδομένων
- Αξιολόγηση Μοντέλου: το εκπαιδευμένο μοντέλο αξιολογείται χρησιμοποιώντας δεδομένα ελέγχου για να εκτιμήσει την απόδοσή του
- Ανάπτυξη και Εφαρμογή: αν το μοντέλο είναι ικανοποιητικό, τότε εφαρμόζεται σε πραγματικά προβλήματα.

Συνολικά, η TN συνδυάζει αλγορίθμους και τεχνικές για να προσομοιώσει ανθρώπινες διαδικασίες σκέψης, παρέχοντας λύσεις σε σύνθετα προβλήματα.

2) Συνελικτικά Νευρωνικά Δίκτυα

Στην παρούσα διπλωματική εργασία γίνεται χρήση συνελικτικών νευρωνικών δικτύων (Convolutional Neural Networks – CNN), μια κατηγορία νευρωνικών δικτύων βαθιάς μάθησης που έχουν σχεδιαστεί ειδικά για ανάλυση και επεξεργασία εικόνων. Τα CNN είναι ικανά να ανιχνεύουν και να μαθαίνουν πρότυπα και χαρακτηριστικά σε εικόνες που περιλαμβάνουν εκατομμύρια εικονοστοιχεία (pixel). Ένα τυπικό CNN που χρησιμοποιείται για την τμηματοποίηση εικόνων περιλαμβάνει τα εξής στάδια:

- Είσοδος Εικόνας: η εικόνα εισάγεται στο CNN ως αρχικό δεδομένο
- Συνέλιξη (Convolution): εφαρμόζει φίλτρα για την εξαγωγή χαρακτηριστικών από διάφορες περιοχές της εικόνας. Αυτά τα χαρακτηριστικά μπορεί να περιλαμβάνουν γραμμές, ακμές, γωνίες ή άλλα βασικά μοτίβα

- Συνάρτηση Ενεργοποίησης (Activation Function): μετά από κάθε συνέλιξη, χρησιμοποιείται μια συνάρτηση ενεργοποίησης, όπως η ReLU ή η σιγμοειδής, για την εισαγωγή μη γραμμικότητας στο δίκτυο, επιτρέποντας την εκμάθηση σύνθετων σχέσεων
- Υποδειγματοληψία (Pooling): μειώνεται η διάσταση των δεδομένων για τη μείωση της πολυπλοκότητας, διατηρώντας τα βασικά χαρακτηριστικά. Συχνά χρησιμοποιούνται τεχνικές όπως η μέγιστη ή η μέση υποδειγματοληψία.
- Επαναλαμβανόμενες Στρώσεις Συνέλιξης και Υποδειγματοληψίας: αυτά τα στάδια επαναλαμβάνονται σε διαδοχικά επίπεδα, με περισσότερα φίλτρα και αυξημένη πολυπλοκότητα, ώστε να εξάγονται πιο λεπτομερή και πολύπλοκα χαρακτηριστικά.
- Αποσυμπίεση (Upsampling): η εικόνα ανακατασκευάζεται στο αρχικό της μέγεθος και χρησιμοποιούνται μέθοδοι, όπως η αντίστροφη υποδειγματοληψία για ανάκτηση της αρχικής χωρικής πληροφορίας
- Τελική Έξοδος (Output Layer): παράγεται ένας χάρτη πιθανοτήτων για κάθε κατηγορία της τμηματοποίησης και χρησιμοποιούνται συνήθως συναρτήσεις όπως η Softmax για την απόδοση των πιθανοτήτων.

Κατά τη διαδικασία εκπαίδευσης, το CNN προσαρμόζει τα βάρη των φίλτρων και των επιπέδων του, αξιοποιώντας δεδομένα που έχουν ήδη επισημανθεί. Με την ολοκλήρωση της εκπαίδευσης, το δίκτυο είναι σε θέση να επεξεργάζεται νέες εικόνες, επιτυγχάνοντας τμηματοποίηση με υψηλή ακρίβεια.

3) Διαίρεση του συνόλου δεδομένων για τις ανάγκες του νευρωνικού δικτύου

Η διαίρεση του συνόλου δεδομένων σε τρία κύρια μέρη - εκπαίδευσης, επικύρωσης και δοκιμής - αποτελεί θεμελιώδες βήμα για την ανάπτυξη αποδοτικών μοντέλων μηχανικής μάθησης. Αυτή η διαδικασία διασφαλίζει ότι το μοντέλο μπορεί να μάθει από τα δεδομένα, να ρυθμίσει τις παραμέτρους του σωστά, και να αξιολογηθεί αντικειμενικά.

– σύνολο δεδομένων εκπαίδευσης (training dataset): Περιλαμβάνει δείγματα που χρησιμοποιούνται για την εκπαίδευση του μοντέλου. Κατά τη διάρκεια της εκπαίδευσης, το μοντέλο μαθαίνει να αναγνωρίζει μοτίβα και σχέσεις στα δεδομένα. Μέσω της σύγκρισης των προβλέψεων με τις πραγματικές τιμές, υπολογίζεται το σφάλμα, και οι αλγόριθμοι, όπως το Gradient Descent, προσαρμόζουν τις παραμέτρους για τη βελτίωση της ακρίβειας. Το μοντέλο "μαθαίνει" να κάνει προβλέψεις ή να λαμβάνει αποφάσεις χωρίς να είναι ρητά προγραμματισμένο για συγκεκριμένες εργασίες.

– σύνολο δεδομένων επικύρωσης (validation dataset): Χρησιμοποιείται για την αξιολόγηση της απόδοσης του μοντέλου σε νέα δεδομένα κατά τη διάρκεια της εκπαίδευσης. Επιτρέπει τη ρύθμιση των παραμέτρων και υπερπαραμέτρων για τη βελτίωση της γενίκευσης του μοντέλου, αποφεύγοντας την υποπροσαρμογή (όπου το μοντέλο δεν μαθαίνει αρκετά) και

την υπερπροσαρμογή (όπου το μοντέλο μαθαίνει μόνο τα συγκεκριμένα μοτίβα των δεδομένων εκπαίδευσης).

– σύνολο δεδομένων δοκιμής (testing dataset): Είναι πλήρως ανεξάρτητο από τα δεδομένα εκπαίδευσης και επικύρωσης και χρησιμοποιείται για την τελική αξιολόγηση του μοντέλου. Παρέχει μια αντικειμενική εικόνα για την απόδοση του μοντέλου σε νέα, άγνωστα δεδομένα, αντανakλώντας την ικανότητά του να λειτουργεί σε πραγματικά περιβάλλοντα.

Τα δεδομένα εκπαίδευσης και επικύρωσης μπορούν να παίξουν πρόσθετους ρόλους στην επιλογή χαρακτηριστικών, επιλέγοντας τις πιο σχετικές ή σημαντικές μεταβλητές για τη βελτίωση της απόδοσης του μοντέλου. Παράλληλα, διευκολύνουν τη ρύθμιση της πολυπλοκότητας του μοντέλου, ώστε να επιτευχθεί ισορροπία μεταξύ ακρίβειας και γενίκευσης. Η προσαρμογή του τελικού μοντέλου είναι αποτέλεσμα του συνδυασμού αυτών των εισροών.

Η τμηματοποίηση εικόνων, ιδιαίτερα στον τομέα της ιατρικής, διαχωρίζει τις εικόνες σε Περιοχές Ενδιαφερόντων (ROI) που είναι πιο εύκολες στην ανάλυση και ερμηνεία. Αυτό συμβάλλει στην παρακολούθηση παθολογιών, όπως οι πολύποδες, αυξάνοντας τη διαγνωστική ακρίβεια και μειώνοντας την ανάγκη για χειρωνακτική παρέμβαση. Ωστόσο, η ακριβής τμηματοποίηση εικόνων αποτελεί πρόκληση λόγω της ποικιλομορφίας στα μεγέθη, τα χρώματα και τις υφές των πολύποδων, καθώς και των δυσδιάκριτων ορίων μεταξύ των περιοχών ενδιαφέροντος και του περιβάλλοντος ιστού.

Οι μέθοδοι σημασιολογικής τμηματοποίησης με βάση εικονοστοιχεία ενσωματώνονται σε συστήματα που είναι ακριβή, γρήγορα και οικονομικά αποδοτικά, προσφέροντας σημαντικές βελτιώσεις στις υπολογιστικά υποστηριζόμενες παρεμβάσεις. Η ανάπτυξη μοντέλων μηχανικής μάθησης περιλαμβάνει την αρχική ρύθμιση παραμέτρων έργου όπως το πρόβλημα που επιδιώκει να λύσει το μοντέλο, ο τύπος των δεδομένων προς επεξεργασία, και οι μετρήσεις απόδοσης που στοχεύει να βελτιστοποιήσει.

Η διαδικασία εκπαίδευσης μοντέλων μηχανικής μάθησης δημιουργεί ένα δίκτυο συνδέσεων μεταξύ επιμέρους στοιχείων του μοντέλου, γνωστών ως "νευρώνες" στα νευρωνικά δίκτυα. Η αρχική ρύθμιση σταθμίσεων μεταξύ των συνδέσεων αυτών δεν είναι στατική, αλλά τροποποιείται κατά τη διάρκεια της εκπαίδευσης, βασιζόμενη στην ανατροφοδότηση από την απόδοση του μοντέλου. Μετά την εισαγωγή δεδομένων εκπαίδευσης, οι προγραμματιστές συγκρίνουν το αποτέλεσμα με τις στοχευόμενες απαντήσεις. Αυτή η σύγκριση καθοδηγεί τη συνάρτηση σφάλματος, η οποία ποσοτικοποιεί την απόδοση του μοντέλου - την απόκλιση μεταξύ των προβλέψεων και των πραγματικών τιμών.

Οι παράμετροι και η λειτουργικότητα του μοντέλου προσαρμόζονται μέσω αλγορίθμων βελτιστοποίησης, με τον Gradient Descent να είναι ένας από τους πιο συνηθισμένους. Συχνά απαιτούνται πολλαπλές "εποχές" ή επαναλήψεις αυτής της διαδικασίας. Ο στόχος είναι να δημιουργηθούν μοντέλα που μπορούν να κάνουν ακριβείς προβλέψεις όταν εκτίθενται σε νέα, άγνωστα δεδομένα.

Για να γενικεύσει καλά ένα μοντέλο μηχανικής μάθησης, πρέπει να επιτευχθεί ισορροπία κατά τη διάρκεια της εκπαίδευσης για να αποφευχθεί η υποπροσαρμογή (όπου το μοντέλο είναι πολύ απλό για να αποτυπώσει το υποκείμενο μοτίβο) ή η υπερβολική προσαρμογή (όπου το μοντέλο είναι υπερβολικά περίπλοκο και καταγράφει τον θόρυβο στα δεδομένα εκπαίδευσης).

Η κατάτμηση εικόνων σε Περιοχές Ενδιαφέροντος (ROI) είναι κρίσιμη για την παρακολούθηση της παθολογίας και τη βελτίωση της διαγνωστικής ικανότητας, αυξάνοντας την ακρίβεια και μειώνοντας τη χειρωνακτική παρέμβαση. Οι μέθοδοι σημασιολογικής τμηματοποίησης βάσει εικονοστοιχείων έχουν τη δυνατότητα να γίνουν μέρος γρήγορων, ακριβών και οικονομικά αποδοτικών συστημάτων.

Η ακριβής κατάτμηση των πολυπόδων αποτελεί πρόκληση για δύο κύριους λόγους: (i) ο ίδιος τύπος πολύποδων εμφανίζει ποικιλία στο μέγεθος, το χρώμα και την υφή, και (ii) το όριο μεταξύ ενός πολύποδα και του περιβάλλοντος.

IV. ΣΤΟΙΧΕΙΑ ΠΕΙΡΑΜΑΤΟΣ

1) Εργαλεία που χρησιμοποιήθηκαν και Περιβάλλον Εργασίας:

Για την συγκεκριμένη διπλωματική εργασία χρησιμοποιήθηκε η πλατφόρμα Google Colaboratory (Colab Pro), με χρήση του περιβάλλοντος Python 3.11.11 (ipyb). Ο πλήρης κώδικας και τα αποτελέσματα του πειράματος είναι διαθέσιμα στο αποθετήριο GitHub: <https://github.com/adouska/Thesis> και στο online notebook:

<https://colab.research.google.com/drive/1qA29e1Bk0dU55kFovvKw-fnl53J7nm3a?usp=sharing>

Για την εκπαίδευση των μοντέλων χρησιμοποιήθηκε μια συνδυασμός των μονάδων επεξεργασίας γραφικών (GPU) A100 και T4, ανάλογα με τις απαιτήσεις των μοντέλων. Η A100 προσφέρει υψηλότερη απόδοση και είναι ιδανική για σύνθετες εφαρμογές μηχανικής μάθησης, ενώ η T4 είναι πιο οικονομική και κατάλληλη για λιγότερο απαιτητικές εφαρμογές τεχνητής νοημοσύνης.

NVIDIA Tesla T4: [54]

Αρχιτεκτονική	Turing
Μνήμη	40 GB HBM2
CUDA Cores	2,560
Τύποι εργασιών	Βασίζεται σε επιταχυντές τεχνητής νοημοσύνης, ιδανική για εντατικές εργασίες εκπαίδευσης μηχανικής μάθησης και deep learning.
Χρησιμοποιείται για	Εκπαίδευση και ανάλυση μοντέλων μηχανικής μάθησης (ML), υποστήριξη για inference και εκτέλεση μοντέλων TensorFlow και PyTorch.
Πλεονεκτήματα	Πολύ καλό για αναγνώριση εικόνας, επεξεργασία βίντεο, και άλλες εργασίες που απαιτούν επιταχυντές GPU.

NVIDIA A100: [55]

Αρχιτεκτονική	Ampere
Μνήμη	16 GB GDDR6
CUDA Cores	6,912
Τύποι εργασιών	Σχεδιασμένη για εντατικές εργασίες AI και deep learning, υψηλής απόδοσης υπολογιστές για training μεγάλης κλίμακας μοντέλων και ανάλυση δεδομένων.
Χρησιμοποιείται για	Υποστηρίζει τόσο training όσο και inference για εξαιρετικά μεγάλα και περίπλοκα μοντέλα τεχνητής νοημοσύνης. Εξαιρετική απόδοση σε έργα HPC (High Performance Computing).
Πλεονεκτήματα	Παρέχει εξαιρετική ταχύτητα και κλιμακωσιμότητα, καθιστώντας την ιδανική για μεγάλα μοντέλα και εντατικά υπολογιστικά φορτία.

Τα δεδομένα αποθηκεύονται στο Google Drive και φορτώνονται στο notebook με κατάλληλες εντολές. Οι παράμετροι (weights) των μοντέλων, μετά την ολοκλήρωση της εκπαίδευσης, αποθηκεύονται επίσης στο Google Drive για μελλοντική αναφορά.

Για την ανάπτυξη του κώδικα χρησιμοποιήθηκαν διάφορα μοντέλα και μετρικές από το MONAI και κυρίως το PyTorch. Το MONAI (Medical Open Network for AI) είναι ένα πλαίσιο ανοιχτού κώδικα (open source) για τη δημιουργία εφαρμογών και μοντέλων τεχνητής νοημοσύνης στον τομέα της ιατρικής εικόνας. Το MONAI παρέχει βελτιστοποιημένα εργαλεία και μοντέλα για τμηματοποίηση, ανίχνευση και ανάλυση ιατρικών εικόνων, υποστηρίζοντας τη σύγχρονη έρευνα και τις εφαρμογές στην ιατρική.

Το PyTorch είναι μια δημοφιλής βιβλιοθήκη ανοιχτού κώδικα για την εκπαίδευση και ανάπτυξη μοντέλων μηχανικής μάθησης και βαθιάς μάθησης. Παρέχει δυναμικά υπολογιστικά γραφήματα, επιτρέποντας στους ερευνητές να αναπτύξουν γρήγορα μοντέλα και να τα εκπαιδεύσουν σε σύγχρονα δεδομένα. Ο συνδυασμός αυτών των εργαλείων επιτρέπει την ευέλικτη και αποδοτική ανάπτυξη μοντέλων για πολύπλοκες εργασίες, όπως η ανάλυση ιατρικών εικόνων.

Στη βαθιά μάθηση χρησιμοποιούνται αλγόριθμοι βελτιστοποίησης (optimizers) που προσαρμόζουν τις παραμέτρους του μοντέλου κατά την εκπαίδευση για την ελαχιστοποίηση της συνάρτησης απωλειών [56]. Επιτρέπουν στα νευρωνικά δίκτυα να μαθαίνουν από τα δεδομένα ενημερώνοντας επαναληπτικά τα βάρη και τις προκαταλήψεις. Στην ανάπτυξη του κώδικα χρησιμοποιήθηκε ο Adam optimizer (Adaptive Moment Estimation) προσφέροντας υπολογιστική αποδοτικότητα και αποτελεσματικότητα.

Ως συνάρτηση κόστους (loss function) χρησιμοποιήθηκε η Dice Loss [57] από την βιβλιοθήκη του Monai λόγω της καταλληλότητας της για τμηματοποίηση εικόνων. Η Dice Loss επικεντρώνεται στη μέγιστη επικάλυψη μεταξύ των προβλεπόμενων και των πραγματικών περιοχών, αντιμετωπίζοντας τη μη ισορροπημένη κατανομή των δεδομένων, που είναι συχνή σε εφαρμογές που η περιοχή ενδιαφέροντος (π.χ., πολύποδες, όγκοι ή άλλες ανωμαλίες) καταλαμβάνει μικρό ποσοστό της συνολικής εικόνας.

