



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Μέθοδοι επεξηγησιμότητας σε εφαρμογές ιατρικής εικόνες

Μελέτη και υλοποίηση

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΣΤΑΝΤΣΙΟΥ ΒΑΣΙΛΗ

Επιβλέπων: Στέφανος Κόλλιας
Ομότιμος Καθηγητής

Αθήνα, Ιούλιος 2024



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Μέθοδοι επεξηγησιμότητας σε εφαρμογές ιατρικής εικόνες

Μελέτη και υλοποίηση

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΣΤΑΝΤΣΙΟΥ ΒΑΣΙΛΗ

Επιβλέπων: Στέφανος Κόλλιας
Ομότιμος Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 23 Ιουλίου 2024.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Στέφανος Κόλλιας
Ομότιμος Καθηγητής

.....
Αθανάσιος Βουλόδημος
Επίκουρος Καθηγητής

.....
Γεώργιος Στάμου
Καθηγητής

Αθήνα, Ιούλιος 2024



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Copyright © – All rights reserved. Με την επιφύλαξη παντός δικαιώματος.

Βασίλης Στάντσιος, 2024.

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.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

.....

Βασίλης Στάντσιος

Διπλωματούχος

Ηλεκτρολόγος Μηχανικός

και Μηχανικός

Υπολογιστών ΕΜΠ

15 Ιουλίου 2024

Περίληψη

Η ανάπτυξη των νευρωνικών δικτύων είναι ραγδαία και η εφαρμογή τους σε καθημερινές δραστηριότητες στις ζωές των ανθρώπων έχει αρχίσει να πραγματοποιείται. Μία από τις βασικότερες εφαρμογές τους αφορά τον ιατρικό τομέα. Πιο συγκεκριμένα η τμηματοποίηση ιατρικής εικόνας, που αποτελεί έναν εξειδικευμένο κλάδο των νευρωνικών δικτύων, είναι ένα χρήσιμο εργαλείο των γιατρών για γρηγορότερες γνωματεύσεις με λιγότερο κόστος διάγνωσης.

Λόγω της πολυπλοκότητας των νευρωνικών δικτύων η ανθρώπινη εποπτεία καθιστάται ανέφικτη, συνεπώς αμφισβητείται και η αξιοπιστία τους. Ειδικά στον κλάδο της ιατρικής, ένας τέτοιος προβληματισμός για την εγκυρότητα των αποτελεσμάτων, όπου κάθε λάθος μπορεί να αποβεί μοιραίο, πρέπει να αντιμετωπιστεί. Η αντιμετώπιση αυτή τίθεται να λύσει η επεξηγησιμότητα, αιτιολογώντας με ποιο τρόπο το δίκτυο κατέληξε στα συγκεκριμένα αποτελέσματα.

Στην εργασία μας χρησιμοποιούνται τρία σύνολα δεδομένων όπου σε αυτά εκπαιδεύονται τα μοντέλα τμηματοποίησης αξιολογώντας τα αποτελέσματα τους τόσο οπτικά όσο και ποσοτικά ενώ παράλληλα προσπαθούμε να κατανοήσουμε πως αυτά τα μοντέλα έφτασαν στα αποτελέσματα τους μέσω διάφορων μεθόδων επεξηγησιμότητας.

Λέξεις Κλειδιά

τμηματοποίηση ιατρικής εικόνας, σημασιολογική τμηματοποίηση, συνελκτικά δίκτυα, επεξηγησιμότητα, δίκτυα κωδικοποιητή-αποκωδικοποιητή, μετασχηματιστές, συνέλιξη, μηχανισμός προσοχής

Abstract

The development of neural networks is happening rapidly and their applications in daily activities in people's lives has begun to take place. One of the most basic applications has to do with the medical field. In particular, medical image segmentation, which is a specialized branch of neural networks, is a useful tool for doctors enabling quicker medical reports at a lower diagnosis cost.

Due to the complexity of neural networks, human supervision becomes unfeasible, therefore their reliability is questioned. Especially in the field of medicine, such a concern about the validity of the results, where any mistake can be fatal, must be addressed. This issue is tackled by explainability, justifying how the network concluded to the specific results.

In our study, three data sets are used where the segmentation models are trained, evaluating their results bot visually and quantitavily, while we simultaneously try to understand how these models reached their results through various explainability methods.

Keywords

medical image segmentation, semantic segmentation, convolutional networks, explainability, encoder-decoder, transformers, convolution, attention

στους γονείς μου

Ευχαριστίες

Θα ήθελα καταρχήν να ευχαριστήσω τον καθηγητή κ. Στέφανο Κόλλια και την κα. Παρασκευή Τζούβελη για την επίβλεψη αυτής της διπλωματικής εργασίας και για την ευκαιρία που μου έδωσαν να την εκπονήσω στο εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης. Επίσης ευχαριστώ ιδιαίτερα την Παρασκευή Θεοφίλου για την καθοδήγησή της και την εξαιρετική συνεργασία που είχαμε. Τέλος θα ήθελα να ευχαριστήσω τους γονείς μου για την καθοδήγηση και την ηθική συμπαράσταση που μου προσέφεραν όλα αυτά τα χρόνια.

Αθήνα, Ιούλιος 2024

Βασίλης Στάντσιος

Περιεχόμενα

Περίληψη	1
Abstract	3
Ευχαριστίες	7
1 Εισαγωγή	17
1.1 Αντικείμενο της διπλωματικής	17
1.2 Οργάνωση του τόμου	18
I Θεωρητικό Μέρος	19
2 Θεωρητικό υπόβαθρο	21
2.1 Μηχανική μάθηση και Νευρωνικά Δίκτυα	21
2.1.1 Μηχανική μάθηση	21
2.1.2 Νευρωνικά Δίκτυα	22
2.1.3 Συνελκτικά Νευρωνικά Δίκτυα CNN	23
2.1.4 Εξ Ολοκλήρου Συνελκτικά Δίκτυα (FCN)	23
2.2 Δίκτυα Κωδικοποιητή-Αποκωδικοποιητή Encoder-Decoder	24
2.3 Μετασχηματιστές Transformers	25
2.4 Τμηματοποίηση ιατρικής εικόνας	26
2.5 Επεξηγησιμότητα στην Τεχνητή Νοημοσύνη	26
2.6 Συναφείς εργασίες	27
3 Θεωρητική ανάλυση επιλεγμένων τεχνικών επεξηγησιμότητας και μοντέλων τμηματοποίησης	29
3.1 Τεχνικές επεξηγησιμότητας	29
3.1.1 Τεχνική Grad-CAM	29
3.1.2 Τεχνική LayerCAM	31
3.1.3 Τεχνική EigenCAM	31
3.1.4 Τεχνική EigenGradCAM	31
3.1.5 Τεχνική Grad-CAM++	32
3.1.6 Τεχνική XGrad-CAM	32
3.2 Μοντέλα τμηματοποίησης	33
3.2.1 SEGFORMER	33
3.2.2 DeepLabV3+	34

3.2.3 UNet	35
3.2.4 Vision Transformer (ViT)	37
3.2.5 Swin	38
II Πρακτικό Μέρος	41
4 Δεδομένα	43
4.1 Σύνολο δεδομένων	43
4.1.1 Kvasir-SEG	43
4.1.2 COVID-QU-EX	45
4.1.3 The Foot Ulcer Segmentation Challenge (FUSeG)	47
4.2 Προεπεξεργασία και επαύξηση δεδομένων	48
5 Μετρικές αξιολόγησης και περιβάλλον εκτέλεσης	49
5.1 Μετρικές αξιολόγησης	49
5.2 Περιβάλλον εκτέλεσης πειράματος	50
6 Πειραματικά αποτελέσματα μοντέλων τμηματοποίησης	51
6.1 Λεπτομέρειες υλοποίησης	51
6.2 Πειραματικά αποτελέσματα μοντέλων τμηματοποίησης	51
6.2.1 Αποτελέσματα για το σύνολο δεδομένων Kvasir-SEG	52
6.2.2 Αποτελέσματα για το σύνολο δεδομένων FUSeG	55
6.2.3 Αποτελέσματα για το σύνολο δεδομένων COVID-QU-EX	59
7 Πειραματικά αποτελέσματα μεθόδων επεξηγησιμότητας	65
7.1 Πειραματικά αποτελέσματα τεχνικών επεξηγησιμότητας	65
III Επίλογος	69
8 Επίλογος	71
8.1 Συμπεράσματα	71
8.2 Μελλοντικές επεκτάσεις	71
Βιβλιογραφία	78

Κατάλογος Σχημάτων

1.1	Ολοκληρωμένο pipeline της παρούσας διπλωματικής εργασίας	18
2.1	Παράδειγμα Συνελκτικού Νευρωνικού Δικτύου.	23
2.2	Παράδειγμα εξ ολοκλήρου συνελκτικού νευρωνικού Δικτύου.[1]	24
2.3	Παράδειγμα αρχιτεκτονικής μετασχηματιστή [2].	25
3.1	Παράδειγμα λειτουργίας της μεθόδου CAM [3].	30
3.2	Παράδειγμα λειτουργίας της μεθόδου Grad-CAM.[4]	31
3.3	Παράδειγμα της μεθόδου Grad-CAM++ για ταξινόμηση αντικειμένου [5].	32
3.4	Παράδειγμα της μεθόδου XGrad-CAM [6].	32
3.5	Η αρχιτεκτονική του SegFormer αποτελούμενη από τον κωδικοποιητή και αποκωδικοποιητή [7].	33
3.6	Παράδειγμα υλοποίησης του μοντέλου τμηματοποίησης DeepLabv3+ [8].	35
3.7	Διαδοχικές μονάδες χωρίς και με atrous convolution [9].	35
3.8	Η αρχιτεκτονική του Unet.[10]	36
3.9	Η αρχιτεκτονική του ViT.[11]	37
3.10	Η αρχιτεκτονική του UperNet [12].	38
3.11	Η αρχιτεκτονική του μετασχηματιστή SWIN [13].	39
3.12	Η αρχιτεκτονική των Block Swins [13].	39
4.1	Παραδείγματα του Kvasir-SEG Dataset [14]	44
4.2	COVID-19 Images	45
4.3	COVID-19 Lung Masks	45
4.4	COVID-19 Infection Masks	45
4.5	Normal Images	46
4.6	Normal Lung Masks	46
4.7	COVID-19 Images	46
4.8	COVID-19 Mask Images with Colour	47
4.9	Παραδείγματα του FUSeg [15]	47

Κατάλογος Εικόνων

6.1	Αποτελέσματα για το μοντέλο τμηματοποίησης Segformer στο σύνολο δεδομένων Kvasir-SEG.	52
6.2	Αποτελέσματα για το μοντέλο τμηματοποίησης Swin στο σύνολο δεδομένων Kvasir-SEG.	53
6.3	Αποτελέσματα για το μοντέλο τμηματοποίησης DeepLabv3+ στο σύνολο δεδομένων Kvasir-SEG.	53
6.4	Αποτελέσματα για το μοντέλο τμηματοποίησης ViT στο σύνολο δεδομένων Kvasir-SEG.	54
6.5	Αποτελέσματα για το μοντέλο τμηματοποίησης Unet στο σύνολο δεδομένων Kvasir-SEG.	54
6.6	Συγκριτικά αποτελέσματα μοντέλων τμηματοποίησης στο σύνολο δεδομένων Kvasir-SEG.	55
6.7	Αποτελέσματα για το μοντέλο τμηματοποίησης Segformer στο σύνολο δεδομένων FUSeg.	56
6.8	Αποτελέσματα για το μοντέλο τμηματοποίησης Swin στο σύνολο δεδομένων FUSeg.	56
6.9	Αποτελέσματα για το μοντέλο τμηματοποίησης DeepLabV3+ στο σύνολο δεδομένων FUSeg.	57
6.10	Αποτελέσματα για το μοντέλο τμηματοποίησης ViT στο σύνολο δεδομένων FUSeg.	57
6.11	Αποτελέσματα για το μοντέλο τμηματοποίησης Unet στο σύνολο δεδομένων FUSeg.	58
6.12	Συγκριτικά αποτελέσματα μοντέλων τμηματοποίησης στο σύνολο δεδομένων FUSeg.	59
6.13	Αποτελέσματα για το μοντέλο τμηματοποίησης Segformer στο σύνολο δεδομένων COVID-QU-Ex.	60
6.14	Αποτελέσματα για το μοντέλο τμηματοποίησης Swin στο σύνολο δεδομένων COVID-QU-Ex.	61
6.15	Αποτελέσματα για το μοντέλο τμηματοποίησης DeepLabV3+ στο σύνολο δεδομένων COVID-QU-Ex.	61
6.16	Αποτελέσματα για το μοντέλο τμηματοποίησης ViT στο σύνολο δεδομένων COVID-QU-Ex.	62
6.17	Αποτελέσματα για το μοντέλο τμηματοποίησης Unet στο σύνολο δεδομένων COVID-QU-Ex.	62

6.18	Συγκεντρωτικά αποτελέσματα μοντέλων τμηματοποίησης στο σύνολο δεδομένων COVID-GU-Ex.	63
7.1	Οπτικά αποτελέσματα τεχνικών επεξηγησιμότητας για το μοντέλο τμηματοποίησης SEGFORMER.	66
7.2	Οπτικά αποτελέσματα τεχνικών επεξηγησιμότητας για το μοντέλο τμηματοποίησης DeepLabV3+.	66
7.3	Οπτικά αποτελέσματα τεχνικών επεξηγησιμότητας για το μοντέλο τμηματοποίησης Swin.	67
7.4	Οπτικά αποτελέσματα τεχνικών επεξηγησιμότητας για το μοντέλο τμηματοποίησης ViT.	67
7.5	Οπτικά αποτελέσματα τεχνικών επεξηγησιμότητας για το μοντέλο τμηματοποίησης UNet.	68

Κατάλογος Πινάκων

6.1	Αποτελέσματα μεθόδων τμηματοποίησης βασισμένα στις μετρικές IoU και Dice για το σύνολο δεδομένων Kvasir-SEG	55
6.2	Αποτελέσματα μεθόδων τμηματοποίησης βασισμένα στις μετρικές IoU και Dice για το σύνολο δεδομένων FUSeg	59
6.3	Αποτελέσματα μεθόδων τμηματοποίησης βασισμένα στις μετρικές IoU και Dice για το σύνολο δεδομένων COVID-QU-Ex	63

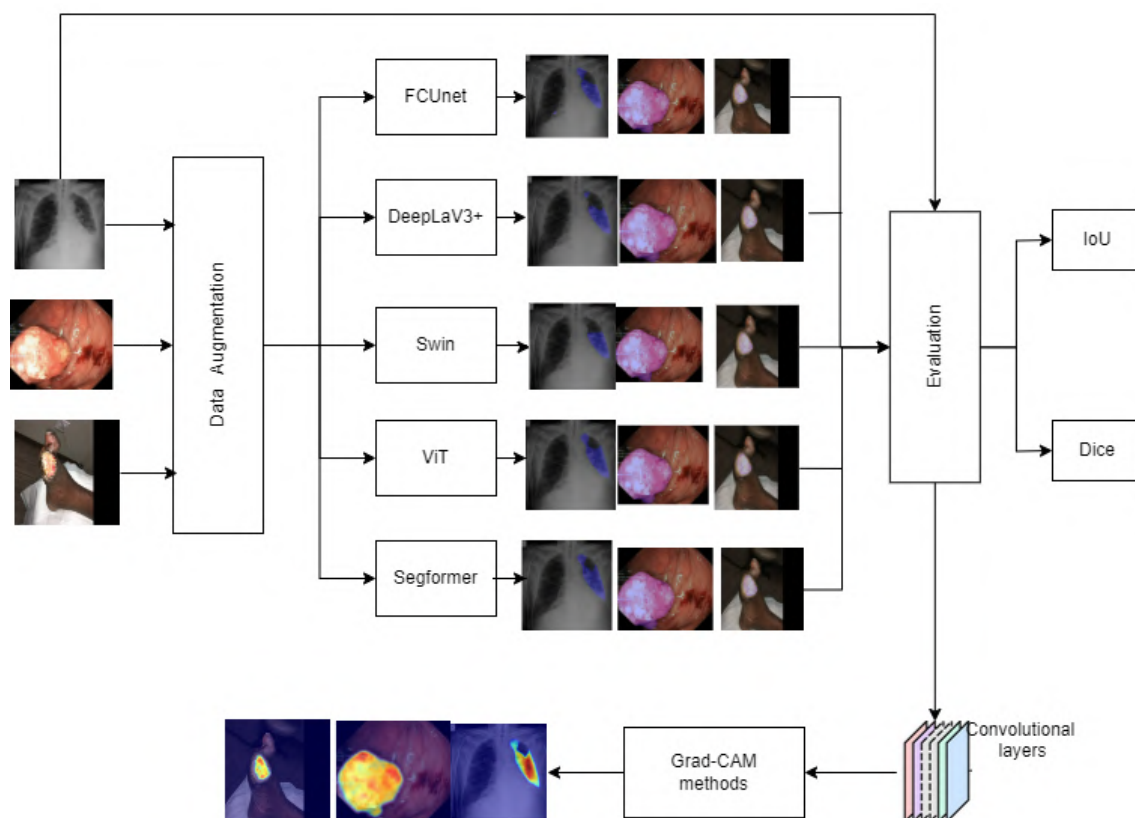
Κεφάλαιο **1**

Εισαγωγή

Η ανάπτυξη των νευρωνικών δικτύων είναι ραγδαία και η εφαρμογή τους σε καθημερινές δραστηριότητες έχει αρχίσει να πραγματοποιείται. Ένας από τους σημαντικότερους κλάδους που έχει εφαρμοστεί και η ανάγκη για εξέλιξη του είναι αναγκαία, αποτελεί ο κλάδος της ιατρικής. Πιο συγκεκριμένα η τμηματοποίηση ιατρικών εικόνων αποτελεί ζωτικής σημασίας έργο για την επιτάχυνση και βελτίωση της διάγνωσης ασθενειών, καθώς και για την εξασφάλιση πιο στοχευμένης και αποτελεσματικής θεραπείας, για ταυτόχρονη εξυπηρέτηση μεγάλο αριθμό ασθενών. Η διαδικασία αυτή εξαλείφει τα λάθη ανθρώπινου παράγοντα και μειώνει το χρόνο της διάγνωσης. Η τμηματοποίηση μέχρι και τα τελευταία χρόνια υλοποιούταν χειροκίνητα, μία υλοποίηση που είναι χρονοβόρα, κουραστική και πρέπει να πραγματοποιηθεί από εξειδικευμένους επαγγελματίες που σχεδιάζουν την περιοχή ενδιαφέροντος ακολουθώντας πρωτόκολλα και έρχοντας σε επαφή με άλλους γιατρούς για να επαληθεύσουν το αποτέλεσμα. Με την άνοδο της μηχανικής μάθησης και του κλάδου της υπολογιστικής όρασης, η δυνατότητα αυτόματης τμηματοποίησης ιατρικών εικόνων είναι πλέον εφικτή. Όσο τα ερωτήματα που γεννιούνται στην σύγχρονη κοινωνία είναι κατά πόσο μπορούμε να εμπιστευτούμε αυτά τα τεχνολογικά επιτεύγματα. Για την γεφύρωση αυτού, μία μέθοδος είναι οι τεχνικές επεξηγησιμότητας που δίνουν οπτικά αποτελέσματα για το πώς παίρνονται οι αποφάσεις αυτών των μοντέλων τμηματοποίησης.

1.1 Αντικείμενο της διπλωματικής

Ορμώμενοι από την ανάγκη της τμηματοποίησης και την επεξήγηση της διαδικασίας αυτής γράφτηκε η παρούσα διπλωματική. Υλοποιήθηκαν 5 μοντέλα τμηματοποίησης, που συγκρίθηκαν μεταξύ τους, καθώς και 6 τεχνικές από την οικογένεια Grad-CAM, με σκοπό την κατανόηση τους. Τα μοντέλα αυτά εκπαιδεύτηκαν πάνω σε τρία διαφορετικά σύνολα δεδομένων, τα οποία δεδομένα προεπεξεργάστηκαν (Data Augmentation) με σκοπό να πάρουμε τα βέλτιστα δυνατά αποτελέσματα. Στο παρακάτω σχήμα δίνεται η διαδικασία που ακολουθήσαμε για την υλοποίηση της διπλωματικής [1.1](#).



Σχήμα 1.1: Ολοκληρωμένο pipeline της παρούσας διπλωματικής εργασίας

1.2 Οργάνωση του τόμου

Η εργασία αυτή είναι οργανωμένη σε 8 κεφάλαια. Το κεφάλαιο 1 αποτελεί την εισαγωγή της παρούσας διπλωματικής. Στο Κεφάλαιο 2 δίνεται το θεωρητικό υπόβαθρο της μελέτης μας καθώς και μία εισαγωγή στα θεμέλια που δημιούργησαν αυτές τις τεχνικές και τα μοντέλα. Στο κεφάλαιο 3 δίνεται η θεωρητική ανάλυση των μοντέλων τμηματοποίησης και των μεθόδων επεξηγησιμότητας που χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία. Στο κεφάλαιο 4 δίνονται τα σύνολα δεδομένων που χρησιμοποιήθηκαν και στο κεφάλαιο 5 η προεπεξεργασία τους. Στο κεφάλαιο 6 παρουσιάζονται οπτικά τα αποτελέσματα που δημιουργήθηκαν από τα επιλεγμένα μοντέλα καθώς και στο κεφάλαιο 7 η επεξηγησιμότητα τους. Τέλος, στο τελευταίο κεφάλαιο δίνονται τα συμπεράσματα της διπλωματικής και οι μελλοντικές επεκτάσεις της.

Μέρος **I**

Θεωρητικό Μέρος

Κεφάλαιο 2

Θεωρητικό υπόβαθρο

Στην ενότητα αυτή παρουσιάζεται μία ανάλυση στην μηχανική μάθηση, στα νευρωνικά δίκτυα καθώς και στα θεωρητικά θεμέλια των μοντέλων τμηματοποίησης και των τεχνικών επεξηγησιμότητας που χρησιμοποιήθηκαν στο πειραματικό μέρος.

2.1 Μηχανική μάθηση και Νευρωνικά Δίκτυα

2.1.1 Μηχανική μάθηση

Η μηχανική μάθηση αποτελεί κομμάτι της Τεχνητής Νοημοσύνης και χρησιμοποιείται για την περιγραφή συστημάτων εκτελώντας εργασίες με εφυσή τρόπους χρησιμοποιώντας την εμπειρία τους και βελτιώνοντας την αποδόση τους χωρίς την ανάγκη επαναπρογραμματισμού ή ανθρώπινης παρέμβασης. Η μηχανική μάθηση ορίζεται από τον Arthur Samuel, αμερικάνος πρωτοπόρος της Τεχνητής Νοημοσύνης, το 1959 ως "Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί". Οι αλγόριθμοι μηχανικής μάθησης δεν βασίζονται σε ένα αυστηρά ορισμένο σύνολο κανόνων που πρέπει να τηρηθούν όπως στις κλασικές μεθόδους αλγορίθμων και προγραμματισμού αλλά στα δεδομένα που τροφοδοτούνται ως είσοδοι και αναγνωρίζοντας μοτίβα και χαρακτηριστικά μαθαίνοντας από αυτά οδηγώντας έτσι στην επίλυση των επιθυμητών στόχων. Η κεντρική ιδέα της μηχανικής μάθησης μπορεί να συγκριθεί με ένα παιδί που εκτίθεται σε διάφορα ερεθίσματα και καταστάσεις και έπειτα μέσω των εμπειριών του μπορεί να πάρει τις δικές του αποφάσεις.

Η μηχανική μάθηση έχει συμβάλει σημαντικά στην σύγχρονη τεχνολογία, τέτοια παραδείγματα βελτίωσης της σημερινής τεχνολογίας αποτελεί η αναγνώριση προσώπων, συστήματα προτάσεων, εντοπισμού αντικειμένων και πιο πρόσφατα την δημιουργία της εφαρμογής chatGPT που αποτελεί πλέον βασικό εργαλείο για όλους τους ανθρώπους. Οι τεχνικές της μηχανικής μάθησης χωρίζονται σε τρεις κατηγορίες:

- Επιβλεπόμενη μάθηση (Supervised Learning)
- Μη επιβλεπόμενη μάθηση (Unsupervised Learning)
- Ενισχυτικής μάθησης (Reinforcement Learning)

Επιβλεπόμενη Μάθηση

Στην κατηγορία της επιβλεπόμενης μάθησης τα δεδομένα που ορίζονται στους αλγόριθμους σαν είσοδο συνοδεύονται από ετικέτες ή διαφορετικά από τα αποτελέσματά τους, για παράδειγμα το όνομα που έχει αποθηκευτεί μία ιατρική εικόνα να είναι η γνωμάτευση της. Αυτά τα δεδομένα είναι γνωστά ως *labeled data*. Σκοπός της συγκεκριμένης κατηγορίας αποτελεί η σωστή και ακριβής προσέγγιση αντιστοίχισης της καινούργιας εισόδου στο μοντέλο με την σωστή πρόβλεψη της εξόδου. Μερικοί αλγόριθμοι μηχανικής μάθησης που ανήκουν στην συγκεκριμένη κατηγορία είναι η κατηγοριοποίηση (*classification*) στην οποία δίνονται οι κατηγορίες στα δεδομένα εισόδου και η παλινδρόμηση (*regression*) στην οποία προσεγγίζεται μία συνεχής τιμή όπως η μελλοντική τιμή ενός αντικειμένου. Αξίζει να σημειωθεί μία υποκατηγορία, η οποία ονομάζεται ημιεπιβλεπόμενη μάθηση η οποία έχει *labeled* δεδομένα στην μειωψηφία της και ακατηγοριοποίητα στην πλειοψηφία της.

Μη Επιβλεπόμενη Μάθηση

Όπως φαίνεται και από το όνομα της κατηγορίας τα δεδομένα εισόδου δεν έχουν καμία ετικέτα ή κατηγοριοποίηση οπότε καλείται να αναγνωρίζει μοτίβα χωρίς προηγούμενη καθοδήγηση ή πληροφορίες. Παραδείγματα αυτής της κατηγορίας αποτελούν η ομαδοποίηση (*clustering*), στην οποία γίνεται ο εντοπισμός εσωτερικών ομοιοτήτων μεταξύ των δεδομένων και εξαιτίας αυτών δημιουργούνται διάφορες ομάδες, και η συσχέτιση (*association*), στην οποία δημιουργούνται κανόνες που συσχετίζουν τα δεδομένα μεταξύ τους.

