



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

**Ανάπτυξη και Αξιολόγηση Μεθόδου Βασισμένης στη
Διαδραστική Μηχανική Μάθηση για την Κατάτμηση
Συμπαγών Όγκων και Οργάνων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Δημήτριος Α. Μπούνιας

Επιβλέπουσα : Κωνσταντίνα Σ. Νικήτα
Καθηγήτρια Ε.Μ.Π.

Αθήνα, Οκτώβριος 2020

Η σελίδα αυτή είναι σκόπιμα λευκή.



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

**Ανάπτυξη και Αξιολόγηση Μεθόδου Βασισμένης στη
Διαδραστική Μηχανική Μάθηση για την Κατάτμηση
Συμπαγών Όγκων και Οργάνων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Δημήτριος Α. Μπούνιας

Επιβλέπουσα : Κωνσταντίνα Σ. Νικήτα
Καθηγήτρια Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή τη 13η Οκτωβρίου 2020.

.....
Κ. Νικήτα
Καθηγήτρια Ε.Μ.Π.

.....
Γ. Στάμου
Αν. Καθηγητής Ε.Μ.Π.

.....
Α. Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2020

.....

Δημήτριος Α. Μπούνιας

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

Copyright © Δημήτριος Μπούνιας, 2020

Με επιφύλαξη παντός δικαιώματος. All rights reserved

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή μέρους αυτής, για εμπορικό ή κερδοσκοπικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Ερωτήματα που αφορούν τη χρήση της εργασίας για εμπορικό-κερδοσκοπικό σκοπό πρέπει να απευθύνονται αποκλειστικά στους συγγραφείς.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτήν την εργασία εκφράζουν τους συγγραφείς και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου συμπεριλαμβανόμενων Σχολών, Τομέων και Μονάδων αυτού.

Περίληψη

Σε αυτήν την εργασία, παρουσιάζεται μία γενικού σκοπού, γρήγορη, ακριβής και συνεπής μέθοδος κατάτμησης ιατρικών εικόνων, βασισμένη στη διαδραστική μηχανική μάθηση (Interactive Machine Learning - IML).

Η μέθοδος χρησιμοποιεί σύντομα σχέδια του χρήστη και προσαρμοστικούς χάρτες γεωδαιτικής απόστασης για να εκπαιδεύσει ένα σύνολο από Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines), παρέχοντας έτσι ένα μοντέλο προσαρμοσμένο στον εκάστοτε ασθενή, το οποίο εφαρμόζεται για τη κατάτμηση των εικόνων της απεικονιστικής εξέτασης. Για την επικύρωση της μεθόδου, χρησιμοποιήθηκαν αναδρομικές μελέτες που περιελάμβαναν 20 ασθενείς με γλοιοβλάστωμα, 50 ασθενείς με καρκίνο του μαστού και 50 ασθενείς με καρκίνο του πνεύμονα, καθώς και 20 τομογραφικές απεικονίσεις σπληνός. Για κάθε απεικονιστική εξέταση ήταν διαθέσιμη σχετική κατάτμηση αναφοράς (ground truth segmentation). Για κάθε μελέτη, χρησιμοποιήθηκαν δύο ειδικοί ακτινολόγοι που πραγματοποίησαν κατάτμηση δύο φορές με την μέθοδο IML και δύο φορές με χειροκίνητο τρόπο. Η αξιολόγηση περιελάμβανε σύγκριση των δύο προσεγγίσεων ως προς την ταχύτητα, τη χωρική επικάλυψη και τη συνέπεια των δύο προσεγγίσεων. Η σημαντικότητα των διαφορών αξιολογήθηκε με χρήση του μη παραμετρικού Wilcoxon test προσημασμένων τάξεων και συσχετισμένων δειγμάτων.

Η μέθοδος IML ήταν ταχύτερη από τη χειροκίνητη κατάτμηση για όλες τις δομές, κατά 53,1%, κατά μέσο όρο. Βρέθηκε στατιστικά σημαντική ($p < 0,001$) διαφορά στην χωρική επικάλυψη σε σχέση με τα δεδομένα αναφοράς για σπλήνα ($Dice_{IML}/Dice_{Χειροκίνητου} = 0,91/0,87$), καρκίνο του μαστού ($Dice_{IML}/Dice_{Χειροκίνητου} = 0,84/0,82$) και πνευμονικούς όζους ($Dice_{IML}/Dice_{Χειροκίνητου} = 0,78/0,83$). Όσον αφορά την συνέπεια μεταξύ διαφορετικών επαναλήψεων του ίδιου ειδικού (intra-rater), βρέθηκε στατιστικά σημαντική ($p = 0,003$) διαφορά μόνο για τον σπλήνα ($Dice_{IML}/Dice_{Χειροκίνητου} = 0,91/0,89$), που ήταν ευνοϊκή για τη μέθοδο IML. Όσον αφορά την συνέπεια της κατάτμησης από διαφορετικούς ειδικούς (inter-rater), βρέθηκαν στατιστικά σημαντικές διαφορές ($p < 0,045$) για τον σπλήνα ($Dice_{IML}/Dice_{Χειροκίνητου} = 0,91/0,87$), τον μαστό ($Dice_{IML}/Dice_{Χειροκίνητου} = 0,86/0,81$), τον πνεύμονα ($Dice_{IML}/Dice_{Χειροκίνητου} = 0,85/0,89$), το περιοχικά μη-ενισχυμένο τμήμα του γλοιοβλαστώματος ($Dice_{IML}/Dice_{Χειροκίνητου} = 0,79/0,67$) και το περιοχικά ενισχυμένο τμήμα του γλοιοβλαστώματος ($Dice_{IML}/Dice_{Χειροκίνητου} = 0,79/0,84$), αποτελέσματα που αποδεικνύουν συνολική υπεροχή της μεθόδου IML σε σχέση με την χειροκίνητη κατάτμηση.

Η ποσοτική αξιολόγηση ως προς την ταχύτητα, τη χωρική επικάλυψη και τη συνέπεια, ανέδειξε την συνολική υπεροχή της προτεινόμενης μεθόδου σε σύγκριση με χειροκίνητη κατάτμηση, για ένα σημαντικό εύρος κλινικών εφαρμογών. Η υλοποίηση της μεθόδου IML είναι ανοικτού κώδικα και διατίθεται μέσω του Cancer Imaging Phenomics Toolkit (CaPTk - <https://www.cbica.upenn.edu/captk>) και ως πρόσθετο για το Medical Imaging Interaction Toolkit (MITK - <https://github.com/CBICA/InteractiveSegmentation>).

Λέξεις Κλειδιά

Μηχανική μάθηση, Μηχανές διανυσμάτων υποστήριξης, C++, Κατάτμηση εικόνας, Ιατρική εικόνα, Μαγνητική τομογραφία, Υπολογιστική τομογραφία, Γλοιοβλάστωμα, Καρκίνος του μαστού, Καρκίνος του πνεύμονα, Κατάτμηση σπληνός

Abstract

In this thesis, a fast, accurate, and consistent method for general-purpose medical image segmentation is presented. The method is based on interactive machine learning (IML).

Utilizing brief user training annotations and the adaptive geodesic distance transform, an ensemble of Support Vector Machines (SVMs) is trained, providing a patient-specific model applied to the whole image. To validate the method, retrospective cohorts were identified comprising 20 brain, 50 breast, and 50 lung cancer patients, as well as 20 spleen scans. Corresponding ground truth annotations were available for each scan. Two experts were recruited to segment each cohort twice with the IML method and twice manually. Evaluation endpoints included speed, spatial overlap, and consistency. Statistical significance was evaluated through the Wilcoxon signed-rank non-parametric test.

The IML method was faster than manual annotation in all structures, by 53.1% on average. A significant ($p < 0.001$) overlap difference was found for spleen ($Dice_{IML}/Dice_{Manual} = 0.91/0.87$), breast tumors ($Dice_{IML}/Dice_{Manual} = 0.84/0.82$), and lung nodules ($Dice_{IML}/Dice_{Manual} = 0.78/0.83$). For intra-rater consistency, a significant ($p = 0.003$) and favorable difference was found only for spleen ($Dice_{IML}/Dice_{Manual} = 0.91/0.89$). For inter-rater consistency significant ($p < 0.045$) differences were found for spleen ($Dice_{IML}/Dice_{Manual} = 0.91/0.87$), breast ($Dice_{IML}/Dice_{Manual} = 0.86/0.81$), lung ($Dice_{IML}/Dice_{Manual} = 0.85/0.89$), the non-enhancing brain tumor sub-region ($Dice_{IML}/Dice_{Manual} = 0.79/0.67$), and the enhancing brain tumor sub-region ($Dice_{IML}/Dice_{Manual} = 0.79/0.84$), which in aggregation favored the IML algorithm.

The quantitative evaluation for speed, spatial overlap, and consistency, revealed the benefits of the proposed IML method when compared with manual annotation, for several clinically relevant problems. The IML method is publicly released through the Cancer Imaging Phenomics Toolkit (CaPTk - <https://www.cbica.upenn.edu/captk>) and as an MITK plugin (<https://github.com/CBICA/InteractiveSegmentation>).

Keywords

Machine learning, Support Vector Machines, C++, Image segmentation, Medical imaging, Magnetic Resonance Imaging, Computational tomography, Glioblastoma, breast cancer, lung cancer, Spleen segmentation

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τα άτομα που συνέβαλαν στην εκπόνηση της παρούσας διπλωματικής εργασίας, ιδιαίτερα την Καθηγήτρια Κωνσταντίνα Νικήτα και τον Καθηγητή Χρήστο Νταβατζίκο, που μου έδωσαν την ευκαιρία να ασχοληθώ με ένα θέμα που βρήκα εξαιρετικά ενδιαφέρον. Με βοήθησαν να εμβαθύνω στο θέμα αυτό και να λάβω πολύτιμη εμπειρία για το πως λειτουργεί ο ακαδημαϊκός χώρος. Επίσης, εκτιμώ πολύ το ότι μου δόθηκε η ευκαιρία να ζήσω έξι μήνες στο εξωτερικό και να εργαστώ με άτομα από διάφορες χώρες, λόγω της συνεργασίας αυτής.

Ευχαριστώ πολύ όλους με τους οποίους συνεργάστηκα στο εργαστήριο CBICA του University of Pennsylvania κατά τη διαμονή μου στις ΗΠΑ. Το κλίμα στο εργαστήριο ήταν άριστο και οι άνθρωποι που εργάζονται εκεί ήταν εξαιρετικοί, τόσο στο επαγγελματικό, όσο και στο φιλικό επίπεδο. Μέσω του εργαστηρίου αυτού, κατάφερα να αποκτήσω την απαραίτητη γνώση, εμπειρία και υποστήριξη για να εκπονήσω τη διπλωματική αυτή, καθώς και να τη παρουσιάσω στο συνέδριο RSNA 2019 στο Chicago.

Τέλος, θέλω ακόμα να ευχαριστήσω την οικογένειά μου και τους φίλους μου που μου συμπαραστάθηκαν και με βοήθησαν στην απόκτηση του διπλώματός μου.

Η σελίδα αυτή είναι σκόπιμα λευκή.

Περιεχόμενα

1. Εισαγωγή	15
2. Εισαγωγικές έννοιες	17
2.1. Ακτινολογία	17
2.2. Ιατρική απεικόνιση	17
2.3. Τεχνολογίες αποθήκευσης και μεταφοράς ιατρικών εικόνων	20
2.3.1. NIFTI	20
2.3.2. DICOM	20
2.3.3. PACS	21
2.4. Στοιχεία ανατομικής και ιστολογίας	22
2.4.1. Καρκίνος	22
2.4.2. Γλοιοβλάστωμα	24
2.4.3. Πνευμονικοί όζοι	26
2.4.4. Καρκίνος του μαστού	26
2.4.5. Κατάτμηση σπληνός	27
2.5. Μηχανές Διανυσμάτων Υποστήριξης	28
2.6. Επισκόπηση των μεθόδων κατάτμησης	29
2.7. Συστήματα CAD	34
2.7.1. Συλλογή δεδομένων	35
2.7.2. Βήμα 1: Προεπεξεργασία	35
2.7.3. Βήμα 2: Κατάτμηση	36
2.7.4. Βήμα 3: Εξαγωγή χαρακτηριστικών	36
2.7.5. Βήμα 4: Επιλογή χαρακτηριστικών	38
2.7.6. Βήμα 5: Ανάλυση	39
2.7.7. Ροή εργασίας συστήματος CAD σε νοσοκομειακό περιβάλλον	39
2.8. Προσαρμοστικοί χάρτες γεωδαιτικής απόστασης	40
2.9. Έννοιες στατιστικής ανάλυσης	42
3. Υλοποίηση	44
3.1. Cancer Imaging Phenomics Toolkit (CaPTk)	44
3.2. Υλοποίηση ως πρόσθετος μηχανισμός MITK	45
3.3. Βιβλιοθήκες λογισμικού που χρησιμοποιήθηκαν στην υλοποίηση	46

3.3.1. ITK	46
3.3.2. OpenCV	46
3.3.3. MITK	47
3.4. Βιβλιοθήκες λογισμικού για στατιστική ανάλυση	47
4. Υλικά και Μέθοδοι	48
4.1. Δεδομένα	48
4.2. Προτεινόμενος αλγόριθμος	49
4.3. Πειραματικός Σχεδιασμός	51
4.3.1. Το πρωτόκολλο αξιολόγησης	51
4.3.2. Πείραμα 1: Αξιολόγηση Συνολικής Απόδοσης	52
4.3.3. Πείραμα 2: Συνέπεια των κατατμήσεων μεταξύ δύο διαφορετικών επαναλήψεων του ίδιου ειδικού	52
4.3.4. Πείραμα 3: Συνέπεια των κατατμήσεων από διαφορετικούς ειδικούς	53
4.3.5. Στατιστική Ανάλυση	53
5. Αποτελέσματα	54
5.1. Πείραμα 1: Αξιολόγηση Συνολικής Απόδοσης	54
5.2. Πείραμα 2: Συνέπεια των κατατμήσεων μεταξύ δύο διαφορετικών επαναλήψεων του ίδιου ειδικού	57
5.3. Πείραμα 3: Συνέπεια των κατατμήσεων από διαφορετικούς ειδικούς	59
6. Συζήτηση	61
7. Συμπέρασμα	65
8. Βιβλιογραφία	66

Ευρετήριο σχημάτων

- **Σχήμα 1:** Εικόνες γλοιοβλαστώματος που αποκτήθηκαν χρησιμοποιώντας διαφορετικά πρωτόκολλα απόκτησης (FLAIR, T1, T1CE και ED). Η ET περιοχή έχει μεγάλη ένταση στο T1CE. Η NE περιοχή είναι η σκούρα περιοχή που περικλείεται από την ET στο T1CE. Η ED περιοχή είναι όλα τα voxels που δεν ανήκουν σε NE και ED, ενώ έχουν υψηλή ένταση στο πρωτόκολλο FLAIR. *Σελίδα 25.*
- **Σχήμα 2:** Πνευμονικός όζος που καταλαμβάνει την περιοχή που είναι επισημασμένη ως κόκκινη στα δεξιά, σε εικόνα αξονικής τομογραφίας. *Σελίδα 26.*
- **Σχήμα 3:** Η μικρή περιοχή με υψηλή ένταση στο κέντρο αποτελεί όγκο του μαστού. *Σελίδα 27.*
- **Σχήμα 4:** Αξονική τομογραφία που απεικονίζει σπλήνα (επισημασμένη περιοχή). *Σελίδα 28.*
- **Σχήμα 5:** Το γραφικό περιβάλλον του CaPTk. *Σελίδα 44.*
- **Σχήμα 6:** Το γραφικό περιβάλλον του MITK Workbench, μαζί με τον πρόσθετο μηχανισμό του αλγορίθμου που αναπτύχθηκε στο πλαίσιο αυτής της εργασίας. *Σελίδα 45.*
- **Σχήμα 7:** Παράδειγμα που δείχνει το αποτέλεσμα να βελτιώνεται ως συνάρτηση του επενδυμένου χρόνου. Στην πρώτη επανάληψη (iteration), ο χρήστης σχεδιάζει γρήγορα πάνω από τις διάφορες περιοχές. Στη δεύτερη επανάληψη, ο χρήστης τοποθετεί επιπλέον ετικέτες για τη διόρθωση αντιπροσωπευτικών εσφαλμένων περιοχών, οι οποίες στη συνέχεια χρησιμοποιούνται για την επανεκπαίδευση του μοντέλου μηχανικής μάθησης. Από αριστερά προς τα δεξιά: (i) αρχική εικόνα (ii) σχέδια του χρήστη (iii) κατάτμηση που παράγει η μέθοδος (iv) κατάτμηση αναφοράς. *Σελίδα 50.*
- **Σχήμα 8:** Παράδειγμα χαρτών AGD. Οι πιο σκοτεινές τιμές δείχνουν χαμηλότερη προσαρμοσμένη γεωδαιτική απόσταση (Adaptive Geodesic Distance - AGD) από τα σχέδια του χρήστη. Στο γλοιοβλάστωμα, το όριο NE και ET είναι σαφέστερο σε T1CE, ενώ το όριο μεταξύ ED και υποβάθρου είναι σαφέστερο στο FLAIR. Από αριστερά προς τα δεξιά: (i) αρχική εικόνα (ii) σχέδια του χρήστη (iii) χάρτης AGD για τις ετικέτες για NE (το περιοχικά μη-ενισχυμένο τμήμα του γλοιοβλαστώματος) (iv) χάρτης AGD για ET (το περιοχικά ενισχυμένο τμήμα του γλοιοβλαστώματος) (v) χάρτης AGD για ED (το περιεσσιακό οίδημα) ετικέτες (vi) χάρτης AGD για το φόντο. *Σελίδα 51.*

- **Σχήμα 9:** Συντελεστής Dice, σε σύγκριση με τις κατατμήσεις αναφοράς, για τις περιπτώσεις όπου: (i) επιμέρους ετικέτες που αντιπροσωπεύουν διαφορετικές περιοχές της δομής υπολογίζονται ως μία, (ii) μεμονωμένες υποπεριοχές του γλοιοβλαστώματος (για τις άλλες δομές δεν υπάρχουν υποπεριοχές). *Σελίδα 55.*
- **Σχήμα 10:** Scatterplots στα οποία τα μπλε σημεία είναι τα ζεύγη των γεωμετρικών όγκων των κατατμήσεων της IML μεθόδου και των γεωμετρικών όγκων των κατατμήσεων αναφοράς. Παρόμοια, τα κόκκινα σημεία είναι ζεύγη των όγκων των χειροκίνητων κατατμήσεων και κατατμήσεων αναφοράς. Η μαύρη γραμμή αντιπροσωπεύει τους όγκους των κατατμήσεων αναφοράς. Τα σχεδιαγράμματα ανήκουν σε (i) διαφορετικές μελέτες όπου όλες οι ετικέτες, που αντιπροσωπεύουν διαφορετικές υποπεριοχές της δομής, υπολογίζονται ως μία και (ii) υποπεριοχές του γλοιοβλαστώματος. *Σελίδα 56.*
- **Σχήμα 11:** Συντελεστής Dice, για αποτίμηση της συνέπειας ανάμεσα σε δύο προσπάθειες του ίδιου ειδικού: (i) για όταν οι επιμέρους ετικέτες που αντιπροσωπεύουν διαφορετικές περιοχές της δομής υπολογίζονται ως μία, (ii) για μεμονωμένες περιοχές του γλοιοβλαστώματος (για τις άλλες δομές δεν υπήρχαν υποπεριοχές). *Σελίδα 58.*
- **Σχήμα 12:** Συντελεστής Dice, για αποτίμηση συνέπειας μεταξύ κατατμήσεων διαφορετικών ειδικών: (i) για όταν οι επιμέρους ετικέτες που αντιπροσωπεύουν διαφορετικές περιοχές της δομής υπολογίζονται ως μία, (ii) για τις υποπεριοχές του γλοιοβλαστώματος (οι άλλες δομές δεν είχαν υποπεριοχές). *Σελίδα 59.*

Ευρετήριο πινάκων

- **Πίνακας 1:** Συνοπτική περιγραφή των δεδομένων που χρησιμοποιήθηκαν στην εργασία. *Σελίδα 49.*
- **Πίνακας 2:** Αποτελέσματα (p-τιμές) του Wilcoxon test με συσχετισμένα δείγματα, για κάθε ειδικό, συγκρίνοντας τα αποτελέσματα του συντελεστή Dice των διαφορετικών προσεγγίσεων (IML και χειροκίνητης κατάτμησης) σε σχέση με τα δεδομένα αναφοράς. Η διαφορά θεωρείται σημαντική όταν $p < 0,05$. *Σελίδα 57.*
- **Πίνακας 3:** Αποτελέσματα και για τα τρία πειράματα. Πείραμα 1: Η απόδοση του αλγορίθμου υπολογίζεται ως η επικάλυψη, που υπολογίζεται μέσω του συντελεστή Dice, των κατατμήσεων με τις κατατμήσεις αναφοράς. Οι τιμές p είναι αποτέλεσμα συζευγμένων συγκρίσεων μεταξύ των IML και χειροκίνητων κατατμήσεων που παρουσίασαν υψηλότερη επικάλυψη για κάθε ειδικό και για κάθε ασθενή. Ο συντελεστής συσχέτισης υπολογίζεται μεταξύ των γεωμετρικών όγκων που υπολογίζονται για τις δύο προσεγγίσεις και των όγκων αναφοράς. “Συντ. Συσχ.” είναι ο συντελεστής συσχέτισης. Πείραμα 2: Οι τιμές δείχνουν επικάλυψη μεταξύ της πρώτης και της δεύτερης προσπάθειας κάθε ειδικού, για κάθε προσέγγιση. Πείραμα 3: Οι τιμές δείχνουν επικάλυψη μεταξύ κατατμήσεων διαφορετικών ειδικών. *Σελίδα 60.*

Η σελίδα αυτή είναι σκόπιμα λευκή.

1. Εισαγωγή

Η κατάτμηση ιατρικών εικόνων, δηλαδή η “σήμανση” περιοχών ενδιαφέροντος, όπως της περιοχής που καταλαμβάνει ένας όγκος, αποτελεί μία σημαντική διαδικασία σε κλινικό και ερευνητικό περιβάλλον [1]–[5], καθώς διευκολύνει επακόλουθες υπολογιστικές αναλύσεις οι οποίες εξαρτώνται από την ακρίβεια της κατάτμησης που παρέχεται ως είσοδος [6]–[8]. Η χειροκίνητη κατάτμηση προτιμάται για τέτοιες διαδικασίες, παρόλο που τείνει να είναι χρονοβόρα και με περιορισμένη αναπαραγωγισιμότητα [3], [9]. Η χειροκίνητη κατάτμηση μιας απεικονιστικής εξέτασης μπορεί να διαρκέσει πολλές ώρες για εικόνες με μεγάλη ανάλυση ή για περίπλοκες ανατομίες, ακόμα και με τη βοήθεια κατάλληλων εργαλείων [4].

Έχει προταθεί μια πληθώρα μεθόδων πλήρως αυτοματοποιημένης κατάτμησης, βασισμένων στη μηχανική μάθηση (machine learning), οι οποίες μπορούν να επιτύχουν εξαιρετικά αποτελέσματα. Όμως, αυτές οι μέθοδοι, αντιμετωπίζουν σημαντικά προβλήματα [10] που εμποδίζουν την υιοθέτησή τους στην κλινική πράξη. Συγκεκριμένα, η δυσκολία γενίκευσης σε άγνωστα δεδομένα και η απουσία ενός μηχανισμού για εκτεταμένες διορθώσεις και βελτιώσεις του μοντέλου και του αποτελέσματος από τους χρήστες [4], [11], αποτελούν τις κυριότερες προκλήσεις.

Η διαδραστική μηχανική μάθηση (Interactive Machine Learning - IML) καλύπτει το κενό ανάμεσα σε χειροκίνητη και πλήρως αυτοματοποιημένη κατάτμηση, επιτρέποντας στον χρήστη να εκπαιδεύσει ένα μοντέλο εστιασμένο στον εκάστοτε ασθενή. Συγκεκριμένα, μέσω γρήγορων και αδρών σχεδίων, ο χρήστης μπορεί να εκπαιδεύσει ένα μοντέλο ώστε να αναγνωρίζει τις διάφορες περιοχές της ανατομίας. Στη συνέχεια, το μοντέλο αυτό χρησιμοποιείται για να πραγματοποιήσει κατάτμηση ολόκληρης της απεικονιστικής εξέτασης [5], [12]–[14]. Ένα σημαντικό πλεονέκτημα των διαδραστικών προσεγγίσεων είναι πως παρέχουν την δυνατότητα για ταχείες βελτιώσεις σε περίπτωση σφαλμάτων, συνήθως μέσω

αξιοποίησης νέων σχεδίων του χρήστη πάνω στις εσφαλμένες περιοχές. Με αυτό τον τρόπο, η τελική κατάτμηση τείνει να πλησιάσει το βέλτιστο αποτέλεσμα ως συνάρτηση του χρόνου που αφιερώνεται στη διαδικασία. Δύο δημοφιλή εργαλεία που προσφέρουν τέτοια λειτουργικότητα είναι το ITK-SNAP [11] και το 3D Slicer [15].

Σε αυτήν την εργασία, προτείνεται μία IML μέθοδος που χρησιμοποιεί προσαρμοστικούς χάρτες γεωδαιτικής απόστασης (Adaptive Geodesic Distance - AGD) [16] μαζί με ένα σύνολο από Μηχανές Διανυσμάτων Υποστήριξης (ΜΔΥ), ώστε να πραγματοποιήσει διαδραστική κατάτμηση ιατρικών εικόνων. Η μέθοδος δεν εξαρτάται από τον τύπο και τις διαστάσεις των εικόνων, ενώ πολυπαραμετρικές εικόνες μπορούν να χρησιμοποιηθούν ως είσοδος. Επίσης, με τη βοήθεια ειδικών, πραγματοποιείται συστηματική αξιολόγηση της απόδοσης της προτεινόμενης μεθόδου έναντι χειροκίνητης κατάτμησης, για διάφορες ανατομίες και απεικονιστικές τεχνικές. Η αξιολόγηση αποσκοπεί στη σύγκριση ως προς την ταχύτητα, τη χωρική επικάλυψη και τη συνέπεια των παραγόμενων κατατμήσεων μεταξύ διαφορετικών χρονικών σημείων και χρηστών.

2. Εισαγωγικές έννοιες

2.1. Ακτινολογία

Η ακτινολογία είναι μια διεπιστημονική ιατρική ειδικότητα που εστιάζει στην ανάλυση ιατρικών εικόνων για τη διάγνωση και τη θεραπεία ασθενειών [17]. Για την απόκτηση των ιατρικών εικόνων χρησιμοποιούνται διάφορες τεχνολογίες όπως η μαγνητική τομογραφία (MRI - Magnetic Resonance Imaging), η αξονική τομογραφία (CT - Computed Tomography), η τομογραφία εκπομπής ποζιτρονίων (PET - Positron Emission Tomography), η ακτινογραφία και ο υπέρηχος. Επειδή η ακτινολογία είναι ένας ευρύς κλάδος που αναπτύσσεται συνεχώς, πολλοί ακτινολόγοι εξειδικεύονται σε ένα μέρος του σώματος [18], ωστόσο υπάρχουν και γενικοί ακτινολόγοι.

Η διαγνωστική ακτινολογία είναι μία υποειδικότητα της ακτινολογίας, που επικεντρώνεται στην χρήση ιατρικών εικόνων με σκοπό τη διάγνωση της αιτίας των συμπτωμάτων ενός ασθενούς, τον έλεγχο για διάφορες ασθένειες και την παρακολούθηση της απόκρισης ενός ασθενούς σε μια θεραπεία [19]. Η διαγνωστική ακτινολογία είναι ένας σημαντικός τομέας της ιατρικής, καθώς είναι ένας ασφαλής και μη επεμβατικός τρόπος για την εκτίμηση της υγείας. Η επεμβατική ακτινολογία επικεντρώνεται στην υποβοήθηση και καθοδήγηση θεραπειών οι οποίες είναι ελάχιστα επεμβατικές, περιορίζοντας την ανάγκη για χειρουργική επέμβαση. Η πρακτική χρησιμοποιείται από γιατρούς για επεμβατικές διαδικασίες όπως η εισαγωγή ενός καθετήρα στις φλέβες [19] ή διάνοιξη μιας αρτηρίας [20]. Άλλα παραδείγματα επεμβατικής ακτινολογίας αποτελούν η βιοψία οργάνων, ο εμβολισμός για τον έλεγχο της αιμορραγίας και η αγγειογραφία.

2.2. Ιατρική απεικόνιση

Ιατρική απεικόνιση είναι η μη επεμβατική διαδικασία παραγωγής εικόνων του εσωτερικού του σώματος για κλινικούς σκοπούς. Οι ιατρικές εικόνες είναι συνήθως ασπρόμαυρες και

έχουν κάποιες ιδιότητες που τις διαφοροποιούν από τις εικόνες που καταγράφονται από φωτογραφικές κάμερες.

Ένα σημαντικό χαρακτηριστικό των ιατρικών εικόνων είναι ότι υπάρχουν πολλές διαφορετικές τεχνολογίες που χρησιμοποιούνται για την απόκτηση αυτών των εικόνων, γνωστές ως απεικονιστικές μέθοδοι (modalities). Κάθε απεικονιστική μέθοδος παράγει αποτελέσματα που έχουν διαφορετική όψη από τις υπόλοιπες και κάθε μέθοδος εξυπηρετεί διαφορετικούς σκοπούς. Δύο δημοφιλείς απεικονιστικές μέθοδοι, που χρησιμοποιούνται και στα πειράματα αυτής της εργασίας, είναι η μαγνητική τομογραφία (MRI - Magnetic Resonance Imaging) και η αξονική τομογραφία (CT - Computed Tomography). Οι εικόνες μαγνητικής τομογραφίας παράγονται χρησιμοποιώντας ραδιοκύματα και μπορούν να επιτύχουν υψηλό επίπεδο λεπτομέρειας. Από την άλλη, οι αξονικοί τομογράφοι χρησιμοποιούν ακτίνες X, κοστίζουν λιγότερο και προτιμώνται για την απεικόνιση οργάνων και οστών.