Όσον αφορά την προετοιμασία των δεδομένων, ενώ συχνά τα δεδομένα χωρίζονται τυχαία σε σύνολα εκπαίδευσης και δοκιμής, στην παρούσα εργασία αποφασίστηκε να ελεγχθεί η επίδραση των μοντέλων σε σταθερά σύνολα δεδομένων. Συγκεκριμένα, τα δεδομένα χωρίστηκαν εξ αρχής με σταθερό τρόπο, ώστε να εξασφαλιστεί η ακριβής σύγκριση των αποτελεσμάτων και να αξιολογηθεί η απόδοση των μοντέλων υπό τις ίδιες συνθήκες.

2) Σύνολο δεδομένων Kvasir-SEG:

Η τμηματοποίηση εικόνας βάσει εικονοστοιχείων (pixels) αποτελεί μια ιδιαίτερα απαιτητική διαδικασία στην ανάλυση ιατρικών εικόνων, και οι σχολιασμένες ιατρικές εικόνες με αντίστοιχες μάσκες κατάτμησης είναι εξαιρετικά σπάνιες. Το Kvasir-SEG είναι ένα σύνολο δεδομένων ανοιχτής πρόσβασης (open source) που περιλαμβάνει 1,000 εικόνες γαστρεντερικών πολύποδων και τις αντίστοιχες μάσκες κατάτμησης, σχολιασμένες και

επαληθευμένες χειροκίνητα από έμπειρους γαστρεντερολόγους από το Vestre Viken Health Trust (Νορβηγία).

Το σύνολο δεδομένων Kvasir-SEG δημιουργήθηκε το 2019 στη Νορβηγία, ως αποτέλεσμα της συνεργασίας των Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen και Håvard D. Johansen, με στόχο τη διευκόλυνση της ανάπτυξης και της συγκριτικής αξιολόγησης μεθόδων υπολογιστικής όρασης και μηχανικής μάθησης στον τομέα της ανάλυσης ενδοσκοπικών δεδομένων από κολονοσκοπήσεις.

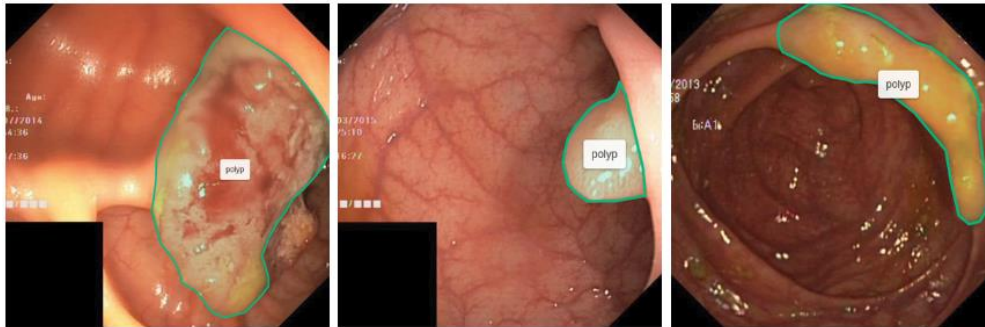


Figure 19: παραδείγματα εικόνων από το σύνολο δεδομένων Kvasir όπου με πράσινο επισημαίνεται το περίγραμμα του ιστού των πολύποδων

Το Kvasir-SEG βασίζεται στο προηγούμενο σύνολο δεδομένων Kvasir [58] το οποίο ήταν το πρώτο σύνολο δεδομένων πολλαπλών τάξεων για την ανίχνευση και ταξινόμηση ασθενειών του γαστρεντερικού συστήματος (GI) και περιλάμβανε 8,000 εικόνες από διάφορες κατηγορίες γαστρεντερικών ευρημάτων. Στο Kvasir-SEG, το επίκεντρο είναι οι πολύποδες, με τις εικόνες να είναι κωδικοποιημένες με συμπίεση JPEG και ανάλυση από 332x487 έως 1920x1072 pixel [59]. Οι μάσκες κατάτμησης δημιουργήθηκαν με τη χρήση εργαλείων όπως το Labelbox [60], όπου ειδικοί σχεδίασαν περιοχές ενδιαφέροντος (ROI) και παράγαγαν δυαδικές μάσκες (binary masks) που απεικονίζουν τον ιστό του πολύποδα με λευκό (foreground) και το υπόλοιπο φόντο με μαύρο (background).

Τα δεδομένα είναι αποθηκευμένα σε ξεχωριστούς φακέλους για τις εικόνες και τις μάσκες, με κοινή ονοματοδοσία για εύκολη συσχέτιση. Επιπλέον, περιλαμβάνεται αρχείο JSON που περιέχει τα σημεία συντεταγμένων για την κατασκευή των масκών. Το συνολικό μέγεθος του συνόλου δεδομένων είναι περίπου 46,2 MB. Το Kvasir-SEG χρησιμοποιείται ευρέως για έρευνα και ανάπτυξη μεθόδων τμηματοποίησης, εντοπισμού, και ταξινόμησης πολυπόδων. Εφαρμογές του περιλαμβάνουν τη δημιουργία αλγορίθμων εκπαίδευσης και επικύρωσης για εικόνες κολονοσκοπικών ευρημάτων, καθώς και τη βελτίωση της ακρίβειας της διάγνωσης. Η χρήση του dataset συμβάλλει επίσης στη μείωση του ποσοστού παράβλεψης πολύποδων και στη βελτίωση της ποιότητας των ενδοσκοπικών εξετάσεων. Το Kvasir-SEG αποτελεί ένα σημαντικό εργαλείο στην επιστημονική κοινότητα, καθώς είναι διαθέσιμο για ερευνητικούς και εκπαιδευτικούς σκοπούς, ενισχύοντας την ανάπτυξη λύσεων αιχμής για ενδοσκοπικές εικόνες.

3) Τροποποιήσεις στη συγκεκριμένη μελέτη:

Στην παρούσα εργασία χρησιμοποιήθηκε μια ελαφρώς τροποποιημένη μορφή του πρωτότυπου συνόλου δεδομένων Kvasir SEG. Στο σύνολο δεδομένων που χρησιμοποιήθηκε όλες οι εικόνες ήταν δισδιάστατες (2D) έγχρωμες (RGB) κωδικοποιημένες με συμπίεση JPG και οι μάσκες δισδιάστατες (2D) ασπρόμαυρες (grayscale) με ίδια συμπίεση και όλες μεγέθους 352x352. Πριν την οργάνωση του σε dataloaders οι μάσκες μετατράπηκαν σε binary ώστε τα pixel των πολυπύδων να έχουν τιμή 1 και τα υπόλοιπα 0. Τα σύνολα εκπαίδευσης (800 εικόνες με τις αντίστοιχες μάσκες), επικύρωσης (100 εικόνες με αντίστοιχες μάσκες) και δοκιμής (100 εικόνες με αντίστοιχες μάσκες) χωρίστηκαν τυχαία εξ'αρχής και χρησιμοποιούνται σταθερά για όλα τα μοντέλα ώστε να αξιολογηθεί η απόδοση κάθε αρχιτεκτονικής υπό τις ίδιες συνθήκες. Μελλοντικά, θα ήταν χρήσιμο να δοκιμαστούν και άλλες διαιρέσεις του συνόλου δεδομένων για να εξασφαλιστεί η τυχαιότητα και η γενίκευση. Η ίδια διαίρεση χρησιμοποιήθηκε και από τους Razvan-Gabriel Dumitru, Darius Peteleaza και Catalin Craciun, σχεδιαστές του Deep Understanding Convolutional Kernel Network (DUCK-Net 2023), και είναι διαθέσιμη στον σύνδεσμο: <https://drive.google.com/drive/folders/1i-tRhznSRWXA33s3rwlQ9hYVvNtrC4Pd>

4) Τεχνικές Προεπεξεργασίας Εικόνων και Μασκών:

Οι τεχνικές προεπεξεργασίας μπορούν να επηρεάσουν σημαντικά την απόδοση και την ακρίβεια των μοντέλων τμηματοποίησης εικόνας. Διαφορετικές μέθοδοι προεπεξεργασίας βοηθούν στη βελτίωση των χαρακτηριστικών των εικόνων, στην κανονικοποίηση των παραλλαγών και καθιστούν το μοντέλο πιο ανθεκτικό σε διαφορετικούς τύπους εισόδου, και διασφαλίζουν ότι οι εικόνες και οι μάσκες είναι ομοιόμορφα προετοιμασμένες για αποτελεσματική και ακριβή εκπαίδευση του μοντέλου. Ακολουθούν οι τεχνικές προεπεξεργασίας που χρησιμοποιήθηκαν στην παρούσα εργασία και πώς επηρέασαν τα δεδομένα εισόδου:

- **Προσθήκη Διάστασης Καναλιού:** με χρήση της EnsureChannelFirst, οι εικόνες και οι μάσκες διασφαλίζεται ότι έχουν τη διάσταση καναλιού [C, H, W], για τις εικόνες οι διαστάσεις μετατρέπονται από [H, W, C] σε [C, H, W] και για τις μάσκες [1, H, W]
- **Αλλαγή Διαστάσεων:** με χρήση της Resize, εξασφαλίζουμε την προσαρμογή όλων των εικόνων και масκών σε σταθερές διαστάσεις 256x256, διασφαλίζοντας ομοιομορφία στα δεδομένα απαραίτητη για τη μαζική επεξεργασία. Ωστόσο, η αλλαγή μεγέθους μπορεί να προκαλέσει παραμόρφωση ή απώλεια σημαντικών χαρακτηριστικών εάν δεν γίνει προσεκτικά.
- **Κανονικοποίηση Εντάσεων (Scaling):** οι εντάσεις των εικόνων κανονικοποιούνται στο διάστημα [0, 1] μέσω της ScaleIntensity διευκολύνοντας την εκπαίδευση του μοντέλου με πιο συγκρίσιμα δεδομένα.
- **Τυχαίοι Μετασχηματισμοί (Data Augmentation)** [61]: εφαρμογή τυχαίων μετασχηματισμών στις εικόνες και τις μάσκες των δεδομένων του training και validation που

βελτιώνει την ευρωστία και τη γενίκευση του μοντέλου οδηγώντας σε καλύτερη απόδοση σε αόρατα (άγνωστα) δεδομένα:

- Τυχαίοι Κατοπτρισμοί (Flip):
 - Οριζόντιος κατοπτρισμός (RandFlip) με πιθανότητα 50%
 - Κάθετος κατοπτρισμός (RandFlip) με πιθανότητα 50%
- Τυχαίες Περιστροφές (Rotate):
 - Περιστροφή κατά τυχαία γωνία έως 30° (RandRotate) με πιθανότητα 50%
- Τυχαία Κλίμακα (Zoom) [62]:
 - Τυχαία μεγέθυνση ή σμίκρυνση με zoom από 0.8x έως 1.2x (RandZoom) με πιθανότητα 50%
- Τυχαία Μετατόπιση Εντάσεων (Shift Intensity):
 - Οι εντάσεις των RGB εικόνων μετατοπίζονται τυχαία κατά ± 0.1 (RandShiftIntensity) με πιθανότητα 50%
- Τυχαία Ρύθμιση Αντίθεσης (Contrast Adjustment):
 - Η αντίθεση RGB εικόνων προσαρμόζεται με τυχαίο γάμμα στο διάστημα (0.7, 1.5) (RandAdjustContrast) με πιθανότητα 50%.

➤ **Μετατροπή σε Τανυστές (Tensors):** οι εικόνες και οι μάσκες μετατρέπονται σε tensors (PyTorch format) με την εντολή `ToTensor` και έτσι οι εικόνες έχουν πλέον τελική μορφή [3, 256, 256] και μάσκες [1, 256, 256]

➤ **Οργάνωση σε DataLoaders:** Δημιουργούνται DataLoaders για το training, validation, και test set. Στον Training Loader εισάγουμε shuffle (ανακατεύουμε) στα δεδομένα για να αυξηθεί η ποικιλία σε κάθε εποχή εκπαίδευσης προάγοντας την γενίκευση του μοντέλου. Αντίθετα στους Validation και Test Loaders δεν εφαρμόζεται shuffle, ώστε τα αποτελέσματα να παραμένουν σταθερά και συγκρίσιμα.

Στο συγκεκριμένο σύνολο δεδομένων, λόγω της προσεκτικής επιλογής και της καλής ανάλυσης και κεντραρίσματος των εικόνων, δεν απαιτήθηκε εκτεταμένη προεπεξεργασία. Ωστόσο, σε άλλες περιπτώσεις, θα μπορούσαν να εφαρμοστούν επιπλέον τεχνικές για τη βελτίωση των χαρακτηριστικών των εικόνων, όπως:

- κανονικοποίηση (normalization) για εξισορρόπηση τιμών χρώματος και φωτεινότητας
- περικοπή (cropping) για εστίαση σε συγκεκριμένες περιοχές ενδιαφέροντος
- αλλαγή χρωμάτων (color space transformation) [63] για αλλαγή χώρου χρώματος πχ. από RGB σε Grayscale
- ισοστάθμιση ιστογράμματος (histogram equalization) για βελτίωση αντίθεσης στις εικόνες
- θάμπωμα (blurring) και εξομάλυνση (smoothing) για μείωση θορύβου
- όξυνση (sharpening) για ενίσχυση ακμών και λεπτομερειών

- χρήση κατωφλίων (thresholding) για απομόνωση συγκεκριμένων χαρακτηριστικών
- κανονικοποίηση φωτισμού (normalization of illumination) για εξομάλυνση διαφορών φωτισμού
- εντοπισμός ακμών (edge detection) για ανίχνευση και ανάδειξη ακμών στις εικόνες
- άλλες μέθοδοι με σκοπό να μπορέσουν να εξαχθούν περισσότερα χαρακτηριστικά από τις εικόνες μέσω μείωσης θορύβου και αλλαγών αντιθέσης και φωτεινότητας

5) Μετρικές Αξιολόγησης:

- Accuracy (Ακρίβεια): ο λόγος των σωστά προβλεπόμενων pixel (true positives και true negatives) προς τον συνολικό αριθμό των pixel.

$$\text{Accuracy} = \frac{\text{TruePositive} + \text{TrueNegative}}{\text{συνολικός αριθμός προβλέψεων}}$$

- Precision (Ευστοχία): ο λόγος των αληθινά θετικών pixel προς το άθροισμα των αληθινά θετικών και των ψευδώς θετικών pixel, δηλαδή όλα τα δείγματα που προβλέφθηκαν ως θετικά.

$$\text{Precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}}$$

- Recall (Ανάκληση): ο λόγος των αληθινά θετικών pixel προς το άθροισμα των αληθινά θετικών και των ψευδώς αρνητικών pixel, δηλαδή όλα τα πραγματικά θετικά δείγματα.

$$\text{Recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}}$$

- F1 Score: Ο αρμονικός μέσος της ακρίβειας και της ανάκλησης. Είναι χρήσιμος όταν υπάρχει ανισοκατανομή στις κλάσεις.

$$\text{F1score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

- Dice Score: μετρική επικάλυψης που χρησιμοποιείται κυρίως σε προβλήματα segmentation. Μετρά πόσο καλά ταιριάζει η προβλεπόμενη περιοχή με την πραγματική. Πολύ χρήσιμο για δυαδικές κατηγορίες σε εικόνες.

$$\text{DiceScore} = \frac{2 \cdot |\text{Επικάλυψη}|}{|\text{Προβλεπόμενη Περιοχή}| + |\text{Πραγματική Περιοχή}|}$$

- Average IoU (Intersection over Union): Μετρά την επικάλυψη μεταξύ της προβλεπόμενης και της πραγματικής περιοχής διαιρούμενη με την ένωσή τους. Ιδανικό για αξιολόγηση segmentation.

$$\text{IoU} = \frac{\text{Επικάλυψη}}{\text{Ένωση}}$$

- AUC (Area Under the Curve): η επιφάνεια κάτω από την καμπύλη ROC, που δείχνει τη σχέση μεταξύ true positive rate και false positive rate. Μετρά την απόδοση ενός μοντέλου για προβλήματα δυαδικής ταξινόμησης, ανεξάρτητα από το κατώφλι.
- Confusion Matrix (Πίνακας Σύγχυσης) Είναι ένας πίνακας που παρουσιάζει τη σύγκριση μεταξύ των προβλεπόμενων και των πραγματικών τιμών. Περιέχει τα True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN).

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Figure 20: υπόδειγμα πίνακα σύγχυσης

Αρχιτεκτονικές Μοντέλων Μηχανικής Μάθησης:

Στην συγκεκριμένη μελέτη χρησιμοποιούνται και συγκρίνονται οι αρχιτεκτονικές:

- **UNet (2015)**

Μια από τις πρώτες αρχιτεκτονικές που χρησιμοποιήθηκαν στην τμηματοποίηση ιατρικών εικόνων, γνωστή για την ικανότητά της να αποδίδει υψηλή ακρίβεια με περιορισμένα δεδομένα εκπαίδευσης.

- **Pyramid Scene Parsing Network – PSP-Net (2017)**

Εισάγει το Pyramid Pooling Module στην δειγματοληψία για βελτιωμένη κατανόηση του περιβάλλοντος εικόνας.

- **DeepLab V3 (2017) με backbone ResNet50**

Χρησιμοποιεί atrous convolutions για την ενίσχυση της ανάλυσης σε πολλαπλές κλίμακες.

- **SegResNet (2018)**

Συνδυάζει την ισχύ των residual blocks με τεχνικές U-Net για βελτιωμένη απόδοση στην τμηματοποίηση.