Ενισχυτική Μάθηση

Στην συγκεκριμένη μάθηση ο αλγόριθμος μαθαίνει αλληλεπιδρώντας με το περιβάλλον του, επιβραβεύεται για κάθε σωστό βήμα που εκτελεί και τιμωρείται για κάθε λανθασμένο. Μέσω αυτής της διαδικασίας, στόχος του προγράμματος είναι να μεγιστοποιήσει την επιβράβευση και να ελαχιστοποιήσει την τιμωρία του χωρίς ανθρώπινη παρέμβαση. Μερικά παραδείγματα της ενισχυτικής μάθησης είναι τα προβλήματα σχεδιασμού πορείας, η βελτιστοποίηση εργασιών και ο έλεγχος ρομπότ.

2.1.2 Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα χωρίζονται σε δύο κατηγορίες, τα βιολογικά νευρωνικά δίκτυα και τα τεχνητά νευρωνικά δίκτυα. Τα βιολογικά νευρωνικά δίκτυα εμπεριέχονται στους εγκεφάλους ζωντανών οργανισμών με νοημοσύνη και κάποιες από τις εργασίες τους, που εκτελούνται εύκολα, είναι η αναγνώριση προτύπων και αντικειμένων. Τα τεχνητά νευρωνικά δίκτυα μιμούνται τα βιολογικά νευρωνικά δίκτυα προσαρμοσμένα στην πραγματικότητα των ηλεκτρονικών υπολογιστών και όχι στον ανθρώπινο εγκέφαλο. Όπως οι νευρώνες στους εγκεφάλους επικοινωνούν μεταξύ τους με ηλεκτρικά σήματα έτσι και οι τεχνητοί νευρώνες επικοινωνούν ανταλλάσσοντας μαθηματικά δεδομένα. Πιο συγκεκριμένα η είσοδος σε κάθε τεχνητό νευρώνα πολλαπλασιάζεται με ένα συγκεκριμένο βάρος το οποίο καθορίζει εάν το

σήμα ενισχύεται ή αποσβάνεται και καθορίζει πόσο σημαντικός είναι για την έξοδο.

2.1.3 Συνελκτικὰ Νευρωνικά Δίκτυα CNN

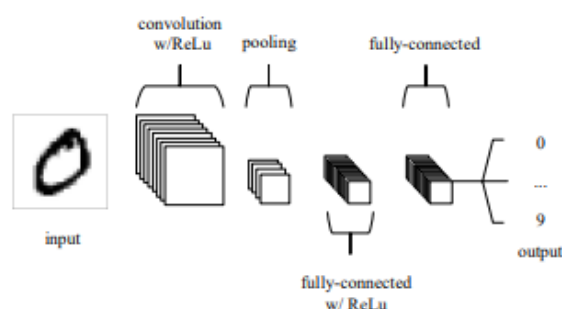
Τα συνελκτικὰ νευρωνικά δίκτυα λαμβάνουν ως είσοδο εικόνες που μέσω από τα βάρη που αναφέρθηκαν πριν εντοπίζουν σε αυτές διάφορα αντικείμενα ή μοτίβα που αφορούν το επιθυμητό αποτέλεσμα. Τα πορίσματα τους προέρχονται από την εξαγωγή χωρικών και χρονικών εξαρτήσεων στις εικόνες με την χρήση διαφόρων φίλτρων, υπεύθυνα για την μείωση των παραμέτρων σε σχέση με προηγούμενες προσεγγίσεις.

Η αρχιτεκτονική τους αποτελείται από τρία νευρωνικά στρώματα, το συνελκτικό στρώμα (convolutional layer), το στρώμα ομαδοποίησης (pooling layer) και το πυκνό πλήρως συνδεδεμένο στρώμα (fully connected layer).

Το συνελκτικό στρώμα συνελκίζει διάφορους πυρήνες στην εικόνα με σκοπό την εξαγωγή χαρακτηριστικών υψηλού επιπέδου όπως οι ακμές και άλλα οπτικά στοιχεία. Για την εξαγωγή βέλτιστων αποτελεσμάτων, τα συγκεκριμένα στρώματα συνεργάζονται μεταξύ τους ώστε το επιθυμητό αποτέλεσμα να είναι η ολική κατανόηση της εικόνας μέσω όλων των χαρακτηριστικών που την ορίζουν όπως οι ακμές, το χρώμα και άλλα οπτικά στοιχεία.

Το στρώμα ομαδοποίησης λαμβάνει το αποτέλεσμα από τα συνελκτικά στρώματα και εκτελεί δειγματοληψία ώστε να μειωθεί το χωρικό μέγεθος του και συνεπώς οι υπολογιστικές απαιτήσεις του δικτύου. Επιπλέον δίνει τα χαρακτηριστικά που παραμένουν αναλλοίωτα είτε λόγω περιστροφής είτε λόγω θέσης. Αξίζει να σημειωθεί ότι υπάρχουν δύο κατηγορίες ομαδοποίησης, average pooling που επιστρέφει τη μέση τιμή που καλύπτει ο πυρήνας πετυχαίνοντας έτσι μείωση των διαστάσεων της εικόνας και max pooling που επιστρέφει την μέγιστη τιμή που εκτός το ότι μειώνει και αυτό τις διαστάσεις, πραγματοποιεί και αποθροβοποίηση. Η ποιότητα εκπαίδευσης του μοντέλου διατηρείται και στις δύο περιπτώσεις.

Το πλήρως συνδεδεμένο στρώμα μαθαίνει μη γραμμικούς συνδυασμούς για τα χαρακτηριστικά υψηλού επιπέδου που λαμβάνονται από το τελικό συνελκτικό στρώμα. Κάθε νευρώνας του συνδέεται με όλους τους νευρώνες του προηγούμενου επιπέδου.

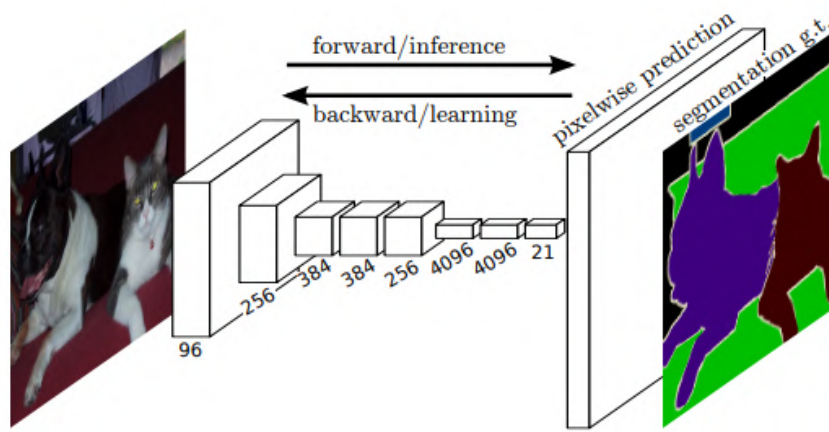


Σχήμα 2.1: Παράδειγμα Συνελκτικού Νευρωνικού Δικτύου.

2.1.4 Εξ Ολοκλήρου Συνελκτικὰ Δίκτυα (FCN)

Στα εξ ολοκλήρου συνελκτικὰ δίκτυα η αρχιτεκτονική αλλάζει καθώς το τελευταίο πυκνό στρώμα από τα απλά συνελκτικὰ δίκτυα αντικαθιστάται από ένα πλήρως συνελκτικό στρώμα. Η διαφορά στην έξοδο τους είναι ότι παράγουν χωρικούς χάρτες τμηματοποίησης

σε όλη την εικόνα και όχι μόνο σε τμήματα της. Για την παραγωγή αυτών των χαρτών υπερδειγματοληπτούν τους πίνακες χαρακτηριστικών από το τελευταίο στρώμα (last layer) και τους συνδυάζουν με τους πίνακες χαρακτηριστικών από τα προηγούμενα στρώματα. Με την δημιουργία αυτών των αναλυτικών χαρτών τμηματοποίησης δημιουργούνται κάποια νέα προβλήματα όπως η μείωση της ανάλυσης των προβλέψεων δηλαδή μη ακριβή περιγράμματα αντικειμένων και η αύξηση του χρόνου εκπαίδευσης. Παρακάτω δίνεται ένα παράδειγμα εξ ολοκλήρου συνεκτικού δικτύου.



Σχήμα 2.2: Παράδειγμα εξ ολοκλήρου συνεκτικού νευρωνικού Δικτύου.[1]

2.2 Δίκτυα Κωδικοποιητή-Αποκωδικοποιητή Encoder-Decoder

Τα δίκτυα Κωδικοποιητή-Αποκωδικοποιητή Encoder-Decoder χρησιμοποιούνται σε πληθώρα εφαρμογών όπως η κατηγοριοποίηση συναισθημάτων που εκφράζεται μέσω σχόλια χρηστών που έχουν δημοσιεύσει στα κοινωνικά δίκτυα ή από κριτικές προϊόντων. Η ανάπτυξη των chat bots απαντώντας στους χρήστες για τα ερωτήματα τους με φυσική γλώσσα, η παραγωγή λεζάντας σε εικόνες (image captioning) ακόμα και η γενετική επεξεργασία καθώς χρησιμοποιούνται στην βιοπληροφορική για την ανάλυση και σύνθεση ακολουθιών DNA, αποτελούν μερικά από τα πολλά παραδείγματα χρήσης δικτύων Encoder-Decoder. Ένα δίκτυο Κωδικοποιητή-Αποκωδικοποιητή χωρίζεται σε δύο κύρια μέρη, τον κωδικοποιητή και αποκωδικοποιητή. Η σύνδεση μεταξύ τους γίνεται μέσω ενός διανύσματος συμπραζομένων (context vector).

Η δουλειά του κωδικοποιητή είναι η επεξεργασία κάθε στοιχείου της ακολουθίας εισόδου και η συγκέντρωση όλων των πληροφοριών σε ένα διάνυσμα σταθερού μήκους. Το διάνυσμα αυτό περιέχει όλη την πληροφορία της ακολουθίας εισόδου και καθιστάται απαραίτητο για τον αποκωδικοποιητή, ο οποίος μέσω αυτού παράγει την ακολουθία εξόδου στοιχείο προς στοιχείο. Η αρχιτεκτονική του Encoder-Decoder έχει διάφορες μορφές αναλόγως για ποια λειτουργία και εφαρμογή χρειάζεται.

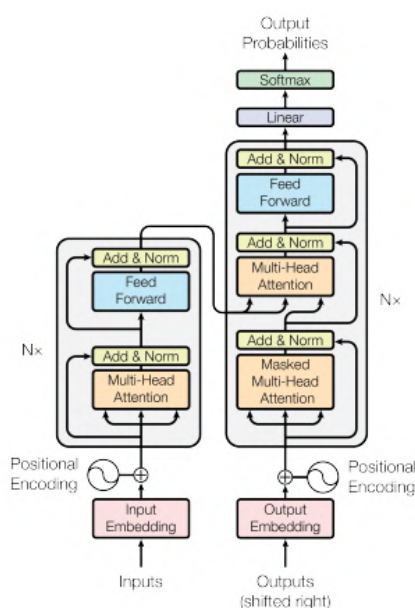
Πιο αναλυτικά όσον αφορά την σημασιολογική τμηματοποίηση, ο κωδικοποιητής χρησιμοποιεί Συνελεκτικά Νευρωνικά Δίκτυα CNNs για την εξαγωγή των χαρακτηριστικών της εικόνας-εισόδου. Μειώνει την ανάλυση της εικόνας μέσω στρωμάτων συνελιξεων και μέσω pooling έχοντας έτσι ως αποτέλεσμα χαρακτηριστικά που περιγράφουν την εικόνα-είσοδο

σε διάφορα επίπεδα αφαίρεσης. Αυτά τα χαρακτηριστικά, που έχουν χαμηλή ανάλυση αλλά περισσότερα κανάλια, δημιουργούν το διάνυσμα συμφραζόμενων (context vector). Ο αποκωδικοποιητής χρησιμοποιώντας τεχνικές όπως upsampling ή deconvolution φτιάχνει την εικόνα εξόδου να έχει την ίδια ανάλυση με την εικόνα εισόδου.

2.3 Μετασχηματιστές Transformers

Οι μετασχηματιστές έχουν παρόμοια λογική με την τεχνική Encoder-Decoder. Η αρχική τους εφαρμογή ήταν στην επεξεργασία φυσικής γλώσσας (NLP) και έπειτα επεκτάθηκαν σε άλλους τομείς όπως επεξεργασία εικόνων (image processing) και αναγνώριση ομιλίας (speech recognition). Η αρχιτεκτονική τους εισάχθηκε από τον Vaswani [2], με σκοπό να χειρίζεται δεδομένα ακολουθιών χωρίς τους περιορισμούς που θέτουν τα αναδρομικά νευρωνικά δίκτυα (RNNs) και τα συνελκτικά νευρωνικά δίκτυα (CNNs). Η διαφορά αυτής της αρχιτεκτονικής σε σχέση με της Encoder-Decoder είναι ότι ενώ βασίζεται σε ακολουθία προς ακολουθία (sequence-to-sequence) που είναι παρόμοια με τα μοντέλα Encoder-Decoder, δεν χρησιμοποιεί συνελίξεις αλλά μηχανισμούς αυτοπροσοχής (self-attention). Ο μηχανισμός αυτός αναπαριστά κάθε token που περιγράφει την ομοιότητα του με τα άλλα tokens. Η θέση των tokens δεν έχει σημασία καθώς τα αποτελέσματα του είναι αναλοιώτα ακόμα και εάν ανακατευτούν.

Για την διαδικασία της τμηματοποίησης εικόνας οι μετασχηματιστές είναι πιο απαιτητικοί καθώς πρέπει να γίνει σωστή επεξεργασία των δεδομένων αφού οι χωρικές πληροφορίες και οι περιοχές ενδιαφέροντος πρέπει να περαστούν σε κάθε token πριν το στάδιο του μηχανισμού προσοχής. Παρακάτω φαίνεται η συνολική αρχιτεκτονική των Transformers.



Σχήμα 2.3: Παράδειγμα αρχιτεκτονικής μετασχηματιστή [2].

2.4 Τμηματοποίηση ιατρικής εικόνας

Η τμηματοποίηση ιατρικών εικόνων αποτελεί ζωτικής σημασίας έργο για την επιτάχυνση και βελτίωση της διάγνωσης ασθενειών, καθώς και για την εξασφάλιση πιο στοχευμένης και αποτελεσματικής θεραπείας για ταυτόχρονη εξυπηρέτηση μεγάλο αριθμό ασθενών. Η διαδικασία αυτή εξαλείφει τα λάθη ανθρώπινου παράγοντα και μειώνει το χρόνο της διάγνωσης. Ανήκει στο πεδίο της σημασιολογικής τμηματοποίησης και στοχεύει σε ιατρικά δεδομένα, όπως δισδιάστατες και τρισδιάστες εικόνες, αξονικές, μαγνητικές υπέρηχους και ακτινογραφίες. Τα αποτελέσματα της βοηθούν στον εντοπισμό φλεγμονών ή άλλων περιοχών ενδιαφέροντος όπως οι πολύποδες στο έντερο ή οι φλεγμονες στους πνεύμονες και άλλα τέτοια παραδείγματα.

Η διαδικασία της τμηματοποίησης μέχρι και τα τελευταία χρόνια υλοποιούταν χειροκίνητα. Η διαδικασία αυτή, όπως είναι φυσικό, είναι χρονοβόρα, κουραστική και πρέπει να πραγματοποιηθεί από εξειδικευμένους επαγγελματίες που σχολιάζουν(annotation) την περιοχή ενδιαφέροντος ακολουθώντας πρωτόκολλα και έρχοντας σε επαφή με άλλους γιατρούς για να επαληθευτούν το αποτέλεσμα. Ο περιορισμένος αριθμός δεδομένων καθώς και οι διαφορετικές διαγνώσεις για τα ήδη υπάρχοντα δεδομένα καθιστά την χειροκίνητη τμηματοποίηση ακόμα πιο δύσκολη. Η ανάγκη λοιπόν για ανάπτυξη αλγορίθμων όρασης υπολογιστών είναι μεγάλη καθώς με την ακρίβεια και την ταχύτητα, χωρίς την ανάγκη ανθρώπινης παρέμβασης, αντιμετωπίζονται οι δυσκολίες που αναφέρθηκαν. Παράδειγμα τέτοιων υλοποιήσεων αποτελούν οι εργασίες [16] [17] όπου παρουσιάζουν τις περιοχές ενδιαφέροντος σε ιατρικές απεικονίσεις ακόμα και την αναγνώριση ασθενειών από ακτινογραφίες χωρίς ανθρώπινη παρέμβαση.

2.5 Επεξηγησιμότητα στην Τεχνητή Νοημοσύνη

Η ραγδαία ανάπτυξη και η πολυπλοκότητα των νευρωνικών δικτύων έχουν οδηγήσει στην ανάγκη της επεξηγησιμότητας (Explainable AI). Τα συνελκτικικά δίκτυα που αναφέρθηκαν πριν θεωρούνται "μαύρα κουτία (black boxes)" αφού είναι δύσκολο για τους χρήστες να κατανοήσουν την μεθοδολογία και να εμπιστευτούν τα πορίσματα τους, πόσο μάλλον όταν αφορούν κλάδους όπως η ιατρική. Την εμπιστοσύνη αυτή έρχεται να χτίσει η επεξηγησιμότητα καθώς εξηγεί πως φτάσαν σε αυτές τις απαντήσεις και γιατί αυτές οι απαντήσεις είναι σωστές. Βασική έννοια explainable AI (XAI) αποτελεί η ερμηνευσιμότητα (Interpretability) δηλαδή η κατανόηση λειτουργίας του συστήματος. Μία άλλη έννοια είναι η επεξηγησιμότητα (Explainability) δηλαδή η παροχή εξηγήσεων για τις αποφάσεις του συστήματος. Η διαφάνεια (Transparency) αποτελεί επίσης σημαντική έννοια καθώς δίνει την δυνατότητα πρόσβασης στα δεδομένα και τις λειτουργίες του μοντέλου. Τέλος έχουμε την Δικαιολογησιμότητα (Justifiability) και την Αμφισβητησιμότητα (Contestability) δηλαδή, την κατανόηση του λόγου πίσω από ένα αποτέλεσμα και την δυνατότητα παροχής στον χρήστη να αμφισβητήσει τις αποφάσεις του μοντέλου αντίστοιχα.

Οι τεχνικές αυτές χωρίζονται σε model-based για απλούστερα μοντέλα και post-hoc για πιο περίπλοκα μοντέλα. Μπορούν να είναι επίσης model-specific ή model-agnostic σχεδιασμένες δηλαδή για συγκεκριμένου ή ανεξάρτητου τύπου μοντέλου αντίστοιχα.

2.6 Συναφείς εργασίες

Το θεωρητικό υπόβαθρο και η κατανόηση των δυνατοτήτων των νευρωνικών μοντέλων σε διάφορους καθημερινούς τομείς, προήλθε από την μελέτη των εργασιών διακεκριμένων συναδέλφων. Αναλύοντας και κατανοώντας μία σειρά από δημοσιεύσεις, από αυτούς τους συναδέλφους, εμπνευστήκαμε για την σημασιολογική τμηματοποίηση και την επεξηγησιμότητα.

Αρχιτεκτονικές βαθιών νευρωνικών δικτύων έχουν υλοποιηθεί και χρησιμοποιηθεί σε διάφορες εφαρμογές από μέλη του Εργαστηρίου Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης του ΕΜΠ. Ειδικότερα επιβλεπόμενες τεχνικές CNN και CNN-RNN έχουν εφαρμοστεί για κατηγοριοποίηση αντικειμένων, στην ιατρική διάγνωση νευροεκφυλιστικών ασθενειών, όπως της νόσου του Πάρκινσον [18], [19], [20], [21], [22] ή της Covid-19 [23], [24], [25], [26], [27], περιλαμβάνοντας κατάτμηση 2-Δ ή 3-Δ εικόνων. Έμφαση έχει δοθεί στην διαφάνεια και στην προσαρμογή των μοντέλων [28], [29], [30], [31] αλλά και στην ανάπτυξη πλέον σύνθετων αρχιτεκτονικών, μπαϋεσιανών, με κάψουλες και αβεβαιότητα [32], [33], [34]. Βαθιές ημι- και αυτο- επιβλεπόμενες 3-Δ νευρωνικές αρχιτεκτονικές, αλλά και αρχιτεκτονικές κωδικοποιητή- αποκωδικοποιητή έχουν εφαρμοστεί στην ανίχνευση βλαβών σε πυρηνικούς αντιδραστήρες [35], [36], στην πρόβλεψη της παραγωγής στον αγροτικό τομέα [37], [38] και στην αναγνώριση και σύνθεση συναισθήματος [39], [40], [41], ενώ άλλες εφαρμόζονται σε προβλήματα αλληλεπίδρασης ανθρώπου- υπολογιστή και ανάλυσης εικόνων [42], [43], [44], [45], [46], [47], [48].

Κεφάλαιο **3**

Θεωρητική ανάλυση επιλεγμένων τεχνικών επεξηγησιμότητας και μοντέλων τμηματοποίησης

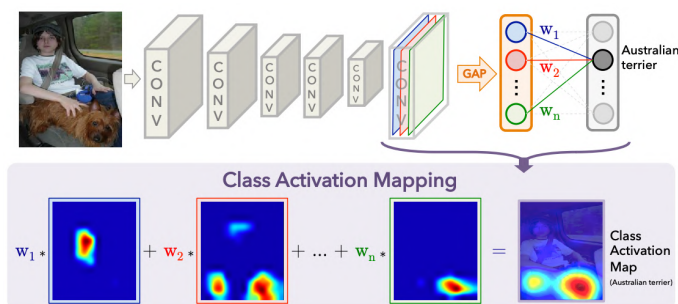
Στο κεφάλαιο αυτό δίνεται το θεωρητικό υπόβαθρο για τις τεχνικές που χρησιμοποιήθηκαν για την τμηματοποίηση των ιατρικών εικόνων καθώς και για τις μεθόδους επεξηγησιμότητας που υλοποιήθηκαν.

3.1 Τεχνικές επεξηγησιμότητας

Στην παρούσα διπλωματική εργασία οι μέθοδοι που επιλέχθηκαν αποτελούν μέρος της οικογένειας Gradient-weighted Class Activation Mapping (Grad-CAM) μεθόδων, οι οποίες δείχνουν που έδωσαν βάση στις εικόνες είσοδοι τα μοντέλα τμηματοποίησης ώστε να βγάλουν τα αποτελέσματα τους. Για την λειτουργία αυτών των μεθόδων είναι απαραίτητη η ύπαρξη κάποιου convolutional layer και συνήθως επιλέγεται το τελευταίο layer καθώς στην πληθώρα των περιπτώσεων είναι πιο αποτελεσματικό. Παρακάτω δίνεται ο αναλυτικός τρόπος λειτουργίας της τεχνικής Grad-CAM καθώς και οι διαφοροποιήσεις των άλλων τεχνικών σε σχέση με αυτήν.

3.1.1 Τεχνική Grad-CAM

Μία τεχνική επεξηγησιμότητας είναι η τεχνική Grad-CAM [4]. Η συγκεκριμένη τεχνική παράγει οπτικές εξηγήσεις για τα αποτελέσματα που προκύπτουν από τα μοντέλα συνελκτικών νευρωνικών δικτύων CNN βελτιώνοντας έτσι τη διαφάνεια τους. Δημιουργούνται χάρτες εντοπισμού που δείχνουν ποιες είναι οι σημαντικές περιοχές στην εικόνα για την πρόβλεψη του επιθυμητού στόχου, αξιοποιώντας την πληροφορία των feature maps [49] του τελευταίου επίπεδου συνέλιξης. Η επιλογή του τελευταίου επιπέδου δεν είναι απαραίτητη καθώς μπορεί να επιλεγεί κάποιο διαφορετικό αλλά τις περισσότερες φορές είναι πιο αποδοτικό. Στην αρχιτεκτονική της CAM, το πρώτο βήμα είναι η εφαρμογή Global Average Pooling στο τέλος του δικτύου εκεί όπου έχουμε για κάθε κλάση ένα feature map. Έπειτα παίρνοντας το spatial average του κάθε feature map από το προηγούμενο βήμα, δημιουργείται ένα διάνυσμα Vector το οποίο γίνεται weighted. Τέλος αυτό εισέρχεται από μία softmax δίνοντας μας το τελικό αποτέλεσμα του δικτύου.



Σχήμα 3.1: Παράδειγμα λειτουργίας της μεθόδου CAM [3].

Η τεχνική Grad-CAM ακολουθεί την ίδια λογική αξιοποιώντας τα feature maps. Έχοντας τις δύο παρακάτω μαθηματικές σχέσεις:

$$a_k^c = \underbrace{\frac{1}{Z} \sum_i \sum_j}_{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A_{ij}^k}}_{\text{gradients via backprop}} \quad (3.1)$$

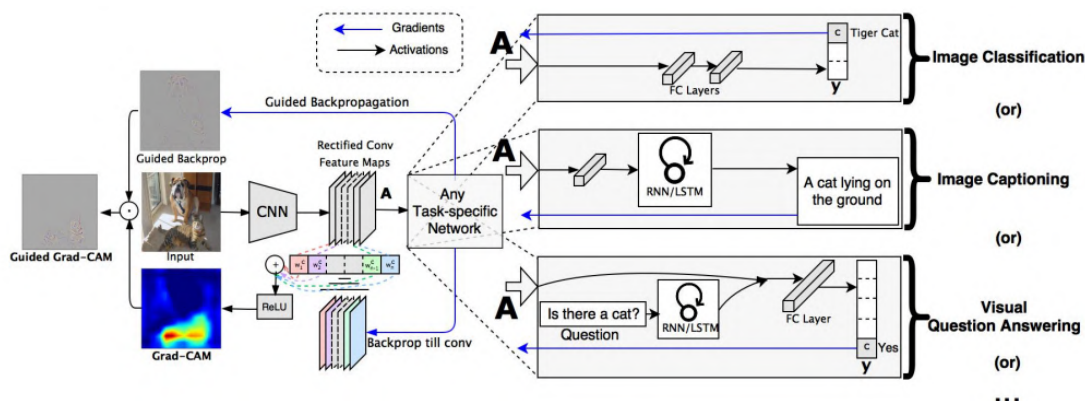
και

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\underbrace{\sum_k a_k^c A^k}_{\text{linear combination}} \right) \quad (3.2)$$

Θα εξηγηθεί η λειτουργία του. Το i, j συμβολίζει το ύψος και το πλάτος της εικόνας εισόδου αντίστοιχα. Το k συμβολίζει τον αριθμό των feature maps A^k και y^c το score κάθε κλάσης πριν την μεταβλητή softmax. Τέλος η σχέση στην εξίσωση 3.2 δίνει την σχέση του συνολικού localization map, δηλαδή το άθροισμα των γραμμικών συνδυασμών των feature maps με τα συγκεκριμένα βάρη και score που αναφέρθηκαν παραπάνω παίρνοντας το από μία συνάρτηση ενεργοποίησης ReLU.