Υπάρχει ένα πλήθος πρωτοκόλλων απόκτησης (acquisition protocols) για κάθε απεικονιστική μέθοδο, τα οποία έχουν διαφορετικές κλινικές εφαρμογές. Εικόνες από την ίδια απεικονιστική μέθοδο αλλά με διαφορετικό πρωτόκολλο απόκτησης μπορούν να ληφθούν κατά την ίδια εξέταση του ασθενούς, χρησιμοποιώντας το ίδιο μηχάνημα. Για παράδειγμα, στα πειράματα απεικόνισης ασθενών με γλοιοβλάστωμα, χρησιμοποιήθηκαν τέσσερα πρωτόκολλα απόκτησης μαγνητικής τομογραφίας: T1 και T2, τα οποία είναι τα βασικά πρωτόκολλα που δείχνουν κυρίως λίπος και ρευστά, αντίστοιχα, T1CE (T1Gd) που περιέχει πληροφορίες σχετικά με την αγγειογένεση και την ακεραιότητα του αιματοεγκεφαλικού φραγμού στον όγκο και απαιτεί τη χορήγηση ενός ακτινοσκιερού υλικού - παράγοντα αντίθεσης (contrast agent) στον ασθενή και το πρωτόκολλο FLAIR, το οποίο τονίζει τον όγκο και την οίδηματώδη περιοχή γύρω του [6], [21]. Οι εικόνες που λαμβάνονται με διαφορετικά πρωτόκολλα απόκτησης μπορούν να ευθυγραμμιστούν, ούτως ώστε να αναφέρονται στο

ίδιο σύστημα συντεταγμένων, μέσω κατάλληλης διαδικασίας ευθυγράμμισης (registration). Ως αποτέλεσμα, αυτές οι ευθυγραμμισμένες εικόνες μπορούν να υποστούν επεξεργασία ως μία εικόνα με πολλά κανάλια, όπως συμβαίνει συνήθως με τα κανάλια R/G/B των κανονικών έγχρωμων εικόνων, επιτρέποντας έτσι την εφαρμογή διαφόρων κοινών μεθοδολογιών του πεδίου της όρασης υπολογιστών.

Μια ιατρική εικόνα έχει συνήθως 2 ή 3 διαστάσεις. Οι τρεις κύριες κατευθύνσεις-διαστάσεις, της τρισδιάστατης εικόνας, ορίζουν ανά δύο τους κυριους προσανατολισμούς επιπέδων τομής (εγκάρσιο, οβελιαίο, στεφανιαίο). Χρησιμοποιώντας δισδιάστατες τομές που ανήκουν σε ένα από του τρεις αυτούς προσανατολισμούς, μπορεί να πραγματοποιηθεί οπτική επιθεώρηση των τρισδιάστατων εικόνων σε οθόνη δύο διαστάσεων. Οι ιατρικές εικόνες μπορούν επίσης να έχουν 4 διαστάσεις, όπου η τέταρτη διάσταση είναι ο χρόνος. Οι 4D εικόνες είναι ουσιαστικά μια σειρά τρισδιάστατων εικόνων που αποκτήθηκαν σε ένα μικρό χρονικό διάστημα και είναι κλινικά χρήσιμες για την παρακολούθηση της λειτουργίας ενός βιολογικού μηχανισμού, όπως η ροή του αίματος ή η οξυγόνωση. Οι τετραδιάστατες εικόνες μπορούν να επιθεωρηθούν οπτικά με τον ίδιο τρόπο με τις τρισδιάστατες, αρκεί να υπάρχει επιπρόσθετα ένας μηχανισμός, όπως ένα γραφικό widget ολίσθησης, για εναλλαγή του χρονικού σημείου που επιθεωρείται.

Οι κατατμήσεις είναι μια ειδική περίπτωση ιατρικών εικόνων, που χρησιμοποιούνται για τη σήμανση περιοχών ενδιαφέροντος σε μία απεικονιστική εξέταση. Είναι τυπικά *ευθυγραμμισμένες* με τις εικόνες που περιγράφουν, ενώ τα pixels/voxels τους έχουν τιμή 0 παντού, εκτός από τα pixels/voxels που καλύπτουν περιοχές ενδιαφέροντος, τα οποία έχουν μια διαφορετική ακέραια τιμή για κάθε περιοχή.

Μια ακόμη σημαντική πτυχή των ιατρικών εικόνων είναι πως η φυσική απόσταση μεταξύ των voxels, που συνήθως αναφέρεται ως "*spacing*", μπορεί να ποικίλλει σε κάθε διάσταση. Αυτό συμβαίνει επειδή η απεικόνιση λαμβάνεται ως μία σειρά από δισδιάστατες εικόνες (slices), οι

οποίες στη συνέχεια συγχωνεύονται σε μία τρισδιάστατη. Η απόσταση μεταξύ των αποκτηθεισών τομών συνήθως είναι διαφορετική από τη φυσική απόσταση μεταξύ των voxels της τομής. Για την αντιμετώπιση των προβλημάτων που προκύπτουν, οι αλγόριθμοι συνήθως μετασχηματίζουν τις εικόνες, ώστε να έχουν την ίδια φυσική απόσταση σε όλες τις διαστάσεις. Μετά την υπολογιστική ανάλυση, οι αλγόριθμοι επαναφέρουν τα αποτελέσματα στον αρχικό χώρο. Η διαδικασία που χρησιμοποιείται για τον μετασχηματισμό του μεγέθους και της φυσικής απόστασης ανά διάσταση μιας εικόνας ονομάζεται επαναδειγματοληψία (resampling).

2.3. Τεχνολογίες αποθήκευσης και μεταφοράς ιατρικών εικόνων

2.3.1. NIFTI

Τα NIFTI-1 και NIFTI-2 (.nii ή .nii.gz) είναι πρότυπα και μορφές αποθήκευσης δεδομένων που χρησιμοποιούνται για την αποθήκευση ιατρικών εικόνων. Είναι απλά πρότυπα με αυστηρούς κανόνες και αυτό έχει συμβάλει στην υψηλή τους υιοθέτηση σε ερευνητικά περιβάλλοντα. Ωστόσο, οι εικόνες που αποθηκεύονται με αυτόν τον τρόπο δεν περιέχουν μετα-πληροφορίες, κάτι που εμποδίζει την υιοθέτησή τους κλινικά.

2.3.2. DICOM

Το DICOM (Digital Imaging and Communications in Medicine - Ψηφιακή απεικόνιση και επικοινωνία στην ιατρική) [22] είναι ένα πρότυπο για την αποθήκευση ιατρικών εικόνων και μετα-πληροφοριών σχετικά με αυτές. Χρησιμοποιείται εξ ολοκλήρου στα κλινικά περιβάλλοντα και υιοθετείται όλο και περισσότερο στην έρευνα. Τα περισσότερα ιατρικά μηχανήματα παράγουν τα αποτελέσματά τους σε μορφή DICOM. Οι μετα-πληροφορίες μπορούν να αποθηκευτούν μέσα στην κεφαλίδα (header) του αρχείου σε ζεύγη κλειδιού/τιμής, που ονομάζονται ετικέτα/χαρακτηριστικό (tag/attribute). Για παράδειγμα, η απεικονιστική μέθοδος αποθηκεύεται στην ετικέτα "(0008,0060)". Άλλες πληροφορίες που μπορούν να αποθηκευτούν είναι το όνομα του ασθενούς, η ημερομηνία απόκτησης, το

όνομα της μελέτης στην οποία ανήκει η εικόνα και το πρωτόκολλο απόκτησης. Οι εικόνες που είναι αποθηκευμένες σε μορφή DICOM μπορούν να είναι εξ ολοκλήρου σε ένα αρχείο ή μοιρασμένες σε πολλά, όπου το καθένα περιέχει μόνο μία τομή. Η τελευταία προσέγγιση ήταν πιο χρήσιμη στο παρελθόν, καθώς επέτρεπε το διαμοιρασμό μιας μεγάλης εικόνας χρησιμοποιώντας πολλαπλές συσκευές αποθήκευσης, όπως τα DVD. Πρέπει επίσης να σημειωθεί ότι πολλές από τις πληροφορίες στην κεφαλίδα αποτελούν PHI (Protected Health Information - Προστατευόμενες πληροφορίες υγείας) και πρέπει να υποστούν επεξεργασία για να ανωνυμοποιηθούν όταν χρησιμοποιούνται για ερευνητικούς σκοπούς. Τέλος, το DICOM δέχεται κριτική επειδή δεν είναι αυστηρό πρότυπο, καθώς δύο εταιρείες ενδέχεται να αποθηκεύουν δεδομένα με διαφορετικό και ασύμβατο τρόπο, παρόλου που και οι δύο αποθηκεύουν τα δεδομένα σε DICOM. Ως αποτέλεσμα, πολλές εφαρμογές, όπως το MITK [23], έχουν ένα σύνολο διαφορετικών reader (αναγνωστών) DICOM, από διαφορετικούς οργανισμούς και προσπαθούν να χρησιμοποιούν αυτόματα τον πιο συμβατό.

2.3.3. PACS

Το PACS (Picture Archiving and Communication System - Σύστημα αρχειοθέτησης εικόνων και επικοινωνίας) είναι μια τεχνολογία που χρησιμοποιείται για την αποθήκευση και ανάκτηση ιατρικών εικόνων σε νοσοκομειακό περιβάλλον, με βάση το πρωτόκολλο DICOM. Αποτελείται από έναν κεντρικό “διακομιστή” (server) στον οποίο αποθηκεύονται όλα τα δεδομένα και διάφορους “πελάτες” (clients), δηλαδή προγράμματα, που τρέχουν στους υπολογιστές των γιατρών. Με αυτό τον τρόπο, οι εικόνες δεν χρειάζεται να μεταφέρονται εντός του νοσοκομείου χρησιμοποιώντας φυσικές συσκευές, αλλά μεταφέρονται μέσω του δικτύου. Κάθε γιατρός, που δικαιούται πρόσβαση σε κάποια δεδομένα, μπορεί να τα εξετάσει μέσω του προγράμματος-πελάτη PACS. Ένα PACS έχει επίσης τη δυνατότητα να εμφανίζει ιατρικές εικόνες, χωρίς να υπάρχει η ανάγκη χρήσης εξωτερικού λογισμικού. Επιπλέον, με τη βοήθεια τεχνολογιών απομακρυσμένης πρόσβασης, όπως το VPN, ένα PACS μπορεί να επιτρέψει την ασφαλή πρόσβαση στις ιατρικές εικόνες σε γιατρούς όταν δεν

βρίσκονται εντός του νοσοκομειακού χώρου. Η ασφάλεια του δικτύου και της υποδομής είναι εξαιρετικά σημαντική, διότι το σύστημα περιέχει προσωπικές πληροφορίες του ασθενούς. Επίσης, δεδομένου ότι είναι ένα πολύ αυτοματοποιημένο σύστημα, ένα PACS μπορεί να μειώσει σημαντικά το κόστος διαχείρισης δεδομένων.

Τα τελευταία χρόνια, έχουν αυξηθεί οι εταιρείες που ειδικεύονται στην ανάπτυξη λογισμικών PACS που αποθηκεύουν τις ιατρικές εικόνες στο cloud της εταιρείας, αντί για έναν διακομιστή εντός του νοσοκομειακού χώρου. Αυτές οι εταιρείες φροντίζουν επίσης το λογισμικό “πελάτη” (client) που τρέχει στο σταθμό εργασίας των γιατρών. Χρησιμοποιώντας ένα PACS που βασίζεται στο cloud, το νοσοκομείο δεν χρειάζεται να απασχολεί ειδικό προσωπικό και να συντηρεί τον κεντρικό διακομιστή, καθώς τέτοιου είδους εταιρείες μπορούν να παρέχουν αυτές τις υπηρεσίες για μια περιοδική αμοιβή. Αυτή η κίνηση θα μπορούσε επίσης να αυξήσει την ασφάλεια των δεδομένων, επειδή μια εταιρεία που ειδικεύεται στην ανάπτυξη PACS διαθέτει προγραμματιστές που ειδικεύονται στην ασφάλεια, που συνήθως δεν συμβαίνει σε ένα νοσοκομείο.

2.4. Στοιχεία ανατομικής και ιστολογίας

Σε αυτή την εργασία, πραγματοποιήθηκαν τέσσερις μελέτες, που αφορούσαν γλοιοβλάστωμα, πνευμονικούς όζους, όγκους του μαστού και εικόνες σπληνός.

2.4.1. Καρκίνος

Ο καρκίνος είναι ένα σύνολο ασθενειών που σχετίζονται με τον ανεξέλεγκτο πολλαπλασιασμό μη φυσιολογικών κυττάρων στο σώμα. Σύμφωνα με το Παγκόσμιο Παρατηρητήριο Καρκίνου (Global Cancer Observatory) του Διεθνούς Οργανισμού Έρευνας για τον Καρκίνο (International Agency for Research on Cancer), υπήρχαν πάνω από 18 εκατομμύρια νέες διαγνώσεις καρκίνου το 2018 παγκοσμίως, ενώ πάνω από 9 εκατομμύρια άνθρωποι έχασαν τη ζωή τους λόγω της νόσου τον ίδιο χρόνο [24].

Οι κύριες αιτίες που προκαλούν καρκίνο είναι η χρήση καπνού που αντιπροσωπεύει το 25-30% των θανάτων, η διατροφή που αντιπροσωπεύει το 30-35%, οι λοιμώξεις που αντιπροσωπεύουν το 15-20% και γενετικά προβλήματα που αντιπροσωπεύουν το 5-10%. Οι υπόλοιπες περιπτώσεις αποδίδονται σε άλλους παράγοντες όπως η ακτινοβολία, το στρες, η σωματική δραστηριότητα και οι περιβαλλοντικοί ρύποι [25]. Δεδομένου ότι ο τρόπος ζωής παίζει τόσο σημαντικό ρόλο, πολλοί από αυτούς τους θανάτους μπορούν να προληφθούν με τη διακοπή του καπνίσματος, τη βελτίωση της διατροφής, τον εμβολιασμό, την σωματική άσκηση και την αποφυγή άμεσης έκθεσης στον ήλιο. Τα είδη του καρκίνου είναι εκατοντάδες, μπορούν όμως να ομαδοποιηθούν στις 5 κύριες κατηγορίες [26] που περιγράφονται παρακάτω.

Το *καρκίνωμα* ξεκινάει στα επιθηλιακά κύτταρα, τα οποία βρίσκονται στο δέρμα ή σε λεπτά στρώματα στα τοιχώματα των αιμοφόρων αγγείων ή στο εξωτερικό των οργάνων. Καλύπτουν επίσης το στήθος και τη κοιλιακή κοιλότητα. Τα περισσότερα καρκινώματα δημιουργούν συμπαγείς όγκους [27]. Ο καρκίνος των πνευμόνων, του μαστού και του προστάτη ανήκουν στα καρκινώματα. Είναι ο πιο κοινός τύπος καρκίνου, αντιπροσωπεύοντας περίπου το 85% των περιπτώσεων στο Ηνωμένο Βασίλειο [26].

Το *σάρκωμα* ξεκινάει σε συνδετικούς ιστούς, όπως οστά, νεύρα και μύες που υποστηρίζουν τα όργανα [27]. Το σάρκωμα αντιπροσωπεύει περίπου το 1% των περιπτώσεων [26]. Υπάρχουν δύο υποκατηγορίες σαρκώματος, το σάρκωμα οστών και το σάρκωμα μαλακών μορίων.

Λευχαιμία ονομάζεται ο καρκίνος των λευκών αιμοσφαιρίων. Προκαλείται όταν ο μυελός των οστών δημιουργεί έναν μεγάλο αριθμό δυσλειτουργικών λευκών αιμοσφαιρίων, τα οποία στη συνέχεια συσσωρεύονται στο αίμα. Η λευχαιμία αντιπροσωπεύει περίπου το 3% των περιπτώσεων, αλλά είναι ο πιο κοινός καρκίνος στα παιδιά [26].

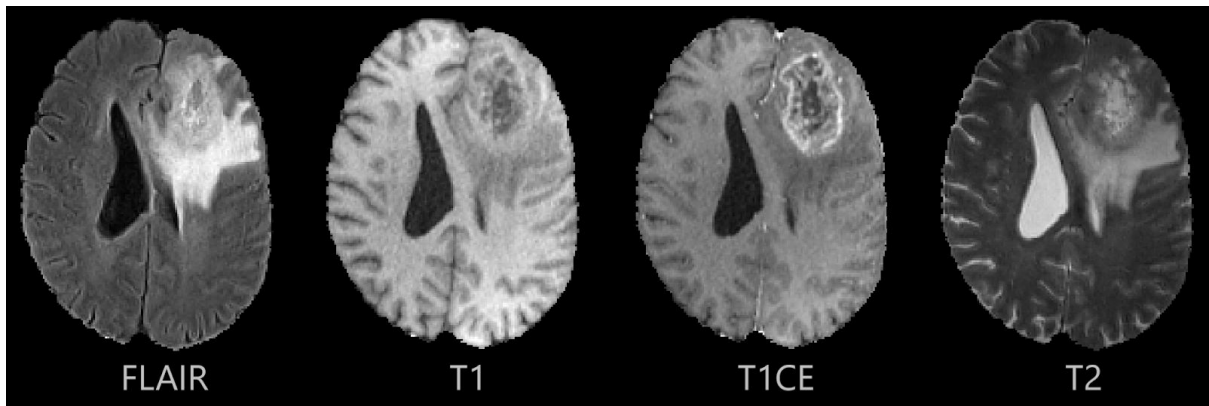
Το *λέμφωμα* είναι καρκίνος που ξεκινά σε κύτταρα του ανοσοποιητικού συστήματος, συγκεκριμένα στα λεμφικά κύτταρα. Το λεμφικό σύστημα αποτελεί ένα δίκτυο αγγείων και αδένων που λειτουργεί ως φίλτρο για τα σωματικά υγρά και καταπολεμά μολύνσεις. Αποτελείται από τους λεμφαδένες, τα λεμφαγγεία και τον σπλήνα. Τα δυσλειτουργικά λεμφοκύτταρα πολλαπλασιάζονται γρήγορα και μπορούν να συσσωρευτούν στο μυελό των οστών και στον σπλήνα και να προκαλέσουν όγκους. Το λέμφωμα αντιπροσωπεύει περίπου το 5% των περιπτώσεων [26].

Οι *καρκίνοι του εγκεφάλου και του νωτιαίου μυελού* ξεκινούν σε κύτταρα του κεντρικού νευρικού συστήματος. Το γλοίωμα, ο πιο κοινός τύπος, αφορά όγκους του εγκεφάλου που αναπτύσσονται από τα γλοιακά κύτταρα. Οι όγκοι του εγκεφάλου είναι ιδιαίτερα επικίνδυνοι επειδή μπορούν να επηρεάσουν μέρη του εγκεφάλου που είναι σημαντικά για τη ζωή και η έγκαιρη ανίχνευσή τους είναι δύσκολη. Αυτού του είδους οι καρκίνοι αντιπροσωπεύουν περίπου το 3% των περιπτώσεων [26].

2.4.2. Γλοιοβλάστωμα

Το γλοιοβλάστωμα είναι μια κατηγορία κακοήθων όγκων του εγκεφάλου. Κατατάσσεται ως “Βαθμού IV” στην κλίμακα ταξινόμησης του ΠΟΥ (Παγκόσμιου Οργανισμού Υγείας) [28]. Για να ταξινομηθεί ως “βαθμού IV”, ένας τύπος όγκου πρέπει να εμφανίζει εκσεσημασμένη αναπλασία, αυξημένη μιτωτική δραστηριότητα με έντονο μικροαγγειακό πολλαπλασιασμό ή / και νέκρωση. Το γλοιοβλάστωμα αναπαράγεται γρήγορα και επεκτείνεται επιθετικά [28], [29]. Σύμφωνα με μια μελέτη στην Αγγλία, η οποία εξέτασε 10.743 ασθενείς με γλοιοβλάστωμα [30], η μέση συνολική επιβίωση ήταν 6,1 μήνες και οι ασθενείς είχαν πιθανότητα 28,4%, 11,5% και 3,4% να επιβιώσουν 1, 2 και 5 χρόνια αντίστοιχα.

Τα γλοιοβλαστώματα είναι μεγάλα και έχουν σύνθετη δομή. Χωρίζονται σε δύο κύριες υποπεριοχές, τον πυρήνα του όγκου (TC - Tumor core) και το περιεστιακό οίδημα (ED -

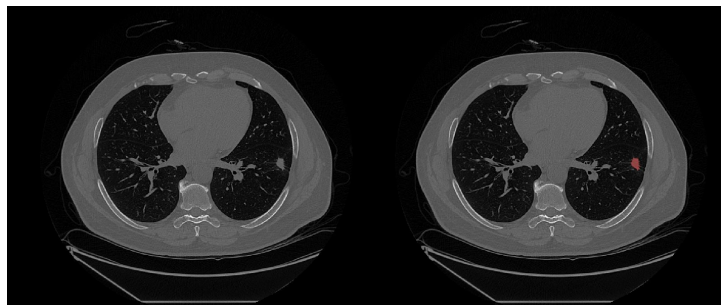


Σχήμα 1: Εικόνες γλοιοβλαστώματος που αποκτήθηκαν χρησιμοποιώντας διαφορετικά πρωτόκολλα απόκτησης (FLAIR, T1, T1CE και ED). Η ET περιοχή έχει μεγάλη ένταση στο T1CE. Η NE περιοχή είναι η σκούρα περιοχή που περικλείεται από την ET στο T1CE. Η ED περιοχή είναι όλα τα voxels που δεν ανήκουν σε NE και ED, ενώ έχουν υψηλή ένταση στο πρωτόκολλο FLAIR.

Peritumoral edema). Ο πυρήνας του όγκου χωρίζεται επίσης στο περιοχικά μη-ενισχυμένο τμήμα (NE - Non-enhancing tumor core) και στο περιοχικά ενισχυμένο τμήμα (ET - Enhancing tumor core) [31], [32]. Τα υποτμήματα του πυρήνα του όγκου παίρνουν τα ονόματά τους από το εάν εμφανίζουν υψηλή ένταση σε εικόνες T1CE (T1Gd), μετά τη χορήγηση ενός παράγοντα αντίθεσης. Υπάρχουν ωστόσο και υποκείμενες διαφορές. Το μη-ενισχυμένο τμήμα αποτελείται κυρίως από προ-νεκρωτικούς και νεκρωτικούς ιστούς του πυρήνα του όγκου, ενώ το ενισχυμένο τμήμα θεωρείται ότι είναι πιο ενεργό [32]. Όμως, το μη-ενισχυμένο τμήμα θεωρείται επίσης ότι είναι διηθητικό σε κάποιο βαθμό [33]. Τέλος, το περιεστιακό οίδημα περιβάλλει τον πυρήνα του όγκου και έχει υψηλή ένταση σε εικόνες T2-FLAIR. Περιγράφει τις περιοχές οίδηματώδους λευκής ουσίας στον φλοιό του εγκεφάλου. Η περιοχή του οιδήματος (ED) μπορεί επίσης να περιέχει προσβεβλημένους ιστούς. Η ποσότητα της διείσδυσης καρκινικών κυττάρων στην οίδηματώδη περιοχή κυμαίνεται ανάλογα με την απόσταση ενός σημείου από τον πυρήνα του όγκου [6]. Για αυτό το λόγο, σε χειρουργικές επεμβάσεις αφαίρεσης του πυρήνα του όγκου, αφαιρείται επίσης ένα μέρος του περιφερικού οιδήματος κοντά στον πυρήνα, ως προφύλαξη.

2.4.3. Πνευμονικοί όζοι

Οι πνευμονικοί όζοι είναι μικρές μάζες που εμφανίζονται στους πνεύμονες και μπορεί να είναι καρκινικοί ή καλοήθεις. Συνήθως έχουν διάμετρο μικρότερη από 30mm και είναι ορατοί σε αξονικές τομογραφίες και ακτινογραφίες [34], [35]. Ο εντοπισμός και η ταξινόμηση τους σε καρκινικούς ή καλοήθεις είναι μία απαιτητική διαδικασία [36], λόγω του μικρού μεγέθους τους και της ομοιότητας μεταξύ των δύο τύπων. Επίσης, λόγω του μικρού μεγέθους τους, λάθη και ανακρίβειες στην κατάτμηση επηρεάζουν αρνητικά την ταξινόμηση και άλλες υπολογιστικές αναλύσεις που χρησιμοποιούν την κατάτμηση ως είσοδο [37], καθώς τέτοια λάθη επηρεάζουν σε μεγάλο βαθμό τον υπολογισμό του γεωμετρικού όγκου και την εξαγωγή χαρακτηριστικών (features).

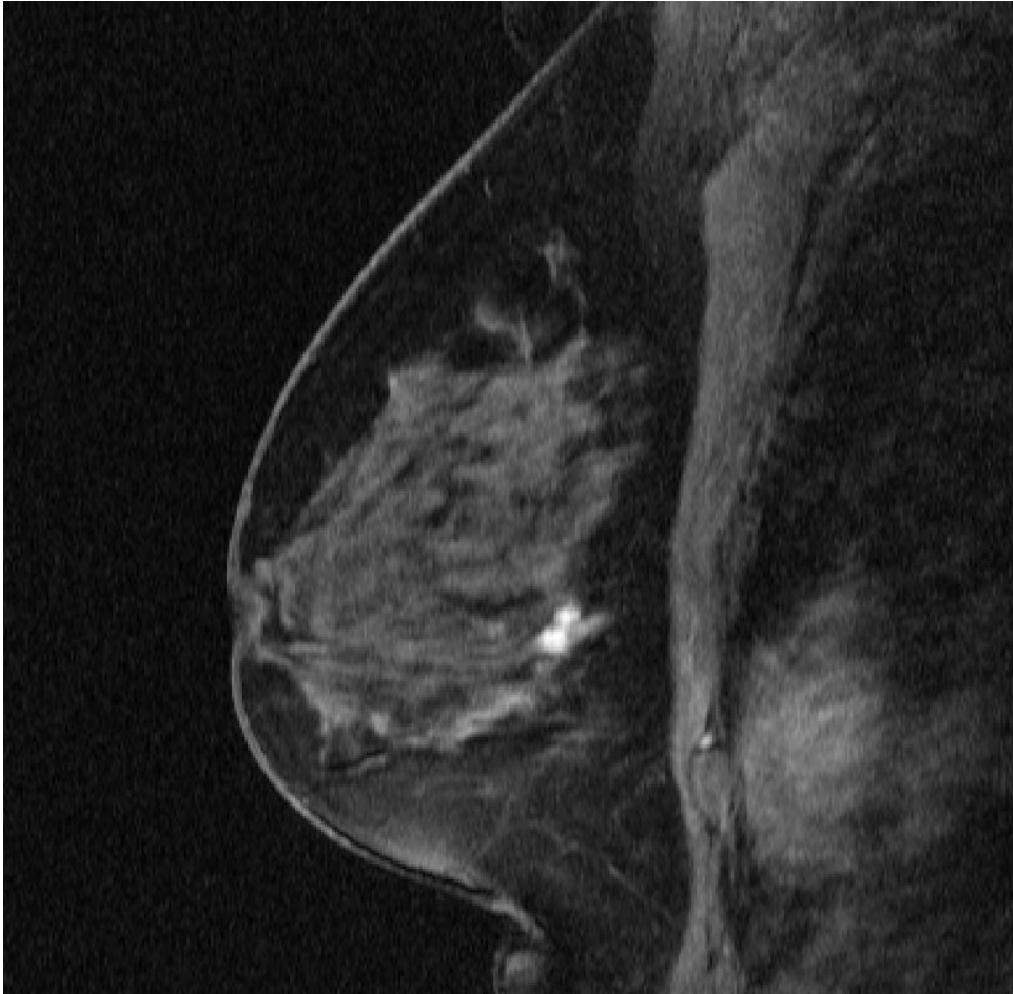


Σχήμα 2: Πνευμονικός όζος που καταλαμβάνει την περιοχή που είναι επισημασμένη ως κόκκινη στα δεξιά, σε εικόνα αξονικής τομογραφίας.

2.4.4. Καρκίνος του μαστού

Ο καρκίνος του μαστού είναι μια ασθένεια που επηρεάζει μεγάλο ποσοστό του γυναικείου πληθυσμού. Το 2019 στις ΗΠΑ, περίπου 41.760 γυναίκες πέθαναν από αυτή την ασθένεια και σε περίπου 268.600 γυναίκες ο καρκίνος ήταν διηθητικός [38]. Η πρόληψη και η έγκαιρη ανίχνευση είναι πολύ σημαντικοί παράγοντες στην καταπολέμηση της νόσου. Σε αυτή την εργασία, αναλύθηκαν εικόνες MRI κακοηθών περιπτώσεων όγκου μαστού. Χρησιμοποιήθηκαν τρεις εικόνες από κάθε απεικονιστική εξέταση, εκ των οποίων η μία

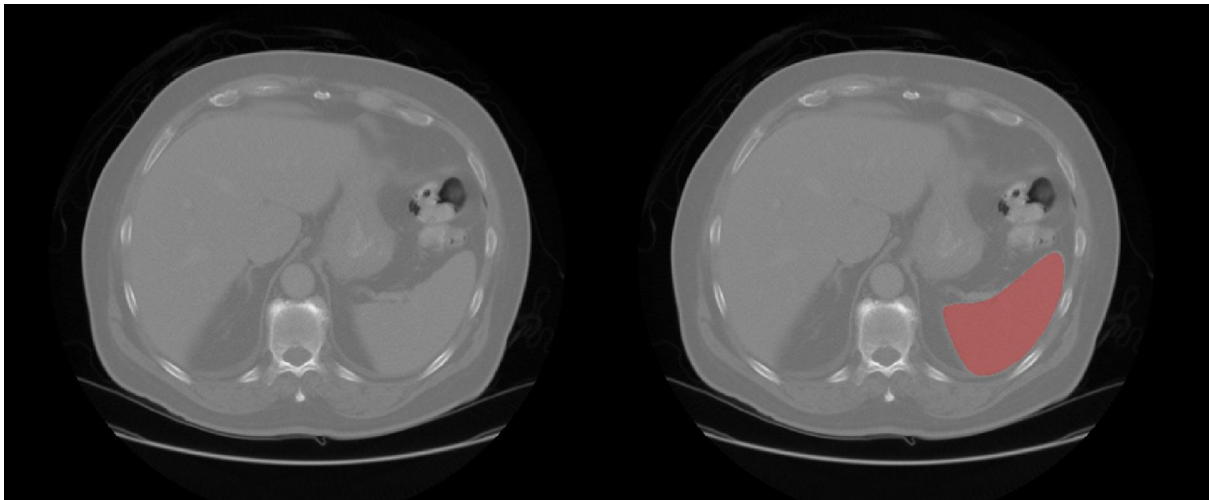
αποκτήθηκε πριν από τη χορήγηση παράγοντα αντίθεσης και οι δύο μετά.



Σχήμα 3: Η μικρή περιοχή με υψηλή ένταση στο κέντρο αποτελεί όγκο του μαστού.

2.4.5. Κατάτμηση σπληνός

Η κατάτμηση σπληνός επιλέχθηκε ως παράδειγμα κατάτμησης οργάνων, καθώς πολλές εφαρμογές απαιτούν την κατάτμηση συμπαγών οργάνων. Ειδικά για σπλήνα, ο γεωμετρικός όγκος του καθώς και αλλαγές στο μέγεθός του με την πάροδο του χρόνου, μπορούν να επηρεάσουν τις αποφάσεις των γιατρών [39], [40]. Γενικά, η κατάτμηση οργάνων διαφέρει σημαντικά ως πρόβλημα από την κατάτμηση όγκων, επειδή τα όργανα τείνουν να είναι μεγαλύτερα και να έχουν ομαλότερα όρια.



Σχήμα 4: Αξονική τομογραφία που απεικονίζει σπλήνα (επισημασμένη περιοχή).

2.5. Μηχανές Διανυσμάτων Υποστήριξης

Η μέθοδος που αναπτύχθηκε σε αυτή την εργασία χρησιμοποιεί ένα σύνολο [41] από Μηχανές Διανυσμάτων Υποστήριξης (ΜΔΥ - Support Vector Machines) [42] για να πραγματοποιήσει προβλέψεις. Τα ΜΔΥ χρησιμοποιούνται για ταξινόμηση (classification) πολλαπλών ετικετών, όπου κάθε pixel/voxel θεωρείται ανεξάρτητο δείγμα. Ωστόσο, τα ΜΔΥ μπορούν να χρησιμοποιηθούν και για παλινδρόμηση (regression), αντί για ταξινόμηση, σε άλλες εφαρμογές.