Οι συγκεκριμένες αρχιτεκτονικές έχουν επιλεγεί γιατί παρουσιάζουν κάποιες ιδιαιτερότητες ή αποτελούν συνδυασμό και βελτίωση παλαιότερων μοντέλων, επομένως θέλουμε να εντοπίσουμε σε τι μπορεί να έχει μεγαλύτερη προσφορά κάθε αρχιτεκτονική. Επίσης κάποιες είναι από τις πρώτες που χρησιμοποιήθηκαν στην τμηματοποίηση ιατρικών εικόνων και είχαν ικανοποιητική απόδοση καλύπτοντας έτσι ένα μεγάλο εύρος στην εξέλιξη του τομέα.

UNet [2015]

Το U-Net αποτελεί μια καινοτόμο αρχιτεκτονική συνελκτικών νευρωνικών δικτύων (CNNs) που σχεδιάστηκε για τμηματοποίηση εικόνων σε επίπεδο pixel, με ιδιαίτερη έμφαση σε

βιοϊατρικές εφαρμογές. Αναπτύχθηκε από τους Olaf Ronneberger, Philipp Fischer και Thomas Brox στα πλαίσια έρευνας του Computer Science Department and BIOS Centre for Biological Signalling Studies του Πανεπιστημίου του Freiburg στην Γερμανία και παρουσιάστηκε το 2015 [64]. Το κύριο πλεονέκτημα του U-Net είναι η ικανότητά του να αποδίδει υψηλή ακρίβεια στην τμηματοποίηση, ακόμη και με περιορισμένα δεδομένα εκπαίδευσης, χαρακτηριστικό που το καθιστά ιδανικό για τη βιοϊατρική, όπου η συλλογή μεγάλων συνόλων δεδομένων είναι συχνά δύσκολη και χρονοβόρα.

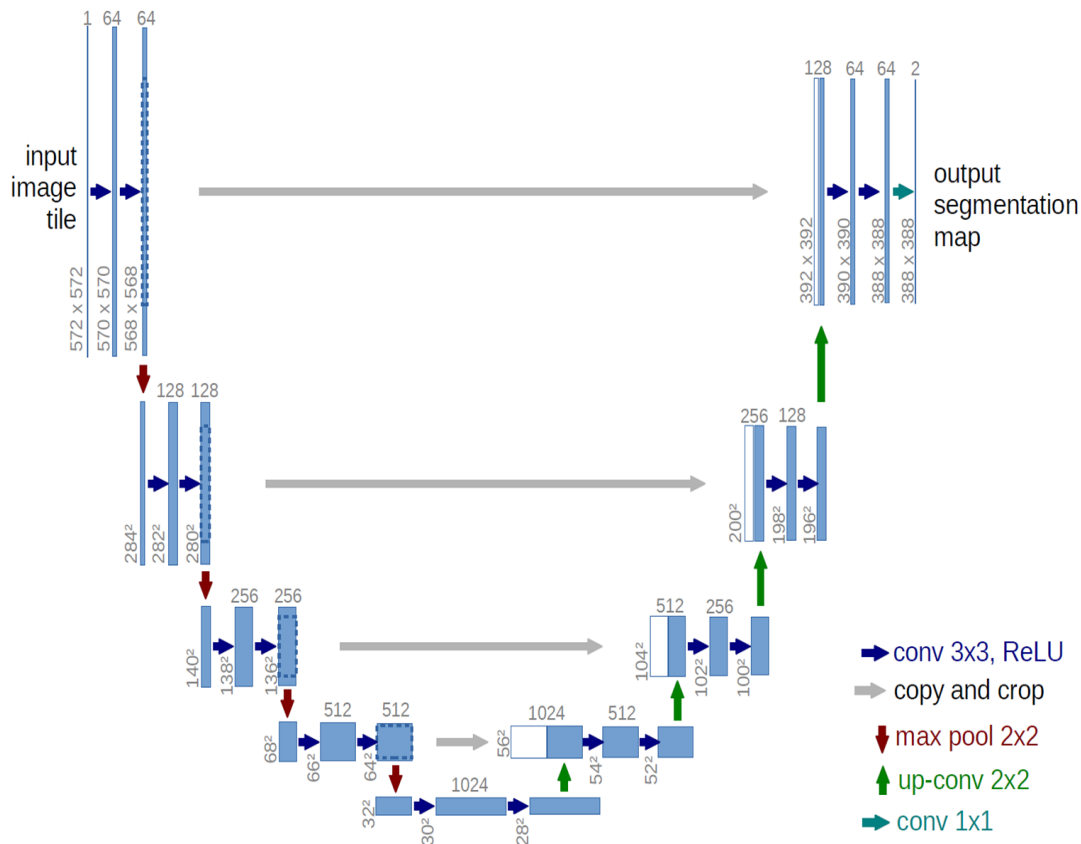


Figure 21: Αρχιτεκτονική U-net (32x32 pixel στη χαμηλότερη ανάλυση).

μπλε πλαίσια: πολυκαναλικοί χάρτες χαρακτηριστικών με αριθμό καναλιών (πάνω) | μέγεθος x-y (κάτω αριστερό άκρο κουτιού) | λευκά πλαίσια: αντιγραφμένοι χάρτες χαρακτηριστικών | βέλη: λειτουργίες

Η δομή του U-Net βασίζεται σε ένα πλήρως συνεκτικό δίκτυο [65] και ακολουθεί μια συμμετρική αρχιτεκτονική σε σχήμα "U". Αυτό το σχήμα αποτελείται από δύο κύρια μέρη τον encoder (συμπίεση) και τον decoder (αποσυμπίεση). Στον encoder η εικόνα εισόδου υφίσταται διαδοχικές συνελίξεις και μέγιστη συγκέντρωση (max-pooling), προκειμένου να εξάγονται τα πιο σημαντικά χαρακτηριστικά, μειώνοντας ταυτόχρονα την ανάλυση της εικόνας. Αντίθετα, στον decoder χρησιμοποιούνται μετασυνελίξεις (up-convolutions) για την αποκατάσταση της ανάλυσης της εικόνας και τη δημιουργία της τελικής τμηματοποίησης. Πιο αναλυτικά ο Encoder εξάγει τα κύρια χαρακτηριστικά από την εικόνα εισόδου μέσω διαδοχικών στρωμάτων συνελίξεων και μείωσης διαστάσεων (downsampling).

- Συνελίξεις (Convolutions): Σε κάθε επίπεδο εφαρμόζονται 2 συνελικτικά φίλτρα 3×3 με συνάρτηση ενεργοποίησης ReLU που αυξάνουν την έκφραση χαρακτηριστικών.

- Max pooling: Μετά από κάθε συνελικτικό μπλοκ, εφαρμόζεται pooling 2×2 για μείωση της ανάλυσης επιτρέποντας την εξαγωγή πιο αφηρημένων χαρακτηριστικών.

- Αύξηση καναλιών: Καθώς η ανάλυση μειώνεται, ο αριθμός των καναλιών αυξάνεται εκθετικά, επιτρέποντας να αποθηκευτούν περισσότερες πληροφορίες σε υψηλού επιπέδου χαρακτηριστικά.

Ο Decoder αποκαθιστά την ανάλυση της εικόνας μέσω διαδοχικών στρωμάτων επέκτασης (upsampling), δημιουργώντας μια έξοδο που έχει ίδια ανάλυση με την εικόνα εισόδου.

- Αντισυνελίξεις (Up-Convolutions): Χρησιμοποιούνται για την επαναφορά της ανάλυσης της εικόνας.

- Συγχώνευση χαρακτηριστικών (Skip connections): Οι χαρακτηριστικές πληροφορίες από τον encoder, που χάθηκαν λόγω pooling, μεταφέρονται στο decoder μέσω συνδέσεων skip. Αυτές οι συνδέσεις πραγματοποιούν αντιγραφή και συγχώνευση (concatenation) των χαρακτηριστικών από το encoder με τα αντίστοιχα επίπεδα στο decoder, βελτιώνοντας την ακρίβεια στην τμηματοποίηση λεπτομερειών.

Μια από τις κύριες καινοτομίες της αρχιτεκτονικής είναι η χρήση συνδέσεων skip. Αυτές οι συνδέσεις μεταφέρουν πληροφορίες από τα πρώιμα στάδια του encoder στα αντίστοιχα στάδια του decoder και διαδραματίζουν βασικό ρόλο στη διατήρηση της χωρικής πληροφορίας που μπορεί να χαθεί κατά τη φάση συμπίεσης επανεισάγοντάς τις στα στάδια αποκωδικοποίησης. Έτσι τα αποτελέσματα της τμηματοποίησης είναι πιο ακριβή. Το U-Net δεν απαιτεί εικόνες συγκεκριμένης ανάλυσης και μπορεί να προσαρμοστεί δυναμικά σε εισόδους διαφορετικών μεγεθών, καθιστώντας το ευέλικτο σε πρακτικές εφαρμογές. Επίσης εκπαιδεύεται αποτελεσματικά με περιορισμένα δεδομένα, καθώς χρησιμοποιεί εντατικά τεχνικές data augmentation [66] και επωφελείται από τις skip connections για καλύτερη γενίκευση.

Υπάρχουν πολλές εφαρμογές του U-Net στη βιοϊατρική τμηματοποίηση εικόνων, όπως η τμηματοποίηση εικόνας εγκεφάλου (BRATS) και η κατάτμηση εικόνας ήπατος (siliver07), καθώς και η πρόβλεψη θέσης δέσμευσης πρωτεΐνης. Πέρα από τη βιοϊατρική, το U-Net έχει βρει εφαρμογές και σε άλλους τομείς επεξεργασίας εικόνων, όπως η δορυφορική απεικόνιση, η τμηματοποίηση χάρτινων δεδομένων, η αναγνώριση αντικειμένων, η ανάλυση φυσικών σκηνών και άλλες περιοχές όπου απαιτείται ακρίβεια σε pixel-level χαρακτηριστικά. Η αρχιτεκτονική συνεχίζει να βελτιώνεται και να προσαρμόζεται σε νεότερες προσεγγίσεις καθώς επίσης έχει αποτελέσει βάση για πολλά πιο σύνθετα μοντέλα. Επιπλέον, η αρχιτεκτονική του αποτελεί τη βάση για πολλές επεκτάσεις.

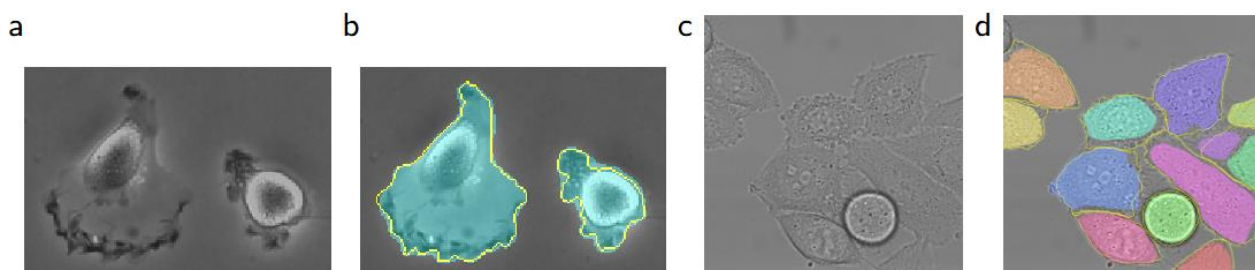


Figure 22: U-net's segmentation success on PhC-U373 (a-b) and DIC-HeLa (c-d) datasets
 (a) original image (b) cyan mask: Unet's segmentation | yellow outline: grand truth mask
 (c) original image (d) colourful masks: Unets segmentation | yellow outlines: grand truth masks

PSPNet (Pyramid Scene Parsing Network) [2017] [67]

Το PSPNet είναι ένα προηγμένο συνελκτικό νευρωνικό δίκτυο που προτάθηκε από τους Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang και Jiaya Jia το 2017 στο Κινέζικο Πανεπιστήμιο του Χονγκ Κονγκ σε συνεργασία με την SenseTime για τη βαθιά εκμάθηση στη σημασιολογική τμηματοποίηση (semantic segmentation) εικόνας, με στόχο τη βελτίωση της κατανόησης σκηνών μέσω της αποτελεσματικής ενσωμάτωσης γενικών (global) πληροφοριών περιβάλλοντος. Το PSPNet έθεσε νέα πρότυπα στον τομέα, επιτυγχάνοντας κορυφαία αποτελέσματα σε πολλαπλές προκλήσεις όπως PASCAL VOC [68], ADE20K, και Cityscapes.

Η βασική αρχή λειτουργίας του έγκειται στη χρήση του Pyramid Pooling Module (PPM), το οποίο επιτρέπει την κατανόηση και ενσωμάτωση πληροφοριών από διαφορετικά επίπεδα χωρικής κλίμακας. Αυτό είναι απαραίτητο για την κατανόηση του περιβάλλοντος (context) μιας εικόνας, ειδικά όταν τα αντικείμενα έχουν μεγάλη ποικιλομορφία σε μέγεθος, θέση ή εμφάνιση.

Η αρχιτεκτονική του PSPNet βασίζεται σε έναν προ-εκπαιδευμένο σκελετό (backbone) όπως το Residual Network (ResNet) [69], για να εξάγει βαθιά χαρακτηριστικά από την εικόνα εισόδου, τα οποία είναι κατάλληλα για την πυραμιδική επεξεργασία.

Το κεντρικό στοιχείο του δικτύου είναι το PPM, το οποίο συλλέγει πληροφορίες από διαφορετικές περιοχές της εικόνας μέσω pooling σε πολλαπλές κλίμακες, συγκεκριμένα εφαρμόζει pooling σε 4 επίπεδα με διαφορετικά μεγέθη παραθύρων, όπως 1×1 , 2×2 , 3×3 και 6×6 , για να συλλέξει global και τοπικές πληροφορίες.

Τα αποτελέσματα αυτών των pooling επιπέδων στη συνέχεια αναδεικνύονται σε ένα κοινό μέγεθος μέσω διγραμμικής παρεμβολής (bilinear interpolation) και συνενώνονται με τα αρχικά χαρακτηριστικά, παρέχοντας έτσι ένα πλούσιο σύνολο πληροφοριών για την τελική πρόβλεψη. Μέσα από επεξεργασία υψηλής ακρίβειας, το PSPNet αποδίδει λεπτομερείς προβλέψεις για κάθε pixel της εικόνας, κατατάσσοντάς το στη σωστή κατηγορία.

Η ενσωμάτωση χαρακτηριστικών πολλαπλών κλιμάκων βοηθά το δίκτυο να αναγνωρίζει με ακρίβεια μεγάλα ή μικρά αντικείμενα, αλλά και να κατανοεί τη συνολική σκηνή. Η απόδοση του PSPNet σε απαιτητικά σύνολα δεδομένων το καθιστά ιδανικό για ποικίλες εφαρμογές, από την τμηματοποίηση σκηνών μέχρι την ιατρική απεικόνιση και η αρχιτεκτονική του μπορεί να προσαρμοστεί εύκολα σε διαφορετικές βασικές δομές (backbones), επιτρέποντας την καλύτερη εκμετάλλευση νέων τεχνολογιών.

Το PSPNet χρησιμοποιείται ευρέως σε τομείς όπως η αυτόνομη οδήγηση [70] (π.χ., αναγνώριση δρόμων, πεζών και οχημάτων), η ανάλυση δορυφορικών εικόνων, η ιατρική απεικόνιση (π.χ., τμηματοποίηση όγκων ή ιστών), η επεξεργασία εικόνων για τη βιομηχανία και την τέχνη.

Το PSPNet εισάγει μια αποτελεσματική προσέγγιση για την ενσωμάτωση παγκόσμιων πληροφοριών περιβάλλοντος μέσω του Pyramid Pooling Module, βελτιώνοντας σημαντικά την ακρίβεια στην κατανόηση σκηνών. Η αρχιτεκτονική του έχει επηρεάσει σημαντικά την εξέλιξη των μεθόδων ανάλυσης εικόνας, προσφέροντας ένα ισχυρό πλαίσιο για μελλοντικές έρευνες και εφαρμογές.