Τα πλεονεκτήματα της τεχνικής Grad-CAM είναι ότι δεν χρειάζεται επανεκπαίδευση ή κάποιου είδους προσαρμογή για την εφαρμογή της. Εάν οι τεχνικές τμηματοποίησης έχουν feature maps (που χρειάζονται για την υλοποίηση των συγκεκριμένων τεχνικών), δηλαδή εάν υπάρχει έστω και ένα συνελκτικό επίπεδο, μπορούν να εφαρμοστούν.

Παρακάτω φαίνεται ένα παράδειγμα ταξινόμησης εικόνων που χρησιμοποιείται ένδιάμεσα η τεχνική Grad-CAM ώστε οι απαντήσεις που θα ληφθούν από τον χρήστη να γίνουν πιο κατανοητές.



Σχήμα 3.2: Παράδειγμα λειτουργίας της μεθόδου Grad-CAM.[4]

Η συγκεκριμένη τεχνική αξιοποιεί τα feature maps με μαθηματικό τρόπο για την διαφάνεια του μοντέλου. Ακολουθώντας την ίδια λογική υπάρχουν και οι παρακάτω τεχνικές που λειτουργούν με παρόμοιο τρόπο.

3.1.2 Τεχνική LayerCAM

Η συγκεκριμένη τεχνική βελτιώνει τα αποτελέσματα όσον αφορά τα χαμηλότερα επίπεδα του μοντέλου καθώς ξεκινάει από το τελευταίο και πηγαίνει προς τα πίσω. Η διαδικασία που υλοποιεί είναι να παίρνει τα feature maps για την στάθμιση των activation maps του προηγούμενου επιπέδου του, τα οποία activation maps αθροίζονται και το άθροισμα αυτό περνάει μέσα από μία συνάρτηση ενεργοποίησης ReLU, δημιουργώντας έτσι το νέο βελτιωμένο συνολικό activation map του προηγούμενου επιπέδου.

3.1.3 Τεχνική EigenCAM

Η τεχνική EigenCAM διαφέρει ως προς την τεχνική GradCAM προσθέτοντας ένα global weight map στο τελευταίο στάδιο του GradCAM, δηλαδή στην τοπική εξήγηση του τελευταίου συνελκτικού δικτύου, παίρνοντας έτσι ένα καλύτερο αποτέλεσμα. Από την PCA ανάλυση στα activation των feature maps του τελευταίου συνελκτικού δικτύου παίρνουμε το first principle component το οποίο χρησιμοποιώντας το παίρνουμε τον global weight map που διαφοροποιεί την τεχνική EigenCAM από την GradCAM. Πιο συγκεκριμένα η EigenCAM για να υπολογίσει τα βάρη του γραμμικού συνδυασμού του feature maps χρησιμοποιεί το γινόμενο των βαρών του τελευταίου, πλήρους συνδεδεμένου επιπέδου(dense) σε συνδυασμό με τα αποτελέσματα του Global Average Pooling(GAP), ενώ η τεχνική GradCAM χρησιμοποιεί τις παραγώγους για την συνεισφορά κάθε feature map.

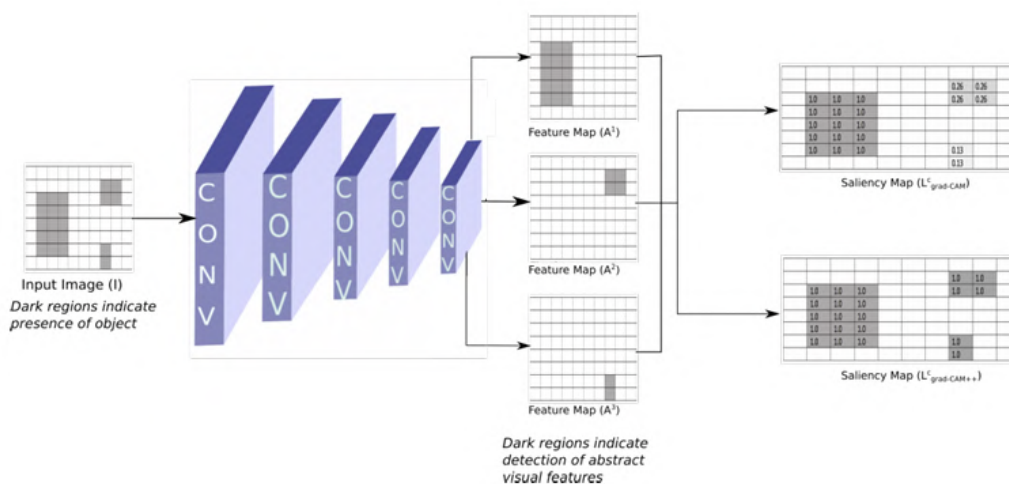
3.1.4 Τεχνική EigenGradCAM

Η τεχνική EigenGradCAM αποτελεί μία εξέλιξη της EigenCAM χρησιμοποιώντας τις λεπτομέρειες που παρέχονται από την Guided Backpropagation. Αναλυτικότερα, χρησιμοποιείται η Guided Backpropagation παίρνοντας μια πιο λεπτομερή οπτικοποίηση των feature maps κρατώντας μόνο τις θετικές τιμές των παραγώγων, καθώς αυτές είναι που αντιπροσωπεύουν τις ενεργοποιήσεις που συμβάλλουν θετικά στην πρόβλεψη του μοντέλου,

και έπειτα αυτή η λεπτομέρης οπτικοποίηση συνδυάζεται με τον global weight map από την τεχνική EigenCAM.

3.1.5 Τεχνική Grad-CAM++

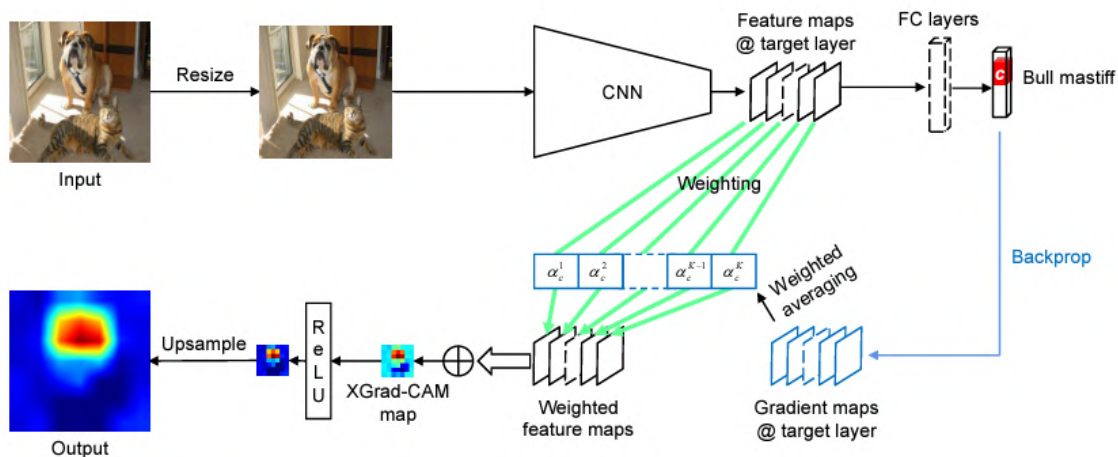
Η τεχνική Grad-CAM++ [5] αποτελεί μια τεχνική επεξηγησιμότητας όπου η διαφορά της με την Grad-CAM βρίσκεται στο ότι χρησιμοποιεί τις δεύτερες παραγώγους σε σχέση με την Grad-CAM όπου χρησιμοποιεί τις πρώτες. Παρακάτω φαίνεται ένα παράδειγμα της δημοσίευσης του Grad-CAM++. Οι αριθμοί στο παράδειγμα αναφέρονται στο πόσο έντονα φαίνονται τα συγκεκριμένα σημεία.



Σχήμα 3.3: Παράδειγμα της μεθόδου Grad-CAM++ για ταξινόμηση αντικειμένου [5].

3.1.6 Τεχνική XGrad-CAM

Η συγκεκριμένη τεχνική διαφοροποιείται και αυτή ως προς τις παραγώγους σε σχέση με την Grad-CAM. Πιο συγκεκριμένα οι παράγωγοι που χρησιμοποιεί είναι scaled, γεγονός που προέκυψε μετά το normalization των activation maps και με βάση αυτά τα αποτελέσματα έγιναν scaled. Παρακάτω δίνεται η συνολική διαδικασία της τεχνικής XGrad-CAM [6].



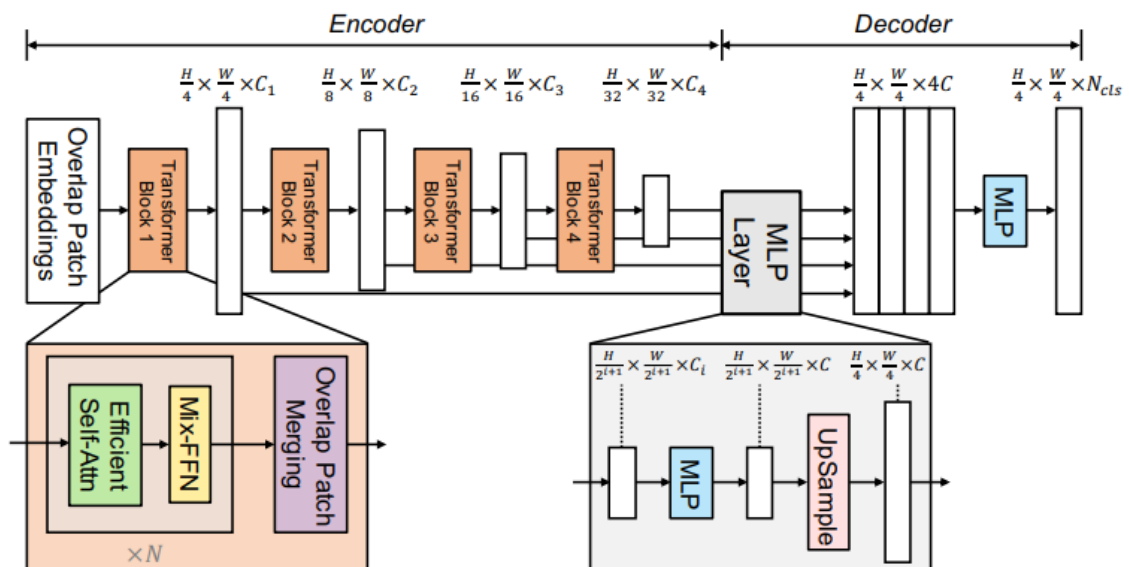
Σχήμα 3.4: Παράδειγμα της μεθόδου XGrad-CAM [6].

3.2 Μοντέλα τμηματοποίησης

Τα μοντέλα τμηματοποίησης που χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία αναλύονται παρακάτω. Η υλοποίηση αυτών των μοντέλων βασίστηκε στο MMSegmentation [50] από την ομάδα OpenMMLab και ο κώδικας για τα μοντέλα που πραγματοποιούν σημασιολογική τμηματοποίηση παρέχεται δωρεάν.

3.2.1 SEGFORMER

Η αρχιτεκτονική του SEGFORMER [7] έχει ως στόχο την εξαγωγή πολυκλιμακών χαρακτηριστικών για την πρόβλεψη μάσκας σημασιολογικής τμηματοποίησης. Αποτελείται από δύο μέρη, το πρώτο είναι ένας ιεραρχικός κωδικοποιητής Transformer και το δεύτερο ένας ελαφρής αποκωδικοποιητής AII-MLP. Η αρχιτεκτονική του φαίνεται στο σχήμα παρακάτω.



Σχήμα 3.5: Η αρχιτεκτονική του SegFormer αποτελείται από τον κωδικοποιητή και αποκωδικοποιητή [7].

Εμβαθύνοντας στην αρχιτεκτονική του, ο κωδικοποιητής ξεκινά με την εικόνα εισόδου όπου διαιρείται σε επικαλυπτόμενα patches. Κάθε patch επεξεργάζεται από ένα συνελικτικό στρώμα, εξαγοντας έτσι patch embeddings, διατηρώντας την τοπική συνέχεια σε αυτά. Έπειτα εφαρμόζεται ο efficient self-attention μηχανισμός χρησιμοποιώντας multi-head self-attention layers. Η εφαρμογή αυτή μειώνει την υπολογιστική πολυπλοκότητα καθιστώντας δυνατή και την χρήση μεγαλύτερων εικόνων με υψηλότερη ανάλυση. Ο κωδικοποιητής περιέχει και τα Mix Feed-Forward Network αντικαθιστώντας έτσι τα 3x3 convolution layers, επιτρέποντας έτσι στο μοντέλο εικόνες εισόδου διαφορετικών αναλύσεων. Όπως φαίνεται και στο σχήμα 3.5 ο Encoder έχει πολλά στάδια stages που το καθένα παράγει πολυκλιμακωτά χαρακτηριστικά. Τα lower stages παράγουν υψηλής ανάλυσης χαρακτηριστικά ενώ τα higher stages χαμηλής ανάλυσης.

Ο αποκωδικοποιητής AII-MLP, το δεύτερο δηλαδή μέρος της αρχιτεκτονικής του SEGFORMER, συνδυάζει αποτελεσματικά χαρακτηριστικά από διάφορα επίπεδα. Αυτό συμβαίνει

καθώς παίρνοντας μέσα από τα MLP layers τα πολυεπίπεδα χαρακτηριστικά από τον Encoder, ενοποιεί έτσι τις διαστάσεις των καναλιών. Στα πολυποίκιλα χαρακτηριστικά που διαμορφώνονται από αυτή την ενοποίηση, εφαρμόζεται η τεχνική upsampling έτσι ώστε όλα τα feature maps να έχουν τις ίδιες διαστάσεις. Αυτά τα συγκεντρωτικά χαρακτηριστικά παίρνουν μέσα από ένα τελικό MLP layer που συνθέτει όλες τις πληροφορίες σε local και global επίπεδο από όλα τα διαφορετικά επίπεδα. Τέλος οι πληροφορίες αυτές παίρνονται από ένα MLP layer παίρνοντας έτσι τις τελικές μας μάσκες για την σημασιολογική τμηματοποίηση. Οι μάσκες που παράχθηκαν έχουν υψηλή ανάλυση στοχεύοντας ακόμα και σε περιπλοκούς στόχους. Μέσω αυτής της αρχιτεκτονικής μπορούν να αντιμετωπιστούν προβλήματα υπολογιστικής πολυπλοκότητας στην είσοδο είτε λόγω της ανάλυσης τους είτε λόγω των παραλλαγών που μπορεί να δεχτεί η εικόνα εισόδου.

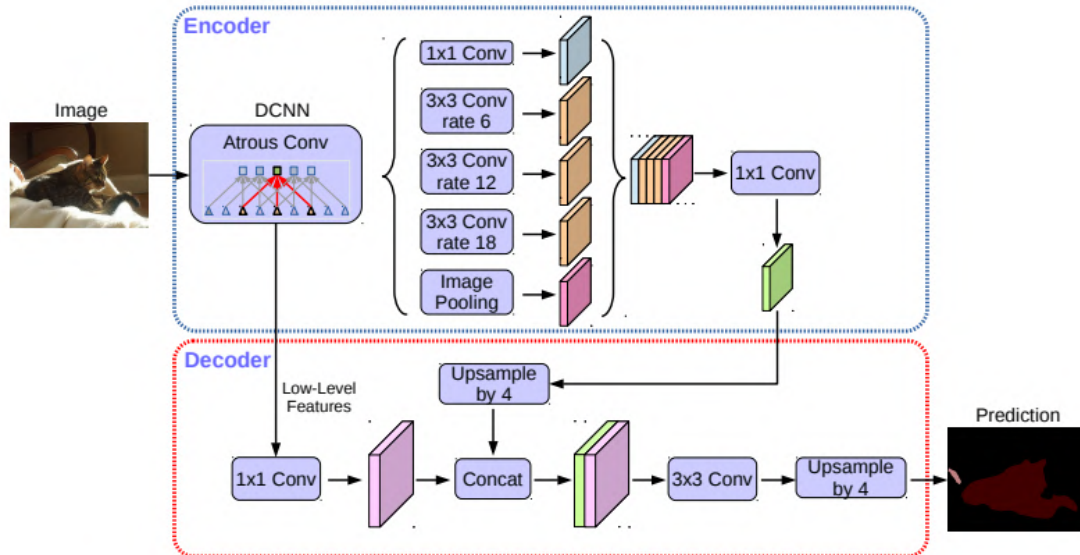
3.2.2 DeepLabV3+

Η αρχιτεκτονική του DeepLabV3+ [8] αποτελεί μία ένωση της αρχιτεκτονικής Encoder-Decoder και της spatial pyramid pooling με σκοπό την βελτίωση των αποτελεσμάτων για την εφαρμογή της σημασιολογικής τμηματοποίησης.

Ο κωδικοποιητής του DeepLabV3+ είναι βασισμένος στο DeepLabV3 [9] και αποτελεί μία εξέλιξη του καθώς περιέχει έναν επιπλέον αποκωδικοποιητή ο οποίος δίνει καλύτερα αποτελέσματα τμηματοποίησης ειδικά κοντά στα περιγράμματα των επιθυμητών αντικειμένων. Το DeepLabV3 χρησιμοποιεί τραχείς συνελίξεις atrous(dilated) convolutions για την εξαγωγή πυκνών χαρακτηριστικών (dense feature maps) ανάλογα και μισούς υπολογιστικούς πόρους που διαθέτει το σύστημα. Οι atrous convolution είναι αυτές που ρυθμίζουν το field-of-view του φίλτρου για την καταγραφή πολλαπλών κλιμάκων πληροφοριών χωρίς να μειωθεί η χωρική ανάλυση. Ο κωδικοποιητής περιέχει επίσης ένα Atrous Spatial Pyramid Pooling(ASPP) module, το οποίο εφαρμόζει παράλληλες atrous συνελίξεις για διαφορετικούς atrous δείκτες. Η προσφορά του έγκεται στην πληθώρα σημασιολογικών πληροφοριών για διαφορετικούς δείκτες που δίνεται από τον κωδικοποιητή. Η καινοτομία του DeepLabV3+ είναι μείωση του υπολογιστικού κόστους και τον αριθμό των παραμέτρων μέσω της χρήσης depthwise separable convolution. Η depthwise convolution εφαρμόζει σε κάθε κανάλι της εισόδου ένα συνελκτικό φίλτρο. Πιο συγκεκριμένα εάν υπάρχουν X αριθμός καναλιών εισόδου θα υπάρχουν και X ξεχωριστές συνελίξεις. Η pointwise αποτελεί μία 1×1 συνέλιξη που συνδυάζει τα αποτελέσματα των depthwise convolutions κατά μήκος των διαφόρων καναλιών ανακατεύοντας τις πληροφορίες μεταξύ ξεχωριστών καναλιών αποτελεσματικά. Ενώνοντας αυτά τα δύο βήματα πετυχαίνουμε το depthwise separable convolution.

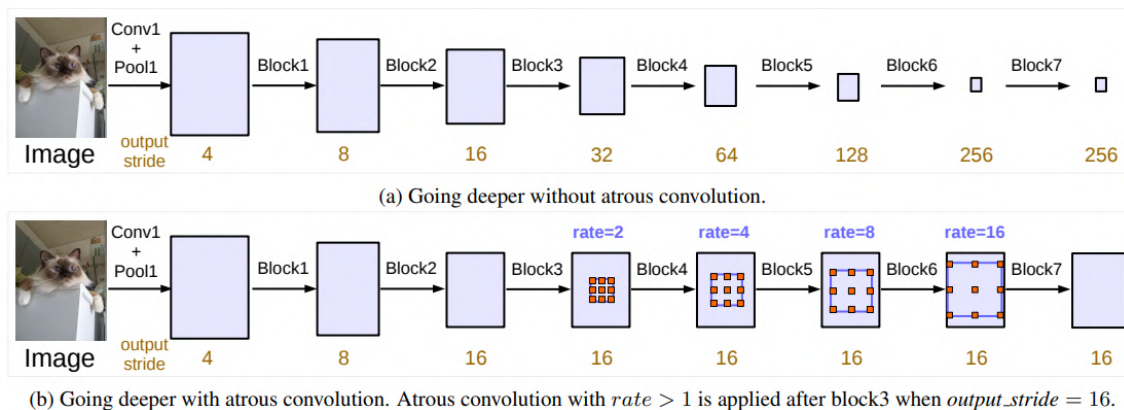
Στον αποκωδικοποιητή όπως αναφέρθηκε τα αποτελέσματα του είναι η καλύτερη τμηματοποίηση ειδικά κοντά στα περιγράμματα των επιθυμητών αντικειμένων. Η διαδικασία του Decoder ξεκινάει εφαρμόζοντας bilinear upsampling στα χαρακτηριστικά του Encoder με συντελεστή 4. Αυτά τα upsampling features συγκεντρώνονται μαζί με τα χαμηλού επιπέδου χαρακτηριστικά από το δίκτυο backbone, που στην δικιά μας περίπτωση αυτό το δίκτυο είναι το ResNet50, που έχουν την ίδια spatial ανάλυση. Για να μην υπάρχει επικάλυψη των χαμηλών επιπέδων χαρακτηριστικών με τα χαρακτηριστικά του Encoder χρησιμοποιείται συνέλιξη 1×1 για την μείωση του αριθμού των καναλιών. Ακολουθούν έπειτα κάποιες 3×3

συνελίξεις που σε συνδυασμό με την τεχνική upsampling δίνεται η σωστή τμηματοποίηση με τις λεπτομέρειες να φαίνονται στο περίγραμμα των επιθυμητών στόχων. Ένα παράδειγμα υλοποίησης του DeepLabV3+ φαίνεται παρακάτω.



Σχήμα 3.6: Παράδειγμα υλοποίησης του μοντέλου τμηματοποίησης DeepLabv3+ [8].

Ενώ η σημασία των atrous convolutions γίνεται πιο κατανοητή στο παρακάτω σχήμα.



Σχήμα 3.7: Διαδοχικές μονάδες χωρίς και με atrous convolution [9].

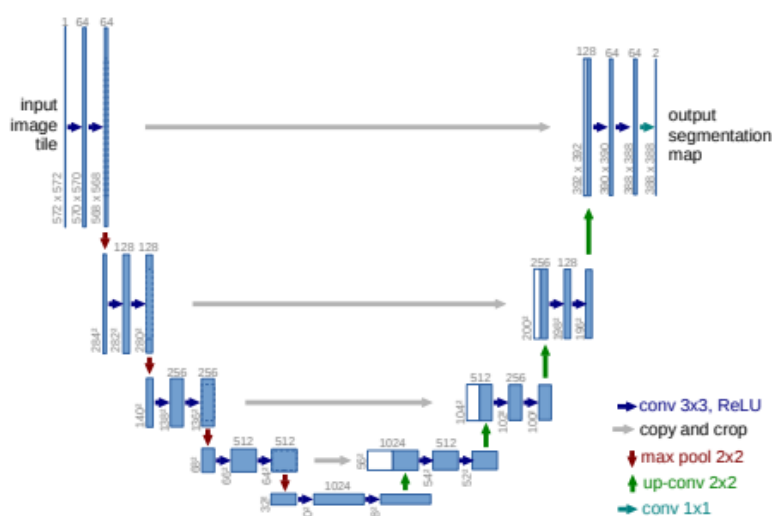
Συνοψίζοντας το DeepLabV3+ αποτελεί ένα σπουδαίο εργαλείο για την σημασιολογική τμηματοποίηση εικόνας συνδυάζοντας την αρχιτεκτονική του Encoder-Decoder και της spatial pyramid pooling. Μέσω την χρήση atrous(dilated) convolutions και depthwise convolution παίρνουμε πληθώρα σημαντικών πληροφοριών για την υλοποίηση της τμηματοποίησης.

3.2.3 UNet

Η τεχνική Unet [10] αποτελεί άλλο ένα βασικό εργαλείο για την τμηματοποίηση ιατρικών εικόνων. Αποτελεί ένα από τα πιο βασικά μοντέλα τμηματοποίησης, ικανό να αντιμετωπίζει το πρόβλημα του μικρού αριθμού δεδομένων αξιοποιώντας τα πιο αποδοτικά.

Η αρχιτεκτονική του βασίζεται στα μοντέλα τύπου Encoder-Decoder. Αποτελείται από μια συστατική διαδρομή (contracting path) που εντοπίζεται στην αριστερή πλευρά του σχήματος και μία συμμετρική επεκτεινόμενη διαδρομή (expansive path) που βρίσκεται στην δεξιά πλευρά. Η συστατική διαδρομή, που ακολουθεί την τυπική αρχιτεκτονική ενός συνελκτικού δικτύου, είναι υπεύθυνη για την αποτύπωση του γενικού πλαισίου και η συμμετρική επεκτεινόμενη διαδρομή για τον ακριβή εντοπισμό (localization). Στην συστατική διαδρομή γίνονται δυο επαναληπτικές συνελίξεις 3x3 η κάθε μία ακολουθούμενη από μία διορθωμένη γραμμική μονάδα (ReLU) και μία διαδικασία max ή mean pooling με βήμα δύο για την υποδειγματοληψία (upsampling). Σε κάθε στάδιο αυτής της υποδειγματοληψίας, ο αριθμός των καναλιών χαρακτηριστικών (feature channels) διπλασιάζεται. Στην δεξιά μεριά, δηλαδή στην επεκτεινόμενη διαδρομή, κάθε βήμα της αποτελεί μια υπερδειγματοληψία (upsampling) των feature maps. Έπειτα εφαρμόζεται συνέλιξη 2x2 που δίνει τα μισά κανάλια μειώνοντας τον αριθμό τους με αυτόν τον τρόπο, μία συνένωση με τον αντίστοιχο περικομένο feature map από την συστατική διαδρομή και ύστερα δύο συνελίξεις 3x3 ακολουθούμενες από διορθωμένη γραμμική μονάδα (ReLU). Τέλος, στο τελευταίο στρώμα είναι μια συνέλιξη 1x1 ώστε να γίνει σωστά η αντιστοίχιση του κάθε διανύσματος χαρακτηριστικών μεγέθους 64 με την επιθυμητό αριθμό κλάσεων.