Η λειτουργία των ΜΔΥ γίνεται ευκολότερα κατανοητή για ταξινόμηση δειγμάτων σε δύο κλάσεις (κατηγορίες). Κάθε δείγμα χαρτογραφείται ως ένα σημείο σε έναν χώρο n διαστάσεων, όπου n ο αριθμός των χαρακτηριστικών (features) του προβλήματος. Τα ΜΔΥ προσπαθούν να διαχωρίσουν τα δείγματα χρησιμοποιώντας ένα $(n-1)$ -διάστατο υπερεπίπεδο, τυπικά επιλέγοντας ένα υπερεπίπεδο που διαχωρίζει καλύτερα τα δεδομένα εκπαίδευσης σε 2 κλάσεις. Κατά την παραγωγή των προβλέψεων, τα δείγματα χαρτογραφούνται στον ίδιο n -διάστατο χώρο και δείγματα που βρίσκονται σε αντίθετες πλευρές του υπερεπιπέδου ανατίθενται σε διαφορετικές κλάσεις. Για παράδειγμα, εάν

υπάρχουν δύο *χαρακτηριστικά*, τότε το n είναι ίσο με 2 και συνεπώς το διαχωριστικό υπερεπίπεδο είναι μια ευθεία γραμμή.

Ο γραμμικός διαχωρισμός είναι γρήγορος, αλλά δεν μπορεί να προσφέρει βέλτιστα αποτελέσματα για τα περισσότερα προβλήματα. Το προηγούμενο παράδειγμα είναι γραμμικός διαχωρισμός, όπου χρησιμοποιείται μια ευθεία γραμμή για διαχωρισμό δειγμάτων με δύο *χαρακτηριστικά*. Υπάρχει, ωστόσο, η ανάγκη για χρήση καμπυλών αντί γραμμών, δηλαδή ανάγκη για μη γραμμικό διαχωρισμό. Για να αντιμετωπιστεί αυτό, τα ΜΔΥ χρησιμοποιούν το “τέχνασμα” πυρήνα (kernel trick). Χρησιμοποιώντας μη γραμμικές συναρτήσεις πυρήνα, που δεν είναι υπερβολικά απαιτητικές υπολογιστικά, ο χώρος χαρακτηριστικών μετατρέπεται σε έναν άλλο χώρο, υψηλότερων διαστάσεων. Σε αυτό τον χώρο, τα αρχικά χαρακτηριστικά μετασχηματίζονται και συνδυάζονται χρησιμοποιώντας την εκάστοτε συνάρτηση πυρήνα. Ως αποτέλεσμα, ο γραμμικός τρόπος διαχωρισμού, που περιγράφεται στην προηγούμενη παράγραφο, μπορεί να χρησιμοποιηθεί στον νέο χώρο χαρακτηριστικών για την εκτέλεση μη γραμμικού διαχωρισμού. Η επιλογή και η παραμετροποίηση της συνάρτησης πυρήνα εξαρτάται από την εκάστοτε εφαρμογή.

2.6. Επισκόπηση των μεθόδων κατάτμησης

Ένα επιτηρούμενο μοντέλο μηχανικής μάθησης εκπαιδεύεται σε δεδομένα εισόδου, για τα οποία υπάρχουν διαθέσιμες οι ορθές σημάνσεις, και παράγει μια πρόβλεψη (έξοδο) για νέα δεδομένα χωρίς σήμανση. Η μη επιτηρούμενη μάθηση δεν χρησιμοποιεί δεδομένα εκπαίδευσης με σημάνσεις, αλλά προσπαθεί να βρει μοτίβα στα δεδομένα και να τα ομαδοποιήσει αναλόγως. Οι ημι-επιτηρούμενες προσεγγίσεις χρησιμοποιούν μερικά δεδομένα εισόδου με σημάνσεις μαζί με μια υψηλή ποσότητα μη επισημασμένων, κάτι που είναι χρήσιμο όταν υπάρχει μεγάλη ποσότητα διαθέσιμων δεδομένων, αλλά λίγα από αυτά είναι επισημασμένα και το κόστος επισήμανσης είναι υψηλό. Οι περισσότερες σύγχρονες

πλήρως αυτόματες μέθοδοι βασίζονται στη βαθιά μάθηση (deep learning). Αυτές οι μέθοδοι συνήθως χρειάζονται μεγάλα σύνολα δεδομένων με σημάνσεις.

Ένας αλγόριθμος που χρησιμοποιείται εδώ και αρκετά χρόνια για επιτηρούμενη κατάτμηση είναι οι k -πλησιέστεροι γείτονες (k -nearest neighbors). Λειτουργεί με την προϋπόθεση ότι παρόμοια δείγματα τείνουν να βρίσκονται κοντά, όταν χαρτογραφούνται σε έναν πολυδιάστατο χώρο χαρακτηριστικών. Είναι ένας αργός αλγόριθμος που απαιτεί από τον χρήστη να επιλέξει το k χειροκίνητα, μέσω πειραματισμού. Κατά την ταξινόμηση, επιλέγονται τα πλησιέστερα, στο δείγμα δοκιμής, δείγματα εκπαίδευσης και ως αποτέλεσμα δίνεται η κλάση που αντιπροσωπεύεται συχνότερα στα k αυτά πλησιέστερα δείγματα εκπαίδευσης. Συνιστάται η επιλογή ενός μονού k για την αποφυγή ισοπαλιών.

Ένας άλλος παρόμοιος αλγόριθμος που χρησιμοποιείται είναι οι ΜΔΥ, που αναλύθηκαν σε ξεχωριστή ενότητα. Οι ΜΔΥ μπορούν επίσης να προσαρμοστούν για να εκτελέσουν ημι-επιτηρούμενη ταξινόμηση, με μεθόδους όπως οι Transductive ΜΔΥ (TSVM) [43], όπου ο αλγόριθμος μπορεί να χρησιμοποιήσει τόσο επισημασμένα δεδομένα εκπαίδευσης όσο και μη επισημασμένα για τη βελτίωση του μοντέλου.

Οι μέθοδοι συνόλων/συνένωσης (ensemble methods) συνδυάζουν πολλαπλά μοντέλα για να επιτύχουν καλύτερα αποτελέσματα από αυτά που παράγει κάθε μοντέλο ξεχωριστά. Συνήθως, οι μέθοδοι συνόλων συνδυάζουν αδύναμους ταξινομητές, δηλαδή μεθόδους που συνήθως έχουν υψηλή πόλωση ή διακύμανση, για να δημιουργήσουν έναν ισχυρό ταξινομητή. Λόγω της αντιστάθμισης μεταξύ πόλωσης και διακύμανσης, τα μεμονωμένα μοντέλα με χαμηλή πόλωση έχουν υψηλή διακύμανση και το αντίστροφο, αλλά μέσω του συνδυασμού μοντέλων η επίδραση ελαχιστοποιείται. Υπάρχουν τρεις βασικοί τρόποι για συνδυασμό των αποτελεσμάτων πολλαπλών ταξινομητών: bagging, boosting και stacking. Στο bagging, οι αδύναμοι ταξινομητές εκπαιδεύονται ανεξάρτητα και το τελικό αποτέλεσμα υπολογίζεται ως ο μέσος όρος όλων των ταξινομητών, μειώνοντας έτσι τη διακύμανση. Για

να εκπαιδευτεί ο κάθε ταξινομητής, επιλέγονται τυχαία δείγματα από τα δεδομένα εκπαίδευσης με αντικατάσταση, πράγμα που σημαίνει ότι δείγματα που επιλέγονται για έναν ταξινομητή έχουν την ίδια πιθανότητα να επιλεγούν και για έναν άλλο. Τα τυχαία δάση (random forests) είναι μια δημοφιλής τεχνική bagging. Σε τεχνικές boosting, τα δείγματα έχουν βάρη, που σημαίνει ότι υπάρχει μια μεταβλητή πιθανότητα να επιλεγούν για την εκπαίδευση ενός ταξινομητή. Αυτό συμβαίνει επειδή η εκπαίδευση σε τεχνικές boosting είναι μια διαδοχική διαδικασία και όχι παράλληλη όπως σε bagging. Όταν ένας ταξινομητής εκπαιδεύεται, τα δείγματα που ταξινομούνται εσφαλμένα λαμβάνουν υψηλότερο βάρος. Κατ' αυτόν τον τρόπο, οι ταξινομητές που εκπαιδεύονται στη συνέχεια, έχουν υψηλότερη πιθανότητα να εκπαιδευτούν στα δείγματα που προηγουμένως ταξινομήθηκαν εσφαλμένα. Αυτή η τεχνική μειώνει την πόλωση. Η τεχνική stacking (στοίβαξη) λειτουργεί με ετερογενείς αδύναμους ταξινομητές, δηλαδή συνδυάζει διαφορετικούς αλγόριθμους. Σε αυτή την περίπτωση, εκπαιδεύεται ένα νέο μοντέλο (meta-learner) που δέχεται ως είσοδο τα αποτελέσματα διαφόρων αδύναμων ταξινομητών.

Το Perceptron είναι ένας αλγόριθμος για δυαδική επιτηρούμενη ταξινόμηση. Δέχεται πολλαπλές εισόδους, όπου η κάθε μία έχει ένα βάρος. Οι σταθμισμένες εισοδοί αθροίζονται και περνούν μέσα από μια συνάρτηση ενεργοποίησης (activation function). Μπορεί επίσης να προστεθεί πόλωση που συμβάλλει στη μετατόπιση του αθροίσματος και στην αλλαγή του αποτελέσματος της συνάρτησης ενεργοποίησης. Η συνάρτηση ενεργοποίησης καθορίζει εάν η έξοδος ταξινομείται ως 1 ή 0. Υπάρχουν δύο τύποι τέτοιων συναρτήσεων, γραμμικές και μη γραμμικές, με τις μη γραμμικές να είναι πιο συνήθεις. Μερικές δημοφιλείς συναρτήσεις είναι οι Sigmoid, Tanh και ReLU (Rectified Linear Unit), με την τελευταία να είναι εξαιρετικά δημοφιλής στη βαθιά μάθηση. Η εκπαίδευση ενός perceptron είναι μια επαναληπτική διαδικασία και πρόκειται για πρόβλημα βελτιστοποίησης, όπου τα βάρη του perceptron ενημερώνονται με στόχο την ελαχιστοποίηση μιας συνάρτησης σφάλματος.

Τα τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks) είναι δομές με τρία επίπεδα/στρώματα. Το πρώτο στρώμα είναι η είσοδος, το τελευταίο η έξοδος και το μεσαίο ονομάζεται κρυφό. Οι νευρώνες (κόμβοι) στο κρυφό στρώμα και στο στρώμα εξόδου είναι perceptrons, ενώ οι κόμβοι στο στρώμα εισόδου περιέχουν την αριθμητική τιμή κάθε χαρακτηριστικού του δείγματος. Όλοι οι νευρώνες στο κρυφό στρώμα έχουν εισόδους από όλους τους νευρώνες του πρώτου στρώματος. Επιπλέον, όλοι οι νευρώνες στο κρυφό στρώμα συνδέονται με όλους τους εξερχόμενους νευρώνες. Δεδομένου ότι μπορούν να υπάρχουν πολλοί νευρώνες εξόδου, είναι πλέον δυνατή η ταξινόμηση με χρήση πολλαπλών ετικετών, κάτι που δεν μπορεί να συμβεί στο perceptron. Τα βάρη των νευρώνων ενημερώνονται κατά την εκπαίδευση, όπου υπολογίζεται μία τιμή σφάλματος μέσω μιας συνάρτησης κόστους, συγκρίνοντας την έξοδο που παράγει το νευρωνικό δίκτυο με την αναμενόμενη τιμή εξόδου για το δείγμα. Μέσω της οπίσθιας διάδοσης (back propagation) του σφάλματος, τα βάρη προσαρμόζονται για να ελαχιστοποιηθεί η συνάρτηση κόστους.

Τα βαθιά νευρωνικά δίκτυα είναι τεχνητά νευρικά δίκτυα που έχουν δύο ή περισσότερα κρυφά στρώματα. Τα βαθιά δίκτυα συνήθως δεν απαιτούν εξαγωγή/επιλογή χαρακτηριστικών πριν εκπαιδευτούν, καθώς μπορούν να μάθουν τα μοτίβα των δεδομένων δημιουργώντας ουσιαστικά τα δικά τους χαρακτηριστικά στα πρώτα στρώματα τους. Η χρήση τους έχει αυξηθεί ραγδαία τις τελευταίες δύο δεκαετίες, λόγω των εξελίξεων στην υπολογιστική ισχύ, τη διαθεσιμότητα δεδομένων και την ανάπτυξη νέων τεχνικών.

Τα συνελκτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNNs) είναι χρήσιμα σε εργασίες αναγνώρισης εικόνας, επειδή μπορούν να δεχτούν ολόκληρη την εικόνα ως είσοδο. Οι ιατρικές εικόνες μπορεί να αποτελούνται από εκατομμύρια pixels/voxels, αλλά τα CNNs καταφέρνουν να διαχειριστούν την πολυπλοκότητα του προβλήματος. Συνήθως ένα CNN έχει τρία στρώματα: ένα συνελκτικό, ένα συγκέντρωσης (pooling) και ένα πλήρως διασυνδεδεμένο. Στο συνελκτικό στρώμα, υπολογίζεται το γινόμενο ενός φίλτρου και των

μητρών όλων των τμημάτων της εικόνας, με σκοπό την εξαγωγή χαρακτηριστικών. Στο στρώμα συγκέντρωσης μειώνεται η διάσταση του πίνακα χαρακτηριστικών, συνδυάζοντας γειτονικά στοιχεία του, συνήθως διατηρώντας το μέγιστο (max pooling) ή το μέσο όρο (average pooling). Μετά από αυτά τα δύο στρώματα, η μήτρα χαρακτηριστικών έχει μειωθεί σημαντικά και χρησιμοποιείται ένα πλήρως διασυνδεδεμένο επίπεδο για να εντοπιστούν οι σχέσεις των δεδομένων, όπως συμβαίνει στο κρυφό στρώμα των τεχνητών νευρικών δικτύου που αναλύθηκαν προηγουμένως.

Τα συνελκτικά νευρωνικά δίκτυα που περιγράφηκαν είναι χρήσιμα για ταξινόμηση, αλλά όχι κατάτμηση. Για να προσαρμοστούν σε προβλήματα κατάτμησης, το U-Net και άλλες παρόμοιες αρχιτεκτονικές έχουν προταθεί και χρησιμοποιηθεί επιτυχώς στην ιατρική απεικόνιση. Η διαφορά είναι ότι εκτός από τη διαδρομή συστολής που μειώνει το μέγεθος των δεδομένων, υπάρχει επίσης μια διαδρομή διαστολής που αυξάνει τις διαστάσεις και επαναφέρει τα δεδομένα στο αρχικό μέγεθος εικόνας (εξαιρουμένων των καναλιών). Αυτό συμβαίνει επειδή στην κατάτμηση πρέπει να εκχωρηθεί μια ετικέτα σε κάθε μεμονωμένο pixel/voxel.

Το transfer learning είναι μια τεχνική μηχανικής μάθησης που χρησιμοποιείται κυρίως για προσαρμογή μοντέλων σε νέους τομείς από αυτούς που εκπαιδεύτηκαν αρχικά. Όσον αφορά την ιατρική απεικόνιση, η τεχνική είναι πιο χρήσιμη στη μαγνητική τομογραφία, επειδή εκεί υπάρχει υψηλή διακύμανση στις εντάσεις και την αντίθεση της εικόνας [44]. Λόγω αυτών των διακυμάνσεων, τα μοντέλα μηχανικής μάθησης δεν αποδίδουν καλά σε εικόνες που αποκτήθηκαν με διαφορετικά μηχανήματα από εκείνα με τα όποια εκπαιδεύτηκαν. Η κύρια ιδέα του transfer learning είναι η χρήση μεγάλων συνόλων δεδομένων για να εκπαιδευτούν ισχυρά μοντέλα και στη συνέχεια η χρήση ενός μικρού αριθμού επισημασμένων εικόνων από το νέο τομέα για προσαρμογή του μοντέλου. Οι *Adaptive MΔΥ* και το *TrAdaBoost* (Transfer AdaBoost) αποδίδουν καλύτερα από τους κανονικούς ταξινομητές MΔΥ στην

κατάτμηση ιατρικών εικόνων [45] ακόμη και όταν εκπαιδεύονται σε μικρό αριθμό εικόνων από τον τομέα προορισμού. Προσαρμογή σε νέο τομέα μπορεί να πραγματοποιηθεί επίσης και σε CNN [44] και να έχει ως αποτέλεσμα υψηλότερη απόδοση. Το transfer learning έχει επίσης επιτύχει στο να προσαρμόσει επιτυχή μοντέλα, που είχαν αρχικά εκπαιδευτεί για μη ιατρικούς σκοπούς, ώστε να πραγματοποιήσουν ιατρική κατάτμηση [46].

Το BraTS (Brain Tumor Segmentation - Κατάτμηση εγκεφαλικών όγκων) [32] είναι ένας ετήσιος διαγωνισμός από διάφορα συνεργαζόμενα ιδρύματα, όπου γίνεται αξιολόγηση και κατάταξη της απόδοσης αλγορίθμων κατάτμησης. Παρόμοιοι διαγωνισμοί υπάρχουν και για άλλες παθήσεις [47]–[49]. Οι αυτοματοποιημένες μέθοδοι που ανταγωνίζονται στο BraTS έχουν εξαιρετική απόδοση, αλλά εξακολουθούν να είναι λιγότερο ακριβείς και συνεπείς από χειροκίνητες κατατμήσεις γιατρών. Ωστόσο, οι διοργανωτές έδειξαν ότι αν γίνει συγχώνευση των αποτελεσμάτων των αλγορίθμων, μέσω ψηφοφορίας για κάθε pixel/voxel, τα αποτελέσματα είναι καλύτερα από όλους τους επιμέρους αλγόριθμους.

2.7. Συστήματα CAD

Ένα σύστημα λογισμικού δεν μπορεί ακόμα να αντικαταστήσει έναν ακτινολόγο στις περισσότερες περιπτώσεις. Κάτι τέτοιο θα απαιτούσε εξαιρετική αποτελεσματικότητα και προσαρμοστικότητα από τον αλγόριθμο. Επιπλέον, οι ασθενείς μπορεί να μην είναι έτοιμοι για μια τόσο θεμελιώδη αλλαγή και οι ασφαλιστικές εταιρείες μπορεί να διστάσουν να καλύψουν τέτοιες θεραπείες. Από την άλλη πλευρά, η διάγνωση από γιατρούς υπόκειται επίσης σε λάθη. Τέτοια σφάλματα μπορούν να προκληθούν από κόπωση, άγχος και άλλους παράγοντες [50], [51]. Ως αποτέλεσμα, ένας γιατρός θα μπορούσε να επωφεληθεί από ένα υπολογιστικό σύστημα που δίνει σωστά αποτελέσματα στις περισσότερες περιπτώσεις, καθώς η αλγοριθμική απόφαση μπορεί να λειτουργήσει ως πρόταση ή εναλλακτική γνώμη, αφήνοντας την τελική απόφαση σε αυτόν [52]–[55]. Ένα τέτοιο σύστημα ονομάζεται CAD,

και ενσωματώνει δύο πεδία: (i) την ανίχνευση με τη βοήθεια υπολογιστή (CADe - Computer Assisted Detection) και (ii) τη διάγνωση με τη βοήθεια υπολογιστή (CADx - Computer Assisted Diagnosis) [56]. Τα συστήματα CADe στοχεύουν στην ανίχνευση ασυνήθιστων δομών στο σώμα, όπως οι όγκοι. Τα συστήματα CADx στοχεύουν στη λήψη αποφάσεων σχετικά με την υγεία του ασθενούς, όπως η ταξινόμηση των όγκων ως κακοήθων ή μη. Παρόλο που τα CADe και CADx μπορούν να υπάρχουν ξεχωριστά, πολλά συστήματα CAD ενσωματώνουν στοιχεία και από τα δύο.

Οι πρώτες προσπάθειες για τη χρήση υπολογιστών με σκοπό την επίλυση ιατρικών προβλημάτων πραγματοποιήθηκαν στη δεκαετία του 1950. Από τη δεκαετία του 1970 είχε καταστεί σαφές ότι δεν είναι εύκολο να επιλυθεί το πρόβλημα με τη χρήση απλών υπολογιστικών τεχνικών και στατιστικής [56]. Όμως, τα τελευταία χρόνια, εξελίξεις στην υπολογιστική ισχύ και στις τεχνικές μηχανικής μάθησης προσφέρουν νέες δυνατότητες και έχουν πυροδοτήσει έντονο ερευνητικό ενδιαφέρον εκ νέου.

2.7.1. Συλλογή δεδομένων

Εάν ένα σύστημα CAD απαιτεί εκπαίδευση όπως απαιτούν τα περισσότερα που βασίζονται στη μηχανική μάθηση, τότε χρειάζεται ένα μεγάλο σύνολο δεδομένων. Αυτό το σύνολο πρέπει να περιέχει εικόνες που λαμβάνονται από πολλαπλά ιδρύματα, διαφορετικά ιατρικά μηχανήματα και διαφορετικές ομάδες ασθενών. Αυτά τα δεδομένα πρέπει να ελεγχθούν, ώστε να εξαιρεθούν ασθενείς για τους οποίους λείπουν κάποιες από τις εικόνες που χρειάζονται. Εναλλακτικά, οι απύσες εικόνες μπορούν να δημιουργηθούν με τεχνικές αύξησης των δεδομένων (*data augmentation*) [57].

2.7.2. Βήμα 1: Προεπεξεργασία

Στο πρώτο βήμα ενός CAD συστήματος, τα δεδομένα πρέπει να κανονικοποιηθούν και ενδεχομένως να εναρμονιστούν εάν έχουν ληφθεί χρησιμοποιώντας διαφορετικά μηχανήματα [58] ή διαφορετικές τεχνικές. Η κανονικοποίηση ελαχίστου-μεγίστου αφορά την κλιμάκωση των δεδομένων σε ένα εύρος (στη μηχανική μάθηση συνήθως χρησιμοποιείται το

[0,1]). Η κανονικοποίηση z-score αφορά τη μετατροπή των δεδομένων χρησιμοποιώντας τη μέση τιμή και την τυπική απόκλιση τους (στη μηχανική μάθηση συνήθως χρησιμοποιείται μηδενική μέση τιμή και τυπική απόκλιση ίση με ένα). Η εναρμόνιση δεδομένων αφορά την ελαχιστοποίηση των διαφορών μεταξύ εικόνων που αποκτήθηκαν από διαφορετικά μηχανήματα, έτσι ώστε το μοντέλο να μπορεί να μάθει τις πραγματικές διαφορές μεταξύ των δειγμάτων. Ανάλογα με την εφαρμογή, μπορεί επίσης να πραγματοποιηθεί επιπρόσθετη επεξεργασία των δεδομένων σε αυτό το βήμα, όπως τεχνικές αποθορυβοποίησης, με στόχο τη διευκόλυνση της μετέπειτα επεξεργασίας. Για εφαρμογές σχετικές με τον εγκέφαλο, συνήθως υπάρχει ένα επιπλέον βήμα, όπου το κρανίο αφαιρείται από την εικόνα, αφήνοντας μόνο τον εγκέφαλο για επεξεργασία.

2.7.3. Βήμα 2: Κατάτμηση

Προκειμένου να απομονωθούν οι δομές που πρέπει να αναλυθούν, απαιτείται κατάτμηση των εικόνων. Ο αλγόριθμος κατάτμησης μπορεί να είναι αυτόματος ή διαδραστικός, αν και ένας αυτόματος αλγόριθμος είναι προτιμητέος, καθώς όλες οι υπόλοιπες διαδικασίες των CAD συστημάτων δεν απαιτούν αλληλεπίδραση με τον χρήστη.

2.7.4. Βήμα 3: Εξαγωγή χαρακτηριστικών

Η κατάτμηση συνήθως ακολουθείται από την εξαγωγή χαρακτηριστικών, μια διαδικασία που στοχεύει στη μείωση των δεδομένων σε κάποιες αριθμητικές τιμές (χαρακτηριστικά) που σχετίζονται με το πρόβλημα που αντιμετωπίζεται [52], [54], [59], [60].

Μια δημοφιλής μέθοδος για την εξαγωγή χαρακτηριστικών ιατρικών εικόνων ονομάζεται *radiomics*. Το "pyradiomics" [61], ένα δημοφιλές python πακέτο για radiomics, διακρίνει τα χαρακτηριστικά που μπορεί να εξαγάγει στις ακόλουθες κατηγορίες: (i) Χαρακτηριστικά "πρώτου βαθμού" (First Order) που σχετίζονται με την κατανομή των εντάσεων των pixels/voxels στην περιοχή που κατατμήθηκε (ii) Χαρακτηριστικά "σχήματος" (Shape) που περιγράφουν το σχήμα της περιοχής ενδιαφέροντος σε 2D ή 3D (iii) Χαρακτηριστικά "GLCM" (Gray Level Co-occurrence Matrix) που σχετίζονται με την από κοινού συνάρτηση

πιθανότητας δεύτερης τάξης της περιοχής ενδιαφέροντος (iv) Χαρακτηριστικά "GLSZM" (Gray Level Size Zone Matrix) που υπολογίζουν τον αριθμό των περιοχών που έχουν pixels/voxels με ίδια ένταση (vi) Χαρακτηριστικά "NGTDM" (Neighboring Gray Tone Difference Matrix) που υπολογίζουν τη διαφορά μεταξύ της έντασης ενός pixel/voxel με τη μέση τιμή έντασης των γειτόνων του σε μια προκαθορισμένη απόσταση και (vii) χαρακτηριστικά "Gray Level Dependence Matrix" που υπολογίζουν τον αριθμό των pixels/voxels που "εξαρτώνται" από καθένα από τα pixels/voxels της εικόνας. Ένα pixel/voxel θεωρείται ότι εξαρτάται από ένα άλλο εάν η απόλυτη τιμή της μεταξύ τους διαφοράς έντασης είναι μικρότερη από μια προκαθορισμένη τιμή.

Μια άλλη δημοφιλής τεχνική εξαγωγής χαρακτηριστικών είναι η *ανάλυση κύριων συνιστωσών* (PCA - Principal Component Analysis). Στο πλαίσιο ενός συστήματος CAD, η PCA χρησιμοποιείται για μείωση των διαστάσεων των πληροφοριών της περιοχής ενδιαφέροντος. Αρχικά, κατασκευάζεται ένας NxN συμμετρικός πίνακας συνδιακύμανσης, όπου N είναι ο αριθμός των pixels/voxels, επειδή απαιτείται μια τιμή για κάθε ζεύγος pixel/voxel. Κάθε τιμή στον πίνακα αντιπροσωπεύει τη συνδιακύμανση του αντίστοιχου ζεύγους. Εάν μια τιμή είναι θετική, τότε υπάρχει θετική συσχέτιση και οι δύο μεταβλητές αυξάνονται/μειώνονται μαζί, ενώ το αντίστροφο ισχύει για τις αρνητικές τιμές. Η PCA κατασκευάζει νέες μεταβλητές, που ονομάζονται κύριες συνιστώσες, που δεν συσχετίζονται μεταξύ τους [62] και είναι ίσες σε αριθμό με τις μεταβλητές εισόδου. Ωστόσο, η PCA περικλείει όσο το δυνατόν περισσότερη πληροφορία στις πρώτες κύριες συνιστώσες, δηλαδή μεγιστοποιεί τη διακύμανση σε αυτές. Ως αποτέλεσμα, επιλέγοντας ως χαρακτηριστικά έναν μικρό αριθμό από τις πρώτες κύριες συνιστώσες, διατηρείται υψηλό ποσοστό της πληροφορίας ενώ οι διαστάσεις του προβλήματος μειώνονται. Η ανάλυση ανεξάρτητων συνιστωσών (ICA - Independent Component Analysis) είναι μια παρόμοια τεχνική που χρησιμοποιείται για το διαχωρισμό ενός σήματος στις υποκείμενες ανεξάρτητες συνιστώσες του. Δεν επικεντρώνεται όμως στη μεγιστοποίηση της διακύμανσης στις πρώτες

συνιστώσες, αλλά εστιάζει στην ανεξαρτησία των συνιστωσών. Μια άλλη τεχνική για να επιτευχθεί μείωση των διαστάσεων του προβλήματος βασίζεται στη χρήση τεχνητού νευρωνικού δικτύου (ANN - Artificial Neural Network) με μη εποπτευόμενο (unsupervised) τρόπο εκπαίδευσης (*autoencoder*). Έχοντας μία συμφόρηση (bottleneck) στο κρυφό επίπεδο, οι πληροφορίες της εικόνας εισόδου συμπιέζονται, επειδή το δίκτυο προσπαθεί να μάθει τις συσχετίσεις των εισόδων.

2.7.5. Βήμα 4: Επιλογή χαρακτηριστικών

Η "κατάρα των πολλών διαστάσεων" (dimensionality curse) είναι ένας όρος που περιγράφει ζητήματα που σχετίζονται με την χρήση δεδομένων μεγάλων διαστάσεων. Ο κύριος λόγος είναι ότι τα δεδομένα υψηλών διαστάσεων τείνουν να είναι αραιά και επηρεάζουν αρνητικά τα αποτελέσματα αλγορίθμων που απαιτούν στατιστική σημαντικότητα. Τα συστήματα CAD αντιμετωπίζουν τέτοια προβλήματα και ως εκ τούτου η *επιλογή χαρακτηριστικών* ακολουθεί συνήθως την εξαγωγή χαρακτηριστικών [53], [57]. Κατ' αυτόν τον τρόπο, επιλέγονται τα πιο σημαντικά χαρακτηριστικά από το χώρο των χαρακτηριστικών και χρησιμοποιούνται για την εκπαίδευση του μοντέλου. Με την επιλογή χαρακτηριστικών, ο χρόνος εκπαίδευσης μειώνεται και υπάρχει χαμηλότερη πιθανότητα υπερεκπαίδευσης (overfitting). Υπάρχουν τρεις κύριες κατηγορίες τεχνικών επιλογής χαρακτηριστικών, οι οποίες περιγράφονται παρακάτω [63].

Στην πρώτη κατηγορία ανήκουν τα *φίλτρα*. Τεχνικές αυτού του τύπου δεν αλληλεπιδρούν με τον ταξινομητή, που δέχεται απλώς τα επιλεγμένα χαρακτηριστικά ως είσοδο. Τα φίλτρα χρησιμοποιούν στατιστικές μεθόδους που είναι ανεξάρτητες από την εφαρμογή, για να κατηγοριοποιήσουν τα χαρακτηριστικά. Μπορούν να διαιρεθούν περαιτέρω σε φίλτρα μιας ή πολλών μεταβλητών. Τα φίλτρα μίας μεταβλητής δεν χρησιμοποιούν τις εξαρτήσεις ανάμεσα στα χαρακτηριστικά στους υπολογισμούς τους, κάτι που μπορεί να οδηγήσει στην επιλογή παρόμοιων χαρακτηριστικών, ενώ τα φίλτρα πολλαπλών μεταβλητών, λαμβάνουν υπόψη τις

εξαρτήσεις αυτές, με κόστος στην ταχύτητα επεξεργασίας. Τα πιο συνηθισμένα φίλτρα βασίζονται σε συσχετισμό Pearson πολλαπλών μεταβλητών.