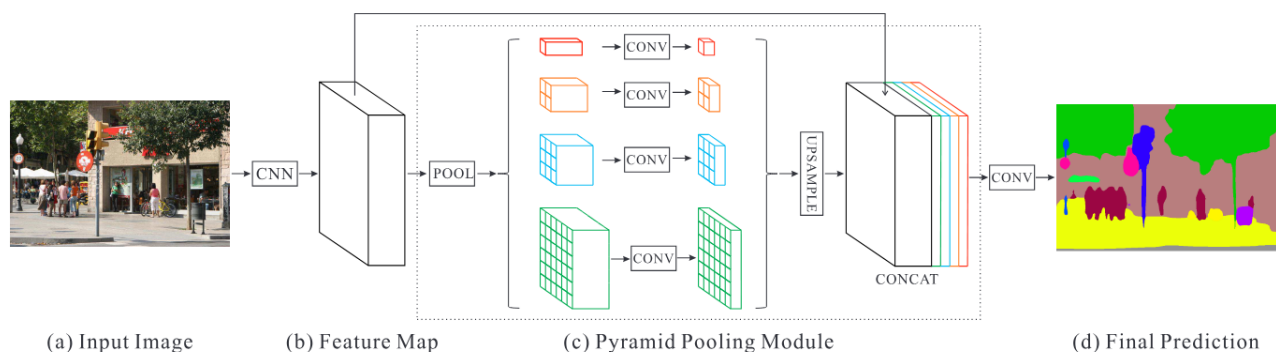


Figure 23: Επισκόπηση του προτεινόμενου PSPNet

(a) εικόνα εισόδου

(b) χάρτης χαρακτηριστικών του τελευταίου convolutional layer που προκύπτει από CNN

(c) εφαρμογή PPM για τη συλλογή διαφορετικών αναπαραστάσεων υποπεριοχής, ακολουθούμενα από επίπεδα upsampling και συνένωσης για να σχηματιστεί η τελική αναπαράσταση χαρακτηριστικών, η οποία μεταφέρει τόσο τοπικές όσο και καθολικές πληροφορίες περιβάλλοντος

(d) τελική πρόβλεψη ανά pixel αφού η αναπαράσταση τροφοδοτείται σε ένα επίπεδο συνέλιξης

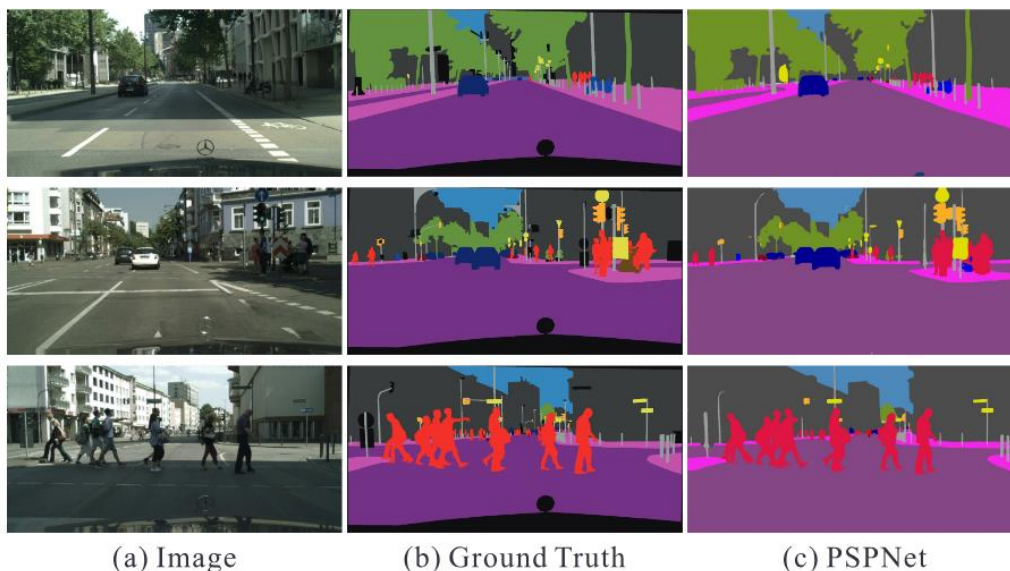


Figure 24: παραδείγματα αποτελεσμάτων PSPNet στο Cityscapes dataset

SegResNet (2018)

Το SegResNet (και συγκεκριμένα αυτό που περιλαμβάνεται στο documentation του MONAI [71]) είναι ένα προηγμένο νευρωνικό δίκτυο σχεδιασμένο ειδικά για τμηματοποίηση ιατρικών εικόνων [72] που αναπτύχθηκε το 2018 από τον Andriy Myronenko (NVIDIA) και κέρδισε την 1η θέση στην πρόκληση BraTS [73], [74]. Βασίζεται στη συνδυαστική χρήση της ισχύος των residual blocks (ResNet) [75] και τεχνικών της κλασικής U-Net αρχιτεκτονικής, προσφέροντας βελτιωμένη απόδοση και ευχέρεια στην εκμάθηση χαρακτηριστικών εικόνας σε διάφορες εφαρμογές επεξεργασίας εικόνας, όπως η ιατρική απεικόνιση και η αυτόματη τμηματοποίηση εικόνας.

Το SegResNet βασίζεται στο ResNet, χρησιμοποιεί υπολειμματικές συνδέσεις (residual connections), που επιτρέπουν τη μεταφορά πληροφοριών από τα βαθύτερα επίπεδα του δικτύου προς τα ανώτερα, αποφεύγοντας έτσι το φαινόμενο vanishing (εξαφάνιση) gradient που επιφέρει μεγάλες δυσκολίες στην εκμάθηση. Επίσης υιοθετεί την αρχή του U-Net, όπου υπάρχει συμμετρική down-sampling (encoder) [76] και up-sampling (decoder) [77] ροή για εξαγωγή χαρακτηριστικών και πρόβλεψη τμημάτων και είναι βελτιστοποιημένο για 3D δεδομένα [78], καθιστώντας το ιδανικό για εικόνες από μαγνητική και αξονική τομογραφία. Βασικά μέρη της αρχιτεκτονικής αποτελούν ο encoder που περιλαμβάνει residual blocks για εξαγωγή χαρακτηριστικών από τα δεδομένα εισόδου, ο decoder που ανακατασκευάζει τις προβλέψεις για κάθε pixel (ή voxel), διατηρώντας τη χωρική πληροφορία σε μια τμηματοποιημένη εικόνα με τεχνικές όπως up-sampling ή deconvolution και τα skip connections που επιτρέπουν τη μετάδοση λεπτομερειών μεταξύ encoder και decoder και βασίζεται σε τεχνικές όπως batch normalization και dropout για καλύτερη γενίκευση. Η καινοτομία του SegResNet έγκειται στην ενσωμάτωση των residual blocks σε κάθε επίπεδο του encoder και του decoder, γεγονός που επιτρέπει στο δίκτυο να έχει μεγαλύτερη ευχέρεια στις παραμέτρους του, μειώνοντας την ανάγκη για περίπολοκη ρύθμιση υπερπαραμέτρων, ενισχύοντας τη γενίκευση του μοντέλου και παρέχοντας ανθεκτικότητα σε ελαττωματικά δεδομένα.

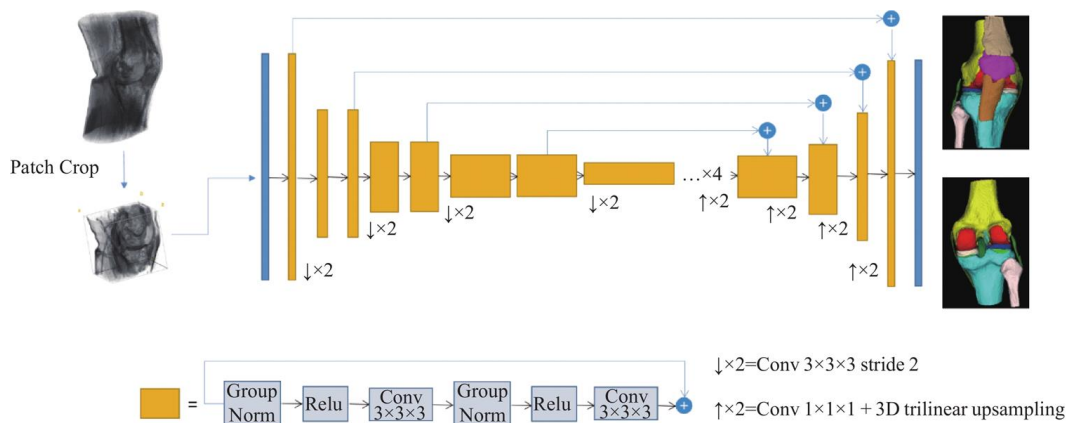


Figure 25: επισκόπηση της αρχιτεκτονικής δικτύου SegResNet με κωδικοποιητή που κάνει down-sample 4 φορές και οι ποσότητες των residual block κάθε σταδίου είναι 1, 2, 2, 2, 4

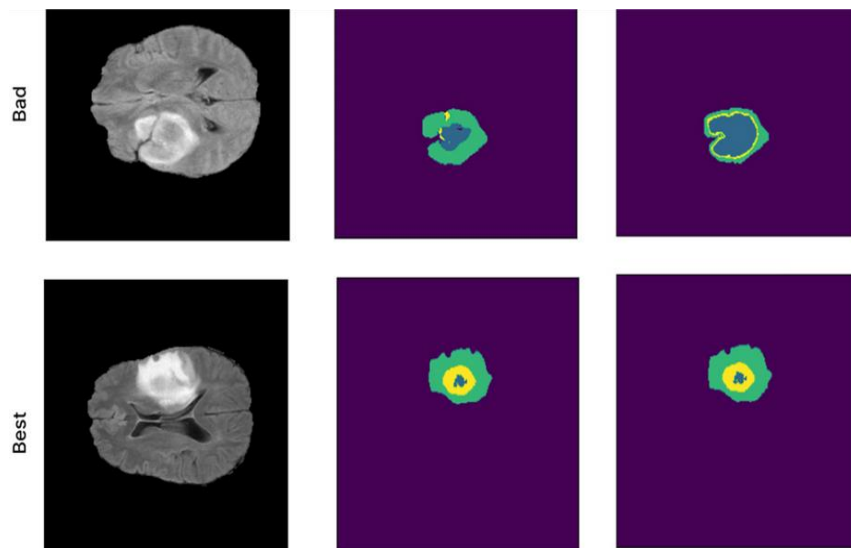


Figure 26: segmentation όγκου σε εικόνα MRI εγκεφάλου με όγκο (αριστερά), πρόβλεψη SegResNet (μέση), μάσκα αναφοράς (δεξιά) | κίτρινο: ET (ενισχυόμενος όγκος), πράσινο: WT (ολόκληρος όγκος), μπλε: TC (πυρήνας όγκου)

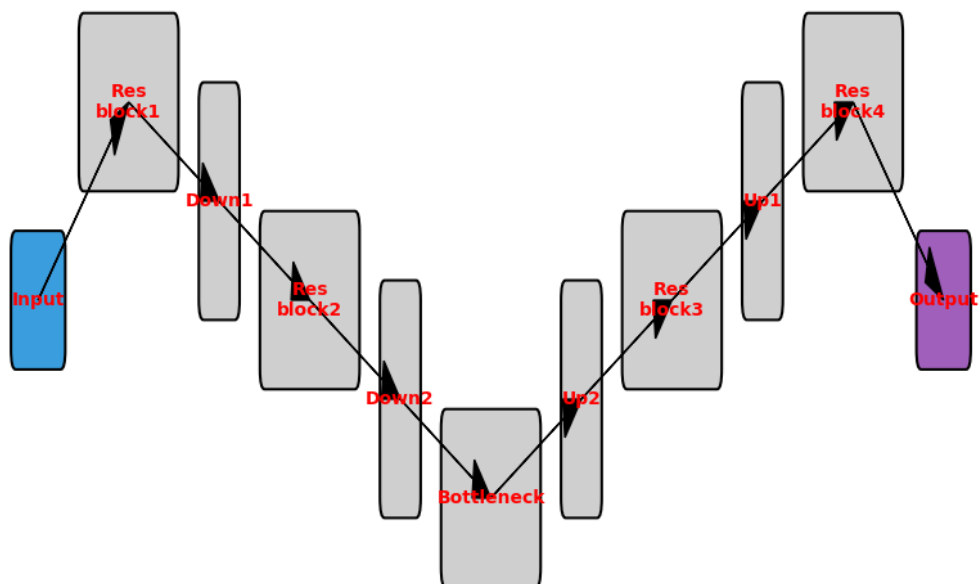


Figure 27: αρχιτεκτονική SegResNet

DeepLabv3 (2017)

Το DeepLabv3 είναι ένα προηγμένο μοντέλο βαθιάς μάθησης που αναπτύχθηκε από τους ερευνητές της Google: Liang-Chieh Chen, George Papandreou, Florian Schroff και Hartwig Adam και παρουσιάστηκε στο άρθρο "Rethinking Atrous Convolution for Semantic Image Segmentation" [79] το 2017. Σκοπός του είναι η βελτίωση της σημασιολογικής τμηματοποίησης εικόνων μέσω της χρήσης καινοτόμων τεχνικών.

Η αρχιτεκτονική του DeepLabv3 βασίζεται σε δύο κύρια στοιχεία [80]:

- Atrous Convolution (Διατεταγμένη Συνελίξη): επιτρέπει την προσαρμογή του πεδίου αντίληψης των φίλτρων χωρίς αύξηση του αριθμού των παραμέτρων ή της υπολογιστικής πολυπλοκότητας. Με αυτόν τον τρόπο, το δίκτυο μπορεί να συλλάβει πολυκλίμακα συμφοραζόμενα, κάτι κρίσιμο για την ακριβή τμηματοποίηση αντικειμένων διαφόρων μεγεθών.
- Atrous Spatial Pyramid Pooling (ASPP): χρησιμοποιεί πολλαπλά atrous convolutions με διαφορετικούς ρυθμούς διασποράς σε παράλληλη διάταξη, επιτρέποντας στο δίκτυο να συλλέξει πληροφορίες σε πολλαπλές κλίμακες. Επιπλέον, το ASPP ενισχύεται με χαρακτηριστικά σε επίπεδο εικόνας που κωδικοποιούν το παγκόσμιο πλαίσιο, βελτιώνοντας περαιτέρω την απόδοση του μοντέλου.

Το DeepLabv3 έχει εφαρμοστεί ευρέως σε διάφορους τομείς που απαιτούν ακριβή σημασιολογική τμηματοποίηση εικόνων όπως η αυτόνομη οδήγηση, η ιατρική απεικόνιση (για την τμηματοποίηση ανατομικών δομών ή ανωμαλιών σε ιατρικές εικόνες), για ανάλυση δορυφορικών εικόνων και παρακολούθηση καλλιέργειών.

Το DeepLabv3 αποτελεί εξέλιξη των προηγούμενων εκδόσεων DeepLab, με σημαντικές βελτιώσεις στην ικανότητα τμηματοποίησης χωρίς την ανάγκη μεταεπεξεργασίας με DenseCRF. Αργότερα, το 2018, παρουσιάστηκε το DeepLabv3+ που συνδυάζει την αρχιτεκτονική του DeepLabv3 με έναν decoder για την περαιτέρω βελτίωση των αποτελεσμάτων, ειδικά στα όρια των αντικειμένων. Συνολικά, το DeepLabv3 αποτελεί ορόσημο στη σημασιολογική τμηματοποίηση εικόνων, προσφέροντας ένα ισχυρό εργαλείο για την κατανόηση και ανάλυση οπτικών δεδομένων σε διάφορες εφαρμογές.

Στην παρούσα εργασία χρησιμοποιήθηκε το deeplabv3-resnet50 από την βιβλιοθήκη [], μια υλοποίηση του μοντέλου DeepLabv3 που χρησιμοποιεί το δίκτυο ResNet-50 ως backbone. Ο σκελετός (backbone) ResNet-50 [81] χρησιμοποιείται για την εξαγωγή χαρακτηριστικών από την εικόνα εισόδου και είναι ένας προκαθορισμένος συνελικτικός νευρωνικός ιστός με 50 επίπεδα, που εισάγει residual connections για καλύτερη εκπαίδευση και αποφυγή εξαφάνισης του σφάλματος κατά την πίσω διάδοση (back propagation). Το κύριο τμήμα του DeepLabv3 είναι το Atrous Spatial Pyramid Pooling (ASPP) module, το οποίο χρησιμοποιεί διατεταγμένες συνελίξεις (atrous convolutions) για τη λήψη πολυκλίμακων χαρακτηριστικών. Περιλαμβάνει συνδυασμό πληροφοριών από διαφορετικές χωρικές κλίμακες και καθολικό πλαίσιο (global context). Έτσι συνδυαστικά τα χαρακτηριστικά που εξάγονται από το ResNet-50

τροφοδοτούνται στο ASPP και η τελική έξοδος είναι ένας χάρτης χαρακτηριστικών που αντιστοιχεί σε κατηγορίες κάθε εικονοστοιχείο (semantic segmentation).