Το σημαντικό χαρακτηριστικό αυτού του μοντέλου είναι ότι περιέχει 23 συνελκτικά στρώματα δηλαδή έχουμε μεγάλο αριθμό καναλιών χαρακτηριστικών στο αριστερό κομμάτι(contracting path) επιτρέποντας έτσι στο δίκτυο να μεταδώσει πληροφορίες με μεγαλύτερη ανάλυση στο γενικό πλαίσιο. Μία ακόμα καθοριστική λειτουργία του είναι τα skip connections που στο παρακάτω σχήμα φαίνονται από τα βελάκια του encoder κατευθείαν στον decoder. Τα skip connections παίρνει κάποια layers από την έξοδο του encoder και τα περνάει κατευθείαν ως είσοδο στα blocks του decoder, δίνοντας μας έτσι μία πιο ολοκληρωμένη εικόνα για τη μάσκα τμηματοποίησης. Η τεχνική Unet στην παρούσα εργασία χρησιμοποιείται ως backbone και ως decoder head έχουμε εξ ολοκλήρου συνελκτικά δίκτυα (FCN) όπου έχει αναλυθεί παραπάνω. Ο συνδυασμός αυτός μας δίνει τα βέλτιστα αποτελέσματα και από τις δύο τεχνικές.

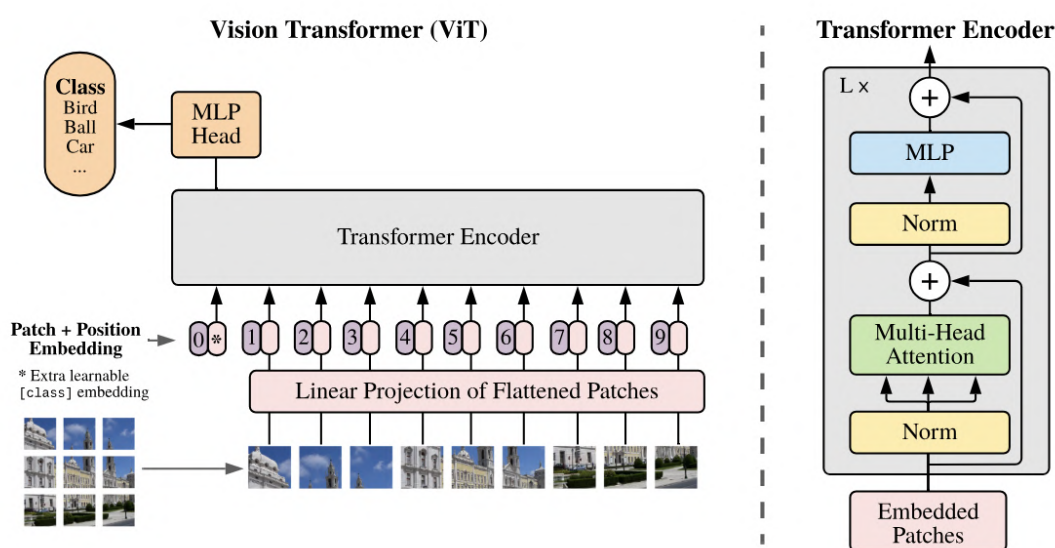


Σχήμα 3.8: Η αρχιτεκτονική του Unet.[10]

3.2.4 Vision Transformer (ViT)

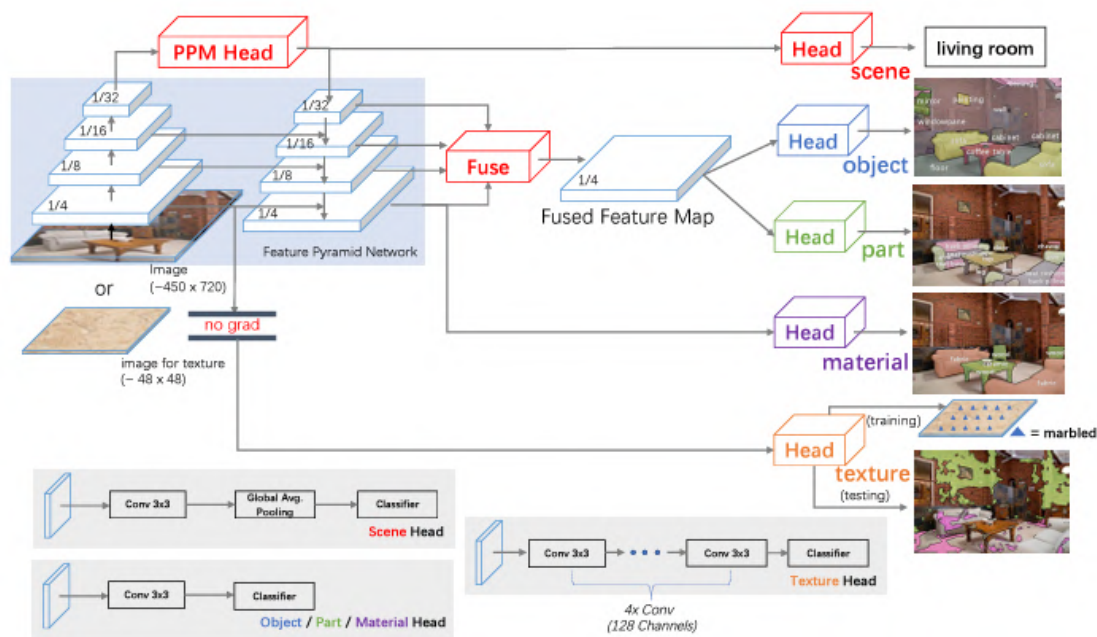
Ο Vision Transformer (ViT) [11] είναι ένα βαθύ νευρωνικό μοντέλο που η αρχιτεκτονική του βασίζεται στους μετασχηματιστές. Αποτελεί σημαντικό εργαλείο για τμηματοποίηση εικόνας και γενικά στο κλάδο της όρασης υπολογιστών και γνώρισε μεγάλη επιτυχία στην επεξεργασία φυσικών γλωσσών NLP. Ο ViT χρησιμοποιείται κυρίως για προβλήματα ταξινόμησης ωστόσο έχει επεκταθεί και στην τμηματοποίηση εικόνων. Η επιτυχία στην ταξινόμηση προήλθε από το τρόπο που δουλεύει καθώς ο διασπάει την εικόνα σε x κομμάτια σταθερού μεγέθους, ενσωματώνονται γραμμικά και έπειτα επεξεργάζονται με σκοπό να δώσουν παγκόσμιες πληροφορίες περιβάλλοντος.

Πιο συγκεκριμένα ο Vision Transformer (ViT) αποτελείται από ένα στρώμα που περνάνε τα κομμάτια από την διάσπαση τις αρχικής εικόνας και από πολλούς διαδοχικούς κωδικοποιητές. Το στρώμα αυτοπροσοχής παίζει σημαντικό ρόλο καθώς μετασχηματίζει την αρχική εικόνα σε τρία μέρη, το διάνυσμα ερωτήματος, κλειδιού και τιμής. Τον βαθμό κάθε ζεύγους διανυσμάτων ανάλογα με το σήμα, αναλαμβάνει το στρώμα προσοχής.



Σχήμα 3.9: Η αρχιτεκτονική του ViT.[11]

Στην παρούσα διπλωματική εργασία το μοντέλο που έχει χρησιμοποιηθεί είναι υβριδικό καθώς χρησιμοποιεί τον συγκεκριμένο μετασχηματιστή ως backbone και ως decode head το UPerNet. Το UPerNet [12] χρησιμοποιεί Pyramid Pooling Module (PPM) και επιπλέον convolutional layers με σκοπό την κατασκευή των χαρτών τμηματοποίησης. Δίνεται το παράδειγμα αυτής της διαδικασίας.



Σχήμα 3.10: Η αρχιτεκτονική του UperNet [12].

3.2.5 Swin

Ο μετασχηματιστής Swin [13] αποτελεί μια καινοτομία σε σχέση με τους κλασικούς μετασχηματιστές αντιμετωπίζοντας σημαντικές προκλήσεις όπως τις πολυποικίλες παραλλαγές που μπορεί να έχουν οι εικόνες-είσοδοι και την πολυπλοκότητα όσον αφορά την υψηλή ανάλυση (resolution) των εικόνων. Η αρχιτεκτονική του Swin βασίζεται σε ένα ιεραρχικό σχεδιασμό ο οποίος διαμορφώνει χάρτες χαρακτηριστικών (feature maps) σε διάφορες κλίμακες με σκοπό τον αποτελεσματικό χειρισμό εικόνων μεταβαλλόμενων αναλύσεων και κλιμάκων. Η αρχική εικόνας εισόδου χωρίζεται σε μικρά μη επικαλυπτόμενα κομμάτια (patches). Καθένα από αυτά τα patches αντιμετωπίζεται ως δείγμα (token) και έπειτα εισάγονται γραμμικά σε ένα μεγαλύτερο διάστασης χώρο, δημιουργώντας έτσι την αρχική είσοδο του μετασχηματιστή.

Η αρχιτεκτονική αυτή χωρίζεται σε διάφορα στάδια, καθένα από αυτά τα στάδια έχει σκοπό να μειώσει τον αριθμό των tokens και να ενισχύσει την αναπαράστασή τους, δηλαδή τις λεπτομέρειες και τα χαρακτηριστικά της αρχικής εικόνας. Στο πρώτο στάδιο, τα tokens που αναφέρθηκαν πριν, διατηρούν τον αριθμό και τις διαστάσεις τους μέσα από διάφορα στρώματα (layers) από τα Swin Transformer blocks που εισήχθησαν. Τα Swin Transformer blocks περιέχουν πολλούς πολυκεφάλους αυτοπροσοχής μηχανισμούς (multi head self attention (MSA) που εντοπίζονται και λειτουργούν μέσα από τοπικά παράθυρα. Περιέχουν επίσης έναν MLP για τον μετασχηματισμό των χαρακτηριστικών. Οι μηχανισμοί αυτοπροσοχής επικεντρώνονται στα παράθυρα που δεν επικαλύπτονται με σκοπό την γραμμική διατήρηση της υπολογιστικής πολυπλοκότητας ώστε να μην αλλάξει το μέγεθος της αρχικής εικόνας.

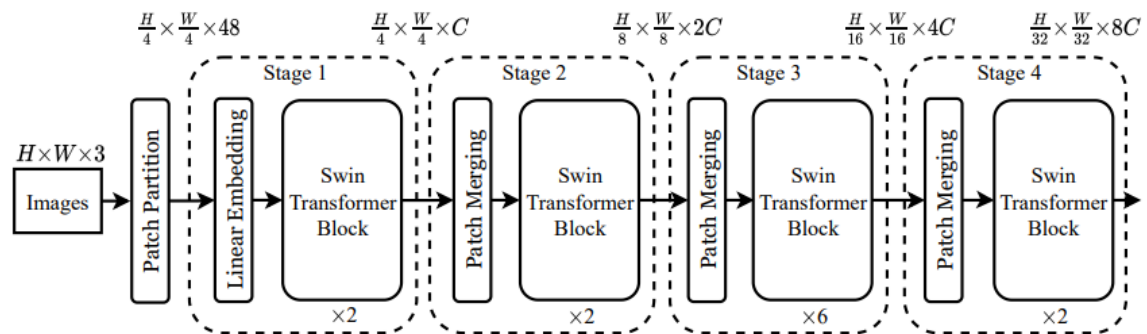
Η καινοτομία του Swin έγκεται στην χρήση τοπικών μετακινούμενων παραθύρων για τη υλοποίηση των μηχανισμών αυτοπροσοχής κάνοντας το σύστημα πιο αποδοτικό σε σχέση με τους απλούς μετασχηματιστές που δεν αποδίδουν καλά σε εικόνες-είσοδους υψηλής ανάλυ-

σης. Για την εύρεση εξαρτήσεων μεταξύ διαφορετικών τμημάτων της εικόνας, το μοντέλο μετακινεί τα παράθυρα μεταξύ διαδοχικών στρωμάτων. Αυτό έχει ως αποτέλεσμα την αύξηση της αναπαραστατικής δύναμης χωρίς επιπλέον υπολογιστικό κόστος καθώς επιτρέπεται, με αυτόν τον τρόπο, η σύνδεση μη επικαλυπτόμενων παραθύρων από τα προηγούμενα στρώματα.

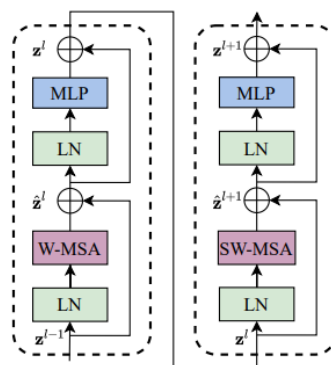
Στο δεύτερο βήμα, ο αριθμός των δειγμάτων μειώνεται μέσω της διαδικασίας ένωσης δειγμάτων (patch merging). Αυτό έχει ως αποτέλεσμα να μειώσει τον αριθμό των δειγμάτων αλλά τα δείγματα που μένουν να είναι πιο πλούσια από άποψη πληροφοριών. Έτσι η υπολογιστική πολυπλοκότητα μένει ίδια, όμως διαπλασιάζονται οι διαστάσεις των χαρακτηριστικών, ενισχύοντας την αναπαραστατική δύναμη. Έπειτα σε αυτά τα δείγματα εφαρμόζονται ξανά τα Swin Transformer blocks βελτιώνοντας ξανά την ιεραρχική αναπαράσταση.

Στο στάδιο τρία και τέσσερα ακολουθείται το ίδιο ακριβώς μοτίβο, δηλαδή σε κάθε στάδιο τα δείγματα ξαναενώνονται μεταξύ τους μειώνοντας τον αριθμό τους και αυξάνοντας τις διαστάσεις τους. Στο τέταρτο στάδιο, το αποτέλεσμα που δημιουργείται, μετά την εφαρμογή των Swin Transformer blocks, είναι μία αφηρημένη αλλά συμπαγής αναπαράσταση της εικόνας εισόδου. Όσο προχωράμε στα στάδια η απόδοση των μηχανισμών προσοχής δεν είναι τόσο αποδοτικοί καθώς μεγαλώνει το μέγεθος των patches που κοιτάνε, όμως η συνεργασία αυτών των 4 σταδίων δίνει τα βέλτιστα αποτελέσματα.

Παρακάτω φαίνεται η αρχιτεκτονική του μοντέλου και δύο διαδοχικά Swin Blocks 3.11 3.12:



Σχήμα 3.11: Η αρχιτεκτονική του μετασχηματιστή SWIN [13].



Σχήμα 3.12: Η αρχιτεκτονική των Block Swins [13].

Όπως και στον ViT έτσι και εδώ, ο μετασχηματιστής Swin βρίσκεται στο backbone δίνοντας μας έτσι τα features από τις εικόνες που έχουμε για είσοδο. Ως decode head χρησιμοποιούμε το UPerNet που η αρχιτεκτονική του έχει αναλυθεί παραπάνω.

Συμπερασματικά, ο μετασχηματιστής Swin συνδυάζει τα θετικά από τους ιεραρχικούς χάρτες και τους μηχανισμούς αυτοπροσοχής, δίνοντας έτσι την δυνατότητα να διαχειριστούν σωστά εικόνες υψηλής ανάλυσης ενισχύοντας έτσι τον κλάδο της όρασης υπολογιστών.

Μέρος **III**

Πρακτικό Μέρος

Κεφάλαιο 4

Δεδομένα

Στο κεφάλαιο αυτό παρουσιάζονται τα σύνολα δεδομένων που χρησιμοποιήθηκαν για την εκπόνηση της διπλωματικής εργασίας, για την επεξεργασία αυτών των συνόλων δεδομένων, για την εκπαίδευση των μοντέλων και τεχνικών που χρησιμοποιήθηκαν καθώς και για την υλοποίησή τους.

4.1 Σύνολο δεδομένων

4.1.1 Kvasir-SEG

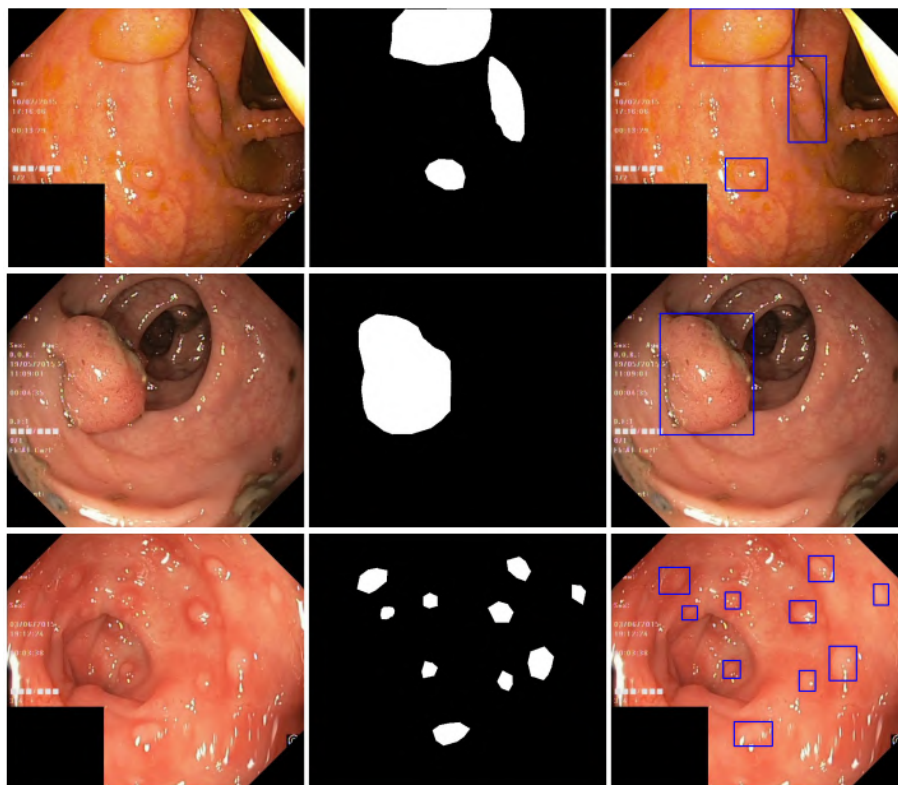
Το Kvasir-SEG [51] είναι ένα σύνολο δεδομένων ανοιχτής πρόσβασης που αποτελείται από εικόνες γαστρεντερικών πολυπόδων με τις αντίστοιχες μάσκες τμηματοποίησης τους, χωρίς αυτόματο σχολιασμό. Οι μάσκες αυτές δίνουν την δυνατότητα στους ερευνητές του τομέα όρασης υπολογιστών να προσφέρουν καινούργιες γνώσεις και καινοτομίες στον τομέα σημασιολογικής τμηματοποίησης πολύποδων και στην αυτόματη ανάλυση από βίντεο κολονοσκόπησης.

Οι πολύποδες αποτελούν ένδειξη για καρκίνο του παχέους εντέρου. Ο καρκίνος στο συγκεκριμένο τμήμα του εντέρου είναι στην δεύτερη θέση πιο κοινού τύπου καρκίνου στις γυναίκες και στην τρίτη για τους άντρες. Η κολονοσκόπηση είναι μία σημαντική εξέταση για την ανίχνευση και αξιολογότητα των πολυπόδων με σκοπό την βιοψία και αφαίρεση τους. Σύμφωνα με τις με τις νεότερες οδηγίες της Αμερικανικής Αντικαρκινικής Εταιρείας (ACS) η εξέτασή αυτή πρέπει να υποβληθεί, κατά κύριο λόγο, στην ηλικία των 45 ετών, καθώς η συχνότητα των πολυπόδων αυξάνεται με την ηλικία. 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 των πολυπόδων. Η έγκαιρη πρόβλεψη τους έχει ως αποτέλεσμα την αύξηση των ποσοστών επιβίωσης από τον καρκίνο του παχέους εντέρου, καθιστώντας δηλαδή την ανίχνευση τους κρίσιμη. Αυτό είναι το κίνητρο πίσω από την ανάπτυξη του συνόλου δεδομένων Kvasir-SEG.

Το σύνολο δεδομένων Kvasir-SEG περιέχει 1000 εικόνες πολυπόδων και τις αντίστοιχες μάσκες τμηματοποίησης τους. Η ανάλυση των εικόνων κυμαίνεται από 332x487 και 1920x1072 εικονοστοιχεία (pixels). Οι εικόνες και οι μάσκες τους είναι αποθηκευμένες σε δύο ξεχωριστούς φακέλους κρατώντας το ίδιο όνομα αρχείου. Τα αρχεία εικόνας κωδικοποιούνται με συμπίεση JPEG και οι συντεταγμένες των πλαισίων οριοθέτησης των πολυπόδων στις εικόνες για τον εντοπισμό τους είναι αποθηκευμένες σε ένα αρχείο JSON.

Οι μάσκες τμηματοποίησης δημιουργήθηκαν μέσω του λογισμικού Labelbox, το οποίο είναι ένα εργαλείο που χρησιμοποιείται για την επισήμανση της περιοχής ενδιαφέροντος ROI σε καρέ εικόνας, στην συγκεκριμένη περίπτωση δηλαδή των περιοχών των πολύποδων. Με την βοήθεια εξειδικευμένων γιατρών υλοποιήθηκαν χειροκίνητα αυτές οι περιοχές των πολύποδων και στις 1000 εικόνες. Τα εικονοστοιχεία που απεικονίζουν τον ιστό των πολύποδων, δηλαδή την περιοχή ενδιαφέροντος, αντιπροσωπεύονται από το προσκίνηιο (άπορη μάσκα), ενώ το φόντο (background) (σε μαύρο) δεν περιέχει θετικά εικονοστοιχεία. Μερικές από τις αρχικές εικόνες του συνόλου δεδομένων μας περιέχουν την εικόνα του καθετήρα σήμανσης θέσης ενδοσκοπίου, ScopeGuide TM, Olympus Tokyo Japan, που βρίσκεται σε μία από τις κάτω γωνίες και φαίνεται ως ένα μικρό πράσινο κουτί. Καθώς αυτές οι πληροφορίες είναι περιττές για την τμηματοποίηση της εικόνας, έχουν αντικατασταθεί με μαύρα κουτιά στο σύνολο δεδομένων Kvasir-SEG.

Η συλλογή των δεδομένων έγινε με την χρήση ενδοσκοπικού εξοπλισμού στο Vestre Viken Health Trust (VV) στην Νορβηγία. Το VV αποτελείται από τέσσερα νοσοκομεία. Τα δεδομένα εκπαίδευσης έχουν συλλεχθεί από το Νοσοκομείο Bærum καθώς διαθέτει ένα μεγάλο γαστρεντερολογικό τμήμα. Επιπρόσθετα οι εικόνες σχολιάστηκαν λεπτομερώς από έναν ή περισσότερους ειδικούς γιατρούς από το VV και το Μητρώο Καρκίνου της Νορβηγίας(CRN). Το CRN προσφέρει νέες γνώσεις για τον καρκίνο μέσω της έρευνας και αποτελεί μέρος της Περιφερειακής Αρχής Υγείας της Νοτιοανατολικής Νορβηγίας. Είναι οργανωμένο ως ανεξάρτητο ίδρυμα εντός του Πανεπιστημιακού Νοσοκομείου του Όσλο. Είναι επίσης αρμόδιο για τα εθνικά προγράμματα προσυμπτωματικού ελέγχου του καρκίνου.



Σχήμα 4.1: Παραδείγματα του Kvasir-SEG Dataset [14]

4.1.2 COVID-QU-EX

Ένα άλλο σύνολο δεδομένων ανοιχτής πρόσβασης αποτελεί το COVID-QU-Ex [52]. Αποτελεί ένα μεγάλο σύνολο δεδομένων με περίπου 8000 στοιχεία. Τα δεδομένα είναι ακτινογραφίες θώρακος και οι αντίστοιχες μάσκες τους. Οι κατηγορίες που χωρίζονται είναι COVID-19 [53], [54] [55] [56] [57] [58] [59], Non-COVID [60] [61] [62], Normal [60] [61] [62]. Τα δεδομένα είναι χωρισμένα σε δύο υποσύνολα, τα δεδομένα για την τμηματοποίηση του πνεύμονα Lung Segmentation Data και τα δεδομένα για την μόλυνση Infection Segmentation Data. Το Lung Segmentation Data περιέχει 33920 εικόνες (11956 COVID-19, 11263 Non-COVID, 10701 Normal) που είναι ήδη μοιρασμένες στους υποφακέλους Train, Val, Test. Η μάσκα της κάθε εικόνας στο συγκεκριμένο υποσύνολο διαχωρίζει την περιοχή των πνευμόνων από την συνολική ακτινογραφία. Το δεύτερο υποσύνολο (Infection Segmentation Data) περιέχει μερικές εικόνες από το προηγούμενο υποσύνολο, συνολικά 5826 εικόνες (1456 Normal, 1457 Non-COVID, 2913 COVID-19). Η διαφορά είναι ότι πέρα από τις μάσκες που διαχωρίζει τους πνεύμονες υπάρχουν και οι μάσκες που διαχωρίζουν την περιοχή της φλεγμονής που βρίσκεται μέσα στην περιοχή των πνευμόνων.