Μια άλλη κατηγορία είναι η *ενσωματωμένη* (embedded) επιλογή χαρακτηριστικών. Αυτές οι μέθοδοι είναι ειδικές για την εκάστοτε τεχνική μηχανικής μάθησης που χρησιμοποιείται και επιλέγουν χαρακτηριστικά κατά την εκπαίδευση του μοντέλου, δηλαδή μπορούν να θεωρηθούν ως μέρος του ταξινομητή. Ένα παράδειγμα επιλογής ενσωματωμένων χαρακτηριστικών είναι η αναδρομική εξάλειψη χαρακτηριστικών (recursive feature elimination) για μηχανές διανυσμάτων υποστήριξης.

Η τελευταία κατηγορία είναι οι *μέθοδοι περιτυλίγματος* (wrapper). Αυτές οι μέθοδοι χρησιμοποιούν την απόδοση του ταξινομητή για να αξιολογήσουν την καταλληλότητα των χαρακτηριστικών. Πρόκειται για μια επαναληπτική διαδικασία που τείνει να είναι πιο αργή και λιγότερο επεκτάσιμη από τις άλλες δύο, αλλά μπορεί να παραγάγει καλά αποτελέσματα για τεχνικές μηχανικής μάθησης που δεν ενσωματώνουν επιλογή χαρακτηριστικών.

2.7.6. Βήμα 5: Ανάλυση

Το τελικό βήμα είναι συνήθως ο κύριος αλγόριθμος του CAD που εκτελεί μια εργασία ταξινόμησης ή παλινδρόμησης. Αυτοί οι αλγόριθμοι έχουν ένα αποτέλεσμα που μπορεί να ερμηνευθεί από τον γιατρό, που αποτελεί μία απόφαση για την υγεία του ασθενούς ή την ύπαρξη ανωμαλιών στο σώμα. Αξίζει να σημειωθεί ότι ορισμένοι αλγόριθμοι βαθιάς μάθησης (deep learning) μπορούν να εκπαιδευτούν χρησιμοποιώντας ολόκληρη την εικόνα ως είσοδο και δεν απαιτούν ξεχωριστή κατάτμηση και εξαγωγή/επιλογή χαρακτηριστικών.

2.7.7. Ροή εργασίας συστήματος CAD σε νοσοκομειακό περιβάλλον

Σε νοσοκομειακό περιβάλλον, ένα σύστημα CAD λαμβάνει τις ιατρικές εικόνες ενός ασθενούς ως είσοδο, πιθανώς παράλληλα με πρόσθετες μετα-πληροφορίες όπως η ηλικία του ασθενούς. Το πρώτο βήμα είναι η προεπεξεργασία των εικόνων, ώστε να πραγματοποιηθεί κανονικοποίηση/εναρμόνιση αυτών με τα δεδομένα εκπαίδευσης. Η

κατάτμηση, η εξαγωγή/επιλογή χαρακτηριστικών και η ανάλυση πραγματοποιούνται σύμφωνα με τα εκπαιδευμένα μοντέλα και το αποτέλεσμα παρουσιάζεται στον γιατρό. Εξαιρουμένης της περίπτωσης όπου το βήμα της κατάτμησης είναι διαδραστικό, η διαδικασία μπορεί να αυτοματοποιηθεί και να πραγματοποιείται χωρίς τους γιατρούς, οι οποίοι δεν χρειάζεται να γνωρίζουν τίποτα για τη διαδικασία. Σε σύγχρονα συστήματα, το λογισμικό CAD μπορεί να εκτελείται απευθείας στο PACS, επιτρέποντας στον γιατρό να έχει πρόσβαση στο αποτέλεσμα της κατάτμησης και της αυτόματης διάγνωσης/ανίχνευσης όταν βλέπει τις ιατρικές εικόνες για πρώτη φορά μέσω του PACS.

2.8. Προσαρμοστικοί χάρτες γεωδαιτικής απόστασης

Οι προσαρμοστικοί χάρτες γεωδαιτικής απόστασης (AGD - Adaptive Geodesic Distance) [16], [64] χρησιμοποιούνται στην προτεινόμενη μέθοδο ως πρόσθετη είσοδος για την εκπαίδευση των μοντέλων. Ο σκοπός τους είναι να υπολογίσουν επιπλέον πληροφορίες σχετικά με τη δομή, που ένα μοντέλο, το οποίο χρησιμοποιεί pixels/voxels ως δείγματα εκπαίδευσης, θα δυσκολευόταν να συναγάγει.

Για την δημιουργία ενός χάρτη AGD, χρειάζονται δύο εικόνες, που έχουν ακριβώς το ίδιο μέγεθος: (i) μία ιατρική εικόνα, (ii) μία δυαδική εικόνα, δηλαδή εικόνα με τιμές 0 ή 1 σε κάθε pixel/voxel. Δεδομένου ότι αυτές οι εικόνες έχουν το ίδιο μέγεθος, κάθε pixel/voxel στη δυαδική εικόνα μπορεί να αντιστοιχιστεί σε ένα συγκεκριμένο pixel/voxel στην ιατρική εικόνα. Οι θετικές τιμές στη δυαδική εικόνα αντιπροσωπεύουν τα σχέδια ενός χρήστη. Πιο συγκεκριμένα, τα θετικά pixels/voxels στη δυαδική εικόνα αντιπροσωπεύουν pixels/voxels στην ιατρική εικόνα που ο χρήστης θεώρησε ότι ανήκουν στη δομή που θέλει να κατατμήσει.

Η γεωδαισία γενικεύει την έννοια των ευθειών γραμμών σε καμπύλους χώρους. Ένα παράδειγμα είναι η προσπάθεια υπολογισμού της απόστασης δύο πόλεων πάνω στη γη. Η ευκλείδεια απόστασή τους υπολογίζεται ως το μήκος μιας ευθείας γραμμής που συνδέει τις

δύο πόλεις, διαπερνώντας τη γη. Αντίθετα, ο υπολογισμός της γεωδαιτικής τους απόσταση θα ακολουθούσε την καμπυλότητα της γης. Το πρόβλημα μπορεί να περιγραφεί αλλιώς ως προσπάθεια εύρεσης της συντομότερης διαδρομής μεταξύ δύο σημείων, ενώ γίνεται κίνηση κατά μήκος μιας επιφάνειας.

Οι χάρτες AGD έχουν επίσης μία τιμή “απόστασης” σε κάθε *pixel/voxel*. Αυτή η “απόσταση” αντιπροσωπεύει τη συντομότερη διαδρομή μεταξύ του εκάστοτε *pixel/voxel* και του πλησιέστερου *pixel/voxel* που έχει επισημανθεί ως θετικό από τον χειριστή. Δεν χρησιμοποιείται όμως η ευκλείδεια απόσταση. Παρόμοια με τη γεωδαιτική, η απόσταση που υπολογίζεται αντιπροσωπεύει την συντομότερη διαδρομή μεταξύ των δύο σημείων ενώ γίνεται κίνηση κατά μήκος μιας επιφάνειας. Ωστόσο, η επιφάνεια καθορίζεται εν μέρει από το προφίλ έντασης της ιατρικής εικόνας [16]. Ως αποτέλεσμα, οι τιμές απόστασης αποτελούν μία σύνθεση της φυσικής απόστασης των δύο *pixel/voxel* και της ύπαρξης διαδρομής ανάμεσα τους με λίγες αυξομειώσεις έντασης. Με άλλα λόγια, έχουν υψηλότερες τιμές τα *pixels/voxels* μακριά από τα σχέδια του χρήστη ή/και τα *pixels/voxels* με αρκετές μεταβολές έντασης ανάμεσα σε αυτά και τα σχέδια.

Οι χάρτες μπορούν να υπολογιστούν γρήγορα, καθώς χρειάζεται επίσκεψη του κάθε *pixel/voxel* μόνο δύο φορές [16]. Το αποτέλεσμα είναι μια εικόνα που έχει τιμές στο $[0, +\infty)$. Οι χαμηλότερες τιμές αντιπροσωπεύουν μεγαλύτερη πιθανότητα το εκάστοτε *pixel/voxel* να ανήκει στην ίδια δομή με εκείνη που σχεδίασε εν μέρει από πάνω ο χρήστης. Οι χάρτες AGD έχουν χρησιμοποιηθεί στο παρελθόν για την εκτέλεση διαδραστικής κατάτμηση μίας δομής (κλάσης), η διαδικασία όμως απαιτούσε από τον χρήστη να επιλέξει ένα κατώφλι τιμών για την παραγωγή της τελικής κατάτμησης από τον χάρτη AGD. Αναλυτικά, επιλέγοντας ένα κατώφλι στο διάστημα $[0, +\infty)$, οι τιμές χαμηλότερες ή ίσες από το επιλεγμένο κατώφλι μετατρέπονται σε 1 στην τελική κατάτμηση, ενώ όλες οι άλλες τιμές σε 0. Αυτή η μέθοδος είναι γρήγορη και λειτουργεί ικανοποιητικά για απλές δομές, ωστόσο αντιμετωπίζει

προβλήματα σε πιο περίπλοκες εφαρμογές, καθώς δεν μπορεί να χειριστεί πολυπαραμετρικές εικόνες ή να εκτελέσει κατάτμηση πολλαπλών κλάσεων. Η επιλογή κατωφλίου είναι επίσης δύσκολη για τους χρήστες, καθώς η βέλτιστη επιλογή τιμής δεν είναι πάντα προφανής. Αυτή η εργασία ενσωματώνει τους χάρτες AGD ως πρόσθετη είσοδο σε ένα μοντέλο μηχανικής μάθησης, επιλύοντας έτσι όλες τις προαναφερθείσες αδυναμίες.

2.9. Έννοιες στατιστικής ανάλυσης

Ο συντελεστής Dice (Dice similarity coefficient - DSC) [65] ποσοτικοποιεί τη χωρική επικάλυψη δύο δυαδικών κατατμήσεων (δηλαδή κατατμήσεις με ετικέτες 0 και 1), λαμβάνοντας υπόψη τον αριθμό των pixel/voxel που έχουν κοινές και διαφορετικές τιμές. Όταν μία κατάτμηση δεν είναι δυαδική, αλλά έχει πολλές ετικέτες που αντιπροσωπεύουν διαφορετικές υποπεριοχές μιας δομής, οι επιμέρους υποπεριοχές μπορούν να μετατραπούν σε δυαδικές κατατμήσεις, ώστε να πραγματοποιηθεί η σύγκριση. Ομοίως, όλες οι ετικέτες, που είναι διάφορες του 0, μπορούν να συγχωνευτούν σε μία, για να ποσοτικοποιηθεί η επικάλυψη ανάμεσα σε δύο κατατμήσεις για ολόκληρη τη δομή. Για τον υπολογισμό του συντελεστή Dice για δυαδικές κατατμήσεις, χρησιμοποιήθηκε ο παρακάτω τύπος για κάθε υποπεριοχή, καθώς και για ολόκληρη τη δομή. TP (*true positive*) είναι ο αριθμός των pixel/voxel που έχουν τιμή 1 και στις δύο κατατμήσεις, FP (*false positive*) ο αριθμός των pixel/voxel που έχουν τιμή 1 στην κατάτμηση που εξετάζεται αλλά 0 στην κατάτμηση αναφοράς και FN (*false negative*) ο αριθμός των pixel/voxel που έχουν τιμή 0 στην κατάτμηση που εξετάζεται αλλά 1 στην κατάτμηση αναφοράς.

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

Τα αποτελέσματα του συντελεστή Dice δεν ακολουθούν την κατανομή Gauss. Επίσης, τα δείγματα στην παρούσα εργασία ήταν οργανωμένα κατά ζεύγη, επειδή για κάθε ασθενή οι

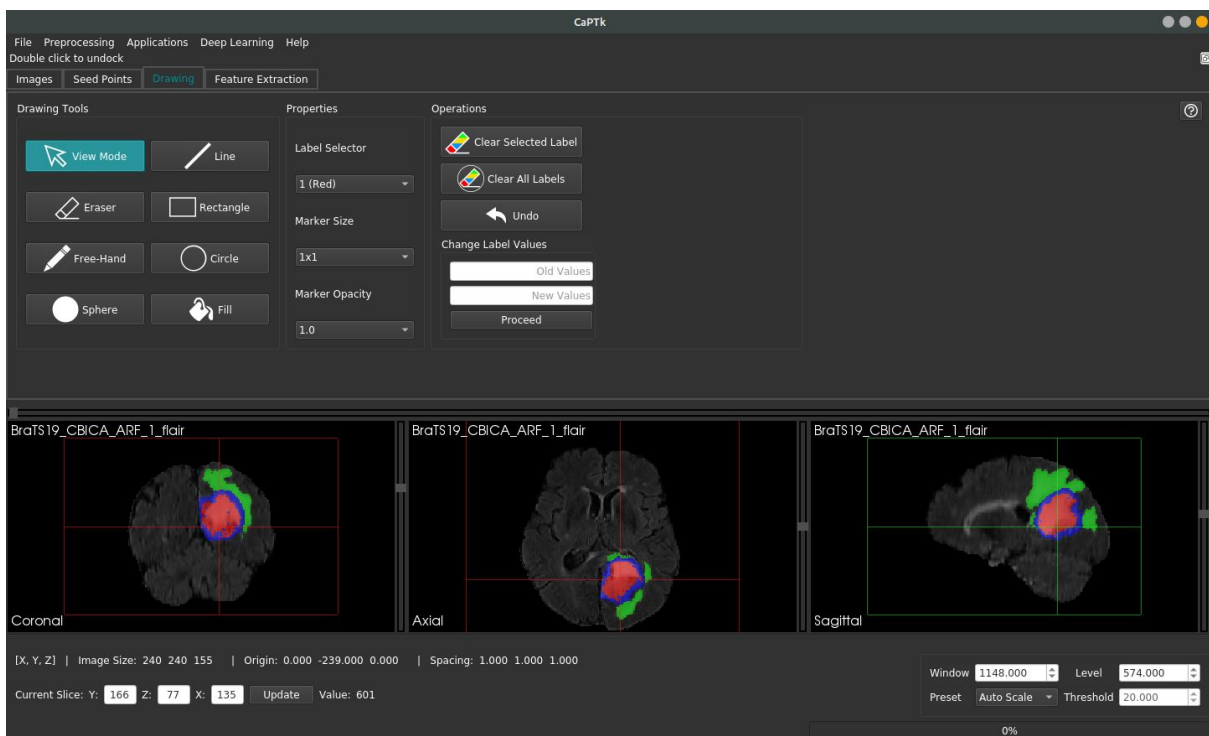
κατατμήσεις που συγκριθηκαν είχαν δημιουργηθεί από διαφορετικούς χειριστές, στάδια του πειράματος ή χρονικά σημεία. Κατά συνέπεια, το μη παραμετρικό τεστ Wilcoxon προσημασμένων τάξεων και συσχετισμένων δειγμάτων χρησιμοποιήθηκε ως κατάλληλη δοκιμασία για την εύρεση της ύπαρξης στατιστικά σημαντικών διαφορών.

Τέλος, ο συντελεστής συσχέτισης Pearson [66] χρησιμοποιήθηκε για τον προσδιορισμό της συσχέτισης μεταξύ των ζευγών γεωμετρικών όγκων των κατατμήσεων που αποκτήθηκαν χρησιμοποιώντας διαφορετικές προσεγγίσεις.

3. Υλοποίηση

3.1. Cancer Imaging Phenomics Toolkit (CaPTk)

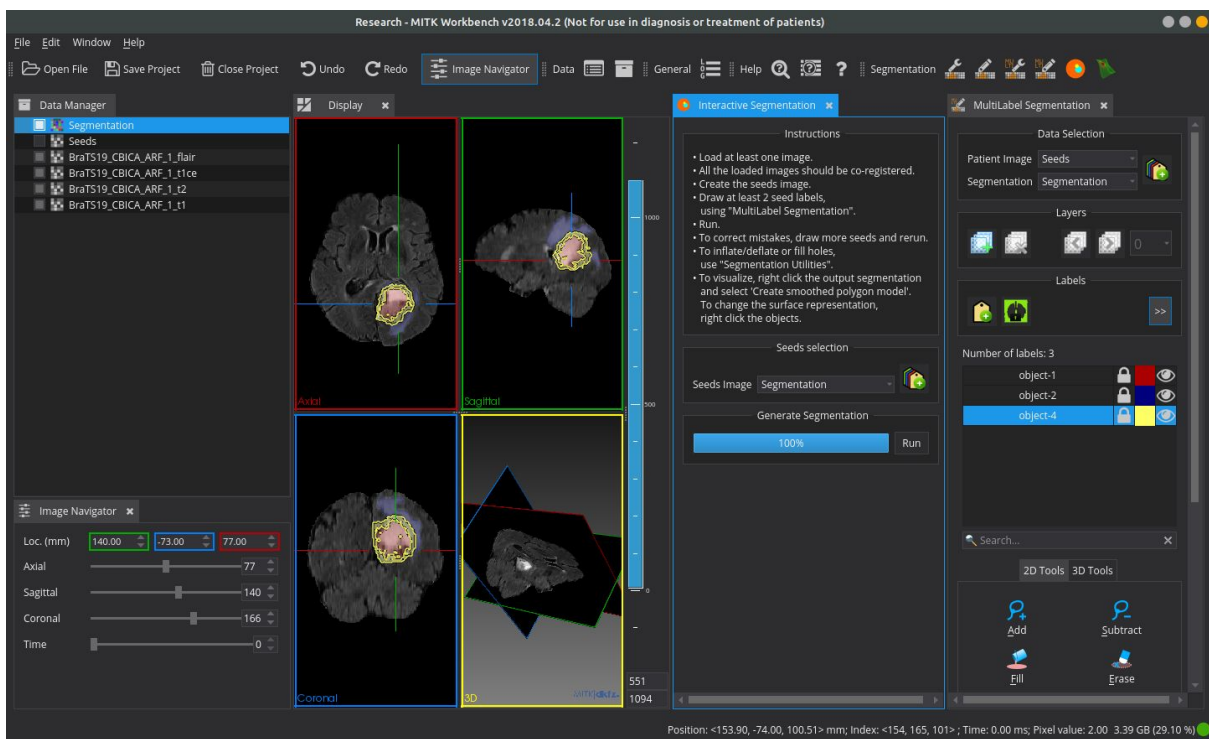
Το CaPTk [67] είναι ένα λογισμικό ανοιχτού κώδικα που μπορεί να λειτουργήσει σε διαφορετικές πλατφόρμες, το οποίο επικεντρώνεται στην έρευνα για τον καρκίνο και έχει αναπτυχθεί από το *Center for Biomedical Image Computing and Analytics (CBICA)* του *University of Pennsylvania*. Παρέχει λειτουργικότητα για προεπεξεργασία εικόνων, κατάτμηση, εξαγωγή χαρακτηριστικών (*feature extraction*) και εκπαίδευση μοντέλων μηχανικής μάθησης, καθώς και εξειδικευμένες εφαρμογές από υπολογιστικές μελέτες που διεξάγονται στο κέντρο. Η υλοποίηση της μεθόδου που παρουσιάζεται σε αυτή την εργασία αποτελεί μέρος της τρέχουσας έκδοσης του CaPTk και μπορεί να ληφθεί μέσω του συνδέσμου: <https://www.cbica.upenn.edu/captk>.



Σχήμα 5: Το γραφικό περιβάλλον του CaPTk.

3.2. Υλοποίηση ως πρόσθετος μηχανισμός MITK

Το MITK [23] είναι ένα *framework* ανοιχτού κώδικα που μπορεί να λειτουργήσει σε διαφορετικές πλατφόρμες, το οποίο χρησιμοποιείται για επεξεργασία ιατρικών εικόνων και έχει αναπτυχθεί από το Γερμανικό Κέντρο Έρευνας για τον Καρκίνο (DKFZ). Το MITK χρησιμοποιείται ευρύτατα τόσο σε ερευνητικά όσο και σε βιομηχανικά περιβάλλοντα. Υποστηρίζει επεκτάσεις μέσω ενός μηχανισμού *προσθέτων (plugins)*, ο οποίος χρησιμοποιήθηκε για μία δεύτερη υλοποίηση του αλγορίθμου. Συγκεκριμένα, μία MITK εφαρμογή (*MITK Workbench*), που περιέχει την μέθοδο που αναπτύχθηκε στο πλαίσιο αυτής της εργασίας ως πρόσθετο μηχανισμό, μπορεί να ληφθεί στη διεύθυνση: <https://github.com/CBICA/InteractiveSegmentation>.



Σχήμα 6: Το γραφικό περιβάλλον του *MITK Workbench*, μαζί με τον πρόσθετο μηχανισμό του αλγορίθμου που αναπτύχθηκε στο πλαίσιο αυτής της εργασίας.

3.3. Βιβλιοθήκες λογισμικού που χρησιμοποιήθηκαν στην υλοποίηση

Και οι δύο υλοποιήσεις του αλγορίθμου είναι σε C++ και το CMake χρησιμοποιήθηκε για τον έλεγχο της διαδικασίας μεταγλώττισης (compilation). Οι υπόλοιπες βιβλιοθήκες λογισμικού που χρησιμοποιήθηκαν περιγράφονται παρακάτω.

3.3.1. ITK

Το ITK [68], [69] είναι μια βιβλιοθήκη λογισμικού ανοιχτού κώδικα που παρέχει εργαλεία για ανάλυση ιατρικών εικόνων. Παρέχει λειτουργικότητα για φόρτωση και αποθήκευση n-διάστατων εικόνων, εύκολη προσπέλαση των pixels/voxels τους και εκτέλεση διαφόρων λειτουργιών επεξεργασίας εικόνων που διατίθενται ως "φίλτρα ITK". Παραδείγματα φίλτρων είναι η δειγματοληψία και η αποθορυβοποίηση.

3.3.2. OpenCV

Το OpenCV [70] είναι μια βιβλιοθήκη λογισμικού ανοιχτού κώδικα για όραση υπολογιστών και μηχανική μάθηση. Η κλάση `cv::ml::SVM` παρείχε τις ΜΔΥ που χρησιμοποιήθηκαν στην υλοποίηση. Για την παροχή εισόδου στο OpenCV, οι εικόνες μετατράπηκαν σε `cv::Mat` (υλοποίηση πινάκων του OpenCV) και στη συνέχεια τα αποτελέσματα μετατράπηκαν ξανά σε ITK εικόνες. Οι τρεις συναρτήσεις πυρήνα που χρησιμοποιήθηκαν ήταν η RBF (Radial Basis Function - Ακτινική συνάρτηση βάσης), η CHI2 (εκθετικό χ^2) και η INTER (Histogram Intersection - Διατομή ιστογράμματος). Η συνάρτηση `trainAuto()` χρησιμοποιήθηκε για την εκτέλεση της διασταυρούμενης επικύρωσης (*cross-validation*), την επιλογή υπερπαραμέτρων και την εκπαίδευση των μοντέλων. Τα προεπιλεγμένα πλέγματα (grids) υπερπαραμέτρων του OpenCV χρησιμοποιήθηκαν για την αναζήτηση των βέλτιστων τιμών, η οποία πραγματοποιείται σε κάθε κύκλο της διασταυρούμενης επικύρωσης για όλες τις παραμέτρους, εκτός από την παράμετρο C των RBF και CHI2, όπου οι τιμές που δοκιμάστηκαν κυμαίνονταν στο [1, 400] με λογαριθμικό βήμα 2,5.

3.3.3. MITK

Το *MITK framework* [23] χρησιμοποιήθηκε για την υλοποίηση του αλγορίθμου ως *πρόσθετου μηχανισμού MITK*. Είναι ένα framework που περιέχει πολλές βιβλιοθήκες που χρησιμοποιούνται συνήθως σε εφαρμογές ιατρικής απεικόνισης, όπως το ITK και το OpenCV. Επιπλέον, το MITK παρέχει πρόσθετες λειτουργίες υψηλού επιπέδου, όπως καλύτερη λειτουργία εισόδου/εξόδου των εικόνων και καλύτερη επεξεργασία των εικόνων μέσω των δομών δεδομένων που προσφέρει. Επίσης παρέχει διάφορα χρήσιμα γραφικά στοιχεία, όπως widgets για διαχείριση των δεδομένων και προβολή των ιατρικών εικόνων.

3.4. Βιβλιοθήκες λογισμικού για στατιστική ανάλυση

Τα πειράματα που πραγματοποιήθηκαν απαιτούσαν πολύπλοκη στατιστική ανάλυση που θα ήταν δύσκολο να εκτελεστεί χωρίς τη χρήση πηγαίου κώδικα, καθώς χρειάστηκαν πολλαπλοί υπολογισμοί για κάθε δομή, υποπεριοχή δομής, χρήστη της μεθόδου, καθώς και μελέτη των αποτελεσμάτων των πολλαπλών επαναλήψεων των πειραμάτων.

Η στατιστική ανάλυση έγινε με Python 3.6. Χρησιμοποιήθηκαν τα ακόλουθα πακέτα (packages): (1) *pandas 1.0.3* [71] για την αποθήκευση/ανάκτηση συντελεστών Dice σε μορφή CSV και την εκτέλεση λειτουργιών υψηλού επιπέδου σε αυτά τα δεδομένα. (2) *numpy 1.18.4* [72] για την εκτέλεση υπολογισμών στα δεδομένα, καθώς μπορεί να εκτελέσει ένα ευρύ φάσμα λειτουργιών χρησιμοποιώντας πίνακες. Επίσης, ο συντελεστής συσχέτισης (correlation coefficient) υπολογίστηκε χρησιμοποιώντας το *numpy* μέσω της συνάρτησης *corrcoef()*. (3) *scipy 1.4.1* [73] για το μη παραμετρικό Wilcoxon τεστ προσημασμένων τάξεων και συσχετισμένων δειγμάτων, καθώς και για υπολογισμό τον μεταβλητών του συντελεστών Dice. (4) *matplotlib 3.1.3* [74] για τη δημιουργία σχεδιαγραμμάτων.

4. Υλικά και Μέθοδοι

4.1. Δεδομένα

Τα πειράματα εγκρίθηκαν από την Επιτροπή Δεοντολογίας (Institutional Review Board - IRB) του University of Pennsylvania (UPenn, Philadelphia, ΗΠΑ). Η αξιολόγηση πραγματοποιήθηκε χρησιμοποιώντας δημόσια και ιδιωτικά κλινικά δεδομένα και διεξήχθησαν 4 αναδρομικές μελέτες (για Σπλήνα (3D-CT, n=20/41, Medical Segmentation Decathlon [75]), Όγκο του μαστού (2D-DCE-MRI, n=50, κλινική δοκιμή καταγραφής πολυπαραμετρικών εικόνων στο UPenn; NIH P01CA85484), Πνευμονικούς όζους (2D-CT, n=50/89, The Cancer Imaging Archive [76]–[78]), Γλοιοβλάστωμα εγκεφάλου (3D-MRI, n=20/335, BraTS'19 [3], [4], [32])). Τα υποσύνολα των δεδομένων που χρησιμοποιήθηκαν επιλέχθηκαν με τυχαίο τρόπο.

Τα δεδομένα για γλοιοβλάστωμα (11 άνδρες / 9 γυναίκες, μέση ηλικία: 62,84 / 64,36, εύρος ηλικίας: 44,82-77,48 / 39,64-77,09) και καρκίνο του μαστού (μόνο γυναίκες, μέση ηλικία: 50,41, εύρος ηλικίας: 32,68-71,97) αποκτήθηκαν τα έτη 2006-2014 και 2002-2006, αντίστοιχα. Τα δεδομένα σπληνός (13 άνδρες / 7 γυναίκες, μέση ηλικία: 63,85 / 58, εύρος ηλικίας: 40-81 / 48-68) και πνεύμονα (34 άνδρες / 16 γυναίκες) αποκτήθηκαν τα έτη 2000-2013 και 2004-2011, αντίστοιχα. Πληροφορίες σχετικά με την ηλικία των ασθενών δεν ήταν διαθέσιμες για τα δεδομένα του πνεύμονα.

Κατατμήσεις αναφοράς, δηλαδή κατατμήσεις που δημιουργήθηκαν χειροκίνητα και θεωρούνται επαρκώς εμπειριστατωμένες, ήταν διαθέσιμες για όλες τις εικόνες, εκτός από τα δεδομένα πνεύμονα, τα οποία δημιουργήθηκαν για τις ανάγκες της έρευνας από μία *fellowship trained* και *board certified* εξειδικευμένη ακτινολόγο θώρακος (Sharyn I. Katz MD/MTR, UPenn, 21 χρόνια εμπειρίας).

Πίνακας 1: Συνοπτική περιγραφή των δεδομένων που χρησιμοποιήθηκαν στην εργασία.