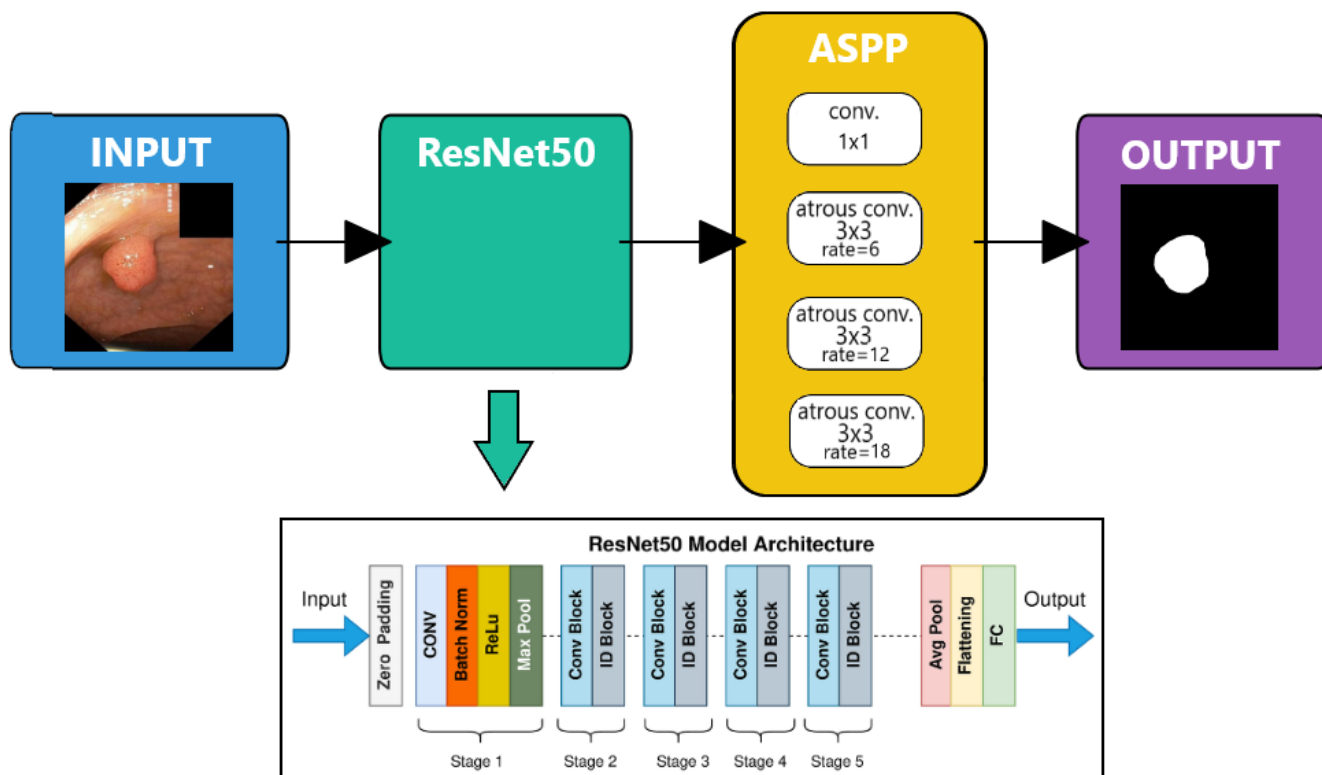


Figure 28: αρχιτεκτονική μοντέλου DeepLabv3 με backbone ResNet50

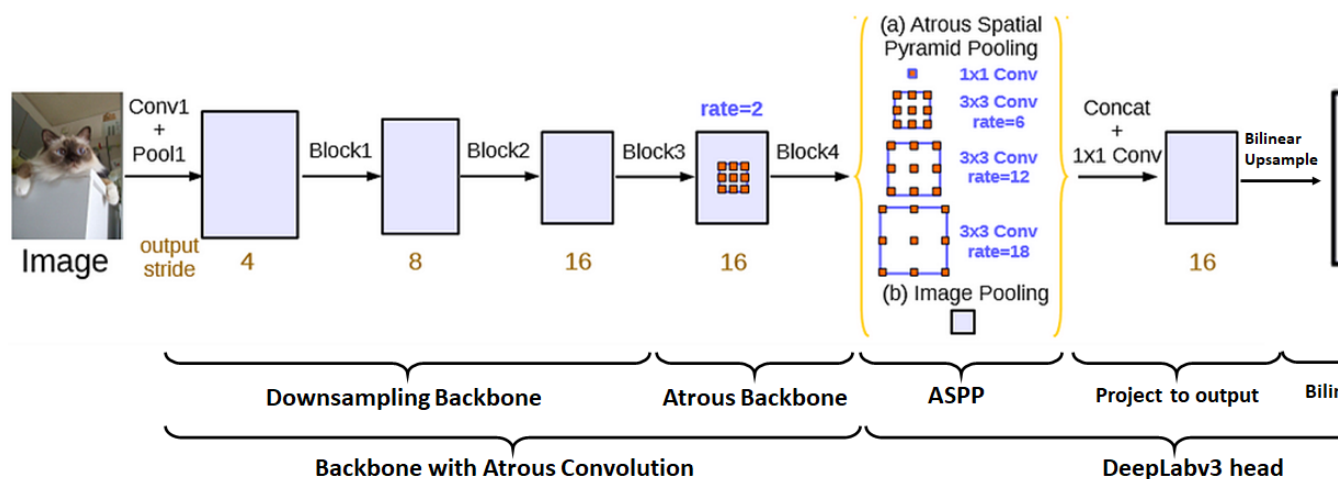


Figure 29: γραφήμα απεικόνισης αρχιτεκτονικής DeepLabv3

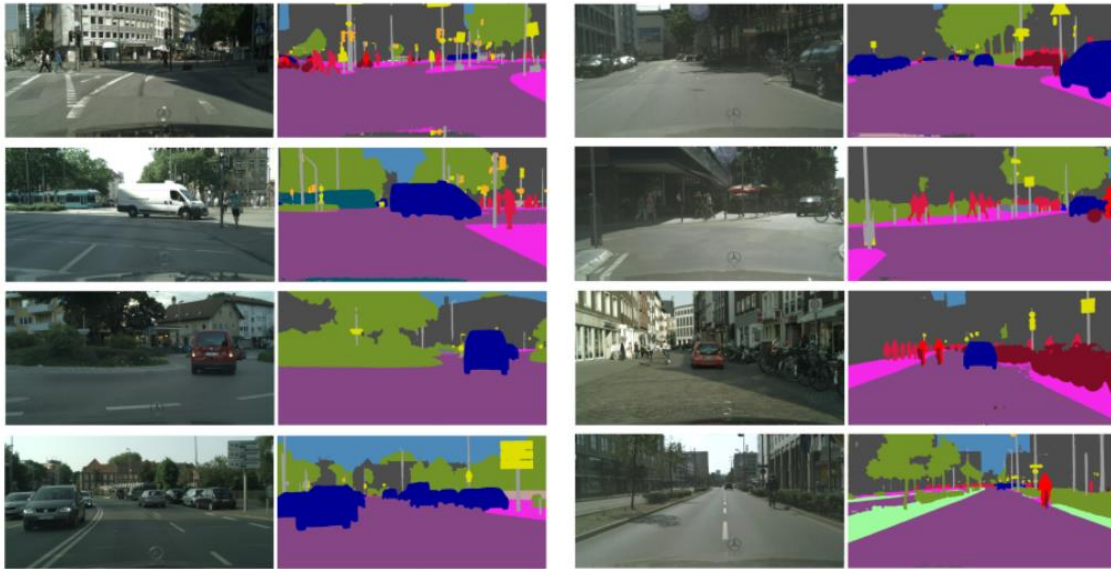


Figure 30: αποτελέσματα οπτικοποίησης στο Cityscapes val set κατά την εκπαίδευση με μόνο training set

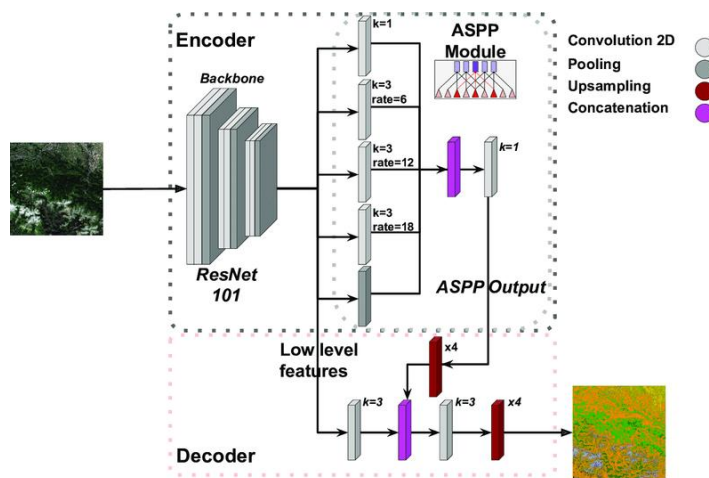


Figure 31: αρχιτεκτονική DeepLabv3+

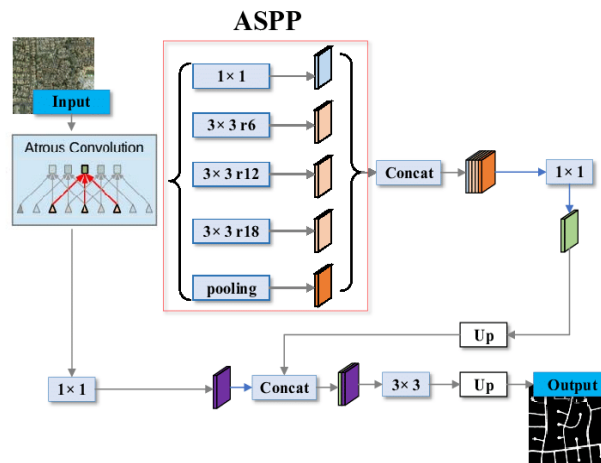


Figure 32: αρχιτεκτονική DeepLabv3

V. ΣΥΓΚΡΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ

Τα αποτελέσματα που παρουσιάζονται προέκυψαν μετά την εκπαίδευση των μοντέλων στο ίδιο σύνολο δεδομένων για 200 εποχές, με επικύρωση κάθε δύο εποχές εκπαίδευσης. Χρησιμοποιήθηκε ο Adam optimizer, η συνάρτηση απώλειας DiceLoss από το MONAI, και ο ρυθμός εκμάθησης καθορίστηκε στο $1e-4$.

Training και Validation Loss ανά εποχή:

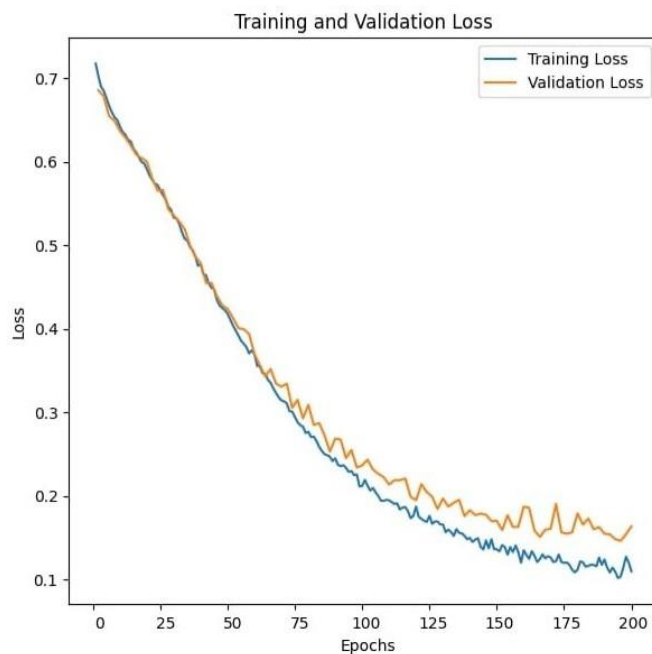


Figure 33: UNet διάγραμμα training και validation loss ανά εποχή

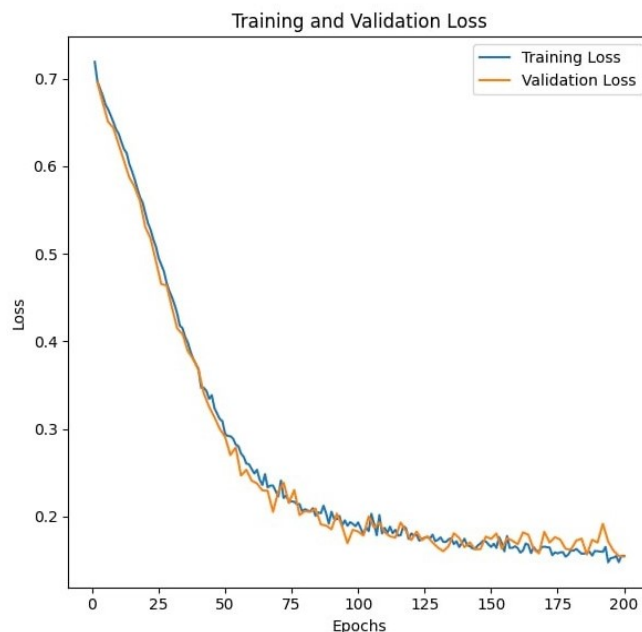


Figure 34: SegResNet διάγραμμα training και validation loss ανά εποχή

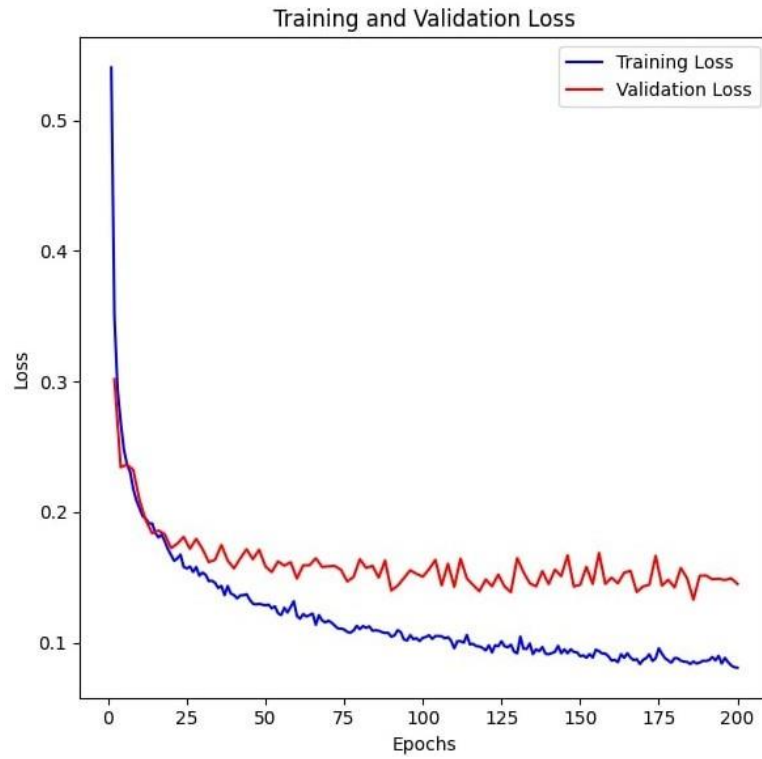


Figure 35: PSPNet διάγραμμα training και validation loss ανά εποχή

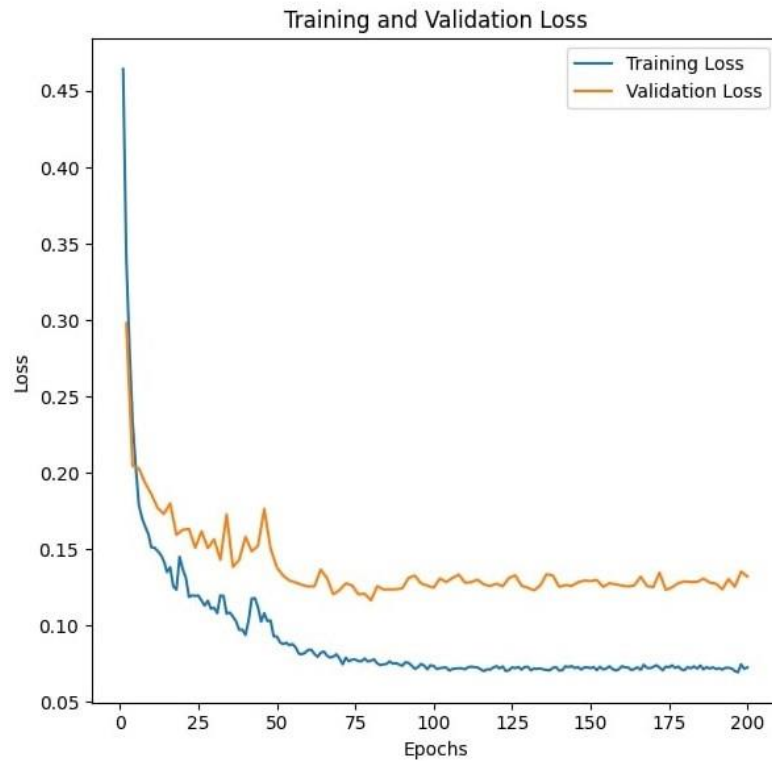


Figure 36: DeepLabv3 διάγραμμα training και validation loss

Η απώλεια εκπαίδευσης (training loss) και επικύρωσης (validation loss) είναι βασικοί δείκτες της απόδοσης και της ικανότητας γενίκευσης ενός μοντέλου βαθιάς μάθησης. Με την προσεκτική παρακολούθηση και αντιμετώπιση των μοτίβων των απωλειών είναι δυνατή η διασφάλιση της ακρίβειας και της ανθεκτικότητας των μοντέλων, μειώνοντας τον κίνδυνο overfitting ή underfitting. Η τελειοποίηση ενός μοντέλου που βασίζεται σε αυτές τις απώλειες διασφαλίζει ότι έχει καλή απόδοση όχι μόνο στα δεδομένα εκπαίδευσης αλλά και σε εφαρμογές πραγματικού κόσμου.

Τα παραπάνω διαγράμματα προέκυψαν μετά από διάφορες δοκιμές μεθόδων όπως προσαρμοστικούς βελτιστοποιητές ρυθμού εκμάθησης (π.χ. Adam optimizer) και συντονισμό υπερπαραμέτρων όπως το μέγεθος παρτίδας (batch size), ο ρυθμός εκμάθησης (learning rate) και η ίδια η αρχιτεκτονική. Ενδεχομένως να πετυχαίναμε καλύτερα αποτελέσματα με rate scheduling, εκπαίδευση σε μεγαλύτερο όγκο δεδομένων ή με cross-validation σε διαφορετικά σύνολα δεδομένων, κάτι το οποίο όμως δεν αποτελεί αντικείμενο της συγκεκριμένης έρευνας.

Παρατηρήσεις:

Όλα τα μοντέλα για το συγκεκριμένο σύνολο δεδομένων και τιμές υπερπαραμέτρων παρουσιάζουν όμαλή πτώση του training και το validation loss ξεκινώντας από μια υψηλή τιμή και τείνοντας προς το 0 χωρίς ιδιαίτερες αυξομειώσεις. Από αυτό μπορούμε να εξάγουμε με ασφάλεια το συμπέρασμα :

- ✓ Κανένα μοντέλο δεν παρουσιάζει ανεπιθύμητη υπερπροσαρμογή (overfitting). Εφόσον η καμπύλη του validation loss δεν παρουσιάζει σημαντική άνοδο μετά από κάποιο σημείο όσο η training loss συνεχίζει να είναι πτωτική.
- ✓ Τα μοντέλα παρουσιάζουν καλή γενίκευση (good generalization). Τόσο το training loss όσο και το validation loss μειώνονται, γεγονός που υποδηλώνει ότι το μοντέλο μαθαίνει και γενικεύεται καλά σε νέα δεδομένα.