Παρακάτω παραθέτουμε κάποια παραδείγματα εικόνων του dataset μας:



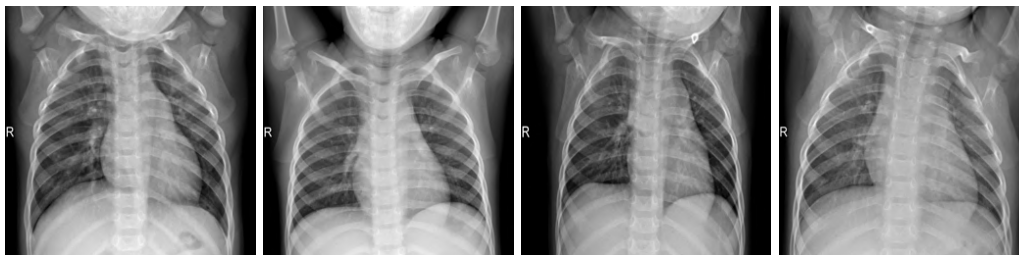
Σχήμα 4.2: COVID-19 Images



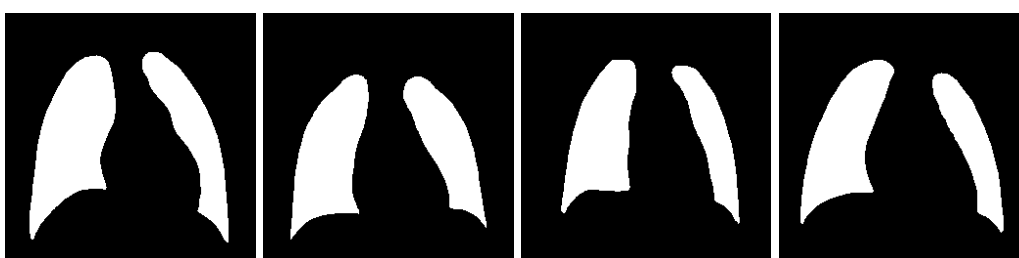
Σχήμα 4.3: COVID-19 Lung Masks



Σχήμα 4.4: COVID-19 Infection Masks

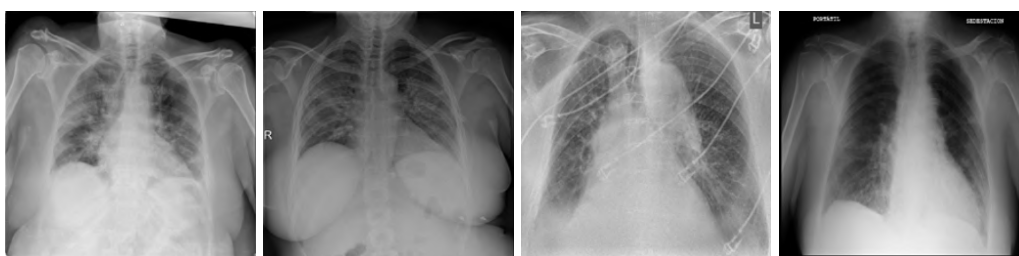


Σχήμα 4.5: *Normal Images*

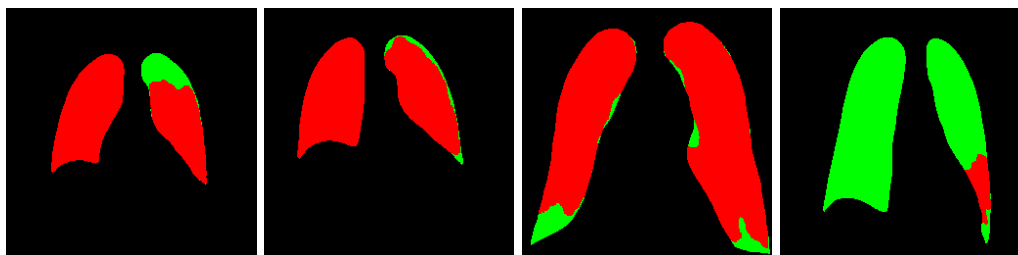


Σχήμα 4.6: *Normal Lung Masks*

Στο σημείο αυτό πρέπει να αναφερθεί ότι χρησιμοποιήθηκε μόνο ένα υποσύνολο αυτών των δεδομένων για λόγους υπολογιστικής απλοποίησης καθώς και για το γεγονός ότι δεν παρέχονται οι μάσκες της υποκατηγορίας Non-COVID οπότε το πρόβλημα μας επικεντρώνεται στην τμηματοποίηση της περιοχής των πνευμόνων καθώς και στην τμηματοποίηση μόνο της φλεγμονής της ασθένειας COVID-19. Επίσης η αρχική μας εικόνα έχει το ίδιο όνομα και με τις δύο μάσκες της και επειδή η τμηματοποίηση της φλεγμονής εμπεριέχεται μέσα στην τμηματοποίηση της περιοχής των πνευμόνων, για το καλύτερο διαχωρισμό, η τμηματοποίηση της περιοχής των πνευμόνων θα έχει το πράσινο χρώμα ενώ η τμηματοποίηση της φλεγμονής θα έχει το κόκκινο χρώμα. Οι δύο αυτές μάσκες ενώθηκαν για πιο ξεκάθαρα οπτικά αποτελέσματα.



Σχήμα 4.7: *COVID-19 Images*



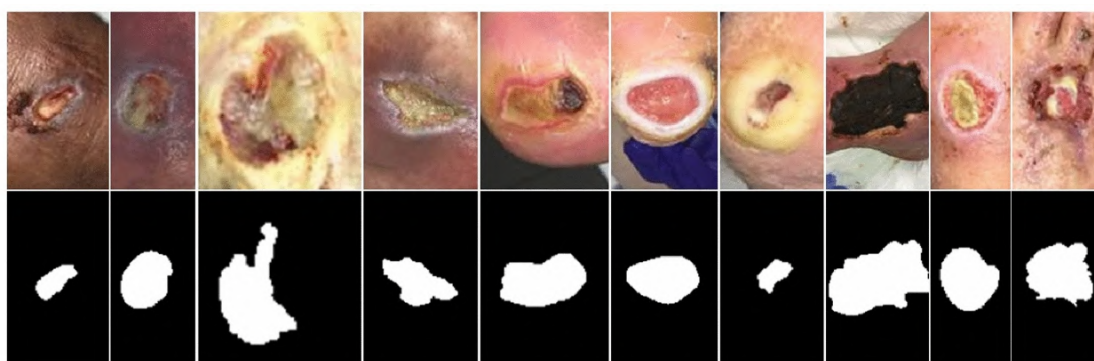
Σχήμα 4.8: COVID-19 Mask Images with Colour

4.1.3 The Foot Ulcer Segmentation Challenge (FUSeG)

Το τελευταίο σύνολο δεδομένων που χρησιμοποιήθηκε στην παρούσα εργασία είναι από την πλατφόρμα του github [15] μέσω της δημοσίευσης Fully automatic wound segmentation with deep convolutional neural networks [63].

Οι χρόνιες και οξείες πληγές των ασθενών επιβαρύνουν οικονομικά το σύστημα υγείας παγκοσμίως. Τα έξοδα για τη φροντίδα αυτών των ασθενών αναμένονται να ξεπεράσουν τα 22 δισεκατομμύρια έως το 2024. Οι εξιδεικευμένοι γιατροί σε αυτά τα τραύματα βασίζονται σε εικόνες για την σωστή διάγνωση και θεραπεία. Η έλλειψη εμπειρίας όμως από νέους γιατρούς σε αυτόν τον τομέα μπορεί να επιφέρει λανθασμένη γνωμάτευση και συνεπώς λανθασμένη θεραπεία. Για αυτόν τον λόγο η αυτόματη τμηματοποίηση αυτών των πληγών καθιστάται αναγκαία, τόσο για τον υπολογισμό της έκτασης όσο και για τις παραμέτρους θεραπείας.

Όσον αφορά τα σύνολα δεδομένων για τμηματοποίηση πληγών (wound), δεν έχουν δημοσιευτεί επαρκή σύνολα δεδομένων για να εκπαιδευτούν τα βαθιά νευρωνικά μοντέλα ώστε να μπορούν να κάνουν τμηματοποίηση πληγής. Αυτό το σύνολο δεδομένων αποτελεί σε συνεργασία με το Κέντρο Πληγών και Αγγειακής Φροντίδας (Advancing the Zenith of Healthcare (AZH)), στο Μιλγουόκι, ένα από τα μεγαλύτερα έως τώρα σύνολα δεδομένων για την τμηματοποίηση πληγής. Τα δεδομένα συλλέχθηκαν σε διάστημα δύο χρόνων στο κέντρο που αναφέρθηκε και περιλαμβάνουν 1109 εικόνες ελκών ποδιού από 889 ασθενείς κατά την διάρκεια πολλαπλών κλινικών επισκέψεων. Οι εικόνες αυτές πάρθηκαν μέσω μιας φωτογραφικής μηχανής Canon SX 620 HS και iPad Pro υπό μη ελεγχόμενες συνθήκες φωτισμού με διάφορα φόντα. Παρακάτω φαίνονται οι εικόνες με τις αντίστοιχες μάσκες τους από το συγκεκριμένο σύνολο δεδομένων.



Σχήμα 4.9: Παραδείγματα του FUSeG [15]

4.2 Προεπεξεργασία και επαύξηση δεδομένων

Για την προεπεξεργασία και αύξηση του συνόλου δεδομένων που χρησιμοποιούνται και για την αύξηση της απόδοσης της διαδικασίας εκπαίδευσης των μοντέλων χρησιμοποιήθηκαν κάποιες τεχνικές επαύξησης δεδομένων. Οι τεχνικές αυτές στην ουσία προσαρμόζουν κάποιες εικόνες τυχαία ώστε να γίνουν τα μοντέλα πιο ανθεκτικά στις αλλαγές και να μπορούν να γενικεύουν καλύτερα σε νέα δεδομένα. Παρακάτω περιγράφονται συνοπτικά οι τεχνικές που χρησιμοποιήθηκαν:

- κλίμακα εικόνας (image scale): Η συγκεκριμένη μεταβλητή δίνει συγκεκριμένη διάσταση στις εικόνες όπου στα δικά μας δεδομένα έχει οριστεί (512, 512).
- Φωτομετρική Παραμόρφωση (PhotoMetricDistortion): Εφαρμόζονται τυχαίες φωτομετρικές παραμορφώσεις στις εικόνες. Πιο συγκεκριμένα αυτές οι παραμορφώσεις απευθύνονται σε αλλαγές στην απόχρωση, την αντίθεση, στο κορεσμό και την φωτεινότητα της εικόνας.
- Αναστροφή (RandomFlip): Η τυχαία αναστροφή κάποιων εικόνων, οριζόντια ή κάθετα, αυξάνει τη ποικιλία των δεδομένων εκπαίδευσης. Ο δείκτης για το RandomFlip έχει οριστεί σε πιθανότητα 0.5 δηλαδή υπάρχει 50% από τις εικόνες που δίνονται να αναστραφούν οριζόντια.
- Περιστροφή (Rotation): Η περιστροφή των εικόνων σε τυχαίες γωνίες βοηθά στην αύξηση των δεδομένων εκπαίδευσης.

Κεφάλαιο 5

Μετρικές αξιολόγησης και περιβάλλον εκτέλεσης

Στο κεφάλαιο αυτό περιγράφεται η επιλογή των μετρικών αξιολόγησης καθώς και το περιβάλλον εκτέλεσης πειράματος.

5.1 Μετρικές αξιολόγησης

Στην παρούσα εργασία οι μετρικές αξιολόγησης που χρησιμοποιήθηκαν είναι ο συντελεστής ομοιότητας (Dice Similarity Coefficient(DSC)) και η διασταύρωση πάνω στην ένωση (Intersection over Union(IoU)).

Η μετρική Dice Similarity Coefficient [64] χρησιμοποιείται για την αξιολόγηση ενός μοντέλου υπεύθυνο για την τμηματοποίηση. Ορίζεται ως ο μέσος (mean) μεταξύ της ακρίβειας (precision) και ευαισθησίας (sensitivity). Με αυτόν τον τρόπο υπολογίζει την επικάλυψη της τμηματοποίησης που δημιουργήθηκε από το μοντέλο με την πραγματική τμηματοποίηση. Η DSC αντιμετωπίζει ισάξια τα ψευδώς θετικά (false positive) και ψευδώς αρνητικά (false negative), θέτοντας έτσι μία ισορροπία ανάμεσα σε αυτά τα δύο που ευθύνονται για την υπερ-τμηματοποίηση (over-segmentation) και την υπο-τμηματοποίηση (under-segmentation, προβλήματα που πρέπει να αξιολογούνται ειδικά όσον αφορά το κομμάτι της τμηματοποίησης στα ιατρικά δεδομένα. Η DSC ορίζεται ως εξής:

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN}$$

όπου TP = True Positive , FN = False Negative και FP = False Positive, δηλαδή αληθώς θετικών, ψευδώς αρνητικών και ψευδώς θετικών εικονοστοιχείων αντίστοιχα. Ένας άλλος τρόπος που μπορεί να οριστεί ισοδύναμα είναι ο εξής:

$$DC = \frac{2|A \cap B|}{|A| + |B|}$$

όπου $|A \cap B|$ ο αριθμός των κοινών στοιχείων στα σύνολα A, B και ο παρανομαστής $|A|$, $|B|$ ο αριθμός των στοιχείων στα δύο αυτά σύνολα αντίστοιχα. Ο αριθμητής του κλάσματος αναφέρεται στις κοινές ενεργοποιήσεις μεταξύ των εικόνων (πραγματικής εικόνας και προβλεπόμενης), ενώ ο παρανομαστής στον αριθμό των ενεργοποιήσεων στις εικόνες ξεχωριστά.

Η Intersection over Union(IoU) [64] αποτελεί ένα άλλο μετρό αξιολόγησης μοντέλων

υπεύθυνα για τη τμηματοποίηση της εικόνας. Μετρά την αναλογία μεταξύ των κοινών ενεργοποιήσεων (επικάλυψη) της προβλεπόμενης από το μοντέλο εικόνας και της πραγματικής εικόνας. Δίνει περισσότερη έμφαση στο πρόβλημα της υπερ-τμηματοποίησης και της υπο-τμηματοποίησης σε σχέση με την μετρική DSC. Η Intersection over Union ορίζεται ως εξής:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

ή ισοδύναμα

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|}$$

όπου ο αριθμητής $|A \cap B|$ ο αριθμός των κοινών στοιχείων στα σύνολα A, B , δηλαδή ο αριθμός των κοινών ενεργοποιήσεων μεταξύ εικόνας που παράγεται από το μοντέλο και πραγματικής εικόνας και ο παρονομαστής $|A \cup B|$ ο αριθμός των στοιχείων σε αυτήν την ένωση.

Συμπερασματικά, όσον αφορά την τμηματοποίηση της εικόνας, οι μετρικές DSC και IoU αποτελούν σημαντικά εργαλεία για την αξιολόγηση των μοντέλων τμηματοποίησης που χρησιμοποιήθηκαν στην παρούσα εργασία.

5.2 Περιβάλλον εκτέλεσης πειράματος

Το πειραματικό μέρος υλοποιήθηκε στην πλατφόρμα Kaggle η οποία παρέχει δωρεάν πρόσβαση σε Virtual Machines εξοπλισμένα με επεξεργαστές vCPUs, 29GB μνήμη (RAM) και 2 κάρτες γραφικών NVIDIA T4 GPU με 15GB μνήμη η κάθε μία. Με βάση αυτές τις περιοχές τα πειράματα υλοποιήθηκαν σε λογικό χρόνο.

Πέρα από την πλατφόρμα του Kaggle χρησιμοποιήθηκε και ένα άλλο εργαλείο, το MM-Segmentation [50]. Το MM-Segmentation έχει αναπτυχθεί από την ομάδα OpenMMLab και ο κώδικας για τα μοντέλα που πραγματοποιούν σημασιολογική τμηματοποίηση παρέχεται δωρεάν. Η χρήση του MM-Segmentation παρέχει επιπλέον μία κλάση που αναφέρεται στο φόντο (background).

Κεφάλαιο 6

Πειραματικά αποτελέσματα μοντέλων τμηματοποίησης

Στο κεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα των πειραμάτων από τα μοντέλα που αναλύθηκαν στο θεωρητικό μέρος. Τα μοντέλα αυτά υλοποιήθηκαν στα σύνολα δεδομένων kvasir-SEG, COVID-QU-Ex, FUSeg που βρίσκονται στο κεφάλαιο 4 έχοντας εφαρμόσει την επαύξηση και προεπεξεργασία των δεδομένων. Τα μοντέλα μας έχουν εκπαιδευτεί για 160k iterations όμως τα περισσότερα έχουν φτάσει. Οι πίνακες που ακολουθούν παρακάτω δίνουν τα αποτελέσματα σε ποσοτική μορφή και οι εικόνες σε ποιοτική.

6.1 Λεπτομέρειες υλοποίησης

Στην παρούσα διπλωματική εργασία, το μέγεθος των εικόνων από τα τρία σύνολα δεδομένων (Kvasir-SEG, COVID-QU-Ex, FUSeg) ορίστηκε σε 512x512 στα μοντέλα τμηματοποίησης. Το batch size και το num size ισούται σε 1 και 4 αντίστοιχα στα περισσότερα μοντέλα μας. Η εκπαίδευση ορίστηκε στα 160 χιλιάδες iterations ωστόσο οι μέγιστες αποδόσεις των μοντέλων επιτεύχθηκαν νωρίτερα, κυρίως για τα μοντέλα τμηματοποίησης Swin και ViT όπου χρησιμοποιήθηκαν pretrained paths. Οι optimizers που επιλέχθηκαν στα μοντέλα ήταν οι Adam και SGD. Στα δύο σύνολα δεδομένων Kvasir-SEG και FUSeg χρησιμοποιήθηκαν τα 80% των δεδομένων στο σύνολο train και το 20% στο σύνολο validation ενώ το υποσύνολο του COVID-QU-Ex ήταν ήδη χωρισμένο στους αντίστοιχους φακέλους (train, test, val). Χρησιμοποιήθηκε κώδικας ανοιχτής πρόσβασης για όλα τα μοντέλα τμηματοποίησης της παρούσας εργασίας.

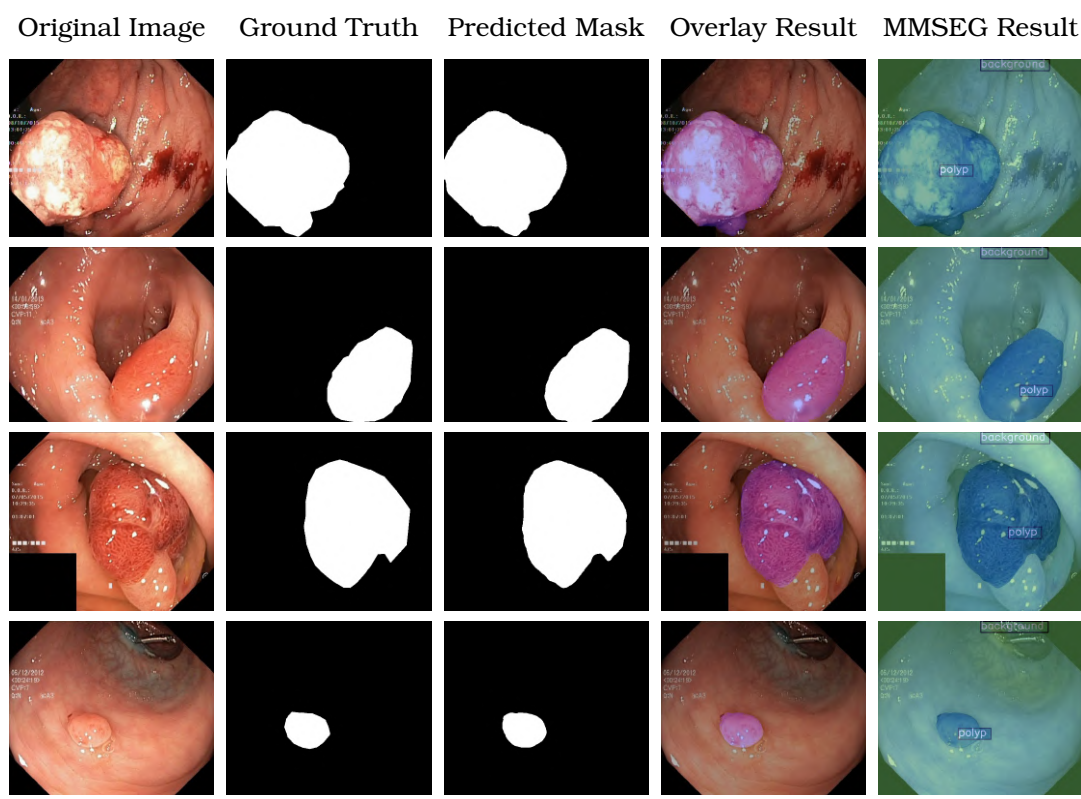
6.2 Πειραματικά αποτελέσματα μοντέλων τμηματοποίησης

Παρακάτω παρουσιάζονται τα πειραματικά αποτελέσματα από τα διάφορα μοντέλα τμηματοποίησης. Η πρώτη στήλη απευθύνεται στην αρχική εικόνα (original image), η δεύτερη για την μάσκα που έχει δημιουργηθεί από τους γιατρούς (ground truth mask), η τρίτη για την εικόνα που προβλέπεται (predicted mask) από τα μοντέλα μας σε δυαδική (binary) μορφή, η τέταρτη την Predicted mask εφαρμοσμένη στην αρχική εικόνα (overlay result) και η τελευταία στήλη δίνει το αποτέλεσμα από το MMsegmentation (MMSEG result), το οποίο βλέπουμε ότι αναφέρει και τα ονόματα τις κάθε κλάσεις. Η εικόνα στην πρώτη στήλη

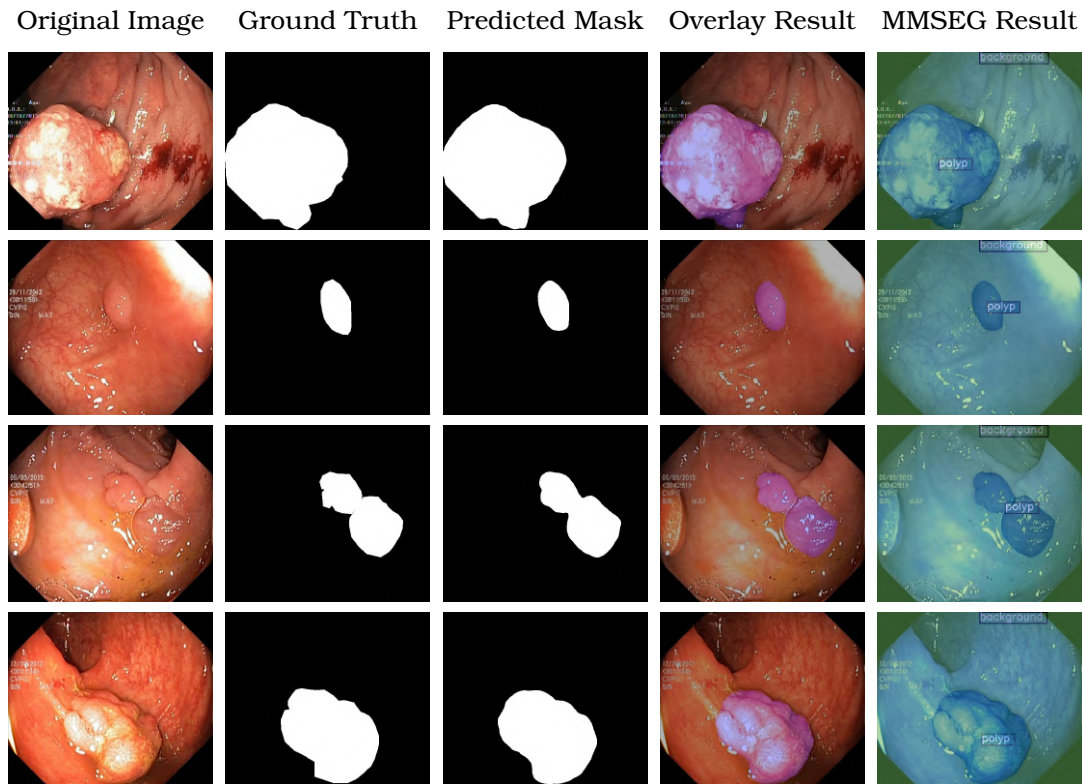
στην αρχή κάθε μοντέλου έχει επιλεχθεί να είναι η ίδια για να μπορούν να συγκριθούν οι διαφοροποιήσεις στα αποτελέσματα των μοντέλων, ενώ οι υπόλοιπες εικόνες έχουν επιλεχθεί τυχαία απο τα αντίστοιχα σύνολα δεδομένων.