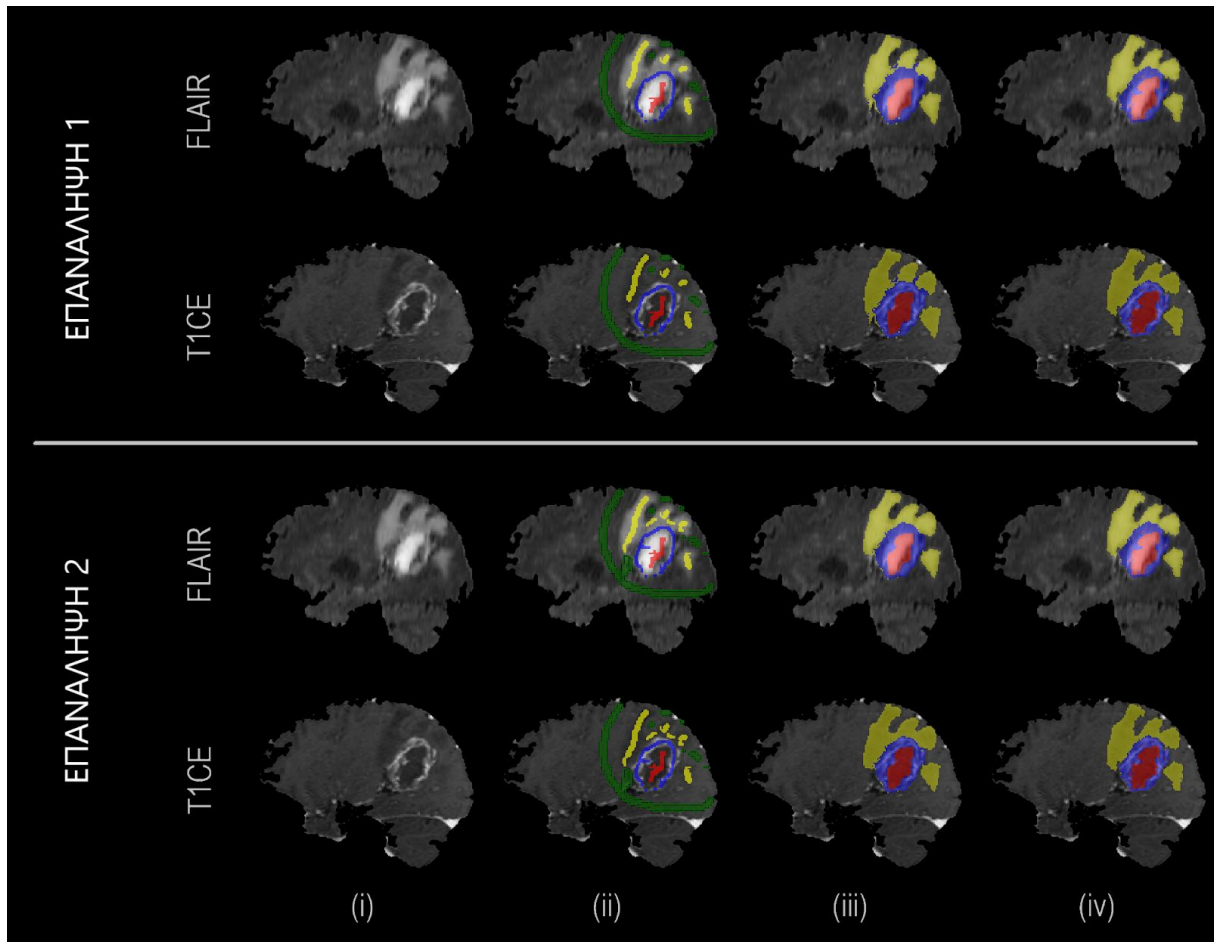
Μελέτη	Απεικονιστική Τεχνική	Αριθμός ασθενών
Σπληνός	3D-CT	20
Μαστού	2D-DCE-MRI: pre-contrast first post-contrast second post-contrast	50
Πνεύμονα	2D-CT	50
Γλοιοβλαστώματος	3D-MRI: T1 T2 T1CE (T1Gd) FLAIR	20

4.2. Προτεινόμενος αλγόριθμος

Ο αλγόριθμος μπορεί να κατατμήσει N περιοχές ενδιαφέροντος (Regions of Interest - ROIs) ταυτόχρονα, αξιοποιώντας $N+1$ σύνολα από *ετικέτες* εισόδου, όπου το επιπρόσθετο σύνολο ετικετών αντιπροσωπεύει πληροφορίες υποβάθρου (ιστοί οι οποίοι δεν είναι ενδιαφέροντος). Στο πρώτο βήμα του αλγορίθμου, ο χρήστης σχεδιάζει γρήγορα και αδρά πάνω από τις διαφορετικές περιοχές ενδιαφέροντος χρησιμοποιώντας τουλάχιστον δύο διαφορετικές ετικέτες (Σχήμα 7).

Για κάθε ζεύγος εικόνας και κλάσης ετικετών, παράγεται ένας χάρτης AGD [16] (Σχήμα 8) που αντικατοπτρίζει μία σύνθεση της χωρικής απόστασης και της διαφοράς έντασης από τα σχέδια του χρήστη, ούτως ώστε να έχουν υψηλότερες τιμές τα *voxels* που βρίσκονται μακριά ή έχουν διαφορετικό προφίλ έντασης (στο σχήμα, τέτοια voxel εμφανίζονται πιο φωτεινά). Επιπροσθέτως, χρησιμοποιούνται τρεις χάρτες “συντεταγμένων”, ένας για κάθε διάσταση της εικόνας, με σκοπό την παροχή περισσότερης χωρικής πληροφορίας στον αλγόριθμο. Οι

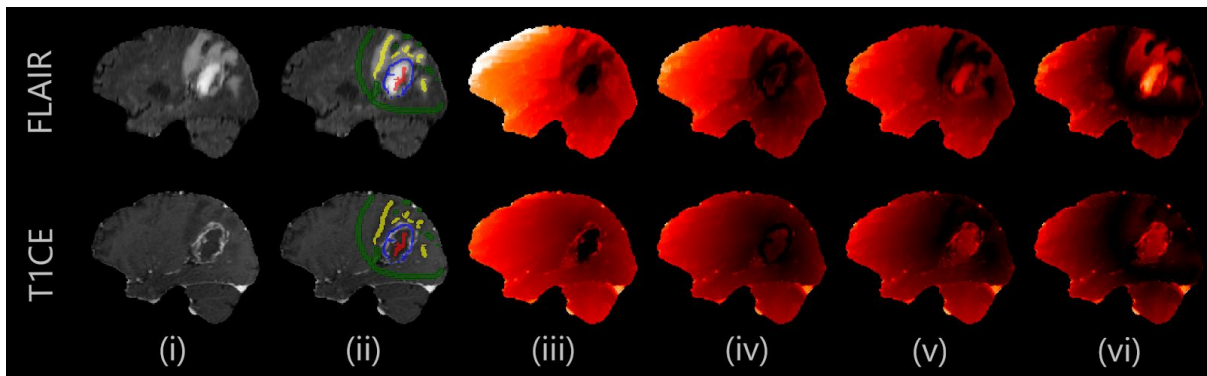
τιμές των χαρτών αυτών κυμαίνονται από μηδέν έως το μέγεθος της εικόνας στην εν λόγω διάσταση.



Σχήμα 7: Παράδειγμα που δείχνει το αποτέλεσμα να βελτιώνεται ως συνάρτηση του επενδυμένου χρόνου. Στην πρώτη επανάληψη (iteration), ο χρήστης σχεδιάζει γρήγορα πάνω από τις διάφορες περιοχές. Στη δεύτερη επανάληψη, ο χρήστης τοποθετεί επιπλέον ετικέτες για τη διόρθωση αντιπροσωπευτικών εσφαλμένων περιοχών, οι οποίες στη συνέχεια χρησιμοποιούνται για την επανεκπαίδευση του μοντέλου μηχανικής μάθησης. Από αριστερά προς τα δεξιά: (i) αρχική εικόνα (ii) σχέδια του χρήστη (iii) κατάτμηση που παράγει η μέθοδος (iv) κατάτμηση αναφοράς.

Ένα σύνολο [41] από ΜΔΥ εκπαιδεύεται σε voxels που ανήκουν στα σχέδια και κατόπιν χρησιμοποιείται για να κατατμήσει το υπόλοιπο μέρος των εικόνων της απεικονιστικής εξέτασης. Κάθε δείγμα εκπαίδευσης (δηλ. voxel) περιγράφεται από τα ακόλουθα χαρακτηριστικά: (i) ένταση σε όλες τις ευθυγραμμισμένες εικόνες, (ii) απόσταση σε όλους τους χάρτες AGD και (iii) τιμή σε όλους τους χάρτες “συντεταγμένων”. Τρία μοντέλα

ταξινόμησης (classification) ΜΔΥ εκπαιδεύονται, με συναρτήσεις πυρήνα: Ακτινικής Βάσης (Radial Basis Function - RBF), εκθετικού χ^2 (Chi-squared) και διατομής ιστογράμματος (Histogram Intersection). Οι υπερπαραμέτροί τους επιλέγονται μέσω της διαδικασίας διασταυρούμενης επικύρωσης. Η τελική πρόβλεψη για κάθε voxel επιτυγχάνεται με συγχώνευση, μέσω ψηφοφορίας, των τριών προβλέψεων από τα μοντέλα, ενώ ο ταξινομητής RBF χρησιμοποιείται για να επιλύσει ισοπαλίες.



Σχήμα 8: Παράδειγμα χαρτών AGD. Οι πιο σκοτεινές τιμές δείχνουν χαμηλότερη προσαρμοσμένη γεωδαιτική απόσταση (Adaptive Geodesic Distance - AGD) από τα σχέδια του χρήστη. Στο γλοιοβλάστωμα, το όριο NE και ET είναι σαφέστερο σε T1CE, ενώ το όριο μεταξύ ED και υποβάθρου είναι σαφέστερο στο FLAIR. Από αριστερά προς τα δεξιά: (i) αρχική εικόνα (ii) σχέδια του χρήστη (iii) χάρτης AGD για τις ετικέτες για NE (το περιοχικά μη-ενισχυμένο τμήμα του γλοιοβλαστώματος) (iv) χάρτης AGD για ET (το περιοχικά ενισχυμένο τμήμα του γλοιοβλαστώματος) (v) χάρτης AGD για ED (το περιεσσιακό οίδημα) ετικέτες (vi) χάρτης AGD για το φόντο.

4.3. Πειραματικός Σχεδιασμός

4.3.1. Το πρωτόκολλο αξιολόγησης

Στην αξιολόγηση της μεθόδου συμμετείχαν οκτώ ειδικοί, δύο για κάθε μελέτη. Ζητήθηκε από τον κάθε ειδικό να κατατμήσει τέσσερις φορές κάθε απεικονιστική εξέταση. Συγκεκριμένα, για κάθε ασθενή ο κάθε ειδικός παρήγαγε δύο χειροκίνητες και δύο IML-υποβοηθούμενες κατατμήσεις. Σημειώνεται ότι όλες αυτές οι κατατμήσεις ήταν επιπρόσθετες των εκτενώς καθορισμένων κατατμήσεων αναφοράς.

Στους ειδικούς δόθηκαν σύντομες οδηγίες για τη χρήση της μεθόδου και τους ζητήθηκε να καταγράψουν τον χρόνο που χρειάστηκαν για να παραγάγουν τις IML και τις χειροκίνητες κατατμήσεις. Για δικαιότερη αξιολόγηση της συνέπειας της κατάτμησης από διαφορετικούς ειδικούς για το γλοιοβλάστωμα, δόθηκε οδηγία στους ειδικούς να εκτελέσουν τη χειροκίνητη κατάτμηση των διαφόρων περιοχών του όγκου (περιοχικά ενισχυμένο τμήμα (enhancing tumor core - ET), περιοχικά μη-ενισχυμένο τμήμα (non-enhancing tumor core - NE) και περιεσσιακό οίδημα (peritumoral edematous/infiltrated tissue - ED) [3], [4]) σε 1 ώρα ή λιγότερο.

4.3.2. Πείραμα 1: Αξιολόγηση Συνολικής Απόδοσης

Αρχικά, αξιολογήθηκε η χωρική επικάλυψη κάθε προσέγγισης σε σχέση με τα δεδομένα αναφοράς, χρησιμοποιώντας Dice Similarity Coefficient (DSC) για να επιλεγεί μία IML-υποβοηθούμενη και μία χειροκίνητη κατάτμηση από κάθε ειδικό, για κάθε ασθενή. Για το γλοιοβλάστωμα, χρησιμοποιήθηκε μόνο η περιοχή *Whole tumor* (WT), δηλαδή η περιοχή που καλύπτεται από την ένωση των ετικετών NE, ET και ED, για αυτές τις επιλογές. Τα DSC των δύο συνόλων συγκρίθηκαν στατιστικά.

Επιπλέον, υπολογίστηκαν οι γεωμετρικοί όγκοι των IML και χειροκίνητων κατατμήσεων και συγκρίθηκαν ποσοτικά με τα δεδομένα αναφοράς. Τέλος, εκτιμήθηκε ο συντελεστής συσχέτισης Pearson [66] για καθένα από τα ζεύγη χωριστά, δηλαδή, (i) IML-υποβοηθούμενη κατάτμηση έναντι αναφοράς και (ii) χειροκίνητη έναντι αναφοράς. Ο μέσος ενεργός χρόνος σχεδίασης συγκρίθηκε για κάθε μελέτη μεταξύ IML και χειροκίνητων κατατμήσεων.

4.3.3. Πείραμα 2: Συνέπεια των κατατμήσεων μεταξύ δύο διαφορετικών επαναλήψεων του ίδιου ειδικού

Υπολογίστηκαν τα DSC μεταξύ των δύο IML κατατμήσεων του κάθε βαθμολογητή, καθώς και μεταξύ των δύο χειροκίνητων κατατμήσεων του (δηλ. $DSC_{IML1/IML2}$, και $DSC_{\text{Χειροκίνητο1}/\text{Χειροκίνητο2}}$), για κάθε ασθενή. Τα DSC των IML κατατμήσεων συγκρίθηκαν στατιστικά με τα χειροκίνητα DSC. Επιπλέον, εξετάστηκε, για κάθε ειδικό ξεχωριστά, η

ύπαρξη στατιστικά σημαντικών διαφορών μεταξύ του DSC των χειροκίνητων και των IML-υποβοηθούμενων κατατμήσεων σε σχέση με τα δεδομένα αναφοράς.

4.3.4. Πείραμα 3: Συνέπεια των κατατμήσεων από διαφορετικούς ειδικούς

Επιλέχθηκαν οι καλύτερες κατατμήσεις του κάθε ειδικού με τα ίδια κριτήρια επιλογής που χρησιμοποιήθηκαν στο πείραμα 1. Τα DSC μεταξύ των καλύτερων IML κατατμήσεων των δύο ειδικών και τα DSC μεταξύ των καλύτερων χειροκίνητων κατατμήσεων τους υπολογίστηκαν για κάθε ασθενή και εκτιμήθηκε το αν υπάρχει στατιστικά σημαντική διαφορά για κάθε μελέτη.

4.3.5. Στατιστική Ανάλυση

Χρησιμοποιήθηκε το παραμετρικό τεστ Wilcoxon προσημασμένων τάξεων και συσχετισμένων δειγμάτων [79] για στατιστικές συγκρίσεις (υποθέτοντας ποσοστό σφάλματος τύπου I ίσο με 0,05). Η επιλογή του τεστ βασίστηκε στο ότι τα δείγματα είναι συσχετισμένα και τα αποτελέσματα τείνουν να μην ακολουθούν κατανομή Gauss.

5. Αποτελέσματα

Σε αυτήν την ενότητα, παρουσιάζονται τα αποτελέσματα της πειραματικής αξιολόγησης για κατάτμηση 20 περιπτώσεων σπληνός, 50 καρκίνου του μαστού, 50 πνευμονικών όζων και 20 περιστατικών γλοιοβλαστώματος. Για κάθε ασθενή, ο κάθε ειδικός παρήγαγε δύο χειροκίνητες και δύο IML-υποβοηθούμενες κατατμήσεις, οι οποίες ήταν επιπρόσθετες των εκτενώς καθορισμένων κατατμήσεων αναφοράς.

5.1. Πείραμα 1: Αξιολόγηση Συνολικής Απόδοσης

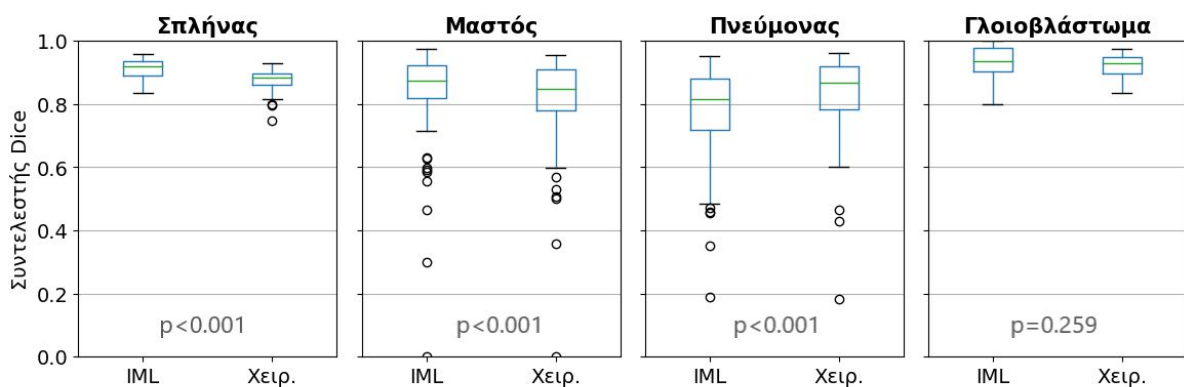
Στο πρώτο πείραμα, αξιολογήθηκε η απόδοση της προτεινόμενης μεθόδου (Πίνακας 3, Σχήμα 9). Για το γλοιοβλάστωμα, οι μη αυτόματες και οι IML-υποβοηθούμενες κατατμήσεις απέδωσαν παρόμοια ζεύγη DSC τόσο για την υποπεριοχή WT (Whole tumor - Συνολική περιοχή που καλύπτουν τα NE, ET και ED), όσο και για μεμονωμένες υποπεριοχές, δείχνοντας έτσι ότι δεν υπάρχει καμία σημαντική διαφορά μεταξύ τους, ενώ το αντίστροφο έδειξε να ισχύει για τις άλλες μελέτες. Η μέθοδος πέτυχε υψηλότερο DSC κατά μέσο όρο στις μελέτες για σπλήνα και καρκίνο του μαστού, αλλά χαμηλότερο για πνευμονικούς όζους, σε σύγκριση με τη χειροκίνητη προσέγγιση. Ωστόσο, η μέθοδος ήταν αρκετά ταχύτερη από τη χειροκίνητη κατάτμηση σε όλες τις μελέτες της αξιολόγησης (Πίνακας 3), κατά 53,1% κατά μέσο όρο.

Ανάλυση του γεωμετρικού όγκου των IML κατατμήσεων, μαζί με των χειροκίνητων κατατμήσεων και αναφοράς (Σχήμα 10, Πίνακας 3 στήλη “Συντ. Συσχ.”) δείχνει ότι τα λάθη που έγιναν με τη μέθοδο ήταν ως επί το πλείστον συστηματικά. Αυτό είναι πιο εμφανές σε εικόνες σπληνός, όπου οι IML κατατμήσεις και οι χειροκίνητες φαίνεται να εμφανίζουν συστηματική υπο-κατάτμηση και υπερ-κατάτμηση, αντίστοιχα. Η μέθοδος έκανε κάποια μη συστηματικά σφάλματα για πνευμονικούς όζους, καθώς και για την υποπεριοχή ET του

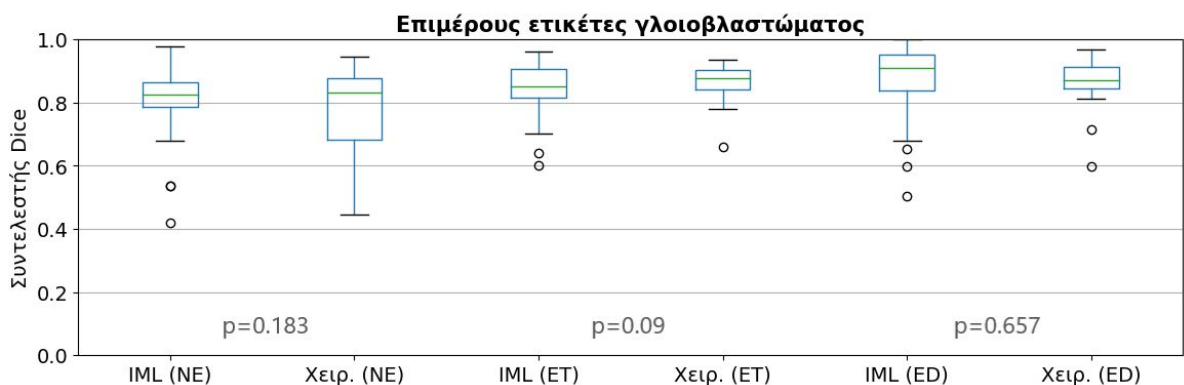
γλοιοβλαστώματος, αλλά αυτές οι δύο περιοχές ήταν επίσης περισσότερο λανθασμένες και στις χειροκίνητες κατατμήσεις. Συγκεκριμένα, η υποπεριοχή ET θεωρείται ως η πιο απαιτητική περιοχή του γλοιοβλαστώματος όσον αφορά την κατάτμηση, επειδή συχνά έχει ασαφή και λεία όρια [3].

Απόδοση

(i) Ολόκληρη η περιοχή ενδιαφέροντος



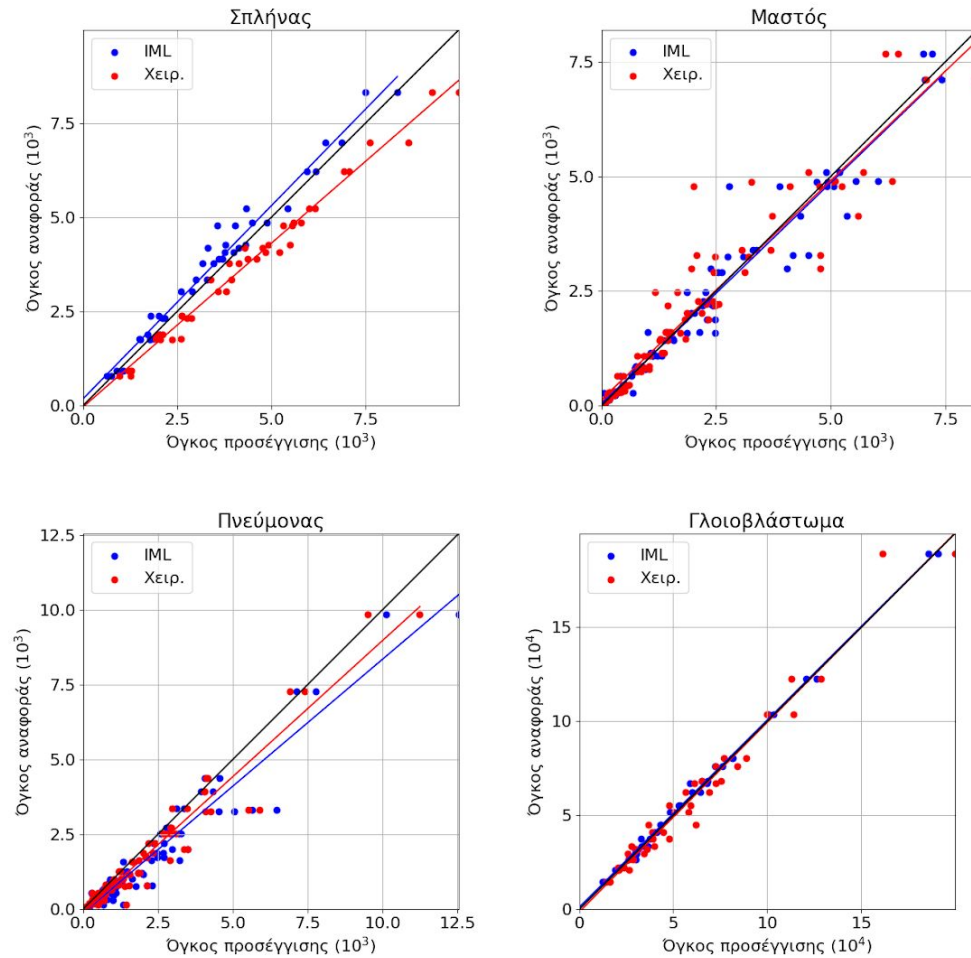
(ii) Επιμέρους ετικέτες



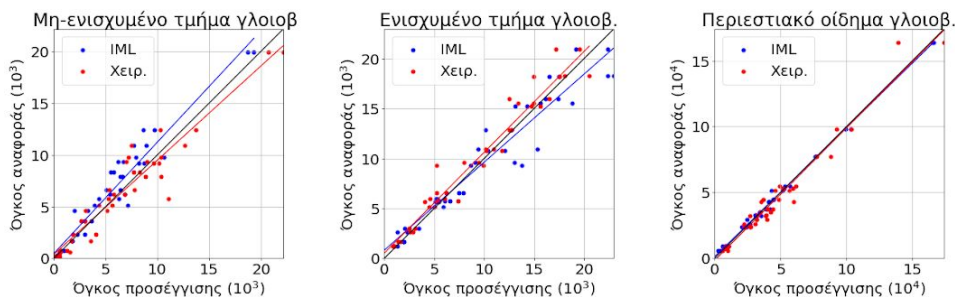
Σχήμα 9: Συντελεστής Dice, σε σύγκριση με τις κατατμήσεις αναφοράς, για τις περιπτώσεις όπου: (i) επιμέρους ετικέτες που αντιπροσωπεύουν διαφορετικές περιοχές της δομής υπολογίζονται ως μία, (ii) μεμονωμένες υποπεριοχές του γλοιοβλαστώματος (για τις άλλες δομές δεν υπάρχουν υποπεριοχές).

Ανάλυση όγκου κατάτμησης

(i) Ολόκληρη η περιοχή ενδιαφέροντος



(ii) Επιμέρους ετικέτες



Σχήμα 10: Scatterplots στα οποία τα μπλε σημεία είναι τα ζεύγη των γεωμετρικών όγκων των κατατμήσεων της IML μεθόδου και των γεωμετρικών όγκων των κατατμήσεων αναφοράς. Παρόμοια, τα κόκκινα σημεία είναι ζεύγη των όγκων των χειροκίνητων κατατμήσεων και κατατμήσεων αναφοράς. Η μαύρη γραμμή αντιπροσωπεύει τους όγκους των κατατμήσεων αναφοράς. Τα σχεδιαγράμματα ανήκουν σε (i) διαφορετικές μελέτες όπου όλες οι ετικέτες, που αντιπροσωπεύουν διαφορετικές υποπεριοχές της δομής, υπολογίζονται ως μία και (ii) υποπεριοχές του γλοιοβλαστώματος.

5.2. Πείραμα 2: Συνέπεια των κατατμήσεων μεταξύ δύο διαφορετικών επαναλήψεων του ίδιου ειδικού

Το δεύτερο πείραμα επιχειρεί να ποσοτικοποιήσει την συνέπεια μεταξύ διαφορετικών επαναλήψεων του ίδιου ειδικού (*intra-rater consistency*), συγκρίνοντας τις κατατμήσεις από τις δύο επαναλήψεις της αξιολόγησης, ξεχωριστά για IML και χειροκίνητη κατάτμηση (Πίνακας 3, Σχήμα 11). Δεν βρέθηκε στατιστικά σημαντική διαφορά μεταξύ των χειροκίνητων και των IML-υποβοηθούμενων κατατμήσεων για οποιαδήποτε από τις μελέτες, εκτός από τη μελέτη σπληνός, όπου οι κατατμήσεις που παρήχθησαν από τη μέθοδο είχαν υψηλότερη μέση χωρική επικάλυψη.

Πίνακας 2: Αποτελέσματα (p-τιμές) του Wilcoxon test με συσχετισμένα δείγματα, για κάθε ειδικό, συγκρίνοντας τα αποτελέσματα του συντελεστή Dice των διαφορετικών προσεγγίσεων (IML και χειροκίνητης κατάτμησης) σε σχέση με τα δεδομένα αναφοράς. Η διαφορά θεωρείται σημαντική όταν $p < 0,05$.

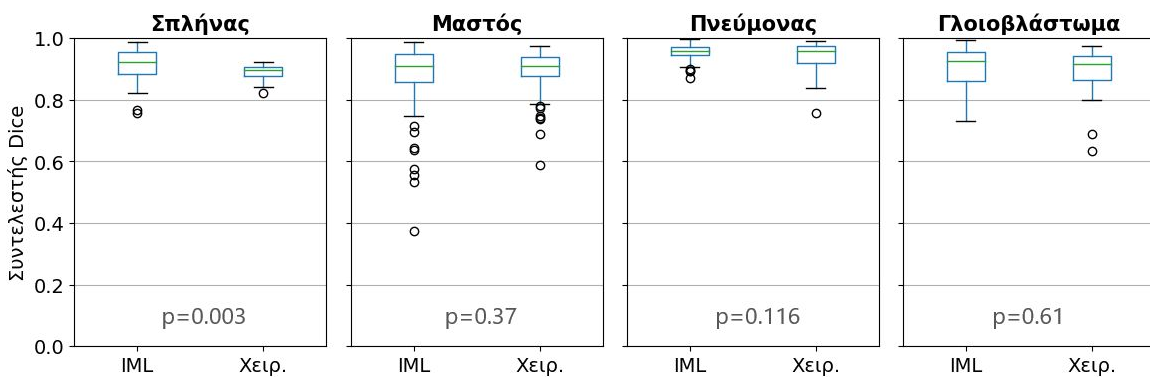
Ετικέτα	Ειδικός 1		Ειδικός 2	
	IML	Χειρ.	IML	Χειρ.
Σπλήνας				
-	0.9405	0.3507	0.1454	0.433
Μαστός				
-	0.2425	0.5116	0.5921	0.1358
Πνεύμονας				
-	0.0422	<0.0001	0.1358	<0.0001
Γλοιοβλάστωμα				
WT	0.156	0.0001	0.8813	0.0008
NE	0.0522	0.0859	0.9405	0.0001
ET	0.1913	0.3317	0.0522	0.0001
ED	0.4781	0.0001	0.6274	0.0008

Επιπρόσθετη ανάλυση του DSC σε σχέση με τις κατατμήσεις αναφοράς (Πίνακας 2) έδειξε στατιστικά σημαντική διαφορά, όσον αφορά την συνέπεια ανάμεσα σε δύο προσπάθειες του

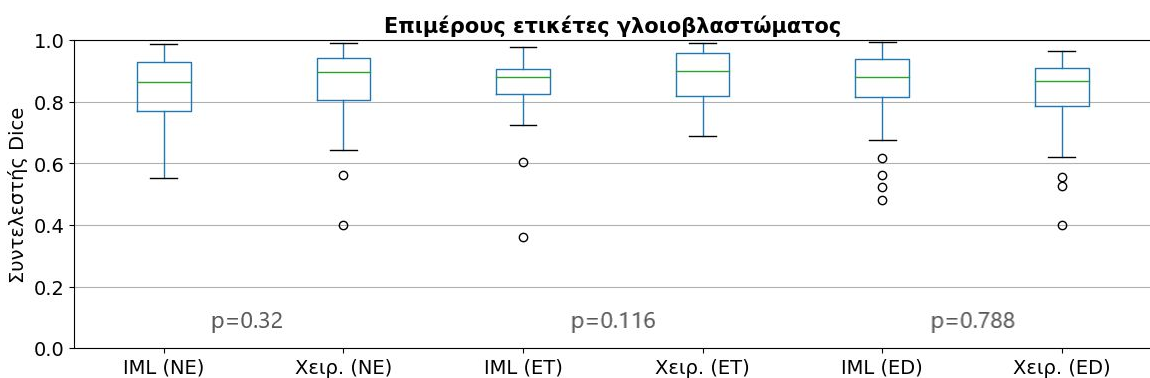
ίδιου ειδικού, για έναν μόνο από τους ειδικούς κατά τη χρήση της μεθόδου IML, ενώ έδειξε σημαντική διαφορά σε 4/8 ειδικούς για χειροκίνητη κατάτμηση. Επιπλέον, δεν υπήρχε σημαντική διαφορά κατά τη χρήση της μεθόδου IML για οποιονδήποτε από τους δύο ειδικούς για όλες τις υποπεριοχές του γλοιοβλαστώματος. Η ίδια ανάλυση για χειροκίνητη κατάτμηση έδειξε σημαντική διαφορά σε όλες τις υποπεριοχές, εκτός από το ET για έναν ειδικό.