Το ιδανικό είναι το training και το validation loss να μειώνονται και να σταθεροποιούνται σε ένα συγκεκριμένο σημείο υποδηλώνοντας βέλτιστη προσαρμογή. Με αυτόν τον τρόπο παρατηρούμε ότι δεν πρόκειται για ένα μοντέλο που δεν ταιριάζει ή που ταιριάζει υπερβολικά και άρα ανταποκρίνεται καλά τόσο στο σύνολο εκπαίδευσης όσο και σε νέα δεδομένα.

Validation Dice Score IoU

Η ερμηνεία μιας καμπύλης validation Dice score (ή IOU - Intersection over Union) ανά εποχή είναι κρίσιμη για την αξιολόγηση της επίδοσης ενός μοντέλου, ιδιαίτερα σε προβλήματα image segmentation. Οι καμπύλες που προέκυψαν παρουσιάζονται ακολούθως:

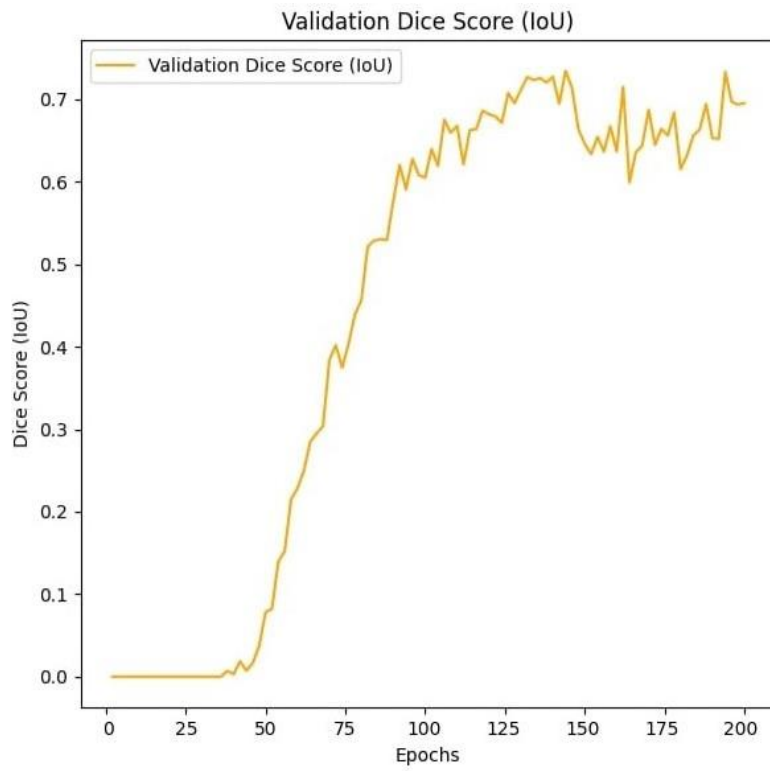


Figure 37: UNet validation dice score (IoU)

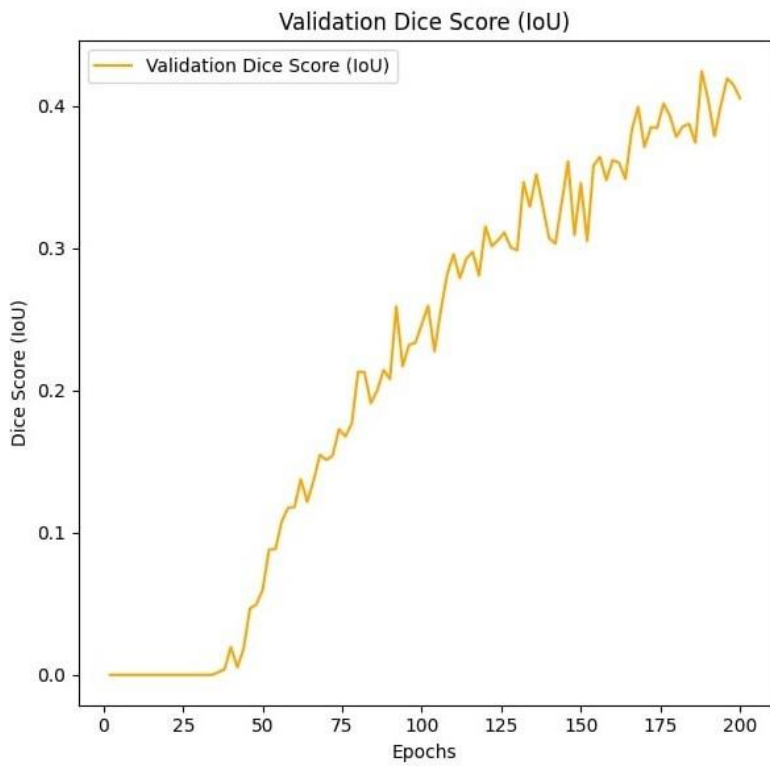


Figure 38: SegResNet validation dice score (IoU)

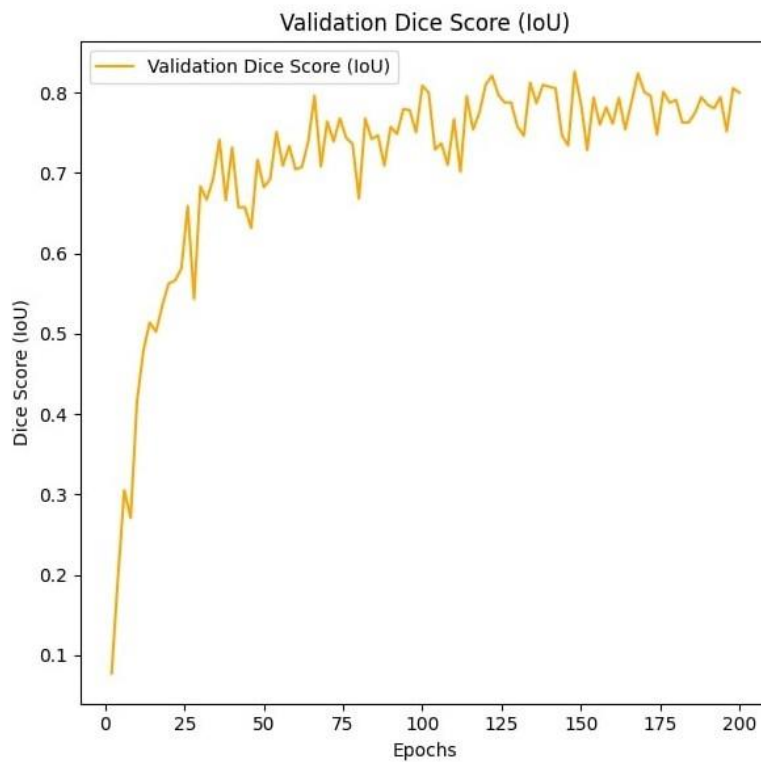


Figure 39: PSPNet validation dice score (IoU)

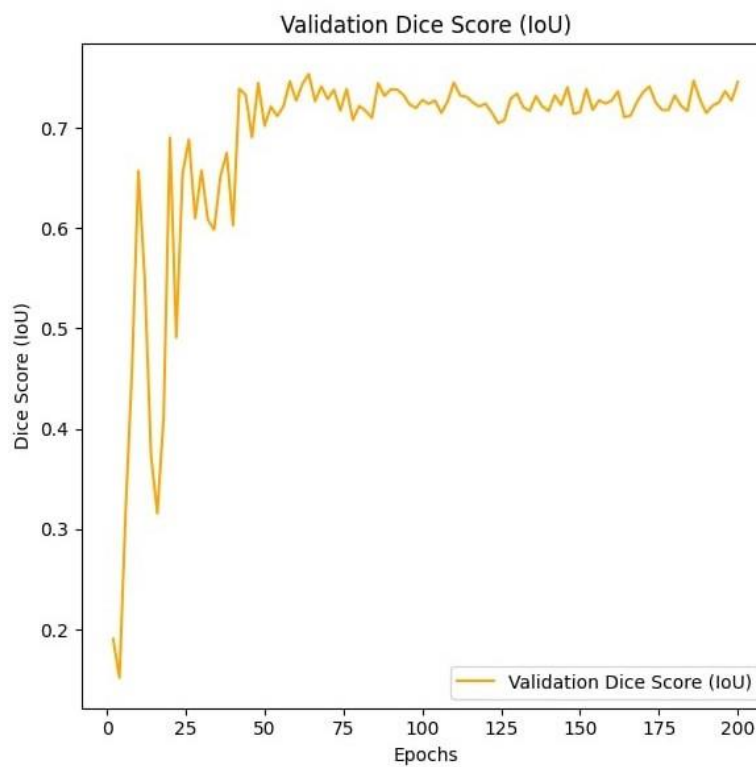


Figure 40: DeepLabv3 validation dice score (IoU)

Παρατηρήσεις:

Για όλα τα μοντέλα οι καμπύλες παρουσιάζουν αρχικά μια απότομη άνοδο αυτό συμβαίνει γιατί στις πρώτες εποχές το validation Dice Score είναι χαμηλό και αυξάνεται επειδή τα μοντέλα μαθαίνουν από τα δεδομένα και δεν έχουν ακόμα “κατανοήσει” καλά τα μοτίβα του προβλήματος αλλά σταδιακά βελτιώνουν την ικανότητά τους να εντοπίζουν και να ταξινομούν τα σωστά εικονοστοιχεία στα δεδομένα validation.

Στην συνέχεια οι καμπύλες σταθεροποιούνται ή/και κορυφώνονται (εξισορρόπηση εκπαίδευσης) αφού τα μοντέλα έχουν μάθει το μέγιστο δυνατό από τα δεδομένα training και validation που τους δόθηκαν. Οι σταθερές καμπύλες με υψηλές τιμές υποδηλώνουν ότι τα μοντέλα γενικεύουν καλά και χωρίς overfitting. Στόχος είναι μια σταθερή ή αργά αυξανόμενη καμπύλη validation Dice score.

Για να έρθουν οι καμπύλες στην επιθυμητή μορφή μπορεί να χρησιμοποιηθούν οι τεχνικές early stopping οπότε η εκπαίδευση σταματά όταν η καμπύλη validation Dice score σταθεροποιηθεί για αρκετές εποχές, αύξηση δεδομένων και data augmentation για περιορισμό του overfitting (στην παρούσα εργασία χρησιμοποιήθηκε data augmentation), ρύθμιση της αρχιτεκτονικής σε περίπτωση που η πολυπλοκότητα του μοντέλου δεν αντιστοιχεί στο πρόβλημα.

Μετά το testing κάθε μοντέλου με φόρτωση των βαρών όπως αυτά έχουν προκύψει από το αντίστοιχο training και validation έχουμε τα εξής δεδομένα:

	UNet	SegResNet	PSPNet	DeepLabv3
Test Average IoU	0.1521	0.1378	0.1450	0.1529
Test Accuracy	0.9581	0.9541	0.9682	0.9749
Precision	0.8345	0.8734	0.9174	0.9152
Recall	0.9284	0.8415	0.8853	0.9332
F1Score	0.8789	0.8571	0.9010	0.9241

Τα αποτελέσματα της αξιολόγησης των μοντέλων βάσει διαφόρων δεικτών απόδοσης παρέχουν πολύτιμες πληροφορίες για την κατανόηση της ικανότητάς τους να επιλύσουν το πρόβλημα της τμηματοποίησης. Για εργασίες segmentation, οι μετρικές όπως το IoU, accuracy, precision, recall και F1Score όσο μεγαλύτερη τιμή μπορούμε να επιτύχουμε τόσο το καλύτερο εφόσον γνωρίζουμε ήδη ότι τα μοντέλα μας δεν κάνουν overfitting λόγω της σταδιακής προόδου των διαγραμμάτων τους από το training όπως μελετήθηκε νωρίτερα.

Συμπεράσματα

Το IoU (Intersection over Union), ως βασικός δείκτης μέτρησης της επικάλυψης μεταξύ της προβλεπόμενης μάσκας (predicted mask) και της πραγματικής (ground truth mask), ανέδειξε ως καλύτερο μοντέλο το **DeepLabv3** με τιμή 0.1529. Η απόδοση αυτή δικαιολογείται από την προηγμένη αρχιτεκτονική του μοντέλου, που βασίζεται στο backbone ResNet50, αξιοποιώντας τόσο το βάθος όσο και την ικανότητά του να εξαγάγει πλούσιες χωρικές και θεματικές πληροφορίες από την εικόνα. Επιπλέον, η χρήση atrous convolutions στο DeepLabv3 επιτρέπει την επεξεργασία χαρακτηριστικών σε διαφορετικές κλίμακες, καθιστώντας το εξαιρετικά κατάλληλο για τη λεπτομερή τμηματοποίηση αντικειμένων. Από τα μοντέλα μας βλέπουμε η σειρά από την χειρότερη προς την καλύτερη απόδοση με κριτήριο το IoU είναι: SegResNet, PSPNet, Unet, DeepLabv3

Η ορθότητα (accuracy) των μοντέλων, που αξιολογεί το ποσοστό των σωστά ταξινομημένων εικονοστοιχείων (τόσο στο background όσο και στο foreground/ROI), κατέδειξε επίσης την ανωτερότητα του DeepLabv3, με ποσοστό 97.49%. Παρόλο που η ορθότητα αποτελεί έναν γενικό δείκτη καθώς μπορεί να μην μπορεί να εντοπίσει ανισορροπίες στο σύνολο δεδομένων, τα αποτελέσματα είναι λογικά, καθώς το DeepLabv3 συνδυάζει την πληροφορία χαμηλού επιπέδου (λεπτομέρειες της εικόνας) με την πληροφορία υψηλού επιπέδου (δομή και περιγράμματα αντικειμένων). Τα υπόλοιπα μοντέλα, όπως το **UNet** και το **SegResNet**, που βασίζονται περισσότερο σε παραδοσιακές αρχιτεκτονικές CNN, εμφανίζουν χαμηλότερη ορθότητα λόγω των περιορισμένων δυνατοτήτων τους να διατηρούν πληροφορίες πολλαπλής κλίμακας. Από τα μοντέλα μας βλέπουμε η σειρά από την χειρότερη προς την καλύτερη απόδοση με κριτήριο το accuracy είναι: SegResNet, Unet, PSPNet, DeepLabv3

Όσον αφορά την ακρίβεια (precision), που εκφράζει την αναλογία των σωστά προβλεπόμενων θετικών εικονοστοιχείων μεταξύ όλων των θετικών προβλέψεων (foreground pixels), το PSPNet υπερτερεί με ποσοστό 91.74%. Αυτό είναι αναμενόμενο, καθώς η αρχιτεκτονική του PSPNet περιλαμβάνει pooling πολλαπλών κλιμάκων, που του επιτρέπει να αναγνωρίζει με ακρίβεια τα θετικά pixels αποφεύγοντας την υπερεκτίμηση (false positives). Ωστόσο, η χαμηλότερη ανάκληση (recall) του PSPNet σε σχέση με το DeepLabv3 (88.53% έναντι 93.32%) υποδεικνύει ότι μπορεί να χάνει ορισμένα θετικά εικονοστοιχεία, κάτι που ενδέχεται να οφείλεται στις διαφορετικές παραδοχές της αρχιτεκτονικής του σχετικά με τις λεπτομέρειες των χαρακτηριστικών.

Η ανάκληση (recall), εκφράζει την ικανότητα του μοντέλου να ανιχνεύει όλα τα θετικά εικονοστοιχεία, μέσω της αναλογίας σωστά προσδιορισμένων θετικών εικονοστοιχείων (predicted foreground) μεταξύ όλων των θετικών εικονοστοιχείων βασικής αλήθειας (ground truth foreground). Η ανάκληση είναι υψηλότερη στο DeepLabv3. Η ισχυρή αυτή απόδοση αποδίδεται στη χρήση atrous convolutions, που βελτιώνουν τη δειγματοληψία χωρίς να μειώνουν την ανάλυση, και στον ισχυρό αποκωδικοποιητή που ανασυνθέτει με ακρίβεια τις λεπτομέρειες της μάσκας.

Η F1-score, η αρμονική μέση της ακρίβειας και της ανάκλησης, είναι επίσης υψηλότερη για το DeepLabv3 (92.41%), επιβεβαιώνοντας τη συνολική ισορροπία του μοντέλου. Το αποτέλεσμα αυτό είναι σύμφωνο με τη φύση του προβλήματος της τμηματοποίησης πολυπόδων από ιατρικές εικόνες, όπου τόσο η ακριβής αναγνώριση των αντικειμένων όσο και η σωστή κάλυψη της περιοχής τους είναι κρίσιμες. Το υψηλό F1-score υποδεικνύει την ισορροπία του μοντέλου μεταξύ της ακρίβειας ανίχνευσης θετικών και του περιορισμού ψευδώς θετικών και αρνητικών.

Αξίζει να σημειωθεί ότι τα αποτελέσματα αυτά ευθυγραμμίζονται με τη θεωρητική κατανόηση των αρχιτεκτονικών. Τα μοντέλα UNet και SegResNet, που είναι λιγότερο περίπλοκα και βασίζονται κυρίως σε συμμετρικές δομές και residual συνδέσεις αντίστοιχα, δυσκολεύονται να ανταγωνιστούν τις πιο προηγμένες αρχιτεκτονικές όπως το DeepLabv3 και το PSPNet. Παρόλα αυτά πιο ελαφριά μοντέλα όπως το UNet μπορεί να είναι προτιμητέα σε περιπτώσεις με περιορισμένους υπολογιστικούς πόρους. Η συνολική ανάλυση αναδεικνύει ότι το DeepLabv3 με backbone ResNet50 είναι η καλύτερη επιλογή για το συγκεκριμένο πρόβλημα, καθώς παρέχει την υψηλότερη ακρίβεια και F1-score.