6.2.1 Αποτελέσματα για το σύνολο δεδομένων Kvasir-SEG

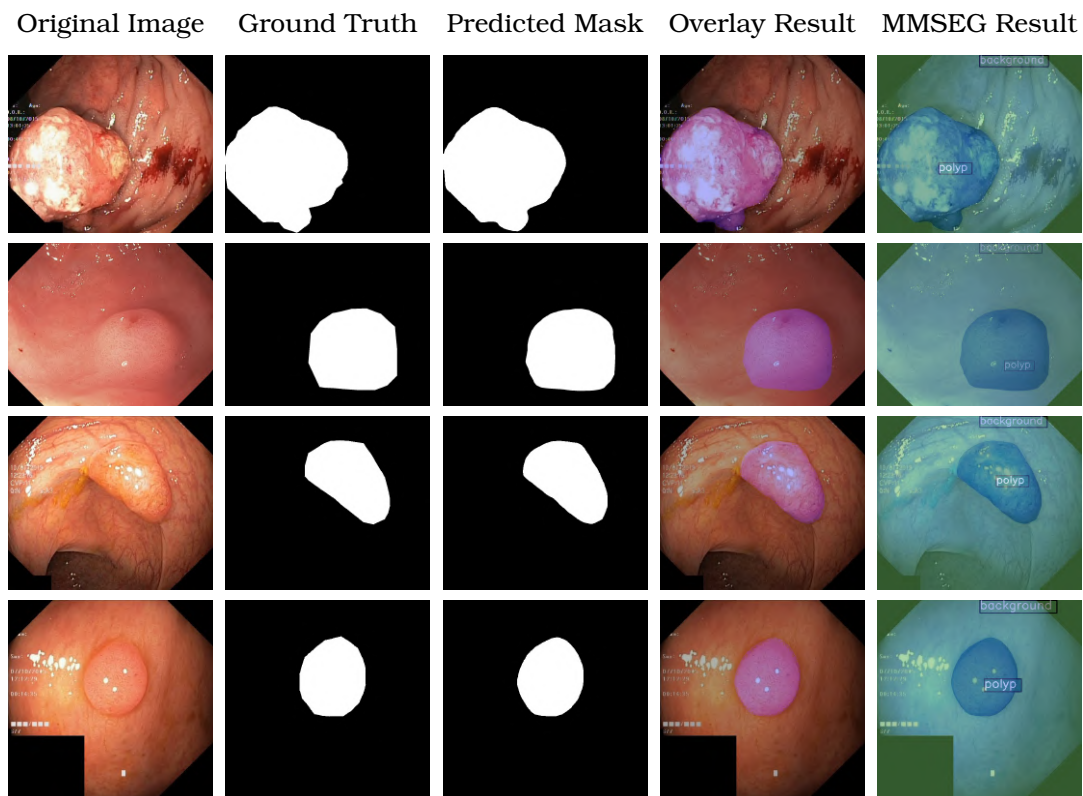
Οι εικόνες που αναφέρονται παρακάτω, καθώς και μέσω των αποτελεσμάτων από τον πίνακα 6.1 μετρικών Dice και IoU στο τέλος της υποενότητας, παρατηρούμε ότι το καλύτερο μοντέλο απο το σύνολο δεδομένων Kvasir-SEG είναι ο Segformer. Τα λιγότερα βέλτιστα αποτελέσματα προήλθαν απο το μοντέλο Unet αυτό φαίνεται και απο την εικόνα 6.5. Η εικόνα 6.6 παρουσιάζει όλα τα μοντέλα τμηματοποίησης για την ίδια εικόνα και οπτικά βλέπουμε ότι φέρνουν καλά αποτελέσματα. Οι μετρικές Dice και IoU το επιβεβαιώνουν καθώς κυμαίνονται σχεδόν όλα τα μοντέλα σε παρόμοιες τιμές.



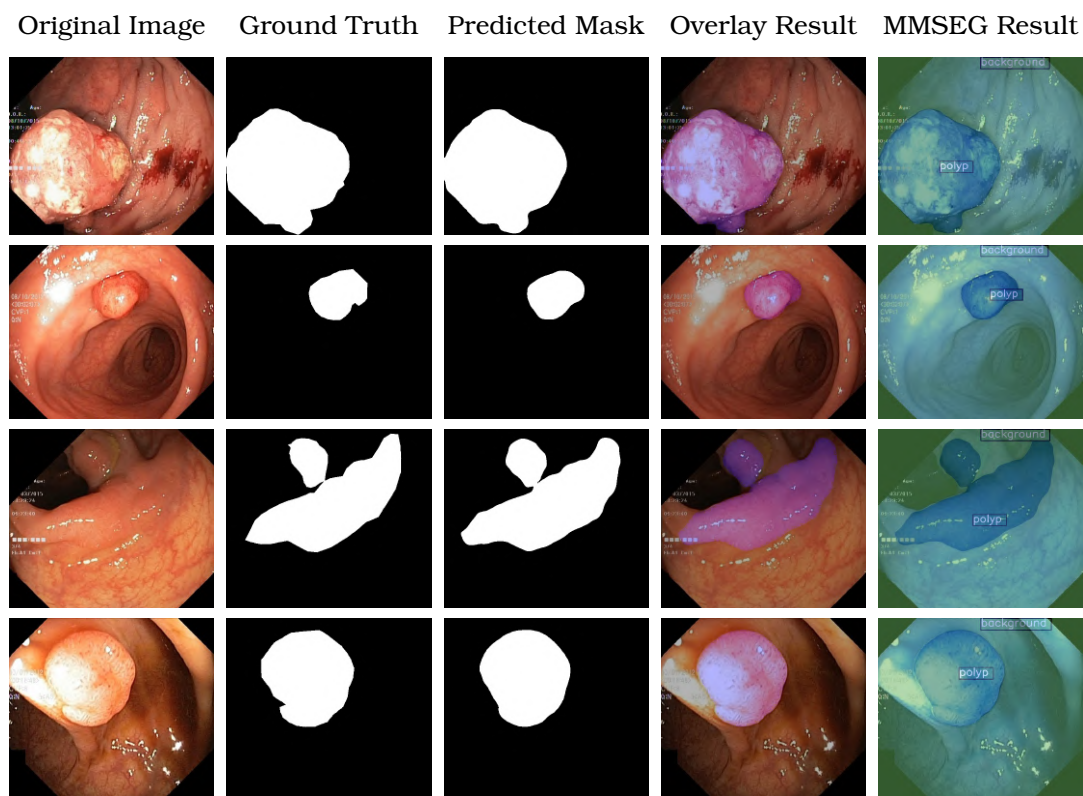
Εικόνα 6.1: Αποτελέσματα για το μοντέλο τμηματοποίησης Segformer στο σύνολο δεδομένων Kvasir-SEG.



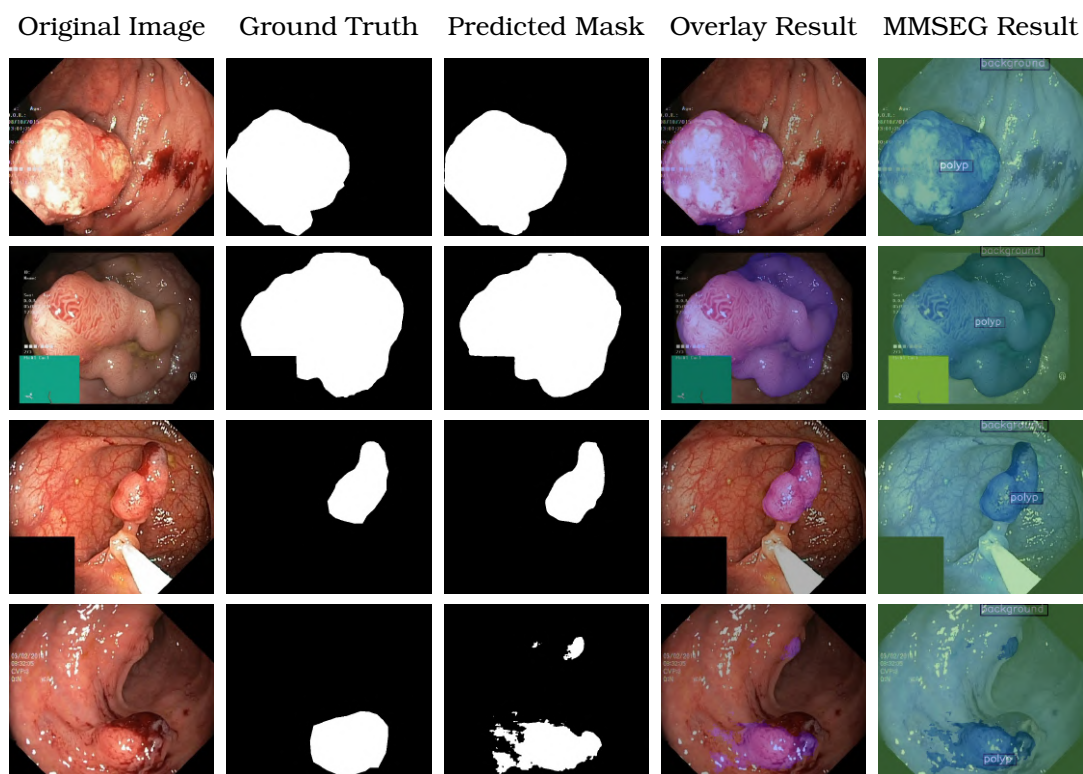
Εικόνα 6.2: Αποτελέσματα για το μοντέλο τμηματοποίησης Swin στο σύνολο δεδομένων Kvasir-SEG.



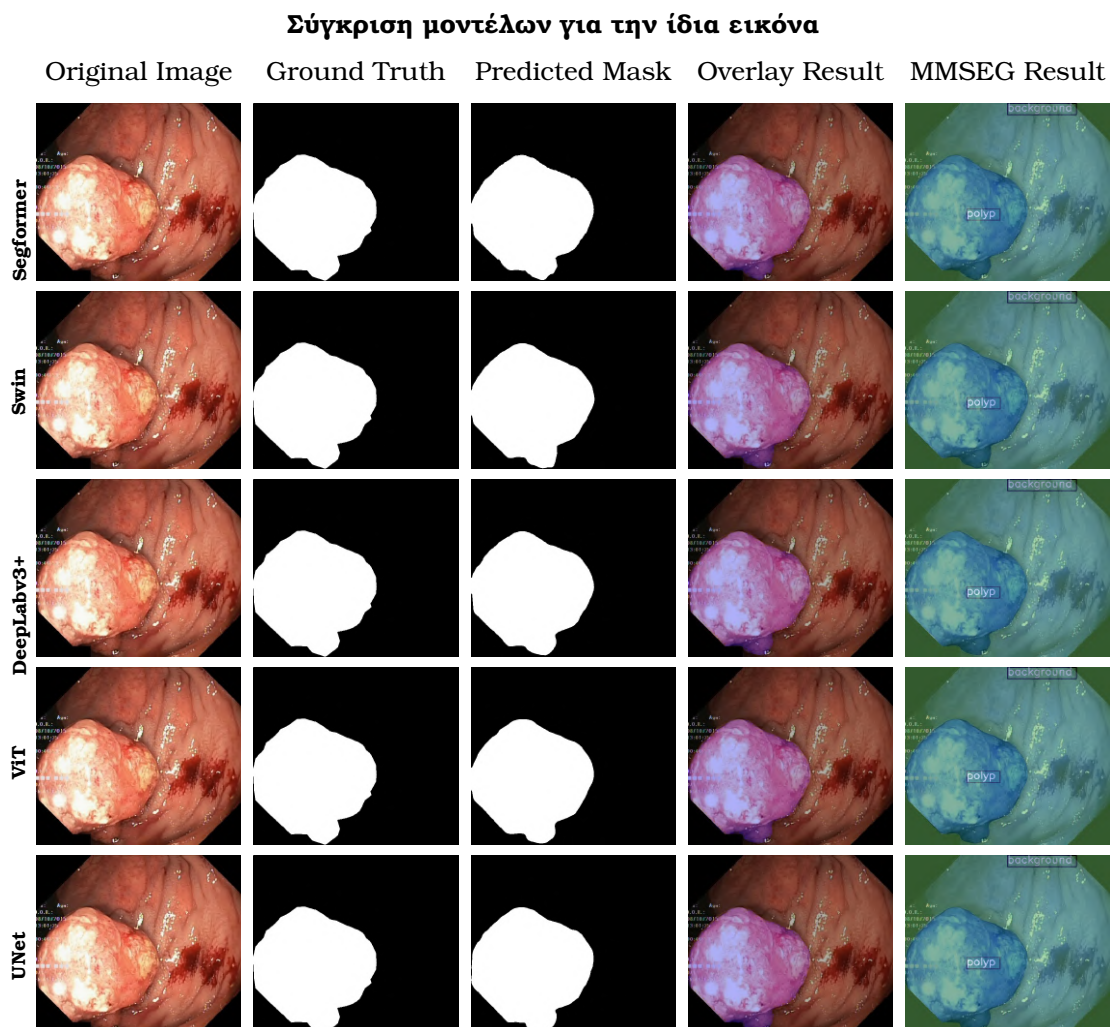
Εικόνα 6.3: Αποτελέσματα για το μοντέλο τμηματοποίησης DeepLabv3+ στο σύνολο δεδομένων Kvasir-SEG.



Εικόνα 6.4: Αποτελέσματα για το μοντέλο τμηματοποίησης ViT στο σύνολο δεδομένων Kvasir-SEG.



Εικόνα 6.5: Αποτελέσματα για το μοντέλο τμηματοποίησης Unet στο σύνολο δεδομένων Kvasir-SEG.



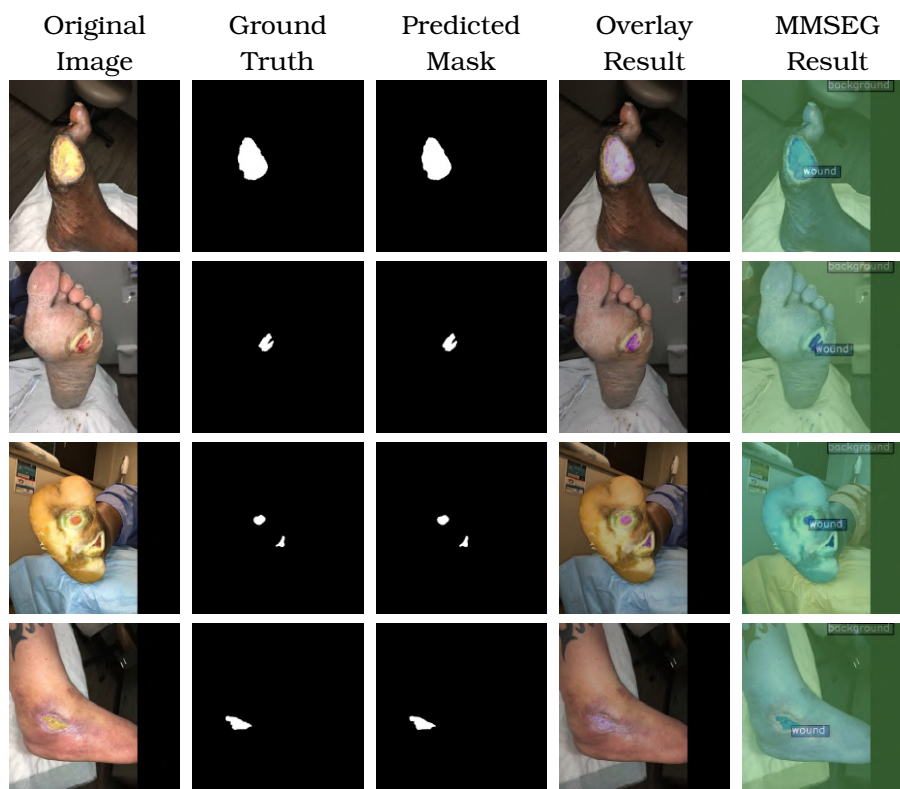
Εικόνα 6.6: Συγκριτικά αποτελέσματα μοντέλων τμηματοποίησης στο σύνολο δεδομένων Kvasir-SEG.

Πίνακας 6.1: Αποτελέσματα μεθόδων τμηματοποίησης βασισμένα στις μετρικές IoU και Dice για το σύνολο δεδομένων Kvasir-SEG

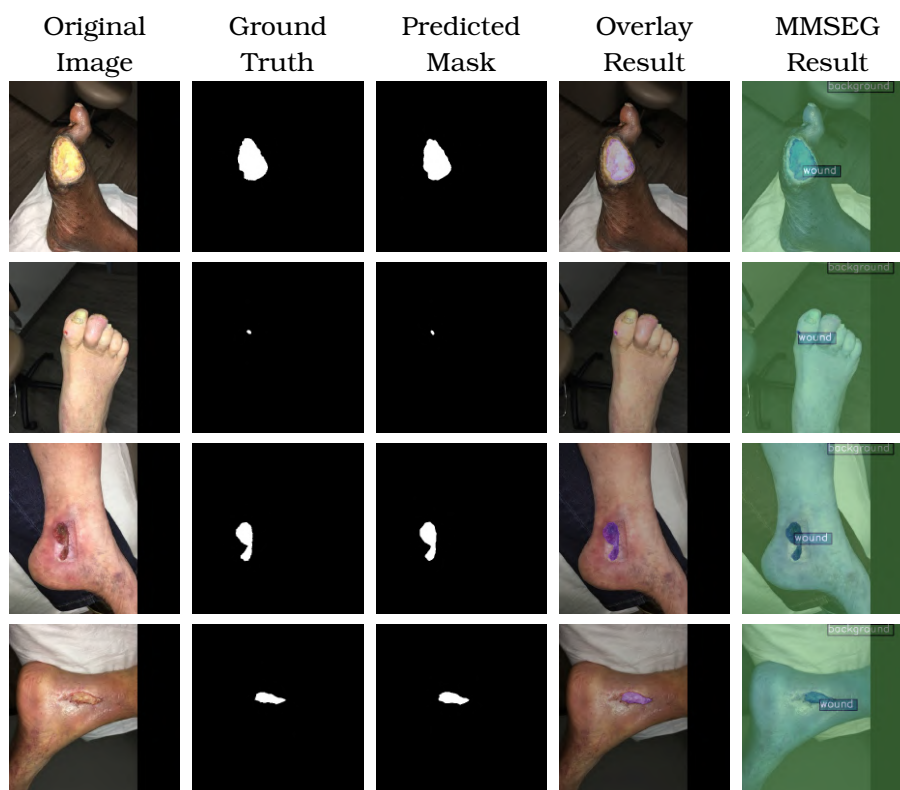
Method	IoU	Dice
Segformer	82.86	90.62
ViT	81.82	89.52
DeepLabv3+	76.72	86.83
Unet	66.44	79.84
Swin	80.98	89.49

6.2.2 Αποτελέσματα για το σύνολο δεδομένων FUSeg

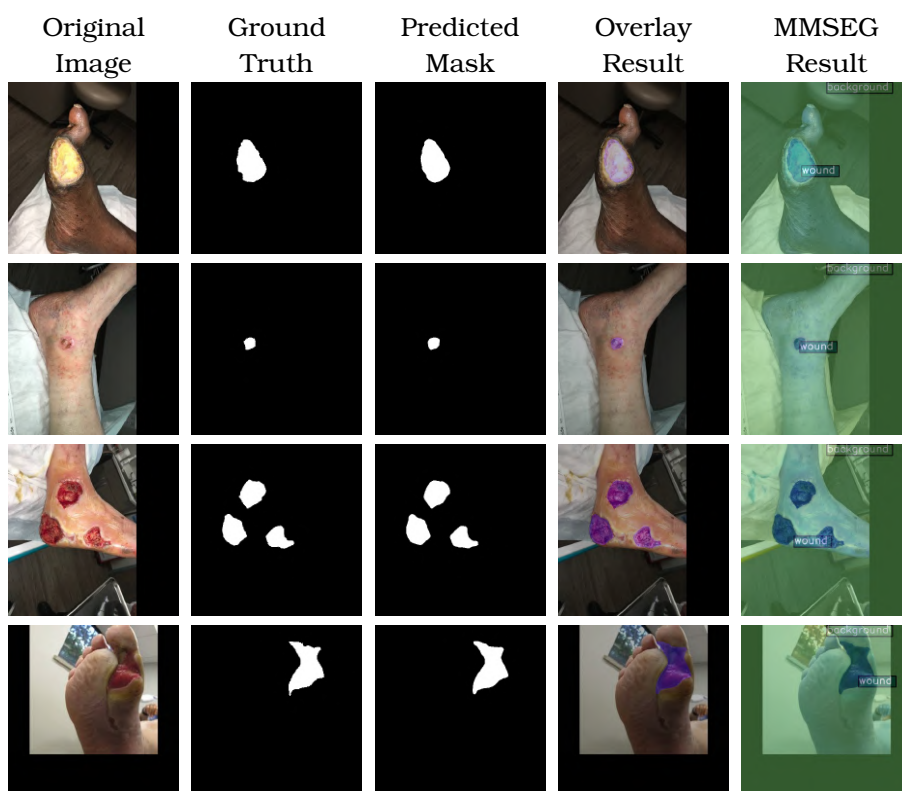
Στο σύνολο δεδομένων FUSeg οι μετρικές Dice και IoU από το πίνακα 6.2 δείχνουν ότι τα μοντέλα τμηματοποίησης μας φέρνουν καλύτερα αποτελέσματα σε σχέση με τα υπόλοιπα σύνολα δεδομένων. Το πιο αποδοτικό μοντέλο είναι ο Segformer ενώ αυτό με την χαμηλότερη απόδοση είναι το Unet. Οπτικά βλέπουμε ότι οι παραγόμενες μάσκες τμηματοποίησης είναι όμοιες με τις ground truth μάσκες από αρχικό σύνολο δεδομένων.



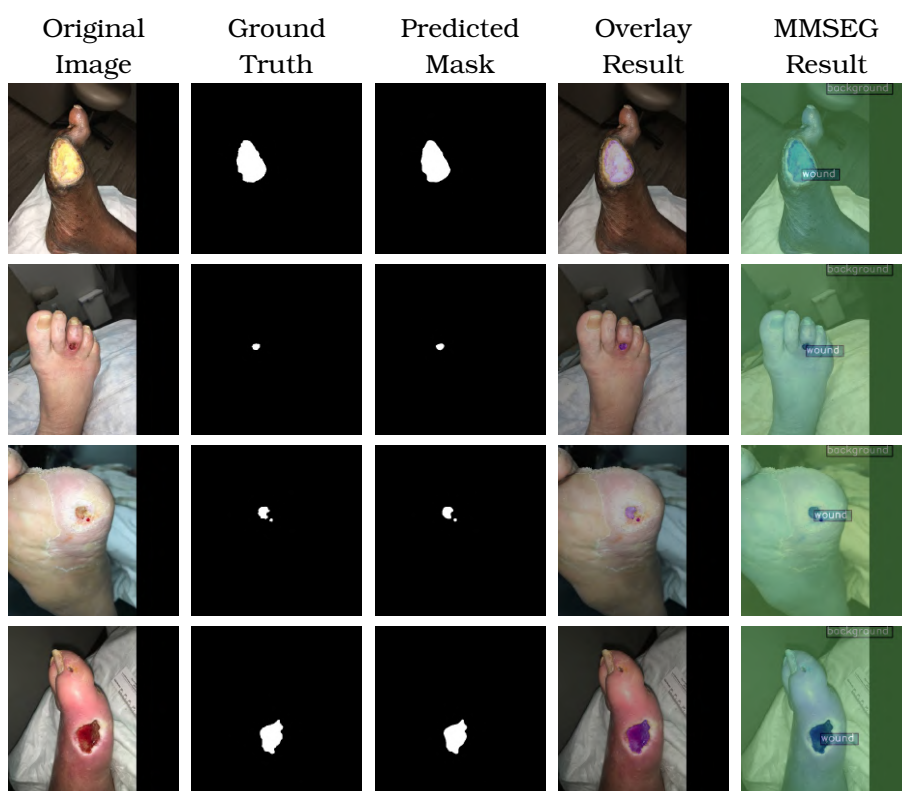
Εικόνα 6.7: Αποτελέσματα για το μοντέλο τμηματοποίησης Segformer στο σύνολο δεδομένων FUSeg.



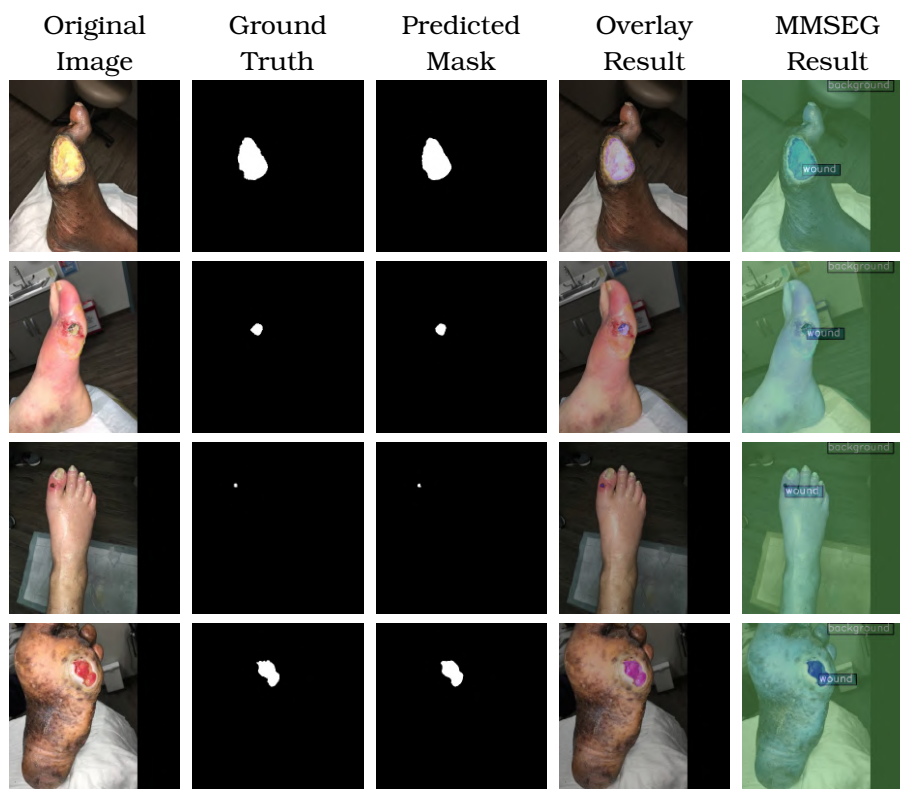
Εικόνα 6.8: Αποτελέσματα για το μοντέλο τμηματοποίησης Swin στο σύνολο δεδομένων FUSeg.