Συνέπεια μεταξύ διαφορετικών επαναλήψεων του ίδιου ειδικού

(i) Ολόκληρη η περιοχή ενδιαφέροντος



(ii) Επιμέρους ετικέτες



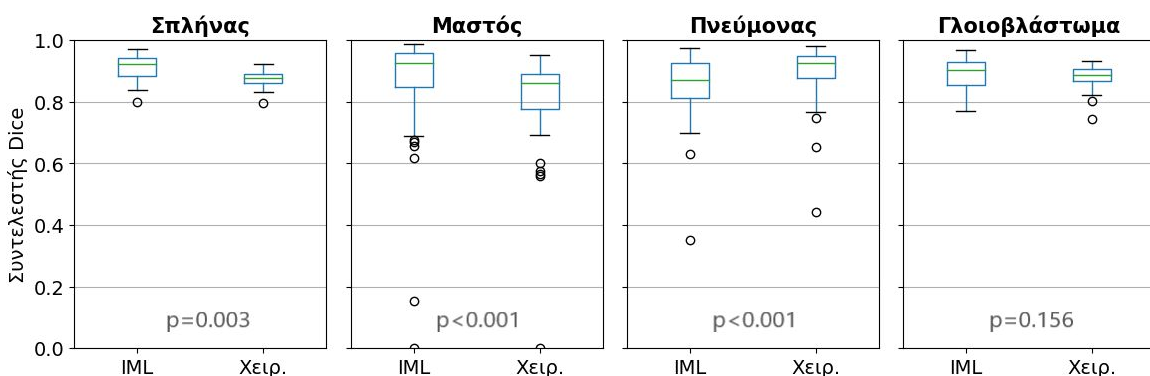
Σχήμα 11: Συντελεστής Dice, για αποτίμηση της συνέπειας ανάμεσα σε δύο προσπάθειες του ίδιου ειδικού: (i) για όταν οι επιμέρους ετικέτες που αντιπροσωπεύουν διαφορετικές περιοχές της δομής υπολογίζονται ως μία, (ii) για μεμονωμένες περιοχές του γλοιοβλαστώματος (για τις άλλες δομές δεν υπήρχαν υποπεριοχές).

5.3. Πείραμα 3: Συνέπεια των κατατμήσεων από διαφορετικούς ειδικούς

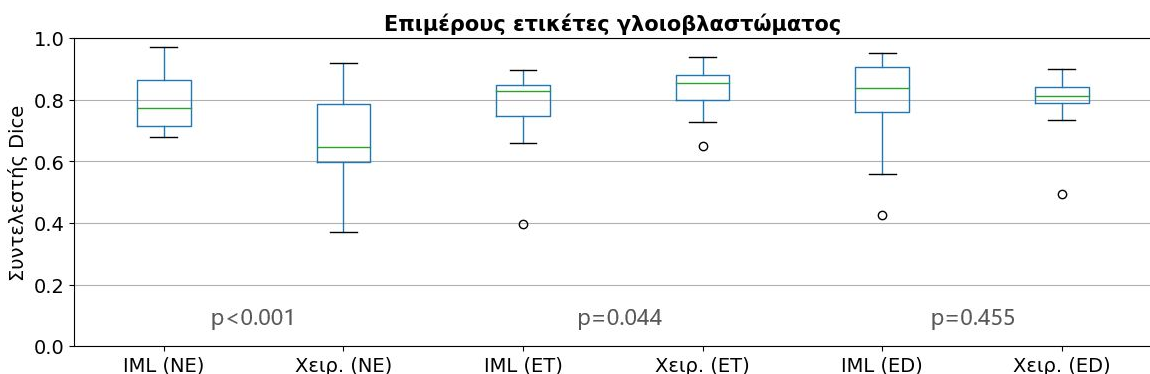
Στο τελευταίο πείραμα, υπολογίστηκε και συγκρίθηκε η συνέπεια μεταξύ IML και χειροκίνητων κατατμήσεων από διαφορετικούς ειδικούς (Πίνακας 3, Σχήμα 12). Υπήρξε σημαντική διαφορά για εικόνες σπληνός, μαστού και πνευμόνων, καθώς και για τις NE και ET υποπεριοχές του γλοιοβλαστώματος. Από τις περιπτώσεις που υπήρχε σημαντική διαφορά, η μέθοδος IML πέτυχε υψηλότερη χωρική επικάλυψη για σπλήνα, καρκίνο του μαστού και την NE υποπεριοχή του γλοιοβλαστώματος. Αντιθετα, οι χειροκίνητες κατατμήσεις έδειξαν υψηλότερη συμφωνία για καρκίνο του πνεύμονα και την υποπεριοχή ET.

Συνέπεια μεταξύ κατατμήσεων από διαφορετικούς ειδικούς

(i) Ολόκληρη η περιοχή ενδιαφέροντος



(ii) Επιμέρους ετικέτες



Σχήμα 12: Συντελεστής Dice, για αποτίμηση συνέπειας μεταξύ κατατμήσεων από διαφορετικούς ειδικούς: (i) για όταν οι επιμέρους ετικέτες που αντιπροσωπεύουν διαφορετικές περιοχές της δομής υπολογίζονται ως μία, (ii) για τις υποπεριοχές του γλοιοβλαστώματος (οι άλλες δομές δεν είχαν υποπεριοχές).

Πίνακας 3: Αποτελέσματα και για τα τρία πειράματα. Πείραμα 1: Η απόδοση του αλγορίθμου υπολογίζεται ως η επικάλυψη, που υπολογίζεται μέσω του συντελεστή Dice, των κατατμήσεων με τις κατατμήσεις αναφοράς. Οι τιμές ρ είναι αποτέλεσμα συζευγμένων συγκρίσεων μεταξύ των IML και χειροκίνητων κατατμήσεων που παρουσίασαν υψηλότερη επικάλυψη για κάθε ειδικό και για κάθε ασθενή. Ο συντελεστής συσχέτισης υπολογίζεται μεταξύ των γεωμετρικών όγκων που υπολογίζονται για τις δύο προσεγγίσεις και των όγκων αναφοράς. “Συντ. Συσχ.” είναι ο συντελεστής συσχέτισης. Πείραμα 2: Οι τιμές δείχνουν επικάλυψη μεταξύ της πρώτης και της δεύτερης προσπάθειας κάθε ειδικού, για κάθε προσέγγιση. Πείραμα 3: Οι τιμές δείχνουν επικάλυψη μεταξύ κατατμήσεων διαφορετικών ειδικών.

		Πείραμα 1						Πείραμα 2			Πείραμα 3			
		Μέση επικάλυψη			Χρόνος		Συντ. Συσχ.		Intra-rater μέση επικ.			Inter-rater μέση επικ.		
Ετικέτα	IML	Χειρ.	ρ	IML	Χειρ.	IML	Χειρ.	IML	Χειρ.	ρ	IML	Χειρ.	ρ	
Σπλήνας														
-	0.91	0.87	<10 ⁻³	66s	100s	0.99	0.99	0.91	0.89	0.003	0.91	0.87	0.003	
Μαστός														
-	0.84	0.82	<10 ⁻³	19s	70s	0.98	0.95	0.88	0.9	0.37	0.86	0.81	<10 ⁻³	
Πνεύμονας														
-	0.78	0.83	<10 ⁻³	93s	125s	0.96	0.97	0.96	0.95	0.116	0.85	0.89	<10 ⁻³	
Γλοιοβλάστωμα														
WT	0.94	0.92	0.259	21m	60m	1	0.98	0.91	0.89	0.61	0.89	0.88	0.156	
NE	0.81	0.79	0.183			0.97	0.95	0.85	0.86	0.32	0.79	0.67	<10 ⁻³	
ET	0.85	0.87	0.09			0.96	0.98	0.85	0.88	0.116	0.79	0.84	0.044	
ED	0.88	0.87	0.657			1	0.98	0.85	0.83	0.788	0.81	0.8	0.455	

6. Συζήτηση

Σε αυτήν την εργασία, παρουσιάστηκε μια IML μέθοδος κατάτμησης, η οποία είναι εύχρηστη, γρήγορη και γενικής χρήσης. Η μέθοδος μπορεί να εφαρμοστεί σε ένα πλήθος ερευνητικών εφαρμογών χωρίς να απαιτείται προσαρμογή στον εκάστοτε τομέα (domain) ή απαιτητική εκπαίδευση χρηστών. Λαμβάνει ως είσοδο *ευθυγραμμισμένες* (co-registered) ιατρικές εικόνες και γρήγορα σχέδια του χρήστη και κατόπιν δημιουργεί χάρτες AGD, ώστε να εκπαιδεύσει ένα σύνολο από ΜΔΥ, οι οποίες χρησιμοποιούνται στη συνέχεια για την κατάτμηση ολόκληρης της απεικονιστικής εξέτασης. Επίσης, αξιολογήθηκε η απόδοση της μεθόδου σε ένα πλήθος συμπαγών δομών προερχόμενο από διάφορες ανατομίες, απεικονιστικές τεχνικές και πρωτόκολλα απόκτησης ιατρικών εικόνων.

Η μέθοδος αξιοποιεί την αποτελεσματικότητα της μηχανικής μάθησης, αλλά ταυτόχρονα αντιμετωπίζει γνωστές αδυναμίες της. Συγκεκριμένα, στην προτεινόμενη μέθοδο δεν υπάρχει η ανάγκη για εκτεταμένη εκπαίδευση μοντέλων χρησιμοποιώντας ογκώδη σύνολα δεδομένων και δεν τίθεται θέμα μείωσης της απόδοσης του μοντέλου σε άγνωστους τομείς. Λόγω της διαδραστικής φύσης των IML μεθόδων, τα μοντέλα κατάτμησης είναι βέλτιστα για τις εικόνες του συγκεκριμένου ασθενή, για τον οποίο γίνεται η ανάλυση. Επιπρόσθετα οφέλη της προτεινόμενης μεθόδου αποτελούν η ικανότητα παραλληλοποίησης της και οι χαμηλές υπολογιστικές απαιτήσεις της. Ωστόσο, το κύριο μειονέκτημα της είναι ότι δεν είναι πλήρως αυτοματοποιημένη.

Η ποσοτική αξιολόγηση της μεθόδου ανέδειξε θετικές προοπτικές για την εφαρμογή της σε ποικίλες ανατομίες, για τις οποίες υπάρχει ιδιαίτερο ενδιαφέρον από την ιατρική έρευνα. Η ακρίβεια της μεθόδου και η συνέπεια μεταξύ των κατατμήσεων από διαφορετικούς χρήστες ήταν συγκρίσιμες με τη χειροκίνητη κατάτμηση, ενώ η συνέπεια μεταξύ διαφορετικών επαναλήψεων του ίδιου χρήστη ήταν υψηλή, γεγονός που δηλώνει ότι η μέθοδος είναι

ευσταθής. Τα ογκομετρικά σφάλματα ήταν κυρίως συστηματικά, υποδηλώνοντας ότι τα αποτελέσματα που αποκτήθηκαν μπορούν να βελτιωθούν μέσω περαιτέρω επαναλήψεων της μεθόδου ή εφαρμογής κοινών ογκομετρικών διαδικασιών, όπως συρρίκνωση/μεγέθυνση. Αποδείχθηκε επίσης ότι η μέθοδος είναι γρήγορη και δεν απαιτεί υπερβολική αλληλεπίδραση με τον χρήστη.

Η μέθοδος που παρουσιάστηκε έχει ένα ευρύ φάσμα κλινικών εφαρμογών, επειδή η πλήρως αυτοματοποιημένη κατάτμηση συναντά δυσκολίες για την ευρεία υιοθέτησή της στην κλινική πράξη. Η εκπαίδευση αυτόματων μεθόδων απαιτεί μεγάλα σύνολα δεδομένων που είναι δύσκολο να δημιουργηθούν και να αποκτηθούν. Επιπλέον, οι μέθοδοι βαθιάς μάθησης (deep learning) συχνά δεν είναι εύκολα προσβάσιμες από τους γιατρούς, επειδή συνήθως απαιτούν εξειδικευμένους υπολογιστές και δεν υπάρχει ακόμα μια τυποποιημένη μέθοδος διανομής τέτοιων λογισμικών. Ακόμη, οι αυτόματες μέθοδοι δεν είναι γενικής χρήσης, αλλά εκπαιδεύονται για συγκεκριμένα προβλήματα. Αυτό σημαίνει ότι θα χρειαστεί αρκετός χρόνος, ακριβές κλινικές μελέτες και σημαντική συνεργασία μεταξύ ιδρυμάτων για την εκπαίδευση αξιόπιστων μοντέλων για κάθε δομή του σώματος. Ορισμένες αυτόματες προσεγγίσεις αντιμετωπίζουν προβλήματα στην κατάτμηση εικόνων που αποκτήθηκαν από μηχανήματα με διαφορετικά προφίλ έντασης και θορύβου από αυτά στα οποία εκπαιδεύτηκαν. Ως αποτέλεσμα, η διαδραστική κατάτμηση παραμένει εξαιρετικά σημαντική και πιθανότατα θα συνεχίσει να είναι στο μέλλον. Τέλος, η κατάτμηση ιατρικών εικόνων είναι μία πολύπλοκη διαδικασία, επειδή η ανάθεση ετικετών συχνά βασίζεται σε προχωρημένες ιατρικές γνώσεις και πληροφορίες για τον εκάστοτε ασθενή, ειδικά για καρκινικούς όγκους. Κατά συνέπεια, οι κατατμήσεις από αυτοματοποιημένες μεθόδους πιθανότατα δεν θα γίνουν ποτέ συνώνυμες με τις κατατμήσεις αναφοράς και θα συνεχίσει να υπάρχει στο μέλλον η ανάγκη για γρήγορες διορθώσεις από ειδικούς για να βελτιωθεί το αποτέλεσμα,

ενδεχομένως με διαδραστικές μεθόδους που χρησιμοποιούν ως είσοδο μια κατάτμηση που παράγεται από έναν αλγόριθμο βαθιάς μάθησης.

Το ITK-SNAP [14] παρέχει επίσης μία μέθοδο διαδραστικής κατάτμησης. Ωστόσο, απαιτεί από τους χρήστες να ακολουθήσουν ένα πιο περίπλοκο πρωτόκολλο για τη διεξαγωγή κατάτμησης με πολλαπλές κλάσεις ετικετών. Ο χρήστης παρέχει πρώτα γρήγορα σχέδια για τις διάφορες περιοχές ενδιαφέροντος και εκπαιδεύει ένα μοντέλο. Στη συνέχεια και ξεχωριστά για κάθε κλάση, ο χρήστης πρέπει να τοποθετήσει “σπόρους” (seeds) και να επιτρέψει την “εξέλιξη” ενός περιγράμματος (contour evolution). Η αξιολόγηση της μεθόδου για γλοιοβλάστωμα υψηλού βαθμού κακοήθειας, επίσης στο σύνολο δεδομένων BraTS, έδειξε χαμηλότερη μέση συμφωνία με τα δεδομένα αναφοράς για το ITK-SNAP για τις περιοχές που αξιολογήθηκαν και στην παρούσα μελέτη, συγκεκριμένα ET ($Dice_{IML}/Dice_{ITK-SNAP}=0.85/0.69$) και WT ($Dice_{IML}/Dice_{ITK-SNAP}=0.94/0.85$). Ο μέσος χρόνος αλληλεπίδρασης του χρήστη ήταν επίσης χαμηλότερος ($Χρόνος_{IML}/Χρόνος_{ITK-SNAP}=21\text{min}/27.8\text{min}$). Επιπλέον, το “grow from seeds” effect (διαδικασία ανάπτυξης περιοχών) του 3D Slicer ακολουθεί μια ροή εργασίας παρόμοια με αυτή που παρουσιάζεται στην παρούσα εργασία, αλλά μπορεί να υποστηρίξει μόνο μία εικόνα ως είσοδο.

Μελλοντική έρευνα μπορεί να βελτιώσει τη μέθοδο. Προηγμένες τεχνικές μηχανικής μάθησης, όπως η ημι-εποπτευόμενη μάθηση (semi-supervised learning), μπορούν να αυξήσουν την ακρίβεια και τη συνέπεια των αποτελεσμάτων. Η τεχνική *transfer learning* θα μπορούσε να επεκτείνει το εύρος των εφαρμογών ώστε να συμπεριλάβει μη-συμπαγή όργανα και όγκους, όπως οι εγκεφαλικές βλάβες (brain lesions). Ακόμα, αν εστιαστεί το ενδιαφέρον σε μία συγκεκριμένη εφαρμογή, μοντέλα προ-εκπαιδευμένα σε μεγάλο αριθμό ασθενών, άτλαντες και εξειδικευμένες τεχνικές προεπεξεργασίας μπορούν να βοηθήσουν στην παραγωγή καλύτερων αποτελεσμάτων. Επιπλέον, ένα μη αναδρομικό σύνολο

δεδομένων, όπου η απόκτηση των εικόνων θα γινόταν με διαφορετικές ρυθμίσεις ανά ασθενή, θα επιβεβαίωνε περαιτέρω την αποτελεσματικότητα της μεθόδου.

Τα αποτελέσματα έδειξαν ότι η μέθοδος έχει ακρίβεια και συνέπεια μεταξύ κατατμήσεων διαφορετικών ειδικών συγκρίσιμες με χειροκίνητη κατάτμηση, σε διάφορες συμπαγείς ανατομικές δομές και τεχνικές απεικόνισης. Επιπλέον, η μέθοδος έδειξε υψηλή συνέπεια μεταξύ διαφορετικών επαναλήψεων του ίδιου ειδικού και μικρή ανάγκη για αλληλεπίδραση με τους χρήστες.

7. Συμπέρασμα

Στο πλαίσιο αυτής της εργασίας συγκρίθηκε η προτεινόμενη μέθοδος διαδραστικής κατάτμησης με χειροκίνητη κατάτμηση σε ένα πειραματικό περιβάλλον που περιελάμβανε κατάτμηση σπληνός, γλοιοβλαστώματος και κακοήθων όγκων του μαστού και των πνευμόνων. Τα πειράματα έδειξαν ότι η μέθοδος είναι συγκρίσιμη με τη χειροκίνητη προσέγγιση ως προς την ακρίβεια και τη συνέπεια, στις περισσότερες περιπτώσεις, ενώ ξεπερνά τη χειροκίνητη κατάτμηση σε ορισμένες άλλες. Επιπλέον, η μέθοδος ήταν πάντα ταχύτερη, χαρακτηριστικό που αναδεικνύει την καταλληλότητα της για εφαρμογές που απαιτούν κατάτμηση μεγάλου πλήθους εικόνων ασθενών.

8. Βιβλιογραφία

- [1] D. L. Pham, C. Xu, and J. L. Prince, “Current methods in medical image segmentation,” *Annu. Rev. Biomed. Eng.*, vol. 2, pp. 315–337, 2000, doi: 10.1146/annurev.bioeng.2.1.315.
- [2] N. Sharma and L. M. Aggarwal, “Automated medical image segmentation techniques,” *J. Med. Phys.*, vol. 35, no. 1, pp. 3–14, Jan. 2010, doi: 10.4103/0971-6203.58777.
- [3] B. H. Menze *et al.*, “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS),” *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015, doi: 10.1109/tmi.2014.2377694.
- [4] S. Bakas *et al.*, “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features,” *Sci Data*, vol. 4, p. 170117, 2017, doi: 10.1038/sdata.2017.117.
- [5] S. Golemati and K. Nikita, Eds., *Cardiovascular Computing-Methodologies and Clinical Applications*. Springer, 2020.
- [6] S. Rathore *et al.*, “Radiomic signature of infiltration in peritumoral edema predicts subsequent recurrence in glioblastoma: implications for personalized radiotherapy planning,” *J Med Imaging (Bellingham)*, vol. 5, no. 2, p. 021219, Apr. 2018, doi: 10.1117/1.Jmi.5.2.021219.
- [7] B. Sahiner *et al.*, *The effect of nodule segmentation on the accuracy of computerized lung nodule detection on CT scans: comparison on a data set annotated by multiple radiologists*, vol. 6514. SPIE, 2007, p. MI.
- [8] S. Golemati, J. Stoitsis, E. G. Sifakis, T. Balkizas, and K. S. Nikita, “Using the Hough transform to segment ultrasound images of longitudinal and transverse sections of the carotid artery,” *Ultrasound Med. Biol.*, vol. 33, no. 12, pp. 1918–1932, Dec. 2007, doi: 10.1016/j.ultrasmedbio.2007.05.021.
- [9] A. S. Panayides *et al.*, “AI in Medical Imaging Informatics: Current Challenges and Future Directions,” *IEEE Journal of Biomedical and Health Informatics*. pp. 1–1, 2020, doi: 10.1109/jbhi.2020.2991043.
- [10] J. H. Thrall *et al.*, “Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success,” *J. Am. Coll. Radiol.*, vol. 15, no. 3, Part B, pp. 504–508, Mar. 2018, doi: 10.1016/j.jacr.2017.12.026.
- [11] P. A. Yushkevich *et al.*, “User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability,” *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006, doi: 10.1016/j.neuroimage.2006.01.015.
- [12] S. Bakas *et al.*, “GLISTRboost: Combining Multimodal MRI Segmentation, Registration, and Biophysical Tumor Growth Modeling with Gradient Boosting Machines for Glioma Segmentation,” *Brainlesion*, vol. 9556, pp. 144–155, 2016, doi: 10.1007/978-3-319-30858-6_1.
- [13] K. Zeng *et al.*, “Segmentation of Gliomas in Pre-operative and Post-operative Multimodal Magnetic Resonance Imaging Volumes Based on a Hybrid Generative-Discriminative Framework,” *Brainlesion*, vol. 10154, pp. 184–194, Oct. 2016, doi: 10.1007/978-3-319-55524-9_18.
- [14] P. A. Yushkevich *et al.*, “User-Guided Segmentation of Multi-modality Medical Imaging Datasets with ITK-SNAP,” *Neuroinformatics*, vol. 17, no. 1, pp. 83–102, Jan. 2019, doi: 10.1007/s12021-018-9385-x.
- [15] A. Fedorov *et al.*, “3D Slicer as an image computing platform for the Quantitative Imaging Network,” no. 1873–5894 (Electronic), 2012.
- [16] B. Gaonkar *et al.*, “Automated Tumor Volumetry Using Computer-Aided Image Segmentation,” *Acad. Radiol.*, vol. 22, no. 5, pp. 653–661, 2015, doi: 10.1016/j.acra.2015.01.005.

- [17] “Diagnostic imaging,” *World Health Organization*.
https://www.who.int/diagnostic_imaging/en/ (accessed Jul. 21, 2020).
- [18] E. Friedberg *et al.*, “Unifying the Silos of Subspecialized Radiology: The Essential Role of the General Radiologist,” *J. Am. Coll. Radiol.*, vol. 15, no. 8, pp. 1158–1163, Aug. 2018, doi: 10.1016/j.jacr.2018.05.016.
- [19] “Imaging and radiology,” *medlineplus.gov*.
<https://medlineplus.gov/ency/article/007451.htm> (accessed Jul. 20, 2020).
- [20] T. P. Murphy, “Introduction to clinical interventional radiology,” *Semin. Intervent. Radiol.*, vol. 22, no. 1, pp. 3–5, Mar. 2005, doi: 10.1055/s-2005-869569.
- [21] T. Kurki, N. Lundbom, and S. Valtonen, “Tissue characterisation of intracranial tumours: the value of magnetisation transfer and conventional MRI,” *Neuroradiology*, vol. 37, no. 7, pp. 515–521, Oct. 1995, doi: 10.1007/BF00593707.
- [22] P. Mildemberger, M. Eichelberg, and E. Martin, “Introduction to the DICOM standard,” *Eur. Radiol.*, vol. 12, no. 4, pp. 920–927, Apr. 2002, doi: 10.1007/s003300101100.
- [23] I. Wolf *et al.*, “The medical imaging interaction toolkit,” *Med. Image Anal.*, vol. 9, no. 6, pp. 594–604, Dec. 2005, doi: 10.1016/j.media.2005.04.005.
- [24] GCO, “All Cancers - GCO,” *International Agency for Research on Cancer*.
<https://gco.iarc.fr/today/data/factsheets/cancers/39-All-cancers-fact-sheet.pdf> (accessed Jul. 27, 2020).
- [25] P. Anand *et al.*, “Cancer is a preventable disease that requires major lifestyle changes,” *Pharm. Res.*, vol. 25, no. 9, pp. 2097–2116, Sep. 2008, doi: 10.1007/s11095-008-9661-9.
- [26] “Types of cancer,” *Cancer Research UK*.
<https://www.cancerresearchuk.org/what-is-cancer/how-cancer-starts/types-of-cancer> (accessed Jul. 27, 2020).
- [27] “What is Cancer?,” *cancer.net*.
<https://www.cancer.net/navigating-cancer-care/cancer-basics/what-cancer> (accessed Sep. 08, 2020).
- [28] A. Gupta and T. Dwivedi, “A Simplified Overview of World Health Organization Classification Update of Central Nervous System Tumors 2016,” *J. Neurosci. Rural Pract.*, vol. 8, no. 4, pp. 629–641, Oct. 2017, doi: 10.4103/jnrp.jnrp_168_17.
- [29] D. N. Louis *et al.*, “WHO classification and grading of tumours of the central nervous system,” *WHO Classification of Tumours of the Central Nervous System. 4th ed. , Revised. Lyon: International Agency for Research Centre*, pp. 12–13, 2016.
- [30] A. Brodbelt, D. Greenberg, T. Winters, M. Williams, S. Vernon, and V. P. Collins, “Glioblastoma in England: 2007–2011,” *Eur. J. Cancer*, vol. 51, no. 4, pp. 533–542, Mar. 2015, doi: 10.1016/j.ejca.2014.12.014.
- [31] L. Sun, S. Zhang, H. Chen, and L. Luo, “Brain Tumor Segmentation and Survival Prediction Using Multimodal MRI Scans With Deep Learning,” *Front. Neurosci.*, vol. 13, p. 810, Aug. 2019, doi: 10.3389/fnins.2019.00810.
- [32] S. Bakas *et al.*, “Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge,” *arXiv*, p. arXiv:1811.02629, 2018, [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2018arXiv181102629B>.
- [33] O. Eidel *et al.*, “Tumor Infiltration in Enhancing and Non-Enhancing Parts of Glioblastoma: A Correlation with Histopathology,” *PLoS One*, vol. 12, no. 1, p. e0169292, Jan. 2017, doi: 10.1371/journal.pone.0169292.
- [34] S. G. Armato 3rd, M. L. Giger, and H. MacMahon, “Automated detection of lung nodules in CT scans: preliminary results,” *Med. Phys.*, vol. 28, no. 8, pp. 1552–1561, Aug. 2001, doi: 10.1118/1.1387272.
- [35] J. Wei, Y. Hagihara, A. Shimizu, and H. Kobatake, “Optimal image feature set for

- detecting lung nodules on chest X-ray images,” in *CARS 2002 Computer Assisted Radiology and Surgery*, 2002, pp. 706–711, doi: 10.1007/978-3-642-56168-9_118.
- [36] M. Woźniak, D. Połap, G. Capizzi, G. L. Sciuto, L. Kośmider, and K. Frankiewicz, “Small lung nodules detection based on local variance analysis and probabilistic neural network,” *Comput. Methods Programs Biomed.*, vol. 161, pp. 173–180, Jul. 2018, doi: 10.1016/j.cmpb.2018.04.025.
- [37] C. A. Owens *et al.*, “Lung tumor segmentation methods: Impact on the uncertainty of radiomics features for non-small cell lung cancer,” *PLoS One*, vol. 13, no. 10, p. e0205003, Oct. 2018, doi: 10.1371/journal.pone.0205003.
- [38] C. E. DeSantis *et al.*, “Breast cancer statistics, 2019,” *CA Cancer J. Clin.*, vol. 69, no. 6, pp. 438–451, Nov. 2019, doi: 10.3322/caac.21583.
- [39] S. Park *et al.*, “Changes in Noninvasive Liver Fibrosis Indices and Spleen Size During Chemotherapy: Potential Markers for Oxaliplatin-Induced Sinusoidal Obstruction Syndrome,” *Medicine*, vol. 95, no. 2, p. e2454, Jan. 2016, doi: 10.1097/MD.0000000000002454.
- [40] F. A. Fasola and A. J. Adekanmi, “HAEMATOLOGICAL PROFILE AND BLOOD TRANSFUSION PATTERN OF PATIENTS WITH SICKLE CELL ANAEMIA VARY WITH SPLEEN SIZE,” *Ann Ib Postgrad Med*, vol. 17, no. 1, pp. 30–38, Jun. 2019, [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/31768154>.
- [41] L. Rokach, “Ensemble-based classifiers,” *Artificial Intelligence Review*, vol. 33, no. 1, pp. 1–39, Feb. 2010, doi: 10.1007/s10462-009-9124-7.
- [42] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27–21, 2011.
- [43] R. Filipovych and C. Davatzikos, “Semi-supervised pattern classification of medical images: Application to mild cognitive impairment (MCI),” *Neuroimage*, vol. 55, no. 3, pp. 1109–1119, Apr. 2011, doi: 10.1016/j.neuroimage.2010.12.066.
- [44] M. Ghafoorian *et al.*, “Transfer Learning for Domain Adaptation in MRI: Application in Brain Lesion Segmentation,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, 2017, pp. 516–524, doi: 10.1007/978-3-319-66179-7_59.
- [45] A. van Oopbroek, M. A. Ikram, M. W. Vernooij, and M. de Bruijne, “Transfer Learning Improves Supervised Image Segmentation Across Imaging Protocols,” *IEEE Trans. Med. Imaging*, vol. 34, no. 5, pp. 1018–1030, May 2015, doi: 10.1109/TMI.2014.2366792.
- [46] H.-C. Shin *et al.*, “Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.
- [47] N. Heller *et al.*, “The KiTS19 Challenge Data: 300 Kidney Tumor Cases with Clinical Context, CT Semantic Segmentations, and Surgical Outcomes,” *arXiv [q-bio.QM]*, Mar. 31, 2019.
- [48] Y. Sun *et al.*, “Multi-Site Infant Brain Segmentation Algorithms: The iSeg-2019 Challenge,” *arXiv [eess.IV]*, Jul. 04, 2020.
- [49] G. Litjens *et al.*, “Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge,” *Med. Image Anal.*, vol. 18, no. 2, pp. 359–373, Feb. 2014, doi: 10.1016/j.media.2013.12.002.
- [50] B. Boyer, C. Balleyguier, O. Granat, and C. Pharaboz, “CAD in questions/answers Review of the literature,” *Eur. J. Radiol.*, vol. 69, no. 1, pp. 24–33, Jan. 2009, doi: 10.1016/j.ejrad.2008.07.042.
- [51] J. Firth-Cozens, “Doctors, their wellbeing, and their stress,” *BMJ*, vol. 326, no. 7391, pp. 670–671, Mar. 2003, doi: 10.1136/bmj.326.7391.670.
- [52] S. G. Mougiakakou, S. Golemati, I. Gousias, A. N. Nicolaidis, and K. S. Nikita, “Computer-aided diagnosis of carotid atherosclerosis based on ultrasound image

- statistics, Laws' texture and neural networks," *Ultrasound in Medicine & Biology*, vol. 33, no. 1. pp. 26–36, 2007, doi: 10.1016/j.ultrasmedbio.2006.07.032.
- [53] M. Gletsos, S. G. Mougiakakou, G. K. Matsopoulos, K. S. Nikita, A. S. Nikita, and D. Kelekis, "A computer-aided diagnostic system to characterize CT focal liver lesions: design and optimization of a neural network classifier," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 3, pp. 153–162, Sep. 2003, doi: 10.1109/titb.2003.813793.
- [54] J. Stoitsis, I. Valavanis, S. G. Mougiakakou, S. Golemati, A. Nikita, and K. S. Nikita, "Computer aided diagnosis based on medical image processing and artificial intelligence methods," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 569, no. 2. pp. 591–595, 2006, doi: 10.1016/j.nima.2006.08.134.
- [55] K. S. Nikita, "Atherosclerosis: the evolving role of vascular image analysis," *Comput. Med. Imaging Graph.*, vol. 37, no. 1, pp. 1–3, Jan. 2013, doi: 10.1016/j.compmedimag.2012.12.001.
- [56] J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," *Expert Syst. Appl.*, vol. 138, p. 112821, Dec. 2019, doi: 10.1016/j.eswa.2019.112821.
- [57] K. Nikita, D. Koutsouris, and S. Pavlopoulos, *Ιατρικά απεικονιστικά συστήματα*. ΤΖΙΟΛΑΣ, 2004.
- [58] N. N. Tsiaparas, S. Golemati, I. Andreadis, J. S. Stoitsis, I. Valavanis, and K. S. Nikita, "Comparison of multiresolution features for texture classification of carotid atherosclerosis from B-mode ultrasound," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 1, pp. 130–137, Jan. 2011, doi: 10.1109/TITB.2010.2091511.
- [59] S. G. Mougiakakou, I. Valavanis, K. S. Nikita, A. Nikita, and D. Kelekis, "Characterization of CT liver lesions based on texture features and a multiple neural network classification scheme," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*, Sep. 2003, vol. 2, pp. 1287–1290 Vol.2, doi: 10.1109/IEMBS.2003.1279504.
- [60] P. Asvestas, G. K. Matsopoulos, and K. S. Nikita, "A Power Differentiation Method of Fractal Dimension Estimation for 2-D Signals," *J. Vis. Commun. Image Represent.*, vol. 9, no. 4, pp. 392–400, Dec. 1998, doi: 10.1006/jvci.1998.0394.
- [61] J. J. M. van Griethuysen *et al.*, "Computational Radiomics System to Decode the Radiographic Phenotype," *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, Nov. 2017, doi: 10.1158/0008-5472.CAN-17-0339.
- [62] S. E. Chatzistergos, I. Andreadis, and K. S. Nikita, "Identification of architectural distortions in mammograms using local binary patterns and radial lengths through an exhaustive evaluation framework," *Expert Syst.*, vol. 35, no. 4, p. e12281, Aug. 2018, doi: 10.1111/exsy.12281.
- [63] V. Bolón-Canedo, N. Sánchez-Marroño, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowl. Inf. Syst.*, vol. 34, no. 3, pp. 483–519, 2013, [Online]. Available: https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1007/s10115-012-0487-8&casa_token=8S2e0NubWboAAAAA:Flb3uuxUkyfq-1gl8RrzXJs9FnHuMs1Q2lumEQHQXJ6-BQBiEhCpK8tWm2fPBHQd8CrSLh4KwY9tDofH.
- [64] B. Gaonkar, L. Shu, G. Hermosillo, and Y. Zhan, "Adaptive geodesic transform for segmentation of vertebrae on CT images," in *Medical Imaging 2014: Computer-Aided Diagnosis*, Mar. 2014, vol. 9035, p. 903516, doi: 10.1117/12.2043527.
- [65] K. H. Zou *et al.*, "Statistical validation of image segmentation quality based on a spatial overlap index," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, Feb. 2004, doi: 10.1016/s1076-6332(03)00671-8.