Είναι σημαντικό να αναγνωριστεί ότι τα υπόλοιπα μοντέλα, όπως το PSPNet, το UNet, και το SegResNet, παρείχαν αξιόλογα αποτελέσματα, ενώ η επιλογή του κατάλληλου μοντέλου εξαρτάται από τις απαιτήσεις εφαρμογής, όπως η ανάγκη για ταχύτητα ή η ευχρηστία σε πιο περιορισμένα υπολογιστικά περιβάλλοντα. Παρόλα αυτά, η επιλογή του κατάλληλου μοντέλου εξαρτάται πάντα από τις απαιτήσεις του εκάστοτε προβλήματος και το συμπέρασμα που βγήκε από την σύγκριση των μοντέλων αφορά το συγκεκριμένο dataset και πρόβλημα και είναι εσφαλμένο να θεωρηθεί ότι το DeepLabv3 είναι πάντα η καλύτερη επιλογή μεταξύ των αναφερθέντων μοντέλων. Για παράδειγμα, αν η ταχύτητα και η αποδοτικότητα είναι σημαντικοί παράγοντες, τότε ελαφρύτερα μοντέλα όπως το UNet ή το SegResNet μπορεί να είναι πιο κατάλληλα, καθώς το DeepLabv3 χρησιμοποιεί πολλούς πόρους και αργεί να εκπαιδευτεί σε σχέση με τα άλλα.

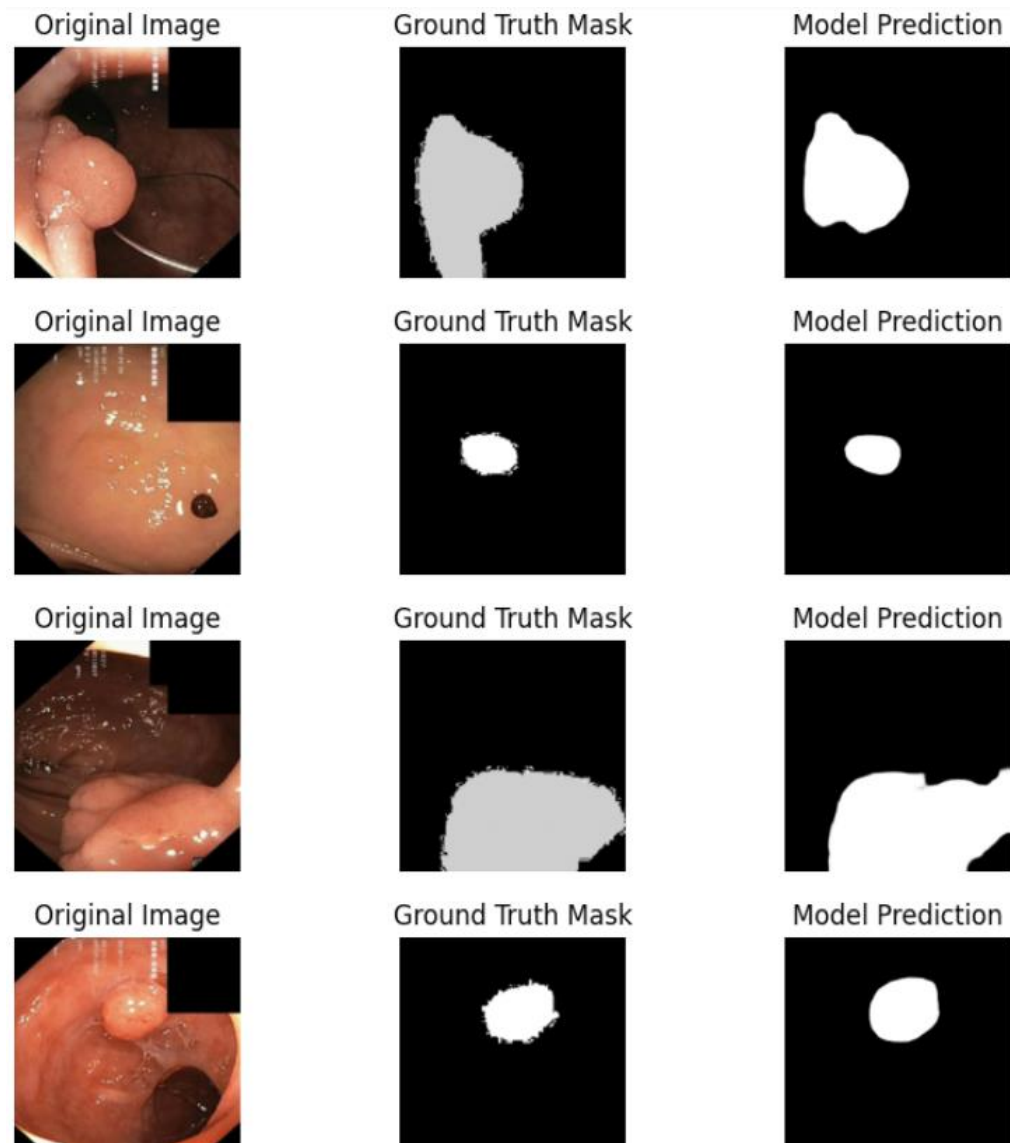


Figure 41: απεικόνιση αποτελεσμάτων DeepLabv3 για εκπαίδευση 200 εποχών.
Οι ground truth masks εμφανίζονται blurry στα όρια λόγω του ότι εμφανίζονται σε binary μορφή με μηδενικό threshold

VI. ΜΕΛΛΟΝΤΙΚΕΣ ΒΕΛΤΙΩΣΕΙΣ/ΠΡΟΕΚΤΑΣΕΙΣ

Η παρούσα μελέτη ανοίγει το δρόμο για μελλοντικές βελτιώσεις και επεκτάσεις που θα μπορούσαν να ενισχύσουν την ακρίβεια, την αποδοτικότητα και την ευελιξία των μεθόδων που χρησιμοποιήθηκαν. Παρακάτω παρατίθενται ορισμένες προτεινόμενες κατευθύνσεις:

1. **Περισσότεροι Πόροι και Εργαλεία:** Η ενσωμάτωση περισσότερων πόρων, όπως προηγμένα μοντέλα και βιβλιοθήκες, μπορεί να ενισχύσει την υποδομή των πειραμάτων. Η χρήση σύγχρονων εργαλείων μπορεί επίσης να διευκολύνει την ανάλυση δεδομένων και τη βελτιστοποίηση των δικτύων.

2. **Ανάπτυξη Νέων Αρχιτεκτονικών:** Η ανάπτυξη αρχιτεκτονικών που συνδυάζουν τα καλύτερα στοιχεία από υπάρχοντα μοντέλα μπορεί να προσφέρει καινοτόμες λύσεις. Για παράδειγμα, η σύνθεση χαρακτηριστικών από CNNs, Transformers και υβριδικές προσεγγίσεις μπορεί να αξιοποιήσει τα πλεονεκτήματα διαφορετικών τεχνολογιών.

3. **Μελέτη Προεκπαιδευμένων Δικτύων (Pretrained Models):** Η αξιοποίηση προεκπαιδευμένων δικτύων μπορεί να μειώσει τον απαιτούμενο χρόνο εκπαίδευσης και να ενισχύσει την απόδοση. Επίσης, η πειραματική χρήση διαφορετικών μορφών δικτύων, όπως Transformers αντί για CNNs, ενδέχεται να αποκαλύψει εναλλακτικές λύσεις για συγκεκριμένα προβλήματα.

4. **Πειράματα Βελτιστοποίησης:** Διεξοδική διερεύνηση των υπερπαραμέτρων (learning rate, batch size, optimizer κ.λπ.) για τη βελτιστοποίηση της απόδοσης και ανάπτυξη προσαρμοσμένων τεχνικών προεπεξεργασίας, λαμβάνοντας υπόψη τις ανάγκες του κάθε δικτύου.

5. **Επέκταση Dataset και Σύγκριση:** Η σύγκριση με περισσότερα datasets μπορεί να ενισχύσει την αξιοπιστία των αποτελεσμάτων. Επιπλέον, η ανάλυση της απόδοσης σε διαφορετικά σύνολα δεδομένων θα αναδείξει τις γενικευτικές ικανότητες των προτεινόμενων αρχιτεκτονικών.

6. **Διερεύνηση Ειδικών Αρχιτεκτονικών:** Η δοκιμή εξειδικευμένων μοντέλων, όπως τα παρακάτω, μπορεί να προσφέρει βελτιώσεις σε συγκεκριμένες εφαρμογές:

- PraNet [82]: Για βελτίωση στην τμηματοποίηση μικρών και λεπτομερών περιοχών.
- HRNetV2 [83]: Για καλύτερη αντιμετώπιση αντικειμένων πολλαπλής κλίμακας [84].
- ResUNet [85] και HarDNet-DFUS [86]: Για συνδυασμό ακρίβειας και ταχύτητας.
- ColonFormer [87]: Για εφαρμογές που απαιτούν μηχανισμούς προσοχής σε λεπτομερή χαρακτηριστικά.
- MSRF-Net [88]: Για ενισχυμένη ανάλυση πολλαπλής κλίμακας.

7. **Εξερεύνηση επιπρόσθετων μοντέλων,** όπως:

- FCN (Fully Convolutional Networks): Απλό και αποδοτικό μοντέλο για βασικές εφαρμογές.
- Transformers: Για την ενσωμάτωση δυναμικών μηχανισμών προσοχής.
- Υβριδικά Δίκτυα: Συνδυασμός CNNs και Transformers για τη βέλτιστη αξιοποίηση των χαρακτηριστικών.

Στο πλαίσιο της διερεύνησης των κατάλληλων αρχιτεκτονικών για την τμηματοποίηση ιατρικών εικόνων κολονοσκόπησης, είναι αξιοσημείωτο να αναφερθεί το μοντέλο DUCK Net. Παρόλο που δεν συμπεριλήφθηκε στην παρούσα ανάλυση, το DUCK Net αναγνωρίζεται ευρέως στη βιβλιογραφία για την υψηλή του απόδοση σε προβλήματα τμηματοποίησης. Η αρχιτεκτονική του συνδυάζει προηγμένες τεχνικές επεξεργασίας εικόνας με βαθιά μάθηση, γεγονός που το καθιστά ιδιαίτερα αποτελεσματικό σε πολύπλοκα ιατρικά δεδομένα.

Η αξιολόγηση του DUCK Net σε μελλοντικές έρευνες θα μπορούσε να προσφέρει σημαντικές συγκριτικές πληροφορίες και να ενισχύσει την κατανόηση των δυνατοτήτων του στην ενίσχυση της διάγνωσης μέσω της τεχνητής νοημοσύνης. Προτείνεται η ενσωμάτωσή του σε μελλοντικές μελέτες, καθώς και η σύγκρισή του με άλλα σύγχρονα μοντέλα, προκειμένου να διαπιστωθεί η συμβολή του στην περαιτέρω βελτίωση της ακρίβειας και της γενίκευσης στην τμηματοποίηση πολυπόδων.

Με τις παραπάνω κατευθύνσεις, η έρευνα μπορεί να επεκταθεί σημαντικά, επιτυγχάνοντας ακόμα μεγαλύτερη ακρίβεια και αποτελεσματικότητα σε σύνθετες εφαρμογές ανάλυσης δεδομένων και τμηματοποίησης.

VII. ΑΝΤΙ ΕΠΙΛΟΓΟΥ

Ήταν μεγάλη μου χαρά που, μέσα από αυτή τη διπλωματική εργασία, είχα την ευκαιρία να έρθω σε επαφή με τη μαγεία της τεχνητής νοημοσύνης και τις εξαιρετικά σημαντικές εφαρμογές της στην ιατρική.

Σε μια εποχή που οι ανάγκες της ανθρωπότητας αλλάζουν ραγδαία, συχνά ξεχνάμε την ανεκτίμητη αξία της έρευνας στην ιατρική, στρέφοντας την προσοχή μας σε πιο "μοντέρνες" επιδιώξεις, οι οποίες συχνά αποσκοπούν μόνο στο κέρδος. Ωστόσο, η ιατρική και η βιοϊατρική τεχνολογία είναι επιστήμες που όχι μόνο διασφαλίζουν την ποιότητα ζωής μας αλλά και την ίδια την ύπαρξή μας. Γι' αυτό, αποτελεί για μένα τιμή να προσφέρω, έστω και μια μικρή συμβολή, σε αυτόν τον τομέα που τόσο θαυμάζω.

Είναι ιδιαίτερα σημαντικό η πρόοδος στους τομείς αυτούς να έχει ως σκοπό την αντιμετώπιση σοβαρών ασθενειών, όπως ο καρκίνος, και άλλων χρόνιων παθήσεων. Παράλληλα, πρέπει να διασφαλίσουμε ότι οι ανακαλύψεις στην υγεία θα είναι προσβάσιμες σε όλους, ανεξαρτήτως κοινωνικής ή οικονομικής κατάστασης. Η έρευνα οφείλει να στοχεύει σε μια δωρεάν, ποιοτική και καθολική παροχή υγείας, προσφέροντας πραγματική αλλαγή στις ζωές των ανθρώπων σε κάθε γωνιά του κόσμου.

Η έρευνα πρέπει να καθοδηγείται από τις ανάγκες της κοινωνίας και να έχει στόχο τη βελτίωση της ζωής των ανθρώπων. Δεν πρέπει να περιορίζεται σε έργα που υπηρετούν αποκλειστικά εταιρικά συμφέροντα, χωρίς ουσιαστική προσφορά στην κοινωνία. Είναι εξαιρετικά σημαντικό οι επιστήμονες και οι μηχανικοί να έχουμε επίγνωση του αντίκτυπου της δουλειάς μας και να παραμένουμε πιστοί στο αξιακό μας σύστημα τόσο στους σκοπούς του εξυπηρετούμε όσο και στα μέσα που χρησιμοποιούμε.

Οφείλουμε να εργαζόμαστε με γνώμονα το κοινό καλό, προσφέροντας στους συνανθρώπους μας και στον εαυτό μας μια καλύτερη ποιότητα ζωής. Με αυτόν τον τρόπο, μπορούμε να συμβάλουμε ουσιαστικά σε έναν κόσμο πιο δίκαιο, ανθρωπινό και βιώσιμο για όλους, όλες και όλα.

VIII. Εξήγηση ξενόγλωσσων όρων και αρκτικόλεξων:

Adam Optimizer: Προσαρμοστικός Βελτιστοποιητής Ρυθμού Μάθησης

ASPP (Atrous Spatial Pyramid Pooling): Διατεταγμένη Χωρική Πυραμιδική Δειγματοληψία

Atrous Convolution: Διατεταγμένη Συνέλιξη

CNN (Convolutional Neural Network): Συνελικτικό Νευρωνικό Δίκτυο

CRC (Colorectal Cancer): Καρκίνος Παχέος Εντέρου και Ορθού

CT (Computed Tomography): Υπολογιστική Τομογραφία

CTC (Computed Tomography Colonography): Αξονική Κολονογραφία

Dice Loss: Συνάρτηση Απώλειας Dice(ζαρι)

FAP (Familial Adenomatous Polyposis): Οικογενής αδενωματώδης πολυποδίαση

FIT (Fecal Immunochemical Test): Ανοσοχημική εξέταση κοπράνων

GCO (Global Cancer Observatory): Παγκόσμιο Παρατηρητήριο Καρκίνου

HDI (Human Development Index): Δείκτης Ανθρώπινης Ανάπτυξης

IoU (Intersection over Union): Τομή επί Ένωση

JPS (Juvenile Polyposis Syndrome): Σύνδρομο νεανικής πολυποδίασης

MAP (MUTYH-Associated Polyposis): Σύνδρομο πολυποδίασης σχετικό με το γονίδιο MUTYH

MRI (Magnetic Resonance Imaging): Μαγνητική Τομογραφία

PHTS (PTEN Hamartoma Tumor Syndrome): Σύνδρομο PTEN (Όγκου Αμαρτώματος)

PJS (Peutz-Jeghers Syndrome): Σύνδρομο Peutz-Jeghers

PPM (Pyramid Pooling Module): Πυραμιδικό Μοντέλο Δειγματοληψίας

ReLU (Rectified Linear Unit): Διορθωμένη Γραμμική Μονάδα

VC (Virtual Colonoscopy): Εικονική Κολονοσκόπηση

WHO (World Health Organization): Παγκόσμιος Οργανισμός Υγείας

ΙΧ. Ιστοσελίδες & διαδικτυακά εργαλεία που χρησιμοποιήθηκαν:

<https://paperswithcode.com/>

<https://qco.iarc.fr/en>

<https://seer.cancer.gov/explorer/>

<https://monai.io/>

<https://www.mednet.gr/>

<https://colab.research.google.com/>

<https://chatgpt.com/>

<https://arxiv.org/>

<https://datasets.simula.no/kvasir-seg/>

<https://github.com/>

X. ΒΙΒΛΙΟΓΡΑΦΙΑ

-
- [1] R. M. H. McMinn, R. T. Hutchings, J. Pegington, and P. H. Abrahams, *Ανατομία του Ανθρώπου - Με Έγχρωμες Εικόνες*, Πασχαλίδης, 2004.
- [2] L. P. Gartner and J. L. Hiatt, "Digestive System," in *Concise Histology*, Saunders Elsevier, 2010, pp. 235–260.
- [3] L. Varytimiadis, *Cold snare polypectomy vs. hot snare polypectomy vs. argon plasma coagulation for small (5-9 mm) left-sided colorectal polyps*, Ph.D. dissertation, School of Medicine, University of Athens, 2022. Available: <https://doi.org/10.12681/eadd/51876>.
- [4] P. Sfikakis *et al.*, *Ασκήσεις Σημειολογίας και Διαφορικής Διαγνωστικής στην Παθολογία (Chapter 2)*, Kallipos, Open Academic Editions, 2015, p. 63. Available: <https://hdl.handle.net/11419/6092>.
- [5] C. P. Spanos, "Rectal cancer," in *Digestive System Malignancies*, Chapter 16, Academic Press, 2022, pp. 73–77. Available: <https://doi.org/10.1016/B978-0-323-98369-3.00018-6>.
- [6] E. E. Torlakovic *et al.*, "Sessile serrated adenoma (SSA) vs traditional serrated adenoma (TSA)," *Am. J. Surg. Pathol.*, vol. 32, pp. 21–29, 2008.
- [7] A. E. Noffsinger, "Serrated polyps and colorectal cancer: New pathway to malignancy," *Annu. Rev. Pathol.*, vol. 4, pp. 343–364, 2009.
- [8] L. A. Torre *et al.*, "Global cancer statistics, 2012," *CA Cancer J. Clin.*, vol. 65, no. 2, pp. 87–108, 2015.
- [9] A. G. Rundle *et al.*, "Colonoscopic screening in average-risk individuals ages 40 to 49 vs 50 to 59 years," *Gastroenterology*, vol. 134, no. 5, pp. 1311–1315, 2008.
- [10] I. Ewing, J. J. Hurley, E. Josephides, and A. Millar, "The molecular genetics of colorectal cancer," *Frontline Gastroenterol.*, vol. 5, pp. 26–30, 2014.
- [11] B. Vogelstein *et al.*, "Genetic alterations during colorectal-tumor development," *N. Engl. J. Med.*, vol. 319, pp. 525–532, 1988.
- [12] L. Varytimiadis *et al.*, "Cold snare polypectomy vs. hot snare polypectomy vs. argon plasma coagulation for small (5–9mm) left-sided colorectal polyps: a prospective randomized trial," *Eur. J. Gastroenterol. Hepatol.*, vol. 33, pp. e909–e915, 2021..
- [13] S. J. Stryker *et al.*, "Natural history of untreated colonic polyps," *Gastroenterology*, vol. 93, pp. 1009–1013, 1987.
- [14] J. Ferlay *et al.*, *Global Cancer Observatory: Cancer Today*. Lyon, France: International Agency for Research on Cancer. Available: <https://gco.iarc.fr/today> (Accessed: Nov. 6, 2025).
- [15] F. Bray, J. Ferlay, I. Soerjomataram, *et al.*, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, pp. 394–424, 2018.
- [16] J. Ferlay, M. Ervik, F. Lam, *et al.*, *Global Cancer Observatory: Cancer Today*, Lyon, France: International Agency for Research on Cancer, 2022. Available: https://gco.iarc.who.int/today/en/dataviz/bars?mode=cancer&key=asr&group_populations=1&types=0_1&sexes=1_2&sort_by=value1&populations=981_982_983_984&multiple_populations=1&values_position=out&cancers_h=41 (Accessed: Jan. 23, 2025).

-
- [17] M. Arnold, M. S. Sierra, M. Laversanne, *et al.*, “Global patterns and trends in colorectal cancer incidence and mortality,” *Gut*, vol. 66, pp. 683–691, 2017.
 - [18] B. K. Edwards, E. Ward, B. A. Kohler, *et al.*, “Annual report to the nation on the status of cancer, 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates,” *Cancer*, vol. 116, pp. 544–573, 2010.
 - [19] S. E. Kim, H. Y. Paik, H. Yoon, *et al.*, “Sex- and gender-specific disparities in colorectal cancer risk,” *World Journal of Gastroenterology*, vol. 21, pp. 5167–5175, 2015.
 - [20] SEERExplorer: An interactive website for SEER cancer statistics* [Online]. Available: <https://seer.cancer.gov/explorer/> (Accessed: Jan. 24, 2025).
 - [21] L. Ricciardiello, D. J. Ahnen, and P. M. Lynch, “Chemoprevention of hereditary colon cancers: Time for new strategies,” *Nature Reviews Gastroenterology & Hepatology*, vol. 13, pp. 352–361, 2016.
 - [22] D. Calva and J. R. Howe, “Hamartomatous polyposis syndromes,” *Surgical Clinics of North America*, vol. 88, pp. 779–817, 2008.
 - [23] M. De Rosa, U. Pace, D. Rega, *et al.*, “Genetics, diagnosis and management of colorectal cancer (Review),” *Oncology Reports*, vol. 34, pp. 1087–1096, 2015.
 - [24] M. W. Lutgens, M. G. van Oijen, G. J. van der Heijden, *et al.*, “Declining risk of colorectal cancer in inflammatory bowel disease: An updated meta-analysis of population-based cohort studies,” *Inflammatory Bowel Diseases*, vol. 19, pp. 789–799, 2013.
 - [25] K. Nottage, J. McFarlane, M. J. Krasin, *et al.*, “Secondary colorectal carcinoma after childhood cancer,” *Journal of Clinical Oncology*, vol. 30, pp. 2552–2558, 2012.
 - [26] D. Desautels, P. Czaykowski, Z. Nugent, *et al.*, “Risk of colorectal cancer after the diagnosis of prostate cancer: A population-based study,” *Cancer*, vol. 122, pp. 1254–1260, 2016.
 - [27] A. Yamada, Y. Komaki, F. Komaki, *et al.*, “Risk of gastrointestinal cancers in patients with cystic fibrosis: A systematic review and meta-analysis,” *Lancet Oncology*, vol. 19, pp. 758–767, 2018.
 - [28] Y. Zhang, H. Liu, L. Li, *et al.*, “Cholecystectomy can increase the risk of colorectal cancer: A meta-analysis of 10 cohort studies,” *PLoS One*, vol. 12, e0181852, 2017.
 - [29] S. Gillessen, A. Templeton, G. Marra, *et al.*, “Risk of colorectal cancer in men on long-term androgen deprivation therapy for prostate cancer,” *Journal of the National Cancer Institute*, vol. 102, pp. 1760–1770, 2010.
 - [30] T. E. Robsahm, B. Aagnes, A. Hjartaker, *et al.*, “Body mass index, physical activity, and colorectal cancer by anatomical subsites: A systematic review and meta-analysis of cohort studies,” *European Journal of Cancer Prevention*, vol. 22, pp. 492–505, 2013.
 - [31] M. Song, W. S. Garrett, and A. T. Chan, “Nutrients, foods, and colorectal cancer prevention,” *Gastroenterology*, vol. 148, pp. 1244–1260.e1216, 2015.
 - [32] J. M. Ordóñez-Mena, V. Walter, B. Schottker, *et al.*, “Impact of prediagnostic smoking and smoking cessation on colorectal cancer prognosis: A meta-analysis of individual patient data from cohorts within the CHANCES consortium,” *Annals of Oncology*, vol. 29, pp. 472–483, 2018.
 - [33] V. Fedirko, I. Tramacere, V. Bagnardi, *et al.*, “Alcohol drinking and colorectal cancer risk: An overall and dose-response meta-analysis of published studies,” *Annals of Oncology*, vol. 22, pp. 1958–1972, 2011.

-
- [34] P. M. Rothwell, M. Wilson, C. E. Elwin, *et al.*, "Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials," *Lancet*, vol. 376, pp. 1741–1750, 2010.
 - [35] N. Thosani, S. N. Thosani, S. Kumar, *et al.*, "Reduced risk of colorectal cancer with use of oral bisphosphonates: A systematic review and meta-analysis," *Journal of Clinical Oncology*, vol. 31, pp. 623–630, 2013.
 - [36] G. A. Makar, J. H. Holmes, and Y. X. Yang, "Angiotensin-converting enzyme inhibitor therapy and colorectal cancer risk," *Journal of the National Cancer Institute*, vol. 106, djt374, 2014.
 - [37] Y. Pang, C. Kartsonaki, Y. Guo, *et al.*, "Diabetes, plasma glucose and incidence of colorectal cancer in Chinese adults: A prospective study of 0.5 million people," *Journal of Epidemiology and Community Health*, vol. 72, pp. 919–925, 2018.
 - [38] F. A. Haggard and R. P. Boushey, "Colorectal cancer epidemiology: Incidence, mortality, survival, and risk factors," *Clinics in Colon and Rectal Surgery*, vol. 22, no. 4, pp. 191–197, 2009.
 - [39] A. Shaikat, *et al.*, "ACG Clinical Guidelines: Colorectal Cancer Screening 2021," *The American Journal of Gastroenterology*, vol. 116, no. 3, pp. 458–479, 2021. Available: <https://doi.org/10.14309/ajg.0000000000001122>.
 - [40] B. K. Edwards, E. Ward, B. A. Kohler, *et al.*, "Annual report to the nation on the status of cancer, 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates," *Cancer*, vol. 116, pp. 544–573, 2010.
 - [41] M. Sideris and S. Papagrigroriadis, "Molecular biomarkers and classification models in the evaluation of the prognosis of colorectal cancer," *Anticancer Research*, vol. 34, pp. 2061–2068, 2014.
 - [42] M. Kekelidze, L. D'Errico, M. Pansini, *et al.*, "Colorectal cancer: current imaging methods and future perspectives for the diagnosis, staging and therapeutic response evaluation," *World Journal of Gastroenterology*, vol. 19, pp. 8502–8514, 2013.
 - [43] D. Colussi, G. Brandi, F. Bazzoli, and L. Ricciardiello, "Molecular pathways involved in colorectal cancer: implications for disease behavior and prevention," *International Journal of Molecular Sciences*, vol. 14, pp. 16365–16385, 2013.
 - [44] N. Hayashi, S. Tanaka, D. G. Hewett, *et al.*, "Endoscopic prediction of deep submucosal invasive carcinoma: validation of the narrow-band imaging international colorectal endoscopic (NICE) classification," *Gastrointestinal Endoscopy*, vol. 78, issue 4, pp. 625–632, 2013.
 - [45] The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002, *Gastrointestinal Endoscopy*, vol. 58, pp. S3–S43, 2003.
 - [46] S. Kudo, R. Lambert, J. I. Allen, *et al.*, "Nonpolypoid neoplastic lesions of the colorectal mucosa," *Gastrointestinal Endoscopy*, vol. 68, no. 4 Suppl, pp. S3–S47, 2008.
 - [47] J. E. G. IJspeert, B. A. J. Bastiaansen, M. E. van Leerdam, *et al.*, "Development and validation of the WASP classification system for optical diagnosis of adenomas, hyperplastic polyps and sessile serrated adenomas/polyps," *Gut*, vol. 65, pp. 963–970, 2016.
 - [48] C. E. Dukes, "The classification of cancer of the rectum," *Journal of Pathology and Bacteriology*, vol. 35, pp. 323–332, 1932.
 - [49] J. R. Jass and B. C. Morson, "Reporting colorectal cancer," *Journal of Clinical Pathology*, vol. 40, no. 9, pp. 1016–1023, 1987.

-
- [50] J. C. Van Rijn, J. B. Reitsma, J. Stoker, *et al.*, “Polyp miss rate determined by tandem colonoscopy: a systematic review,” *The American Journal of Gastroenterology*, vol. 101, no. 2, p. 343–350, 2006.
- [51] M. F. Kaminski, P. Wieszczy, M. Rupinski, *et al.*, “Increased rate of adenoma detection associates with reduced risk of colorectal cancer and death,” *Gastroenterology*, vol. 153, no. 1, pp. 98–105, 2017.
- [52] D. L. Pham, C. Xu, and J. L. Prince, “Current methods in medical image segmentation,” *Annual Review of Biomedical Engineering*, vol. 2, no. 1, pp. 315–337, 2000.
- [53] J. Riegler, G. Lux, C. Griwodz, *et al.*, “Multimedia and Medicine: Teammates for Better Disease Detection and Survival,” *ACM Multimedia*, 2016.
- [54] Z. Jia, M. Maggioni, J. Smith, and D. P. Scarpazza, “Dissecting the Nvidia Turing T4 GPU via microbenchmarking,” *arXiv preprint arXiv:1903.07486*, 2019.
- [55] J. Choquette and W. Gandhi, “NVIDIA A100 GPU: Performance & Innovation for GPU Computing,” *2020 IEEE Hot Chips 32 Symposium (HCS)*, Palo Alto, CA, USA, pp. 1–43, 2020, doi: 10.1109/HCS49909.2020.9220622.
- [56] Analytics Vidhya, “A Comprehensive Guide on Deep Learning Optimizers,” Oct. 2021. [Online]. Available: <https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-deep-learning-optimizers/> (Accessed: Jan. 20, 2025).
- [57] MONAI, *Monai 1.4.0 Documentation, Loss Functions, Segmentation Loss*, 2025. [Online]. Available: <https://docs.monai.io/en/stable/losses.html> (Accessed: Jan. 25, 2025).
- [58] K. Pogorelov, K. Randel, C. Griwodz, *et al.*, “Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection,” *Proceedings of the Multimedia Systems Conference (MMSYS)*, ACM, pp. 164–169, 2017.
- [59] D. Jha, P. H. Smedsrud, M. A. Riegler, *et al.*, “Kvasir-SEG: A segmented polyp dataset,” *arXiv preprint arXiv:1911.07069*, 2019.
- [60] M. Sharma, D. Rasmuson, B. Rieger, D. Kjelkerud, *et al.*, “Labelbox: The best way to create and manage training data,” *LabelBox, Inc.*, 2019. [Online]. Available: <https://www.labelbox.com/>. (Accessed: Jan. 17, 2025).
- [61] S. Muhammad and J. Zhang, “Segmentation of liver tumors by Monai and PyTorch in CT images with deep learning techniques,” *Applied Sciences*, vol. 14, p. 5144, 2024. [Online]. Available: <https://doi.org/10.3390/app14125144>.
- [62] M. J. Cardoso, W. Li, R. Brown, *et al.*, “MONAI: An open-source framework for deep learning in healthcare,” *arXiv preprint arXiv:2211.02701*, 2022.
- [63] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. New York: Pearson Education, 2018, pp. 1022.
- [64] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham: Springer, 2015, pp. 234–241.
- [65] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of CVPR*, pp. 3431–3440, 2015.
- [66] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of CVPR*, pp. 770–778, 2016.

-
- [67] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2881–2890. Available: <https://arxiv.org/pdf/1612.01105>.
 - [68] M. Everingham, L. J. V. Gool, C. K. I. Williams, J. M. Winn, and A. Zisserman, "The PASCAL Visual Object Classes VOC challenge," *Int. J. Comput. Vis. (IJCV)*, vol. 88, no. 2, pp. 303–338, 2010.
 - [69] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
 - [70] M. Cordts et al., "The Cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 3213–3223.
 - [71] MONAI, *MONAI: The Medical Open Network for AI*, 2020. Available: <https://github.com/Project-MONAI/MONAI> (Accessed: Feb. 5, 2025).
 - [72] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization", NVIDIA, Santa Clara, CA., 2018. Available: <https://arxiv.org/abs/1810.11654>.
 - [73] S. Bakas et al., "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," *arXiv:1811.02629*, 2018.
 - [74] B. H. Menze et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
 - [75] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 2818–2826. Available: <https://arxiv.org/pdf/1512.00567>.
 - [76] L. Shen et al., "A feasibility study of knee joint semantic segmentation on 3D MR images," *CT Theory Appl.*, vol. 31, no. 5, pp. 531–542, 2022, doi: 10.15953/j.ctta.2022.091.
 - [77] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *arXiv:1802.02611*, 2018.
 - [78] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2961–2969.
 - [79] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 1204–1213.
 - [80] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *arXiv:1606.00915*, 2016.
 - [81] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv*, Dec. 2015. Available: <https://arxiv.org/abs/1512.03385>.
 - [82] D. P. Fan et al., "PraNet: Parallel Reverse Attention Network for Polyp Segmentation," in *Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Lecture Notes in Computer Science, vol. 12266, Springer, Cham, 2020, pp. 263–272. Available: https://doi.org/10.1007/978-3-030-59725-2_26.
 - [83] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 5693–5703.
 - [84] K. Sun et al., "High-resolution representations for labeling pixels and regions," *CoRR*, vol. abs/1904.04514, 2019. Available: <http://arxiv.org/abs/1904.04514>.

-
- [85] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, 2020. Available: <https://doi.org/10.1016/j.isprsjprs.2020.01.013>.
- [86] T. Y. Liao et al., "HarDNet-DFUS: An Enhanced Harmonically-Connected Network for Diabetic Foot Ulcer Image Segmentation and Colonoscopy Polyp Segmentation," *arXiv:2209.07313*, 2022.
- [87] N. T. Duc, N. T. Oanh, N. T. Thuy, T. M. Triet, and V. S. Dinh, "ColonFormer: An efficient transformer-based method for colon polyp segmentation," *IEEE Access*, vol. 10, pp. 80575–80586, 2022. Available: <https://doi.org/10.1109/ACCESS.2022.3195241>.
- [88] A. Srivastava et al., "MSRF-Net: A Multi-Scale Residual Fusion Network for Biomedical Image Segmentation," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 5, pp. 2252–2263, 2022. Available: <https://doi.org/10.1109/JBHI.2021.3138024>.