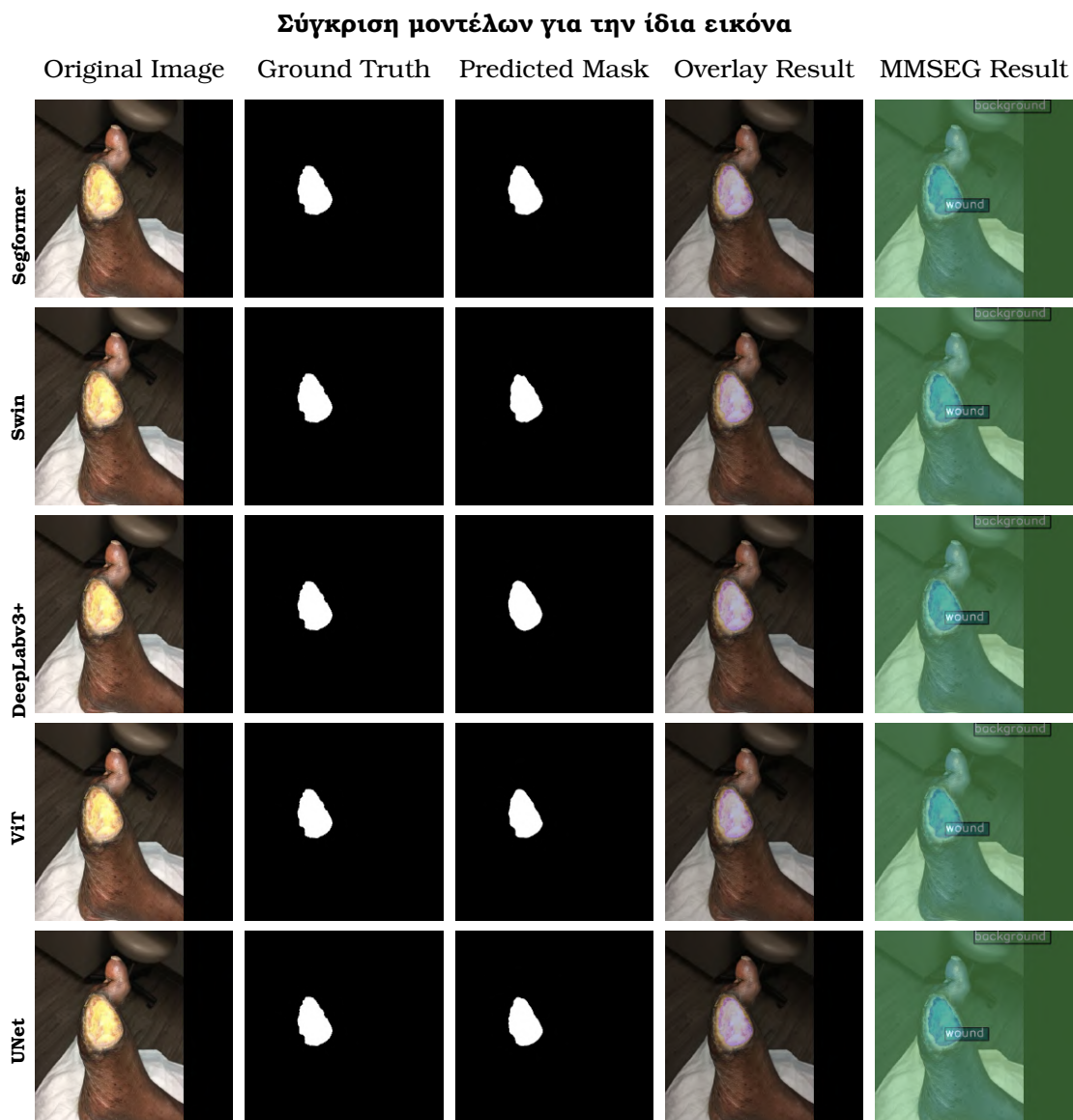
Εικόνα 6.9: Αποτελέσματα για το μοντέλο τμηματοποίησης *DeerLabV3+* στο σύνολο δεδομένων *FUSeg*.



Εικόνα 6.10: Αποτελέσματα για το μοντέλο τμηματοποίησης *ViT* στο σύνολο δεδομένων *FUSeg*.



Εικόνα 6.11: Αποτελέσματα για το μοντέλο τμηματοποίησης *Unet* στο σύνολο δεδομένων *FUSeg*.



Εικόνα 6.12: Συγκριτικά αποτελέσματα μοντέλων τμηματοποίησης στο σύνολο δεδομένων FUSeg.

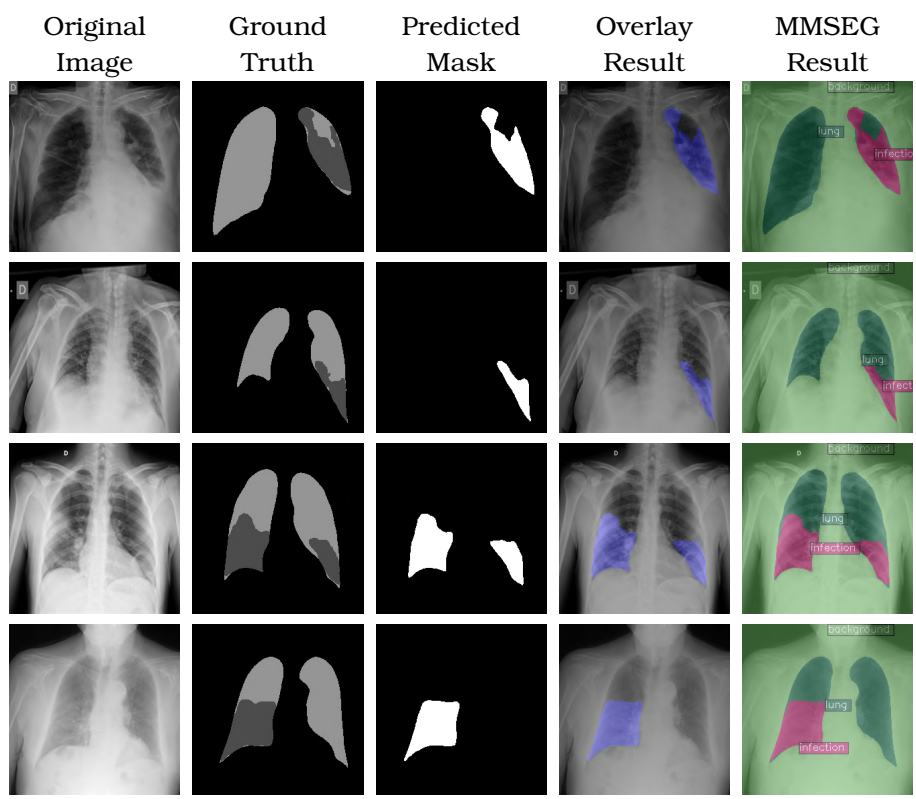
Πίνακας 6.2: Αποτελέσματα μεθόδων τμηματοποίησης βασισμένα στις μετρικές IoU και Dice για το σύνολο δεδομένων FUSeg

Method	IoU	Dice
Segformer	85.4	92.12
ViT	83.19	90.83
DeepLabv3+	83.6	91.07
Unet	75.94	86.32
Swin	83.2	90.84

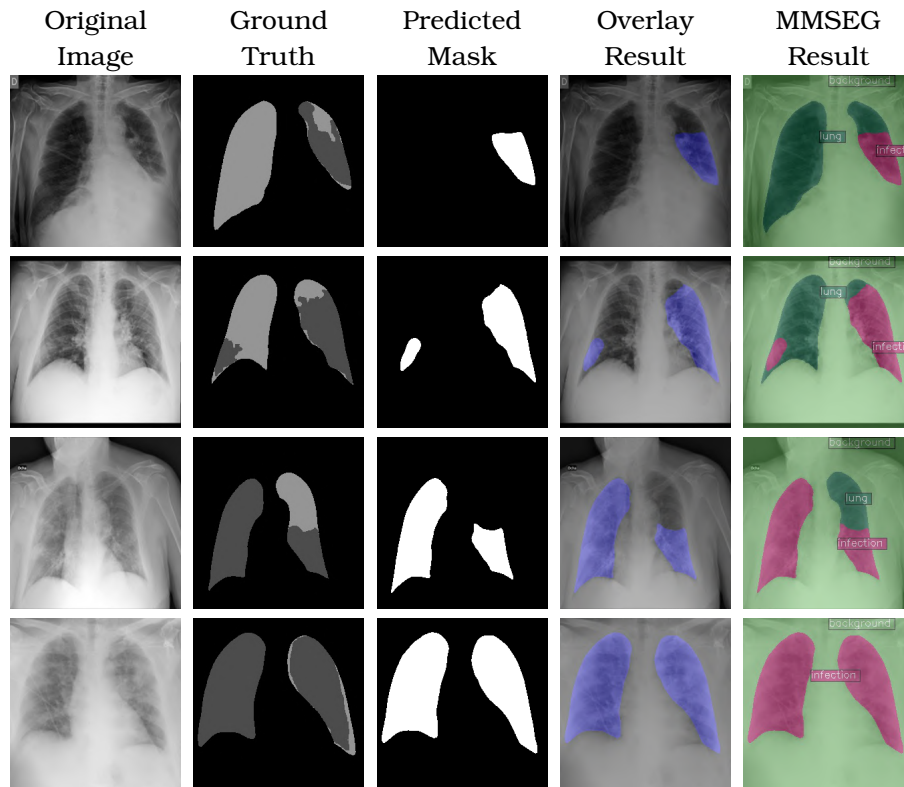
6.2.3 Αποτελέσματα για το σύνολο δεδομένων COVID-QU-Ex

Το σύνολο δεδομένων COVID-QU-Ex έχει τα λιγότερα βέλτιστα αποτελέσματα οπτικά σε σχέση με τα άλλα δύο σύνολα δεδομένων. Η φύση του προβλήματος λόγω της ανάλυσης των

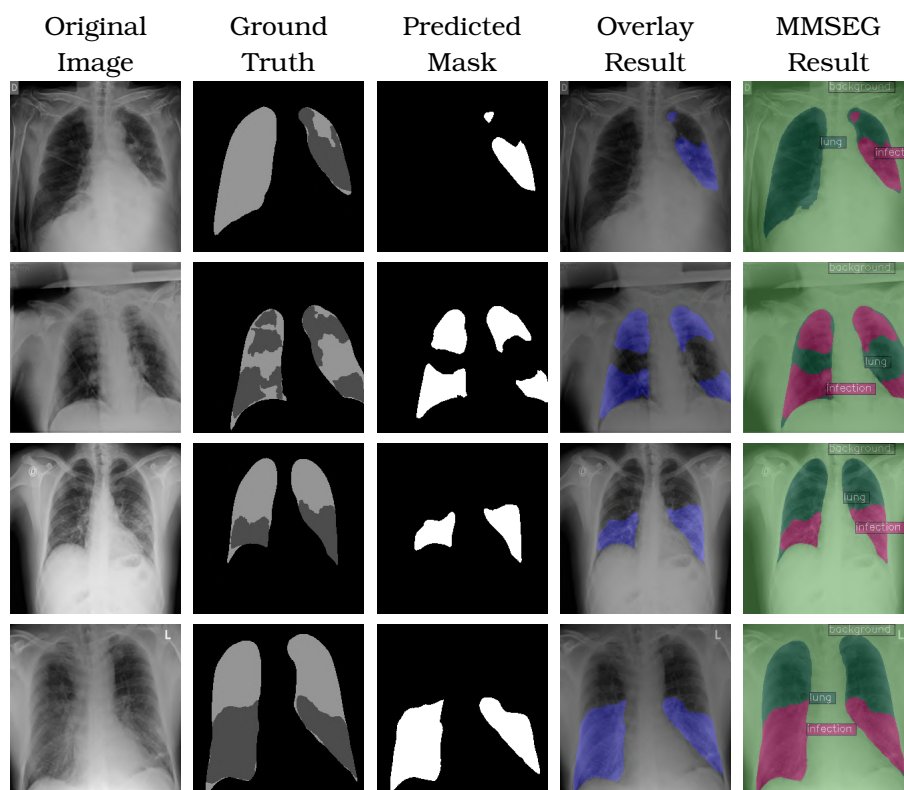
ακτινογραφιών καθώς και η ένταξη της κλάσης των πνευμόνων που καθιστά την εκπαίδευση πιο σύνθετη από άποψη πολυπλοκότητας, αποτελούν μερικές από τις αιτίες που τα οπτικά αποτελέσματα δεν είναι τόσο αποδοτικά όσο στα άλλα δύο σύνολα δεδομένων. Το καλύτερο μοντέλο τμηματοποίησης, με βάση τις μετρικές από το πίνακα 6.3 για το COVID-QU-Ex είναι το Swin.



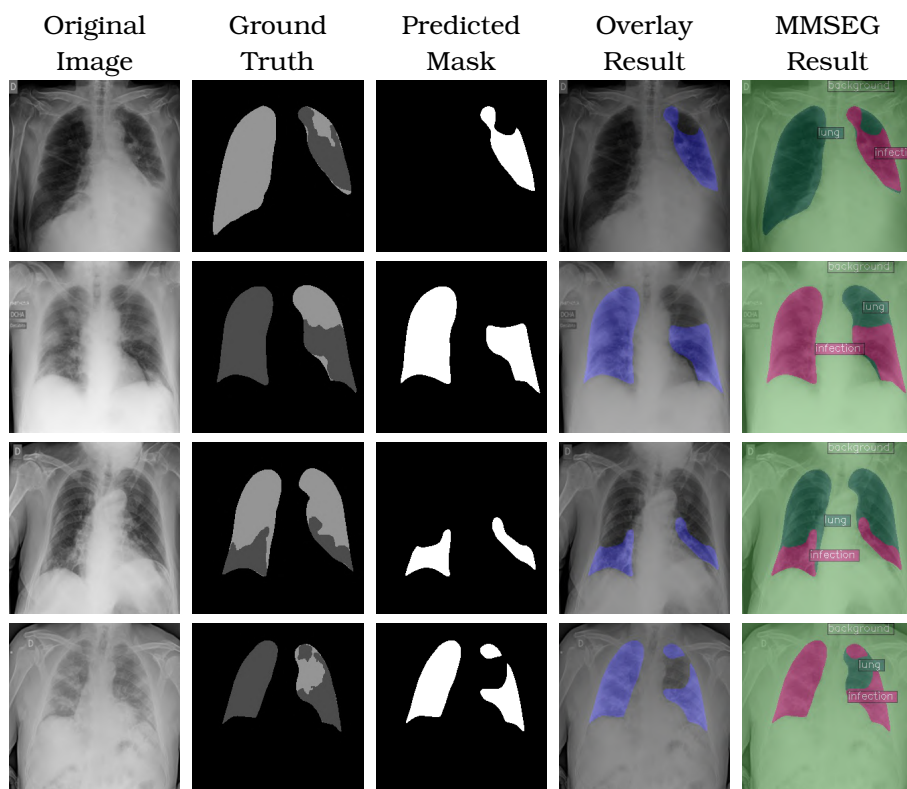
Εικόνα 6.13: Αποτελέσματα για το μοντέλο τμηματοποίησης Segformer στο σύνολο δεδομένων COVID-QU-Ex.



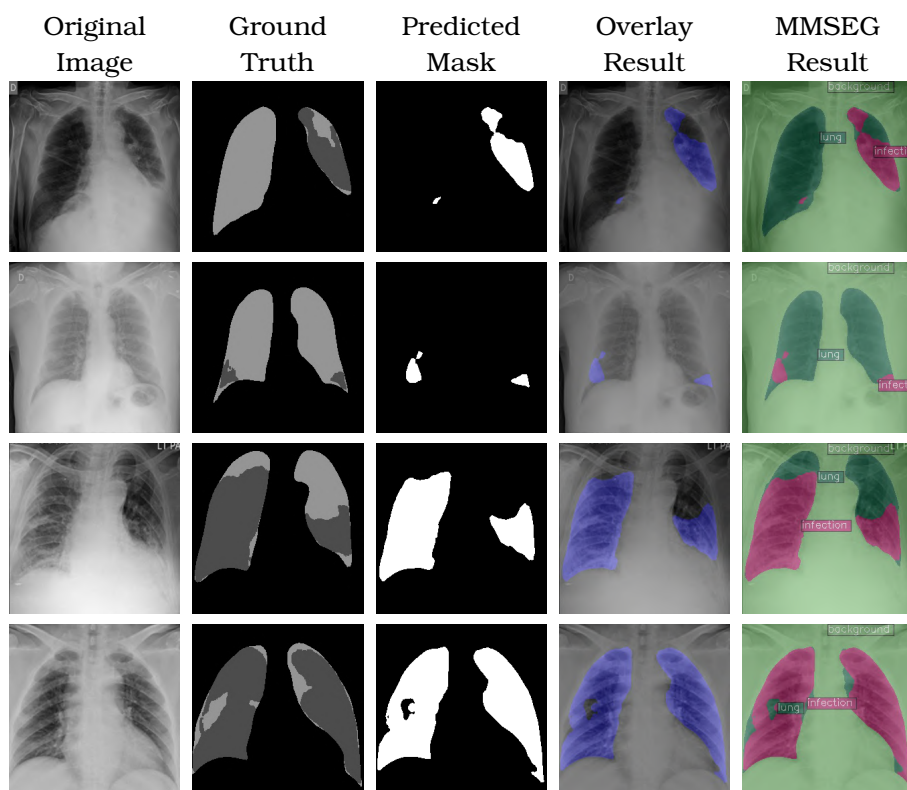
Εικόνα 6.14: Αποτελέσματα για το μοντέλο τμηματοποίησης Swin στο σύνολο δεδομένων COVID-QU-Ex.



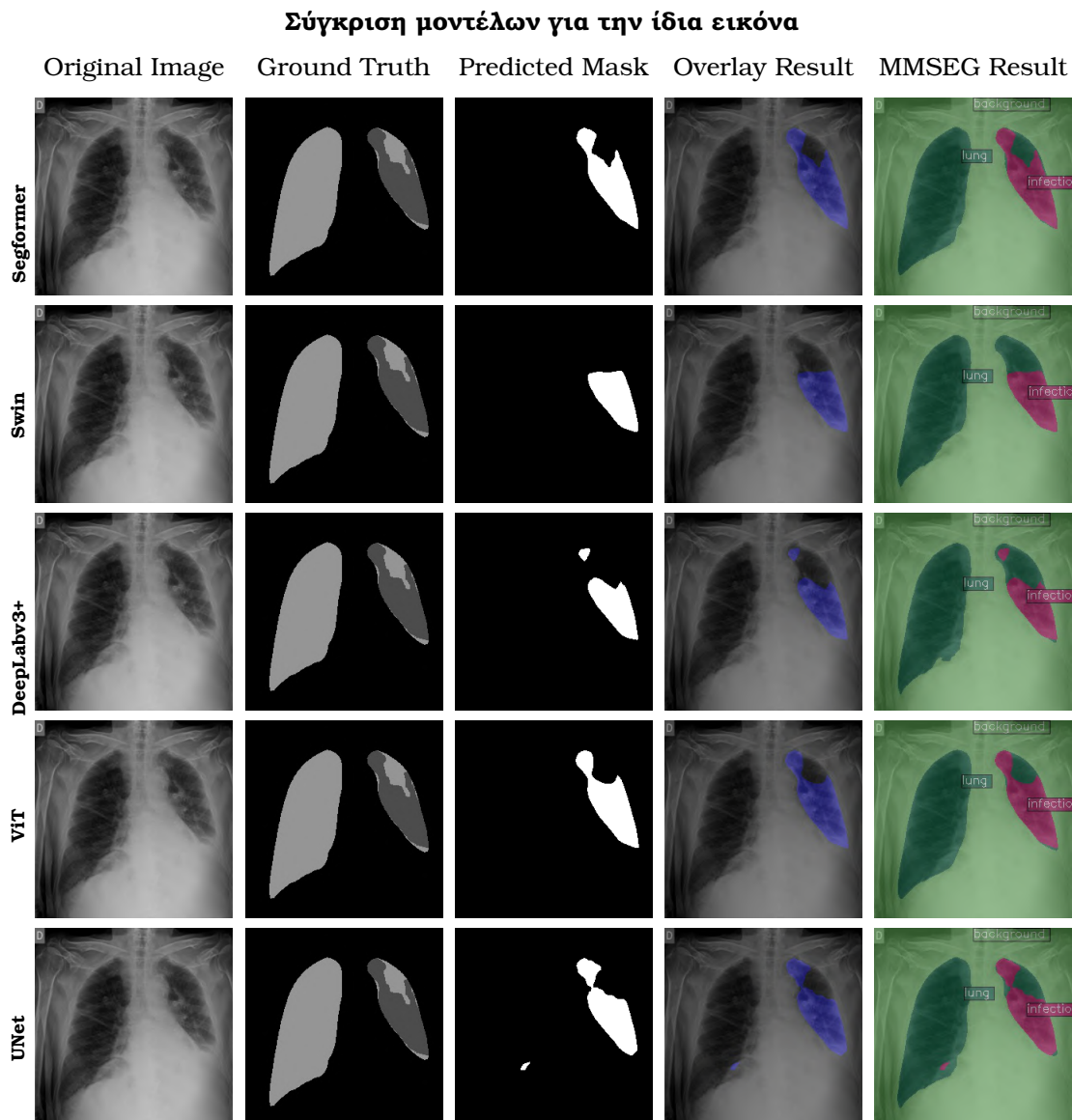
Εικόνα 6.15: Αποτελέσματα για το μοντέλο τμηματοποίησης DeepLabV3+ στο σύνολο δεδομένων COVID-QU-Ex.



Εικόνα 6.16: Αποτελέσματα για το μοντέλο τμηματοποίησης ViT στο σύνολο δεδομένων COVID-QU-Ex.



Εικόνα 6.17: Αποτελέσματα για το μοντέλο τμηματοποίησης Unet στο σύνολο δεδομένων COVID-QU-Ex.



Εικόνα 6.18: Συγκεντρωτικά αποτελέσματα μοντέλων τμηματοποίησης στο σύνολο δεδομένων COVID-QU-Ex.

Πίνακας 6.3: Αποτελέσματα μεθόδων τμηματοποίησης βασισμένα στις μετρικές IoU και Dice για το σύνολο δεδομένων COVID-QU-Ex

Method	IoU	Dice
Segformer	74.87	85.63
ViT	75.84	86.26
DeepLabv3+	75.16	85.82
Unet	70.48	82.68
Swin	76.56	86.72

Κεφάλαιο 7

Πειραματικά αποτελέσματα μεθόδων επεξηγησιμότητας

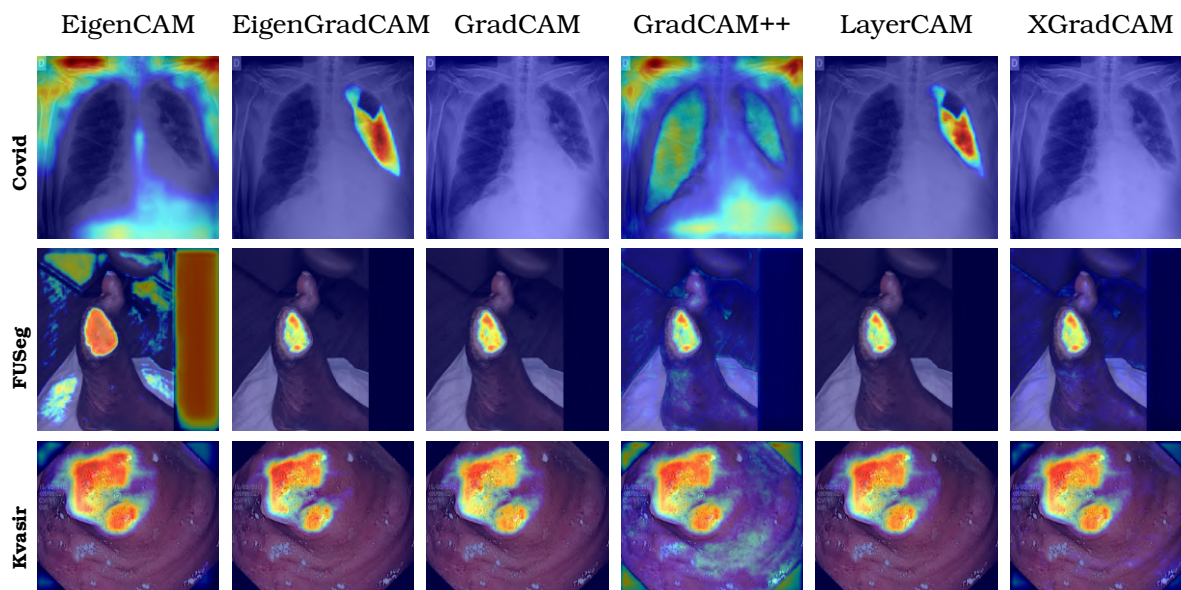
Στην συγκεκριμένη ενότητα θα παρουσιαστούν τα παραδείγματα μεθόδων επεξηγησιμότητας από τα τρία σύνολα δεδομένα που εκπαιδεύτηκαν τα μοντέλα τμηματοποίησης.

7.1 Πειραματικά αποτελέσματα τεχνικών επεξηγησιμότητας

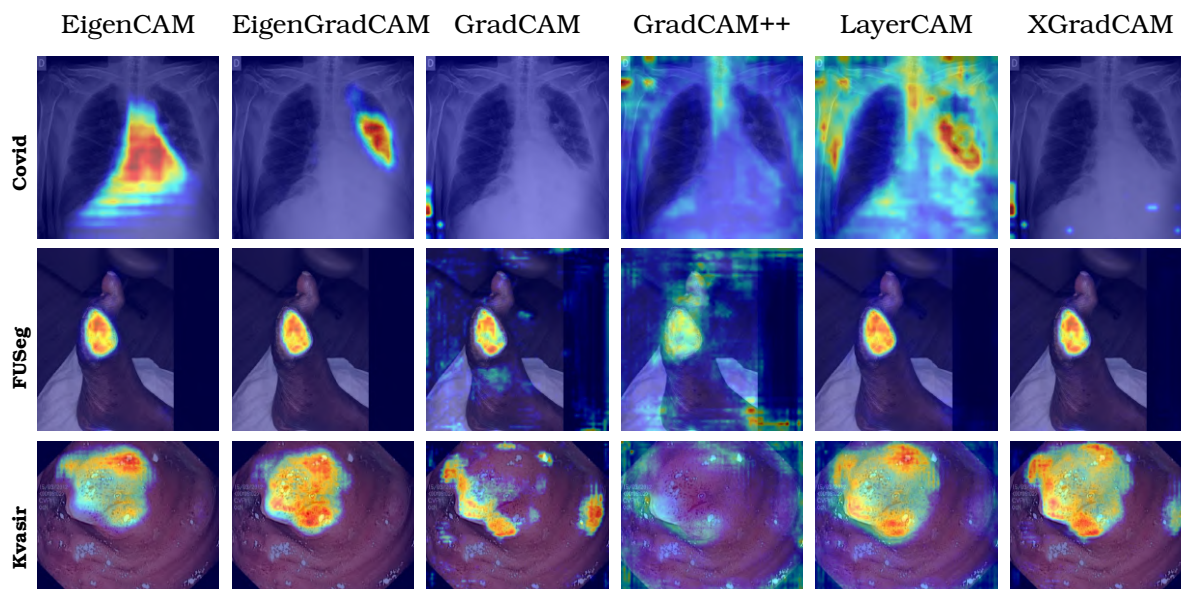
Για την υλοποίηση των τεχνικών επεξηγησιμότητας επιλέχθηκε το τελευταίο layer, αφού δοκιμάστηκαν και τα υπόλοιπα, αφού αποδείχθηκε το πιο αποδοτικό. Στα μοντέλα τμηματοποίησης Swin, ViT επειδή δομούνται από self attention layers και όχι από convolutional, που είναι απαραίτητα για την υλοποίηση αυτών των τεχνικών, χρησιμοποιήθηκε το προτελευταίο layer από την κεφαλή κατονοώντας και πάλι έτσι την εσωτερική λειτουργία αυτών των τεχνικών. Η υλοποίηση των τεχνικών επεξηγησιμότητας της οικογένειας Grad-CAM φαίνεται παρακάτω και οι εικόνες που έχουν επιλεχθεί, είναι οι εικόνες που δείχνουν την σύγκριση μεταξύ όλων των μοντέλων τμηματοποίησης.