- [66] S. M. Stigler, "Francis Galton's Account of the Invention of Correlation," *Stat. Sci.*, vol. 4, no. 2, pp. 73–79, 1989.
- [67] C. Davatzikos *et al.*, "Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome," *J Med Imaging (Bellingham)*, vol. 5, no. 1, p. 011018, Jan. 2018, doi: 10.1117/1.Jmi.5.1.011018.
- [68] M. McCormick, X. Liu, J. Jomier, C. Marion, and L. Ibanez, "ITK: enabling reproducible research and open science," *Front. Neuroinform.*, vol. 8, p. 13, Feb. 2014, doi: 10.3389/fninf.2014.00013.
- [69] T. S. Yoo *et al.*, "Engineering and algorithm design for an image processing Api: a technical report on ITK--the Insight Toolkit," *Stud. Health Technol. Inform.*, vol. 85, pp. 586–592, 2002, [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/15458157>.
- [70] BRADSKI and G, "The OpenCV library," *Dr Dobb's J. Software Tools*, vol. 25, pp. 120–125, 2000, Accessed: Jun. 16, 2020. [Online]. Available: <https://ci.nii.ac.jp/naid/10028167478/>.
- [71] W. McKinney and Others, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, 2010, vol. 445, pp. 51–56, [Online]. Available: <http://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf>.
- [72] T. E. Oliphant, *A guide to NumPy*, vol. 1. Trelgol Publishing USA, 2006.
- [73] P. Virtanen *et al.*, "Author Correction: SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nat. Methods*, vol. 17, no. 3, pp. 352–352, Mar. 2020, doi: 10.1038/s41592-020-0772-5.
- [74] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, May 2007, doi: 10.1109/MCSE.2007.55.
- [75] A. L. Simpson *et al.*, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," *arXiv*, p. arXiv:1902.09063, 2019, [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2019arXiv190209063S>.
- [76] H. J. Aerts *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nat. Commun.*, vol. 5, p. 4006, 2014, doi: 10.1038/ncomms5006.
- [77] H. J. W. L. Aerts *et al.*, "Data From NSCLC-Radiomics [Data set]," *The Cancer Imaging Archive.*, 2019.
- [78] K. Clark *et al.*, "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *J. Digit. Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013, doi: 10.1007/s10278-013-9622-7.
- [79] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945, doi: 10.2307/3001968.



NATIONAL TECHNICAL UNIVERSITY OF ATHENS

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

DEPARTMENT OF INFORMATION TRANSMISSION SYSTEMS
AND MATERIAL TECHNOLOGY

Development and Evaluation of an Interactive Machine Learning-based Method for Segmentation of Solid Tumors and Organs

DIPLOMA THESIS

Dimitrios A. Bounias

Supervisor : Konstantina S. Nikita
Professor at N.T.U.A.

Athens, October 2020

Page left intentionally blank.



NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING
DEPARTMENT OF INFORMATION TRANSMISSION SYSTEMS
AND MATERIAL TECHNOLOGY

Development and Evaluation of an Interactive Machine Learning-based Method for Segmentation of Solid Tumors and Organs

DIPLOMA THESIS

Dimitrios A. Bounias

Supervisor : Konstantina S. Nikita
Professor at N.T.U.A.

Approved by the three-member examination committee on October 13, 2020.

.....
K. Nikita
Professor at N.T.U.A.

.....
G. Stamou
Assoc. Professor at N.T.U.A.

.....
A. Stafylopatis
Professor at N.T.U.A.

Athens, October 2020

.....

Dimitrios A. Bounias

Electrical and Computer Engineering Graduate, NTUA

Copyright © Dimitrios Bounias, 2020 – All rights reserved

It is prohibited to copy, store and distribute this work, in whole or in part, for commercial or for-profit purposes. It is permissible to reprint, store and distribute for non-profit, educational or research purposes, provided that the source is mentioned and the present message is maintained.

Questions concerning the use of this work for commercial- profit purposes should be addressed exclusively to the authors.

The views and conclusions contained in this work express the authors and it should not be construed that they represent official positions of the National Technical University of Athens, including its Schools, Departments, and Units.

Abstract

In this thesis, a fast, accurate, and consistent method for general-purpose medical image segmentation is presented. The method is based on interactive machine learning (IML).

Utilizing brief user training annotations and the adaptive geodesic distance transform, an ensemble of Support Vector Machines (SVMs) is trained, providing a patient-specific model applied to the whole image. To validate the method, retrospective cohorts were identified comprising 20 brain, 50 breast, and 50 lung cancer patients, as well as 20 spleen scans. Corresponding ground truth annotations were available for each scan. Two experts were recruited to segment each cohort twice with the IML method and twice manually. Evaluation endpoints included speed, spatial overlap, and consistency. Statistical significance was evaluated through the Wilcoxon signed-rank non-parametric test.

The IML method was faster than manual annotation in all structures, by 53.1% on average. A significant ($p < 0.001$) overlap difference was found for spleen ($Dice_{IML}/Dice_{Manual} = 0.91/0.87$), breast tumors ($Dice_{IML}/Dice_{Manual} = 0.84/0.82$), and lung nodules ($Dice_{IML}/Dice_{Manual} = 0.78/0.83$). For intra-rater consistency, a significant ($p = 0.003$) and favorable difference was found only for spleen ($Dice_{IML}/Dice_{Manual} = 0.91/0.89$). For inter-rater consistency significant ($p < 0.045$) differences were found for spleen ($Dice_{IML}/Dice_{Manual} = 0.91/0.87$), breast ($Dice_{IML}/Dice_{Manual} = 0.86/0.81$), lung ($Dice_{IML}/Dice_{Manual} = 0.85/0.89$), the non-enhancing brain tumor sub-region ($Dice_{IML}/Dice_{Manual} = 0.79/0.67$), and the enhancing brain tumor sub-region ($Dice_{IML}/Dice_{Manual} = 0.79/0.84$), which in aggregation favored the IML algorithm.

The quantitative evaluation for speed, spatial overlap, and consistency, revealed the benefits of the proposed IML method when compared with manual annotation, for several clinically relevant problems. The IML method is publicly released through the Cancer Imaging Phenomics Toolkit (CaPTk - <https://www.cbica.upenn.edu/captk>) and as an MITK plugin (<https://github.com/CBICA/InteractiveSegmentation>).

Keywords

Machine learning, Support Vector Machines, C++, Image segmentation, Medical imaging, Magnetic Resonance Imaging, Computational tomography, Glioblastoma, breast cancer, lung cancer, Spleen segmentation

Acknowledgments

I would like to thank everyone who contributed to this diploma thesis, especially Professor Konstantina Nikita and Professor Christos Davatzikos, who gave me the opportunity to work on a topic that interests me a lot. They enabled me to delve deeper into the subject and gain valuable experience on how academia works. I also really appreciate that I was given the opportunity to live abroad for six months and work with people from various countries because of this collaboration.

I would like to also thank everyone with whom I worked with in the CBICA lab of the University of Pennsylvania during my stay in the US. The workplace climate was great and people working there were excellent, both as coworkers and as individuals. By working in CBICA, I was able to gain the necessary knowledge, experience, and support to prepare this diploma thesis, as well as to present it at the RSNA 2019 conference in Chicago.

Lastly, I also want to thank my family and friends who supported me and helped me throughout my studies.

Table of contents

1. Introduction	12
2. Introductory concepts	13
2.1. Radiology	13
2.2. Medical imaging concepts	13
2.3. Medical imaging technologies	15
2.3.1. NIfTI	15
2.3.2. DICOM	16
2.3.3. PACS	16
2.4. Clinical background	17
2.4.1. Cancer	17
2.4.2. Glioblastoma	18
2.4.3. Lung nodules	20
2.4.4. Breast cancer	20
2.4.5. Spleen segmentation	21
2.5. Support Vector Machines	22
2.6. Overview of segmentation methods	23
2.7. CAD Systems	27
2.7.1. Data collection	28
2.7.2. Step 1: Preprocessing	28
2.7.3. Step 2: Segmentation	28
2.7.4. Step 3: Feature extraction	28
2.7.5. Step 4: Feature selection/elimination	30
2.7.6. Step 5: Analysis	31
2.7.7. Workflow of a CAD in a hospital environment	31
2.8. Adaptive geodesic distance maps	32
2.9. Statistical analysis concepts	33
3. Implementation	35
3.1. Cancer Imaging Phenomics Toolkit (CaPTk)	35
3.2. MITK Plugin	36
3.3. Software libraries used in the implementation	36

3.3.1. ITK	37
3.3.2. OpenCV	37
3.3.3. MITK	37
3.4. Software libraries for statistical analysis	38
4. Materials and Methods	39
4.1. Data	39
4.2. Proposed Segmentation Algorithm	39
4.3. Experimental Design	42
4.3.1. Evaluation protocol	42
4.3.2. Experiment 1: Overall performance evaluation	42
4.3.3. Experiment 2: Intra-rater segmentation consistency	42
4.3.4. Experiment 3 - Inter-rater segmentation consistency	43
4.3.5. Statistical Analysis	43
5. Results	44
5.1. Experiment 1: Overall Performance Evaluation	46
5.2. Experiment 2: Intra-rater segmentation consistency	47
5.3. Experiment 3: Inter-rater segmentation consistency	49
6. Discussion	50
7. Conclusion	53
8. References	54

List of figures

- **Figure 1:** Glioblastoma images, as seen through different MRI acquisition protocols. ET appears bright in T1CE, NE is the dark area enclosed by ET in T1CE. ED is all voxels that are bright in FLAIR, that are not part of NE and ET. *Page 19.*
- **Figure 2:** Lung nodule present in a CT image. The nodule is highlighted with red on the right. *Page 20.*
- **Figure 3:** The small and bright area at the center of the mammogram is a breast tumor. *Page 21.*
- **Figure 4:** Image of spleen in CT (highlighted area). *Page 22.*
- **Figure 5:** The graphical interface of CaPTk. *Page 35.*
- **Figure 6:** The graphical interface of MITK Workbench, featuring the developed plugin. *Page 36.*
- **Figure 7:** Example showcasing the result improving as a function of invested time. In the first iteration the user quickly draws over the different areas. In the second iteration, the user places few additional labels to correct representative misclassified areas, which are then used to retrain the machine learning model. From left to right: (i) Anatomical image (ii) User annotations (iii) result segmentation (iv) ground truth segmentation. *Page 41.*
- **Figure 8:** Example of AGD maps. Darker values indicate lower adaptive geodesic distance from the user drawings. In glioblastomas, NE and ET boundaries are clearer in T1CE, while the boundary between ED and background is clearer in FLAIR. From left to right: (i) Anatomical image (ii) User annotations (iii) AGD map for NE (Non-enhancing tumor core) annotation (iv) AGD map for ET (Enhancing tumor core) (v) AGD map for ED (peritumoral edema) (vi) AGD map for background. *Page 41.*

- **Figure 9:** Dice coefficient, compared to ground truth, of: (i) all individual labels representing different areas of the structure counted as one, (ii) the individual areas of glioblastomas. *Page 44.*
- **Figure 10:** Scatterplots in which blue points are the pairs of volume of IML method and volume of ground truth and red are the pairs of manual segmentation volume and ground truth. The black line represents the ground truth's volume. The plots belong to (i) different cohorts where all individual labels, representing different areas of the structure, are counted as one, (ii) the sub-regions of glioblastomas. *Page 45.*
- **Figure 11:** Dice coefficient, for intra-rater consistency, between the first and second round of the raters of: (i) all individual labels representing different areas of the structure counted as one, (ii) the individual areas of glioblastomas. *Page 46.*
- **Figure 12:** Dice coefficient, for inter-rater consistency, between segmentations of different raters, for: (i) all individual labels representing different areas of the structure counted as one, (ii) the individual areas of glioblastomas. *Page 48.*

List of tables

- **Table 1:** Outline of the datasets used. *Page 40.*
- **Table 2:** Results (p-values) of a paired Wilcoxon test for each rater, comparing the dice coefficient results of the different approaches relative to ground. $p < 0.05$ indicates a significant difference. *Page 47.*
- **Table 3:** Results for all three experiments. Experiment 1: Performance is calculated as dice coefficient relative to ground truth. p-values are a result of paired comparisons between the highest scoring IML-assisted and manual segmentations for each rater and each case. Correlation coefficient is calculated between the resultant and ground truth volumes. Experiment 2: Values indicate overlap between the first and second cycle of each rater. $p < 0.05$ indicates a significant difference between the results of IML and manual segmentations. Experiment 3: Values indicate overlap between segmentations of different raters. $p < 0.05$ indicates a significant difference in the inter-rater results for the respective cohort. *Page 49.*

1. Introduction

Medical image segmentation, i.e., annotation of regions of interest such as the area where a tumor resides, is an important task in clinical and research environments [1]–[5], facilitating subsequent computational analyses, which depend on the accuracy of the segmentation given as input [6]–[8]. Manual expert annotations are currently considered the gold standard, but tend to be tedious, time-consuming, and the results often have limited reproducibility [3], [9]. Even with the assistance of various tools [4], manual annotation can take multiple hours for a large scan or complex anatomies.

A plethora of fully automatic machine learning (ML) methods that can achieve state-of-the-art results have been proposed, but tend to face various challenges [10] that hinder clinical translation. Some of the most important challenges are generalization to unseen datasets and need for extensive expert corrections and refinements [4], [11].

Interactive Machine Learning (IML) methods fill the void between manual and automatic approaches by allowing an operator to train a patient-specific model via quick and rough drawings, which then automatically segments the entire scan [5], [12]–[14]. IML approaches provide the option for expedited refinements, and the final segmentation tends to get closer to the desired result as a function of the invested time. Two popular tools offering such functionality are ITK-SNAP [11] and 3D Slicer [15].

In this thesis, an IML method is proposed, leveraging adaptive geodesic distance (AGD) [16] maps alongside an ensemble of support vector machines (SVMs) that is agnostic to image type/dimensionality and supports multiparametric input images. The performance of the proposed method against manual expert segmentation was systematically evaluated across different anatomical structures and image modalities. Evaluation endpoints comprised speed, spatial overlap agreement, and consistency between different time-points and raters.

2. Introductory concepts

2.1. Radiology

Radiology is a multidisciplinary medical discipline devoted to the analysis of medical images for use in the diagnosis and treatment of diseases [17]. Examples of technologies used to acquire the medical images are MRI (Magnetic Resonance Imaging), CT (Computed Tomography), PET (Positron Emission Tomography), X-Ray, and ultrasound. As technologies advance and new ones enter the field rapidly, subspecialization is increasing in prevalence [18], although general radiologists are still common.

Diagnostic radiology is a field that focuses on using medical images for the purpose of diagnosing the causes of patient symptoms, screening for different diseases, and monitoring treatment response [19]. Diagnostic radiology is an important field in medicine, as it provides a non-interventional view of the body, which can be invaluable in assessing the health of a patient, with virtually no associated risk.

Interventional radiology focuses on assisting and guiding minor procedures undertaken in the human body, substituting the need for surgery. The practice can be used to insert instruments into the body, such as a catheter [19], and for cleaning blockages in the arteries [20]. Other examples of interventional radiology include needle biopsies of organs, embolization to control bleeding, and angiography.

2.2. Medical imaging concepts

Medical imaging is the process of noninvasively producing images of the interior of the body for clinical purposes. Medical images are typically greyscale and have some distinct properties that differentiate them from traditional images.

An important distinction of medical images is that there are a lot of different technologies, known as *modalities*, that are used to acquire them. Each modality produces results that look substantially different than the others and serves different purposes. Two well known modalities, that are used in the experiments hereafter, are MRI (Magnetic Resonance Imaging) and CT (Computed Tomography). MR scanners use radio waves and can achieve a high level of detail. CT scans are produced with x-rays, are typically less expensive, and are preferred for imaging organs and bones.

A multitude of acquisition protocols exist for each modality, each having different clinical interests. Images captured with the same modality but different acquisition protocols can be acquired during the same scan of the patient, using the same machine. For example, in the experiments regarding patients with glioblastoma, four MRI acquisition protocols were used: T1 and T2, which are standard protocols that mostly highlight fat and fluids respectively, T1CE (T1Gd) which contains information regarding angiogenesis and the integrity of the blood–brain barrier in the tumor and requires administering a contrast agent to the patient, and FLAIR which highlights the tumor and the peritumoral edematous region [6], [21]. Images taken with different acquisition protocols can be aligned to follow the same coordinate system through a process called “*registration*”. As a result, these *co-registered* images can be processed as one image that has multiple channels, as it typically happens with the R/G/B channels of normal color images, enabling the application of various common computer vision methodologies.

A medical image usually has 2 or 3 dimensions. The 3 dimensions of a 3D image, define 3 planes when paired in groups of 2. These 3 planes are named sagittal, coronal, and axial. Visual inspection of 3D images can happen in a 2-dimensional monitor, by displaying 2D slices belonging to one of the three aforementioned planes. Medical images can also be 4-dimensional, where the fourth dimension is time, and are displayed in the same way as 3D

images, with an additional slider that controls time progression. 4D images are essentially a series of 3D images acquired over a small period of time and are clinically useful for inspecting how well a biological mechanism, like blood flow or oxygenation, operates in certain areas of the patient.

Segmentations are a special case of medical images used to delineate regions of interest in a scan. They are typically *co-registered* with the images they are describing and their pixels/voxels have 0 as a value everywhere except for the pixels/voxels covering the regions of interest, which have a distinct integer value for each region.

One other important aspect of medical images is that the physical distance between voxels, commonly referred to as "*spacing*", can be variable in each dimension. That is required because scans are acquired in 2D *slices*, which are then merged together. The space between the slices acquired can be, and usually is, different to the physical space between voxels in the slice. To alleviate issues associated with this, algorithms usually transform the images to have *isotropic* spacing, i.e., to have the same spacing in all dimensions, and subsequently transform the results back to the original space after the computation. The process used to transform the size and spacing of an image is called "*resampling*".

2.3. Medical imaging technologies

2.3.1. NIfTI

NIfTI-1 and NIfTI-2 (.nii or .nii.gz) are data formats used for storing medical images. They are adopted in research because they store data in a straightforward way and the standard is strict. However, images stored this way lack a lot of meta-information which hinders their clinical adoption.

2.3.2. DICOM

DICOM (Digital Imaging and Communications in Medicine) [22] is a standard for storing medical images and meta-information about them. It is used in clinical environments and it is becoming increasingly adopted in research too. Most medical scanners produce their results in the DICOM format. Meta-information can be stored inside the header in key/value (called tag/attribute) pairs. For instance, image modality is stored in tag "(0008,0060)". Other examples of information that can be stored include the name of the patient, date of acquisition, study in which the image belongs to, and acquisition protocol. Images stored in the DICOM format can be a single file, or multiple files where each contains only one of the slices. The latter was more useful in the past since it enabled sharing a large image using multiple storage devices, such as DVDs. It should be noted that a lot of the information in the header falls under PHI (Protected Health Information) and should be edited to be anonymized/randomized when used for research purposes. DICOM however has received criticism because it is not a strict standard and different companies might store data in a different way. As a result, a lot of applications, like MITK [23], have a set of different DICOM readers, from different organizations, and use the most compatible one automatically.

2.3.3. PACS

PACS (Picture Archiving and Communication System) is a technology used for storing and retrieving medical images across a hospital system that uses the DICOM format. It is composed of a central server where all the data are stored and various clients, i.e., computer programs, that run on the clinicians' computers. In that way, images don't have to be transported using physical devices across the hospital, but can be accessed instantly through the PACS client by anyone that has been granted access to them. A PACS also has the ability to display medical images directly, without the need of external software. Furthermore, with the help of remote access technologies, such as VPNs, a PACS can allow secure off-site access to the images for the clinicians. Security of the network and the

infrastructure is highly important because private patient information can be transmitted and accessed. Additionally, given that it is a highly automated and centralized system, a PACS can greatly reduce administrative costs.

In recent years, there has been an increase in PACS servers hosted in the cloud by dedicated companies. These companies also take care of the client software that runs on the practitioners workstation. By using a cloud based PACS, the hospital offloads the maintenance of the system to a company in exchange for a periodic fee. This move could also increase the security of the data, because a company that specializes in developing PACS has a dedicated security department that a hospital could not afford.

2.4. Clinical background

In this thesis, four cohorts were analyzed, consisting of glioblastomas, lung nodules, breast tumors, and images of spleen.

2.4.1. Cancer

Cancer is a group of diseases involving uncontrolled growth of cells in the body. According to the Global Cancer Observatory of the International Agency for Research on Cancer, there were over 18 million new cancer diagnoses in 2018 worldwide, while over 9 million people lost their lives due to the disease in the same year [24].

The main causes of cancer are tobacco use which accounts for 25-30% of cancer related deaths, diet which accounts for 30-35%, infections which account for 15-20%, genetics which account for 5-10%, and the remaining cases are attributed to other factors such as radiation, stress, physical activity, and environmental pollutants [25]. Since lifestyle is a main contributing factor, a lot of deaths could be prevented by quitting smoking, improving diet, vaccination, exercising, and avoiding direct exposure to sunlight. There are hundreds of types of cancer, but they can get grouped into the 5 main categories [26] described below.

Carcinoma starts in epithelial cells, which can be found on the skin or in thin layers in blood vessels or the exterior of organs. They also cover cavities, such as the chest and abdominal cavities. Most carcinomas form solid tumors [27]. Lung, breast, and prostate cancer are carcinomas. It is the most common type of cancer, accounting for about 85% of cancers in the UK [26].

Sarcoma starts in connective tissue, such as bones, nerves, muscles that support organs [27]. It accounts for about 1% of cases [26]. It can be further divided into bone sarcoma and soft tissue sarcoma.

Leukaemia is the cancer of white blood cells. It is caused by the bone marrow, which creates malfunctioning white blood cells in excess, causing them to build up in the bloodstream. Leukaemia accounts for about 3% of cases, but is the most common cancer in children [26].

Lymphoma is a cancer that begins in cells of the immune system, particularly in lymphatic cells. The lymphatic system is a network of vessels and glands which filters body fluids and fights infection; it consists of the lymph glands, the lymphatic vessels, and the spleen. The malfunctioning lymphocytes that are rapidly dividing can build up in the bone marrow and spleen and cause tumors. Lymphomas account for about 5% of cases. Myelomas start in plasma cells, which are a type of white blood cells used to produce antibodies and cause them to multiply rapidly. They account for about 1% of cases [26].

Brain and spinal cord cancers start in cells of the central nervous system. Glioma, the most common type, is a brain tumor that develops from glial cells. Brain tumors are particularly dangerous because they can impact parts of the brain that are important for life and early detection is hard. These cancers account for about 3% of cases [26].

2.4.2. Glioblastoma

Glioblastomas are a class of malignant brain tumors. They are classified as “Grade IV” on the WHO (World Health Organization) classification scale [28]. To be classified as “Grade

IV” a tumor type has to show anaplasia, mitotic activity with microvascular proliferation, and/or necrosis. Glioblastomas are known to reproduce rapidly and expand aggressively [28], [29]. According to a study in England that examined 10,743 patients with glioblastoma [30], median overall survival was 6.1 months, and the patients had 28.4%, 11.5% and 3.4% chance to survive 1, 2, and 5 years respectively.

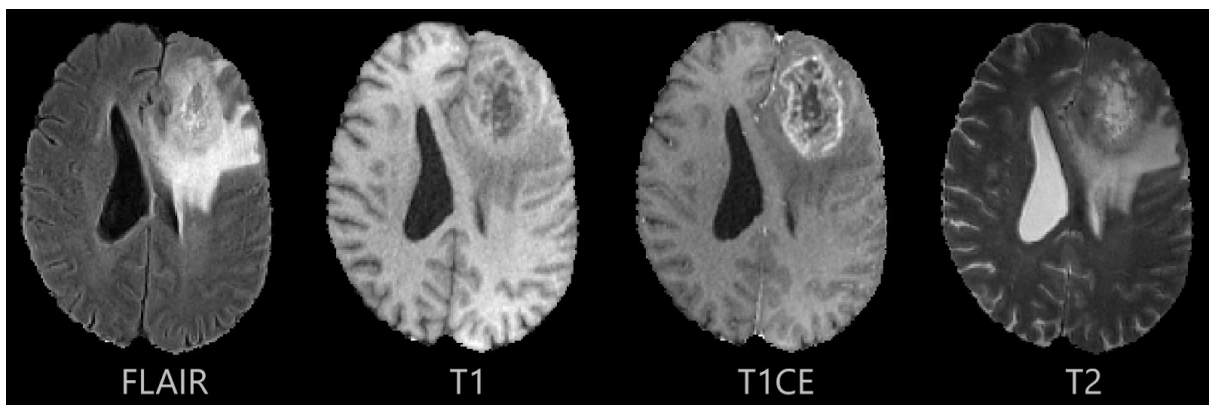


Figure 1: Glioblastoma images, as seen through different MRI acquisition protocols. ET appears bright in T1CE, NE is the dark area enclosed by ET in T1CE. ED is all voxels that are bright in FLAIR, that are not part of NE and ET.

Glioblastomas are large and have a complex structure. They are divided into two main sections, the tumor core (TC) and the peritumoral edematous/invaded tissue (ED). The tumor core is also divided into the non-enhancing (NE) and the enhancing (ET) parts [31], [32]. The tumor core subregions get their names from whether or not they show high intensity in the T1CE modality, after the administration of a contrast agent, however there are also underlying differences. The non-enhancing part consists of mostly pre-necrotic and necrotic regions of the tumor core, while the enhancing part is considered to be more active [32]. Although, the non-enhancing part is also considered to be infiltrative to some extent [33]. Lastly, the peritumoral edematous/invaded tissue surrounds the tumor core and has a hyper-intense signal in T2-FLAIR. It describes the regions of edematous white matter into the subcortex of the gyri. The ED region can also be infiltrated. The amount of infiltration can

be variable throughout the ED region and depends on the distance of an area to the tumor core [6]. As a result, sections of the peritumoral edematous tissue also get removed as a precaution, in case of tumor core removal surgery.

2.4.3. Lung nodules

Lung nodules are small masses of tissue in the lungs, which can be either cancerous or benign. They are typically less than 30mm in diameter and are visible in either CT scans or X-Rays [34], [35]. It is difficult to detect lung nodules and classify them as cancerous or not [36], because of their size and the similarity between malignant and benign nodules. Also, due to their small size, errors and inaccuracies in their segmentation heavily impact further computational analysis [37], as it can impact volume calculation and feature extraction.

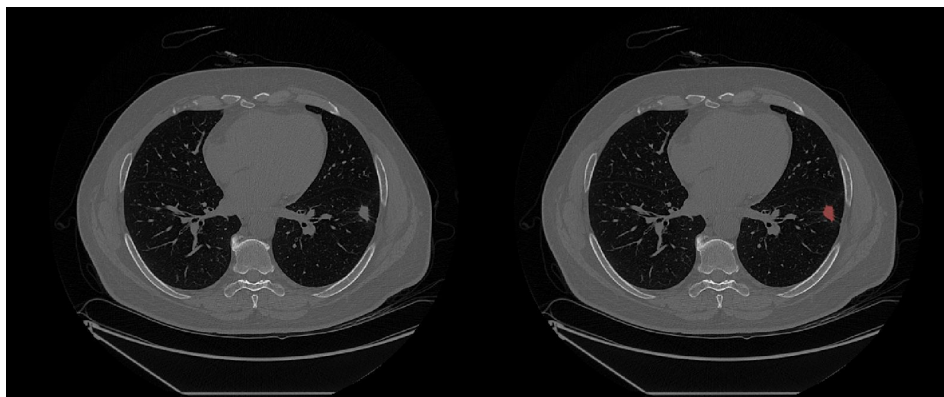


Figure 2: Lung nodule present in a CT image. The nodule is highlighted with red on the right.

2.4.4. Breast cancer

Breast cancer is a disease that affects a large percentage of the female population. In the US, approximately 41,760 women died from it in 2019 and around 268,600 women had an invasive case [38]. Prevention and early detection are very important in the fight against the disease. In this thesis, MRI images of malignant breast tumor cases were analyzed. There

were three images used for each case, one acquired before the administration of contrast agent and two after.