Τα αποτελέσματα στο σύνολο δεδομένων COVID-QU-Ex παρατηρούμε ότι δεν επικεντρώνονται σε όλες τις τεχνικές στην επιθυμητή περιοχή της έκτασης του covid, γεγονός που οφείλεται στην πολυπλοκότητα του προβλήματος και στην φύση του λόγο ανάλυσης των ακτινογραφιών. Επίσης οι τεχνικές EigenCAM και EigenGradCAM, κυρίως στο μοντέλο ViT, επικεντρώνονται στο φόντο (background) και όχι στην κλάση. Πιθανά αίτια αποτελούν ο τρόπος λειτουργίας τους καθώς αναγνωρίζουν ότι αυτά είναι τα dominant features και επιστρέφουν μόνο το first principle vector ή τα RGB channels έχουν οριστεί ανάποδα.

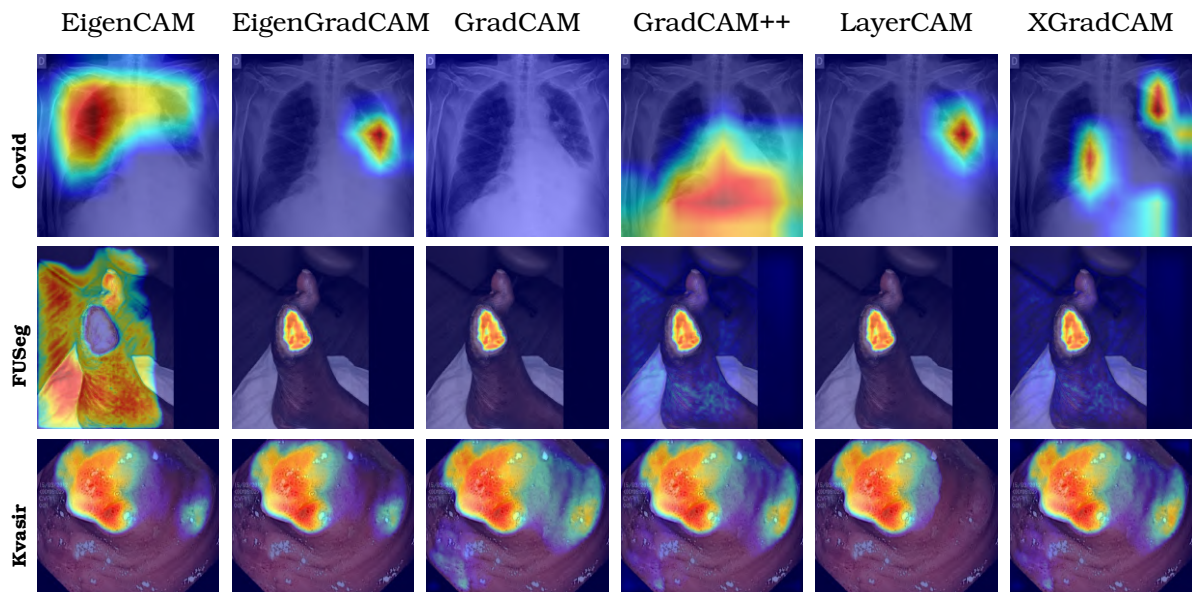
Στο μοντέλο τμηματοποίησης Unet χρησιμοποιήθηκαν 4 τεχνικές. Οι τεχνικές EigenCAM και EigenGradCAM δεν υλοποιήθηκαν καθώς δεσμεύουν περισσότερη μνήμη από αυτήν που διατίθεται στην πλατφόρμα του Kaggle.



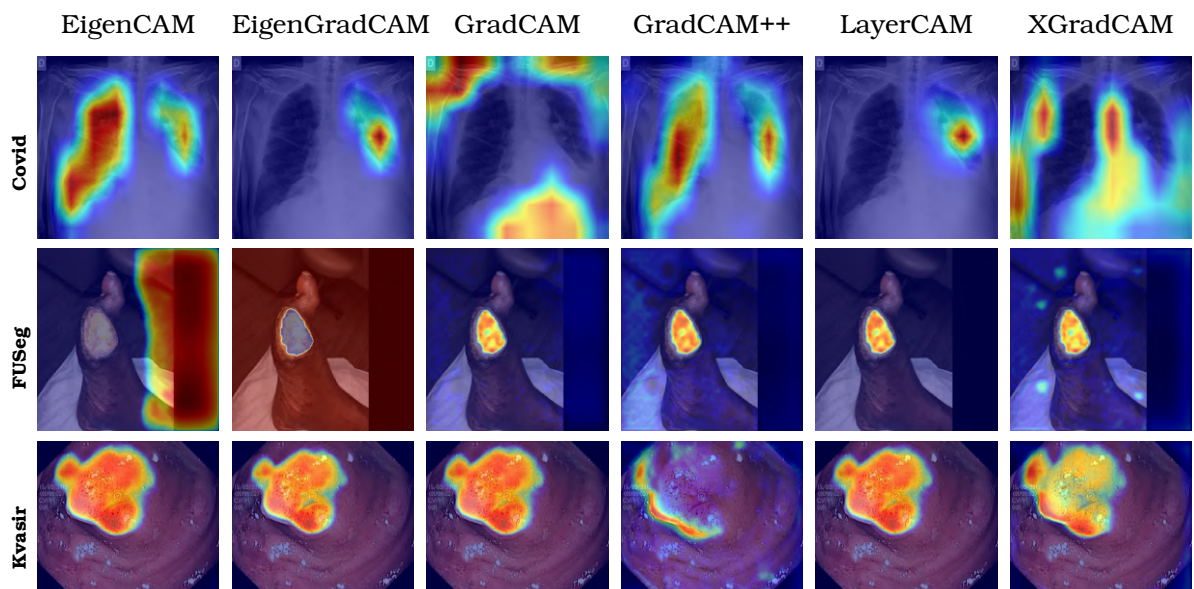
Εικόνα 7.1: Οπτικά αποτελέσματα τεχνικών επεξηγησιμότητας για το μοντέλο τμηματοποίησης SEGFORMER.



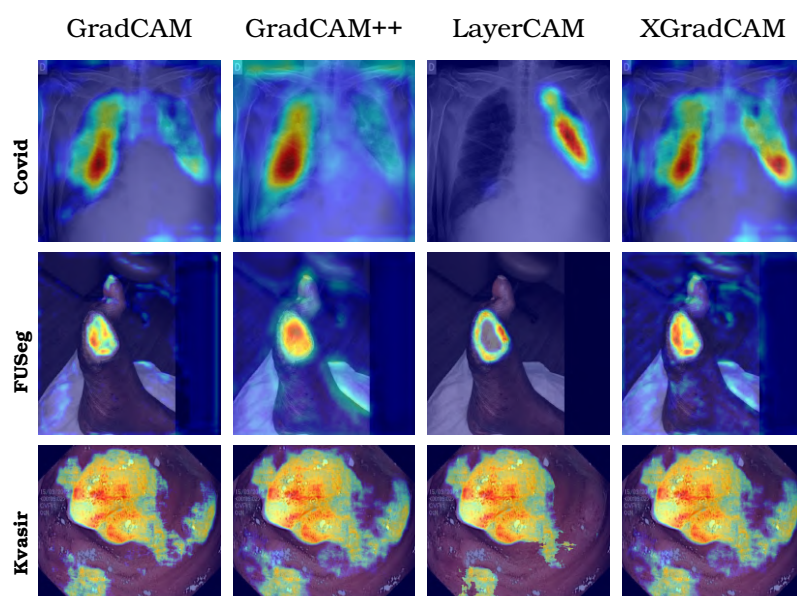
Εικόνα 7.2: Οπτικά αποτελέσματα τεχνικών επεξηγησιμότητας για το μοντέλο τμηματοποίησης DeepLabV3+.



Εικόνα 7.3: Οπτικά αποτελέσματα τεχνικών επεξηγησιμότητας για το μοντέλο τμηματοποίησης Swin.



Εικόνα 7.4: Οπτικά αποτελέσματα τεχνικών επεξηγησιμότητας για το μοντέλο τμηματοποίησης ViT.



Εικόνα 7.5: Οπτικά αποτελέσματα τεχνικών επεξηγησιμότητας για το μοντέλο τμηματοποίησης UNet.

Μέρος **III**

Επίλογος

Επίλογος

8.1 Συμπεράσματα

Στα πλαίσια της παρούσας διπλωματικής εξετάστηκαν οι αποδόσεις 5 μοντέλων τμηματοποίησης και 6 μεθόδων επεξηγησιμότητας. Παρατηρούμε ότι οι transformers που έχουν χρησιμοποιηθεί παρουσιάζουν καλύτερα αποτελέσματα στα σύνολα δεδομένων μας σε σχέση με τα υπόλοιπα μοντέλα και πιο συγκεκριμένα ο Segformer στα σύνολα δεδομένων που αφορούν τους πολύποδες (kvasir-SEG) και τα τράυματα (FUSeG) ενώ ο Swin όσον αφορά το σύνολο δεδομένων του Covid (Covid-QU-Ex). Η καλή απόδοση αυτών των μοντέλων επιβεβαιώνεται και οπτικά από τα παραδείγματα που έχουν δοθεί. Το μοντέλο Unet δίνει τα χειρότερα αποτελέσματα και στα τρία σύνολα δεδομένων, γεγονός που οφείλεται και στην αρχιτεκτονική του σε σχέση με τα υπόλοιπα προηγμένα μοντέλα. Τα ισχυρά μοντέλα παρουσιάζουν σχεδόν ίδιες μάσκες με αυτές που έχουν δημιουργηθεί από τους γιατρούς με εξαίρεση το σύνολο δεδομένων COVID-QU-Ex, στο οποίο η φύση των δεδομένων δεν επιτρέπεται τόσο ακρίβεια οπτικά.

Οι τεχνικές επεξηγησιμότητας που χρησιμοποιήθηκαν έδωσαν ικανοποιητικά αποτελέσματα για την εξήγηση των μοντέλων τμηματοποίησης που χρησιμοποιήθηκαν. Αυτές οι τεχνικές μας έδειξαν, μέσω των οπτικών αποτελεσμάτων, που βασίστηκαν οι αποφάσεις των μοντέλων. Έχουν παρόμοια αποτελέσματα και οι έξι τεχνικές της οικογένειας GradCAM στο σύνολο δεδομένων Kvasir-SEG και FUSeG. Τα λιγότερα καλά αποτελέσματα εμφανίζονται στο σύνολο δεδομένων COVID-QU-Ex, καθώς ο εντοπισμός της επιθυμητής περιοχής τμηματοποίησης αποδείχθηκε αρκετά απαιτητικός, γεγονός που οφείλεται στην πολυπλοκότητα του προβλήματος, καθώς περιέχεται και η κλάση των πνευμόνων και στα ίδια τα δεδομένα.

Τέλος, παρατηρούμε ότι ενώ τα μοντέλα μας έχουν υψηλά αποτελέσματα, γεγονός που φαίνεται και από τις μετρικές αξιολόγησης, αδυνατούν να αιχμαλωτήσουν κάποιες λεπτομέρειες όπως οι αιχμηρές γωνίες κάποιων ground truth μασκών.

8.2 Μελλοντικές επεκτάσεις

Στην παρούσα διπλωματική εργασία χρησιμοποιήθηκαν αρκετά μοντέλα τμηματοποίησης με διαφορετικές αρχιτεκτονικές και έξι τεχνικών επεξηγησιμότητας σε τρία σύνολα δεδομένων, αντιπροσωπεύοντας έτσι επαρκώς την ερευνητική περιοχή του αντικειμένου της εργασίας. Ωστόσο υπάρχουν πολλές επεκτάσεις που μπορούν να υλοποιηθούν. Μια από τις

πιθανές επεκτάσεις της παρούσας διπλωματικής αποτελεί η δοκιμή περισσότερων τεχνικών επεξηγησιμότητας. Πιο συγκεκριμένα θα μπορούσαν να χρησιμοποιηθούν οι τεχνικές LIME και SHAP καθώς και τεχνικές που αφορούν τα attention layers επειδή οι τεχνικές της οικογένειας Grad-CAM χρειάζονται convolutional layers για να πραγματοποιηθούν. Επίσης η υλοποίηση και άλλων μοντέλων τμηματοποίησης ή η δοκιμή διάφορων υπερ-παραμέτρων και συναρτήσεων απώλειας στα ήδη υπάρχοντα μοντέλα θα έδιναν μία πιο ολοκληρωμένη εικόνα.

Βιβλιογραφία

- [1] Jonathan Long, Evan Shelhamer και Trevor Darrell. *Fully Convolutional Networks for Semantic Segmentation*, 2015.
- [2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser και Illia Polosukhin. *Attention is All you Need*. *Advances in Neural Information Processing Systems*. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan και R. Garnett, επιμελητές, τόμος 30. Curran Associates, Inc., 2017.
- [3] Anh Pham Thi Minh. *Overview of Class Activation Maps for Visualization Explainability*, 2023.
- [4] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh και Dhruv Batra. *Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization*. *International Journal of Computer Vision*, 128(2):336-359, 2019.
- [5] Aditya Chattopadhyay, Anirban Sarkar, Prantik Howlader και Vineeth N Balasubramanian. *Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks*. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018.
- [6] Ruigang Fu, Qingyong Hu, Xiaohu Dong, Yulan Guo, Yinghui Gao και Biao Li. *Axiom-based Grad-CAM: Towards Accurate Visualization and Explanation of CNNs*, 2020.
- [7] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez και Ping Luo. *SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers*, 2021.
- [8] Liang Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff και Hartwig Adam. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*, 2018.
- [9] Liang Chieh Chen, George Papandreou, Florian Schroff και Hartwig Adam. *Rethinking Atrous Convolution for Semantic Image Segmentation*, 2017.
- [10] Olaf Ronneberger, Philipp Fischer και Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*, 2015.

- [11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit και Neil Houlsby. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*, 2021.
- [12] Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang και Jian Sun. *Unified Perceptual Parsing for Scene Understanding*, 2018.
- [13] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin και Baining Guo. *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*, 2021.
- [14] Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Pål Halvorsen, Thomasde Lange, Dag Johansen και Håvard D. Johansen. *Kvasir-SEG: A Segmented Polyp Dataset*, 2019.
- [15] Universityof Wisconsin-Milwaukee Big Data Group. *Foot Ulcer Segmentation Challenge Dataset*. <https://github.com/uwm-bigdata/wound-segmentation/tree/master/data/Foot%20Ulcer%20Segmentation%20Challenge>, 2021. Αρχειοθετημένο: 2024-06-30.
- [16] Georgios Sapountzakis, Paraskevi Antonia Theofilou και Paraskevi Tzouveli. *Covid-19 Detection From X-Rays Images Using Deep Learning Methods*. *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, σελίδες 1–5, 2023.
- [17] Natalia Salpea, Paraskevi Tzouveli και Dimitrios Kollias. *Medical Image Segmentation: A Review of Modern Architectures*, σελίδες 691–708. 2023.
- [18] Dimitrios Kollias, Athanasios Tagaris, Andreas Stafylopatis, Stefanos Kollias και Georgios Tagaris. *Deep neural architectures for prediction in healthcare*. *Complex & Intelligent Systems*, 4(2):119–131, 2018.
- [19] Athanasios Tagaris, Dimitrios Kollias και Andreas Stafylopatis. *Assessment of Parkinson’s disease based on deep neural networks*. *International Conference on Engineering Applications of Neural Networks*, σελίδες 391–403. Springer, 2017.
- [20] Athanasios Tagaris, Dimitrios Kollias, Andreas Stafylopatis, Georgios Tagaris και Stefanos Kollias. *Machine learning for neurodegenerative disorder diagnosis—survey of practices and launch of benchmark dataset*. *International Journal on Artificial Intelligence Tools*, 27(03):1850011, 2018.
- [21] Ilianna Kollia, Andreas Georgios Stafylopatis και Stefanos Kollias. *Predicting Parkinson’s disease using latent information extracted from deep neural networks*. *2019 International Joint Conference on Neural Networks (IJCNN)*, σελίδες 1–8. IEEE, 2019.
- [22] James Wingate, Ilianna Kollia, Luc Bidaut και Stefanos Kollias. *Unified deep learning approach for prediction of Parkinson’s disease*. *IET Image Processing*, 14(10):1980–1989, 2020.

- [23] Dimitrios Kollias, Anastasios Arsenos, Levon Soukissian και Stefanos Kollias. *Miacov19d: Covid-19 detection through 3-d chest ct image analysis*. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, σελίδες 537–544, 2021.
- [24] Anastasios Arsenos, Dimitrios Kollias και Stefanos Kollias. *A Large Imaging Database and Novel Deep Neural Architecture for Covid-19 Diagnosis*. *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, σελίδες 1–5. IEEE, 2022.
- [25] Dimitrios Kollias, Anastasios Arsenos και Stefanos Kollias. *AI-MIA: Covid-19 detection and severity analysis through medical imaging*. *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, σελίδες 677–690. Springer, 2023.
- [26] Anastasios Arsenos, Andjoli Davidhi, Dimitrios Kollias, Panos Prassopoulos και Stefanos Kollias. *Data-Driven Covid-19 Detection Through Medical Imaging*. *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (I-CASSPW)*, σελίδες 1–5, 2023.
- [27] Dimitrios Kollias, Anastasios Arsenos και Stefanos Kollias. *Domain Adaptation Explainability & Fairness in AI for Medical Image Analysis: Diagnosis of COVID-19 based on 3-D Chest CT-scans*. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, σελίδες 4907–4914, 2024.
- [28] Dimitrios Kollias, Miao Yu, Athanasios Tagaris, Georgios Leontidis, Andreas Stafylopatis και Stefanos Kollias. *Adaptation and contextualization of deep neural network models*. *2017 IEEE symposium series on computational intelligence (SSCI)*, σελίδες 1–8. IEEE.
- [29] D Kollias, N Bouas, Y Vlaxos, V Brillakis, M Seferis, I Kollia, L Sukissian, J Wingate και S Kollias. *Deep Transparent Prediction through Latent Representation Analysis*. *arXiv preprint arXiv:2009.07044*, 2020.
- [30] Dimitrios Kollias, Anastasios Arsenos και Stefanos Kollias. *A deep neural architecture for harmonizing 3-D input data analysis and decision making in medical imaging*. *Neurocomputing*, 542:126244, 2023.
- [31] Dimitris Kollias, Y Vlaxos, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate και S Kollias. *Transparent adaptation in deep medical image diagnosis*. *International Workshop on the Foundations of Trustworthy AI Integrating Learning, Optimization and Reasoning*, σελίδες 251–267. Springer, 2020.
- [32] Fabio De Sousa Ribeiro, Francesco Caliv, Mark Swainson, Kjartan Gudmundsson, Georgios Leontidis και Stefanos Kollias. *Deep bayesian self-training*. *Neural Computing and Applications*, 32(9):4275–4291, 2020.

- [33] Fabio De Sousa Ribeiro, Georgios Leontidis και Stefanos Kollias. *Capsule routing via variational bayes*. *Proceedings of the AAAI Conference on Artificial Intelligence*, τόμος 34, σελίδες 3749–3756, 2020.
- [34] Fabio De Sousa Ribeiro, Georgios Leontidis και Stefanos Kollias. *Introducing routing uncertainty in capsule networks*. *Advances in Neural Information Processing Systems*, 33:6490–6502, 2020.
- [35] Francesco Caliva, Fabio Sousa De Ribeiro, Antonios Mylonakis, Christophe Demazière, Paolo Vinai, Georgios Leontidis και Stefanos Kollias. *A deep learning approach to anomaly detection in nuclear reactors*. *2018 International joint conference on neural networks (IJCNN)*, σελίδες 1–8. IEEE, 2018.
- [36] Stefanos Kollias, Miao Yu, James Wingate, Aiden Durrant, Georgios Leontidis, Georgios Alexandridis, Andreas Stafylopatis, Antonios Mylonakis, Paolo Vinai και Christophe Demaziere. *Machine learning for analysis of real nuclear plant data in the frequency domain*. *Annals of Nuclear Energy*, 177:109293, 2022.
- [37] Bashar Alhnaity, Stefanos Kollias, Georgios Leontidis, Shouyong Jiang, Bert Schamp και Simon Pearson. *An autoencoder wavelet based deep neural network with attention mechanism for multi-step prediction of plant growth*. *Information Sciences*, 560:35–50, 2021.
- [38] Bashar Alhnaity, Simon Pearson, Georgios Leontidis και Stefanos Kollias. *Using deep learning to predict plant growth and yield in greenhouse environments*. *International Symposium on Advanced Technologies and Management for Innovative Greenhouses: GreenSys2019 1296*, σελίδες 425–432, 2019.
- [39] Andreas Psaroudakis και Dimitrios Kollias. *MixAugment & Mixup: Augmentation Methods for Facial Expression Recognition*. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, σελίδες 2367–2375, 2022.
- [40] Dimitrios Kollias, Stefanos Zafeiriou, Irene Kotsia, Abhinav Dhall, Shreya Ghosh, Chunchang Shao και Guanyu Hu. *7th abaw competition: Multi-task learning and compound expression recognition*. *arXiv preprint arXiv:2407.03835*, 2024.
- [41] Dimitrios Kollias, Viktoriia Sharmanska και Stefanos Zafeiriou. *Distribution matching for multi-task learning of classification tasks: a large-scale study on faces & beyond*. *Proceedings of the AAAI Conference on Artificial Intelligence*, τόμος 38, σελίδες 2813–2821, 2024.
- [42] *Face detection in color images and video sequences*. *2000 10th Mediterranean Electrotechnical Conference. Information Technology and Electrotechnology for the Mediterranean Countries. Proceedings. MeleCon 2000 (Cat. No. 00CH37099)*, τόμος 2, σελίδες 498–502. IEEE, 2000.

- [43] Phivos Mylonas, Evaggelos Spyrou, Yannis Avrithis και Stefanos Kollias. *Using visual context and region semantics for high-level concept detection*. *IEEE Transactions on Multimedia*, 11(2):229-243, 2009.
- [44] Th Athanasiadis. *Image indexing and retrieval using expressive fuzzy description logics*. *Signal, Image and Video Processing*, 2:321-335, 2008.
- [45] Yannis Avrithis, Nicolas Tsapatsoulis και Stefanos Kollias. *Color-based retrieval of facial images*. *2000 10th European Signal Processing Conference*, σελίδες 1-4. IEEE, 2000.
- [46] Manolis Wallace, Ilias Maglogiannis, Kostas Karpouzis, George Kormentzas και Stefanos Kollias. *Intelligent one-stop-shop travel recommendations using an adaptive neural network and clustering of history*. *Information Technology & Tourism*, 6(3):181-193, 2003.
- [47] G Caridakis, A Raouzaïou, K Karpouzis και S Kollias. *Synthesizing gesture expressivity based on real sequences*.
- [48] *Optimal filter banks for signal reconstruction from noisy subband components*. *IEEE transactions on signal processing*, 44(2):212-224, 1996.
- [49] Hyungsik Jung και Youngrock Oh. *Towards Better Explanations of Class Activation Mapping*, 2021.
- [50] MMSegmentation Contributors. *MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark*. <https://github.com/open-mmlab/msegmentation>, 2020.
- [51] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Paal Halvorsen, Thomasde Lange, Dag Johansen και Haavard D Johansen. *Kvasir-seg: A segmented polyp dataset*. *International Conference on Multimedia Modeling*, σελίδες 451-462. Springer, 2020.
- [52] Anas M. Tahir, Muhammad E. H. Chowdhury, Yazan Qiblawey, Amith Khandakar, Tawsifur Rahman, Serkan Kiranyaz, Uzair Khurshid, Nabil Ibte haz, Sakib Mahmud και Maymouna Ezeddin. *COVID-QU-Ex*. Kaggle, 2021.
- [53] *QaTa-COV19 Database*. <https://www.kaggle.com/aysendegerli/qatacov19-dataset>, 2021. Accessed 14 March 2021.
- [54] *Covid-19-image-repository*. <https://github.com/ml-workgroup/covid-19-image-repository/tree/master/png>, 2021. Accessed 14 March 2021.
- [55] *Eurorad*. <https://www.eurorad.org/>, 2021. Accessed 14 March 2021.
- [56] *Covid-chestxray-dataset*. <https://github.com/ieee8023/covid-chestxray-dataset>, 2021. Accessed 14 March 2021.
- [57] *COVID-19 DATABASE*. <https://www.sirm.org/category/senza-categoria/covid-19/>, 2021. Accessed 14 March 2021.

- [58] Kaggle. *COVID-19 Radiography Database*. <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>, 2020. Accessed 14 March 2021.
- [59] GitHub. *COVID-CXNet*. <https://github.com/armiro/COVID-CXNet>, 2020. Accessed 14 March 2021.
- [60] *RSNA Pneumonia Detection Challenge*. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data>, 2021. Accessed 14 March 2021.
- [61] *Chest X-Ray Images (Pneumonia)*. <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>, 2021. Accessed 14 March 2021.
- [62] *Medical Imaging Databank of the Valencia Region. PadChest: A large chest x-ray image dataset with multi-label annotated reports*. <https://bimcv.cipf.es/bimcv-projects/padchest/>, 2021. Accessed 14 March 2021.
- [63] C. Wang, D.M. Anisuzzaman, V. Williamson και others. *Fully automatic wound segmentation with deep convolutional neural networks*. *Scientific Reports*, 10:21897, 2020.
- [64] Dominik Müller, Iñaki Soto-Rey και Frank Kramer. *Towards a Guideline for Evaluation Metrics in Medical Image Segmentation*, 2022.