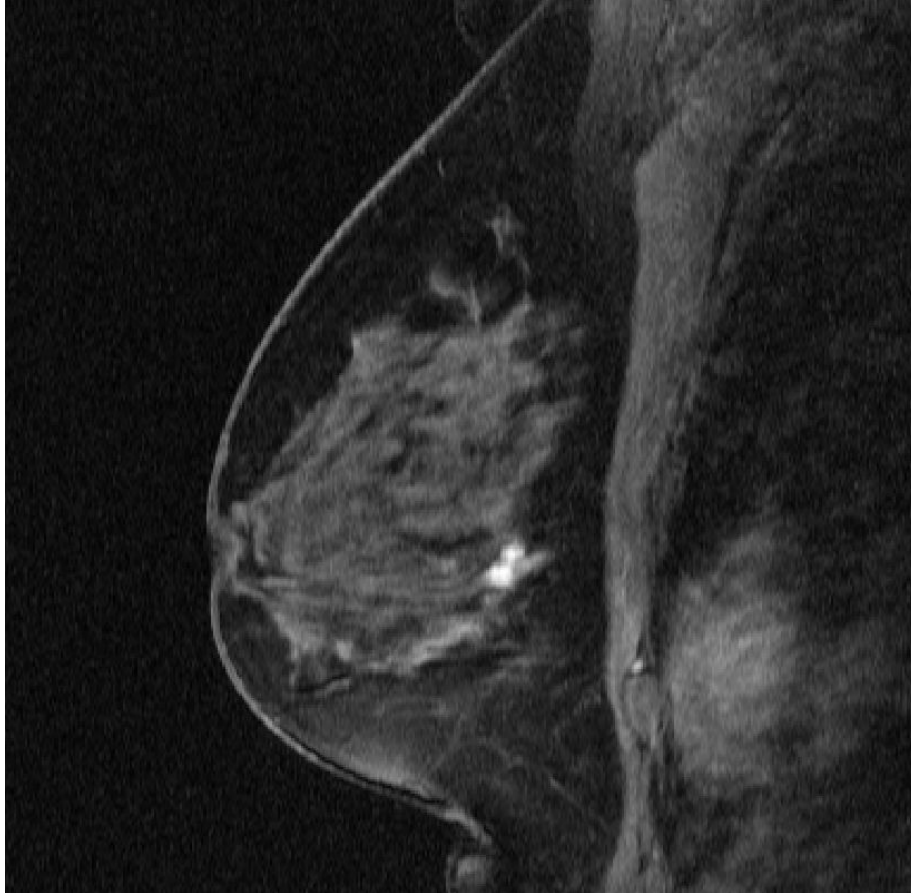


Figure 3: The small and bright area at the center of the mammogram is a breast tumor.

2.4.5. Spleen segmentation

Spleen segmentation was performed as an example of organ segmentation, which has a lot of applications. Specifically for spleen, its volume as well as changes in its size, can impact clinical decisions [39], [40]. It should also be noted that organ segmentation differs significantly from the segmentation of tumors, because organs tend to be larger with smoother boundaries.

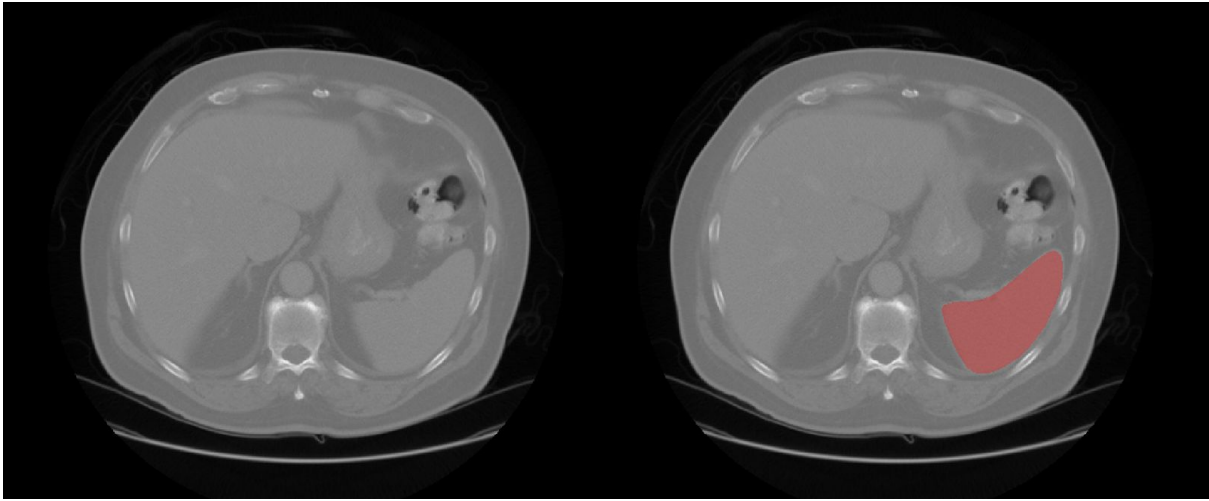


Figure 4: Image of spleen in CT (highlighted area).

2.5. Support Vector Machines

The method described in this thesis uses an ensemble [41] of Support Vector Machines (SVMs) [42] to make predictions. SVMs are used to perform multi-class supervised classification of individual pixels/voxels. However, SVMs can also be used for regression in other tasks.

The underlying theory of SVMs is best understood for 2-class classification. Each sample is mapped as a point in a n -dimensional space, where n is the number of features. SVMs try to separate the samples using a $(n-1)$ -dimensional hyperplane, typically by choosing a hyperplane that best separates the training data into 2 classes. During inference, samples are mapped in the same n -dimensional space and samples that reside in opposite sides of the hyperplane get assigned to different classes. As an example, if there are two features, the separating hyperplane would be a line.

Linear separation is fast but not powerful enough to provide optimal results for most problems. Such separation happens in the example above, where a line is used to separate samples with two features. There is however a need for using curves instead of lines, i.e., non-linear separation. To address this, SVMs utilize the “kernel trick”. Using non-linear

kernel functions that are not excessively computationally heavy, the feature space can be transformed to another, higher-dimensional, space where the features are transformed and combined using the kernel function. As a result, the linear way of separation described in the previous paragraph can be used on the new feature space to perform non-linear separation. The choice and parameterization of the kernel function is application-specific.

2.6. Overview of segmentation methods

A supervised ML model trains on labeled input data and produces a prediction (output) on new unlabeled data. Unsupervised learning uses no labeled input data, but rather tries to find patterns in the data and group them accordingly. Semi-supervised approaches use a few labeled input data alongside a large amount of unlabeled ones, which is useful when there is a large amount of available data, but few of them are manually annotated and annotating them would be costly. Most state-of-the-art fully automatic methods are based on deep learning. These methods typically need large annotated datasets.

An algorithm that has traditionally been used for supervised segmentation is k-nearest neighbors (kNN). It operates on the assumption that similar objects tend to be close in a multidimensional feature space. It is a slow algorithm that requires the user to select k, through experimentation. At classification, the k nearest (to the test sample) training samples are selected and the most often represented class in the k near training samples is selected. Selecting an odd k is recommended to avoid ties.

Another similar algorithm that has been used are SVMs, which were analyzed in a separate section. SVMs could also be adapted to perform semi-supervised classification, such as the Transductive Support Vector Machine (TSVM) [43], where the algorithm can utilize both labeled training data and unlabeled ones to improve the model.

Ensemble methods combine multiple models to achieve better results than what each one of them would have produced individually. Usually, ensemble methods combine weak learners, i.e., methods which typically have high bias or variance, to create a strong learner. Because of the bias-variance tradeoff, individual models with low bias have high variance, and vice versa, but through combining some the effect is minimized. There are three main ways to combine the results of the multiple classifiers: bagging, boosting, and stacking. In bagging, the weak classifiers are trained independently and the final result is the average of all classifiers. To train each classifier, random samples of the training data are chosen with replacement, meaning that samples chosen for one classifier have the same probability to be chosen for another. Random forests are a popular bagging technique. Bagging reduces variance. In boosting, samples are weighted, meaning that there is a variable probability for them to be chosen to train a classifier. That is because training in boosting is a sequential procedure and not a parallel one like bagging. When a classifier is trained in boosting, the samples that were misclassified get assigned a higher weight. Thus, classifiers trained afterwards have a higher chance to be trained on the previously misclassified samples, which reduces bias. Stacking works with heterogeneous weak learners, i.e., combining different algorithms. In that case a new model (meta-learner) is trained that accepts the results of the different weak learners as input.

Perceptron is an algorithm for binary supervised classification. It accepts multiple inputs, which are subsequently individually weighted. The weighted inputs are summed and passed through an activation function. A bias could also be added to help shift the outcome of the activation function. The activation function determines if the output is classified as 1 or 0. There are two types of activation functions, linear and nonlinear. Usually non-linear activation functions are selected. Some popular ones comprise the Sigmoid, Tanh, and ReLU (Rectified Linear Unit) functions, with the latter being popular in deep learning.

Training a perceptron is an iterative procedure. It is an optimization problem, where the weights of the perceptron are updated to minimize an error function.

Artificial Neural Networks (ANN) are structures with three layers. The first layer is the input, the last one is the output, and the middle one is called a hidden layer. Neurons (nodes) in the hidden and output layer are perceptrons, while nodes in the input layer contain the numerical value for each feature of the sample. All neurons in the hidden layer have weighted input from all the input neurons. Additionally, all neurons in the hidden layer are connected to all output neurons. Since multiple output neurons can exist, multi-label classification is now possible, which was not the case for the perceptron. The weights of the neurons are updated in training. An error value is calculated through a cost function, according to the output value produced and the expected value. Through back-propagation, weights are adjusted to minimize the cost function.

Deep neural networks are artificial neural networks that have two or more hidden layers. Deep networks typically don't require any feature extraction beforehand and can learn the patterns of the data by essentially generating their own features. Their adoption has exploded in the last two decades, because of advances in computing power, data availability and ML techniques.

Convolutional Neural Networks (CNNs) are useful in image recognition tasks, because they can accept a whole image as input, which for medical imaging can account to millions of pixels, but manage to keep the complexity of the network from getting high. Usually a CNN has three layers: a convolutional, a pooling, and a fully-connected layer. In the convolutional layer, the dot product of a filter and the matrix of an image portion is calculated, with the aim of extracting features. The pooling layer decreases the dimensionality of the feature matrix, by combining squares of the feature matrix, usually by keeping their maximum (max-pooling) or their average (average-pooling). After these two layers, the feature matrix has been

reduced significantly and a fully-connected layer is used to learn the relationships of the data, like the hidden layer in an artificial neural network.

The convolution neural network described above is useful for classification, but not segmentation. To adapt it to segmentation, the U-Net and other similar architectures have been proposed and used successfully in medical imaging. The difference is that apart from the contracting path that reduces the size of the data, there is also an expansive path that increases it and brings it back to the original image size (excluding channels). That is because in segmentation a label needs to be assigned to each individual pixel/voxel.

Transfer learning is a machine learning technique used primarily for domain adaptation. It is more useful in MRI, because there is more variation in the image intensities and contrast [44]. Thus, models don't perform well on images acquired with different machines/protocols than the one they were trained on. The main idea of transfer learning is to use large legacy datasets for training models and then use a small number of annotated scans from the target domain to adapt the model. Adaptive SVM and TrAdaBoost (Transfer AdaBoost) have been shown to outperform traditional SVM classifiers in medical image segmentation [45] even when trained on a small number of images from the target domain. Domain adaptation has been shown to work well with CNNs too [44]. Transfer learning could also work with models originally trained for non-medical tasks, then adapted to medical segmentation [46].

BraTS (Brain Tumor Segmentation) [32] is a yearly multi-institute effort to acquire medical images of brain tumors and use them to assess the robustness of segmentation algorithms in the form of a challenge. Similar challenges exist for other segmentation tasks too [47]–[49]. Automated methods used in BraTS have great performance, but are still less robust than segmentations obtained manually from clinicians, with less accuracy and inter-rater agreement. However, fusing the results of multiple algorithms, through label

voting, can obtain better results than the individual algorithms and inter-rater agreement that is comparable to the experts.

2.7. CAD Systems

A software system can't replace the radiologist in most situations yet. This would require exceptional sensitivity, specificity, and adaptability from the algorithm. Additionally, the general population might not be ready for such a fundamental change and insurance companies might be hesitant to cover such treatments. On the other hand, physicians are subject to mistakes too. Such mistakes could be triggered by fatigue, stress, and other factors [50], [51]. Thus, a physician could benefit from a software solution that is not perfect but correct enough in most situations, as the algorithmic decision could act as a suggestion or an alternative opinion, assisting the radiologist but leaving the final decision to them [52]–[55]. Such a system is called a CAD, which is an umbrella term that encapsulates computer-aided detection (CADe) and computer-aided diagnosis (CADx) [56]. CADe systems aim to detect unusual structures in the body, such as tumors. CADx systems aim to make intelligent decisions about the patient's health, such as classifying tumors as malignant or not. While the two types can exist separately, a lot of CAD systems encapsulate both CADe and CADx components.

While the first attempts in using computers to solve medical problems occurred in the 1950s, by the 1970s it had become clear that it is not an easy problem to solve using traditional computer techniques and statistics [56]. In more recent times, advances in computer power and machine learning techniques have made the exploration of the field plausible again and have spawned a lot of research on the subject.

2.7.1. Data collection

If a CAD system requires training, as most intelligent ones do, a large dataset is required. This dataset should preferably contain images taken from multiple institutions, various machines, and different cohorts. This data needs to be checked to exclude scans with missing images and outliers. Alternatively, missing scans can be populated with data augmentation techniques [57]. Training data need to be normalized and potentially harmonized to assist the model in learning the actual differences.

2.7.2. Step 1: Preprocessing

In the first step of a CAD system, data should be normalized/standardized and potentially harmonized if they have been acquired using different scanners/techniques. Normalization pertains to scaling the data in the [0,1] range and standardization to transforming the data to have zero mean and a standard deviation of one. Data harmonization pertains to minimizing differences between images acquired from different scanners [58] so that the model can learn the real differences between cases. Additional processing can occur in this step depending on the application, that aims to make the data easier to be processed later, such as denoising techniques. For brain applications there might be an additional *skull-stripping* step, where the skull is removed from the image, leaving only the brain to be processed.

2.7.3. Step 2: Segmentation

Segmentation of the images is needed, in order to isolate the structures that need to be analyzed. The segmentation algorithm can be automatic or interactive, although automatic methods are preferable if available for the specific problem, since no other CAD step typically requires human interaction.

2.7.4. Step 3: Feature extraction

Feature extraction typically follows segmentation, a process that aims to reduce the input data into values (features) that are relevant to the problem at hand [52], [54], [59], [60].

A popular method for feature extraction of medical images is called radiomics. "pyradiomics" [61], a popular radiomics python package, categorizes the features it can extract into the following categories: (i) "First Order" features which are related to the distribution of pixel/voxel intensities in the segmented area (ii) "Shape" features which describe the shape of the segmented area in 2D/3D (iii) "Gray Level Co-occurrence Matrix" (GLCM) features which are related to the second-order joint probability function of the segmented area (iv) "Gray Level Size Zone Matrix" (GLSZM) features which calculates the number of gray level zones in the image, i.e., areas that have the same intensity values (v) "Gray Level Run Length Matrix" (GLRLM) features which calculate the length of consecutive pixels/voxels with the same intensity value (vi) "Neighbouring Gray Tone Difference Matrix" (NGTDM) features which calculate the difference between the intensity of a pixel/voxel and the average intensity value of its neighbours within a predefined distance and (vii) "Gray Level Dependence Matrix" features which calculate the number of pixels/voxels that are "dependent" on a pixel/voxel; a pixel/voxel is dependent on another if the absolute of their intensity difference is less than a predefined value.

One other popular feature extraction technique is Principal Component Analysis (PCA). In this context, PCA is used for dimensionality reduction of the information of the segmented area. A $N \times N$ symmetric covariance matrix is constructed, where N is the number of pixels/voxels, because a value is needed for each pixel/voxel pair. Each value in the matrix represents the covariances of the respective pair. If a value is positive then there is a positive correlation and the two variables increase/decrease together, while the reverse is true for negative values. PCA constructs new variables, called principal components, that are not correlated [62] and are the same in number as the input variables. However, PCA encloses as much information as it can in the first principal components, i.e., maximizes variance in the first components. As a result, by selecting a small number of the first principal components as features, most information can be kept while the dimensionality of the

problem reduces. Independent component analysis (ICA) is a similar technique used to separate a signal into underlying independent components. It doesn't focus on maximizing variance in the first components, but focuses on making the components independent. Another technique is the autoencoder, which is an unsupervised way to achieve dimensionality reduction by using an artificial neural network (ANN). By having a "bottleneck" in the hidden layer the information of the input image is compressed, because the network learns correlations of the input.

2.7.5. Step 4: Feature selection/elimination

"Dimensionality curse" is a term that describes issues pertaining to having data arranged in high-dimensional spaces. The main reasoning is that high-dimensional data tend to be sparse and that negatively impacts algorithms that need statistical significance. CAD systems can face such problems and as a result feature selection usually follows feature extraction [53], [57]. The most important features are selected from the feature space and they are used to train the model. With feature selection, training time gets lower too and there is less chance of overfitting. There are three main feature selection technique categories that are described below [63].

The first category is called a filter and these types of techniques don't interact with the classifier, that simply accepts the selected features as input. Filters use application-independent statistical methods to categorize the features. They can be further subdivided in univariate filters, that don't use feature dependencies in the calculations which can result in selecting similar features, and multivariate filters, that take notice of feature dependencies at the cost of processing speed. The most common filters are multivariate Pearson correlation-based.

Another category is embedded feature selection. These types of methods are specific to certain machine learning techniques and select features during training of the model. They

can be thought of as part of the classifier. An example of embedded feature selection is Recursive Feature Elimination for Support Vector Machines.

The last category is wrapper methods. These methods use the performance of the classifier to evaluate the suitability of the features. This is an iterative procedure that tends to be slower and less scalable than the other two, but can produce better results with machine learning techniques where embedded feature selection does not exist. An example is backward/forward elimination.

2.7.6. Step 5: Analysis

The final step is usually the main algorithm of the CAD that performs a classification or regression task. These algorithms have an output that can be interpreted by the clinician and usually is a decision about the health of the patient. It should be noted that some deep learning algorithms might be trained using the whole image as input and not require separate segmentation and feature extraction/selection.

2.7.7. Workflow of a CAD in a hospital environment

During usage in a hospital setting, a CAD system takes the medical images of a patient as input, potentially alongside additional meta-information such as the age of the patient. The first step is preprocessing the images to normalize/harmonize them with the training data. Segmentation, feature extraction/selection and analysis happen in accordance to the trained models and the result is presented to the clinicians. Excluding the case where the segmentation step is interactive, this process can be automated and transparent to the clinician, who does not need to be aware of the process. In modern systems, CAD software can run directly in the PACS, enabling the clinician to have the segmentation and the generated diagnosis/detection available for him when he sees the scans for the first time.

2.8. Adaptive geodesic distance maps

Adaptive geodesic distance (AGD) [16], [64] maps are utilized in the proposed method as additional input to the model, alongside medical images. Their purpose is to convey additional information related to the structure that a model, which uses pixels/voxels as training samples, would struggle to deduce.

To produce an AGD map, two images, having exactly the same size, are needed: (i) a medical image, (ii) a binary image, i.e., an image with values 0 or 1 at each pixel/voxel. Since these images have the same size, each pixel/voxel in the binary image can be directly mapped to a particular pixel/voxel in the medical image. The positive values in the binary image represent the drawings of a human operator. More precisely, the positive pixels/voxels in the binary image represent samples of pixels/voxels in the medical image that the operator deemed as belonging to the structure that needs to be segmented.

Geodesics generalize the concept of straight lines to curved spaces. An example of geodesics would be trying to calculate the distance of two cities on earth. Their euclidean distance would be calculated as the length of a straight line connecting the two cities that passes through earth, while their geodesic distance would follow the curvature of the earth. The problem can be thought of as trying to find the shortest path between two points while moving along a surface.

AGD maps have a “distance” value at each pixel/voxel. This “distance” represents the shortest path between the relevant voxel and the closest pixel/voxel marked as positive by the operator. This distance is not euclidean; similar to geodesics it represents the shortest path between two points while moving along a surface. However, the surface is defined by the intensity profile of the medical image [16]. As a result, the distance values are a composite of physical distance and intensity change. In other words, pixels/voxels far away from the drawings and/or with different intensity have higher values.

The maps can be calculated rather quickly, by visiting each pixel/voxel twice [16]. The result is an image that has values in $[0, +\infty)$. Lower values represent a higher probability that the pixel/voxel belongs to the same structure as the ones that the operator drew upon. AGD maps have been used in the past to perform single-class interactive segmentation, which required the operator to threshold the map to produce the final segmentation. In detail, by choosing a value in $[0, +\infty)$, the values lower or equal than the chosen threshold would be transformed to 1 in the output segmentation, while all the other values to 0. This method was quick and worked well for simple structures, however it faced challenges on more complex applications, as it couldn't handle multiparametric images or perform multi-class segmentation. The thresholding step was also difficult for operators, as the correct value selection was not always obvious. This thesis incorporates AGD maps as additional input to a machine learning model, thus solving all the aforementioned shortcomings.

2.9. Statistical analysis concepts

Dice coefficient [65] quantifies the spatial overlap of two binary segmentations, by identifying the pixels/voxels that have a similar value in both segmentations and the ones that do not. When a segmentation is not binary, but has multiple labels representing different sub-regions of a single structure, the individual sub-regions can be converted to binary segmentations and compared. Similarly, all labels can also be merged into one to quantify overlap for the whole structure. For the dice coefficient calculation, the following formula was used for each sub-region as well as the whole structure. TP is the number of true positive pixels/voxels, FP are false positive, and FN are false negative.

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

The dice coefficient results did not follow a Gaussian distribution. The samples were also paired, because for each patient there were segmentations to compare from different operators, cycles, or time points. As a result, the non-parametric signed-rank paired Wilcoxon test was used to evaluate significance.

Finally, Pearson's Correlation Coefficient [66] was used to quantify the correlation between paired sets of *volumes* of segmentations acquired with different means.

3. Implementation

3.1. Cancer Imaging Phenomics Toolkit (CaPTk)

CaPTk [67] is an open-source cross-platform software tool focusing on cancer research, developed by the *Center for Biomedical Image Computing and Analytics (CBICA)* of the *University of Pennsylvania*. It provides functionality for preprocessing, segmentation, feature extraction, and training of machine learning models, as well as advanced specialized applications of computational studies conducted in the center. The implementation of the method is part of current releases of CaPTk and can be obtained through this link: <https://www.cbica.upenn.edu/captk>.

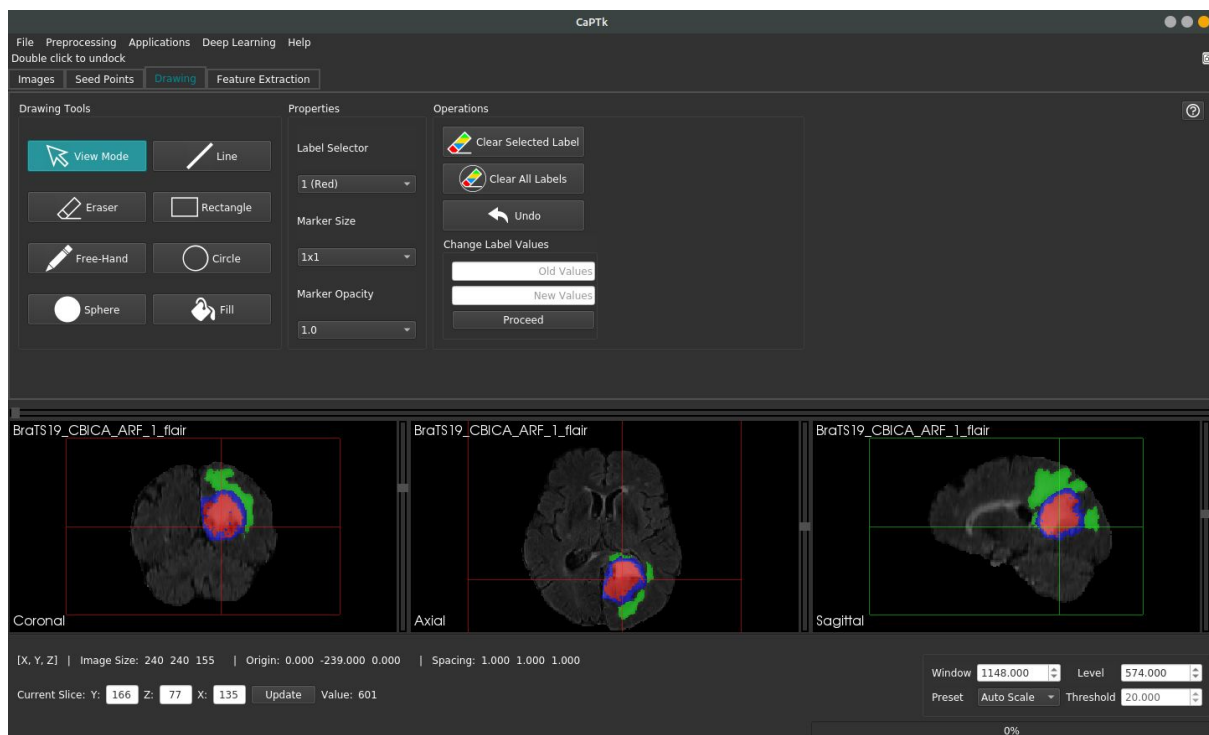


Figure 5: The graphical interface of CaPTk.

3.2. MITK Plugin

MITK [23] is an open-source cross-platform application framework for medical image processing, developed by the *German Cancer Research Center (DKFZ)*, that has seen success in both research and industrial environments. MITK supports extensions via a plugin mechanism, which was used to implement the algorithm. An MITK application (*MITK Workbench*) that contains the developed method as a plugin can be found at: <https://github.com/CBICA/InteractiveSegmentation>.

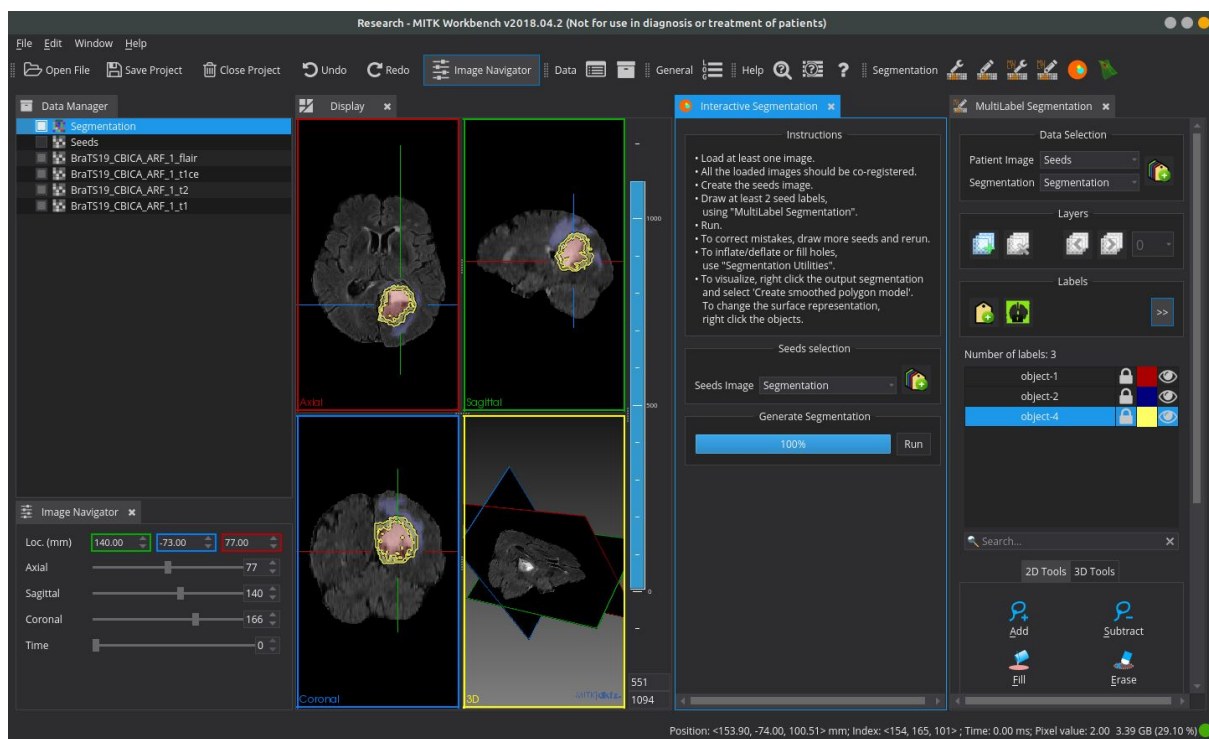


Figure 6: The graphical interface of *MITK Workbench*, featuring the developed plugin.

3.3. Software libraries used in the implementation

Both algorithm implementations are in C++ and CMake was used to control the software compilation process. Additional software libraries that were used are detailed below.

3.3.1. ITK

ITK [68], [69] is an open-source software library that provides tools for medical image analysis. It provides functionality for loading and saving n-dimensional *itk::Image(s)*, iterating over their pixels/voxels, and performing multiple operations which are implemented as “ITK filters”. Such relevant filters are resampling, normalization, standardization, and image denoising. ITK version 4.0+ can be used with the source code of the algorithm.

3.3.2. OpenCV

OpenCV [70] is an open-source software library for computer vision and machine learning. The class “*cv::ml::SVM*” was used to implement the SVMs used in the implementation. To provide input to OpenCV, the images are transformed to *cv::Mat*, OpenCV’s matrix implementation, and the results are transformed back to *ITK images* after the computation. The three kernels used comprised *RBF* (Radial Basis Function), *CHI2* (exponential χ^2), and *INTER* (Histogram Intersection). The *trainAuto()* function was used to perform cross-validation, hyperparameter tuning, and training. The default parameter grids of OpenCV were used for the grid search performed in each cycle of cross-validation, except for the C parameter of *RBF* and *CHI2* where the values tested ranged [1,400] with a logarithmic step of 2.5.

3.3.3. MITK

MITK [23] was used to develop the MITK plugin of the algorithm. It is a plugin-based application framework that contains various libraries commonly used in medical imaging tasks, such as ITK and OpenCV. MITK also provides additional high-level functionality, such as better image I/O functions and image data containers, as well as various graphical components like an image viewer and a data manager.

3.4. Software libraries for statistical analysis

The experiments required a complex statistical analysis that would have been difficult to perform without a dedicated codebase, given that multiple calculations were needed for each cohort, rater, segmentation cycle, and structure sub-region.

Python 3.6 was used to perform the statistical analysis. The following packages were used:

(1) *pandas* 1.0.3 [71] for storing/retrieving dice coefficients in CSV form and performing high-level operations on them. (2) *numpy* 1.18.4 [72] for performing a vast range of array operations using the data. Correlation coefficient was also calculated using *numpy* through the `corrcoef()` function. (3) *scipy* 1.4.1 [73] for non-parametric paired wilcoxon tests, as well as some operations of Dice calculation. (4) *matplotlib* 3.1.3 [74] for creating the plots.

4. Materials and Methods

4.1. Data

Experiments were approved by the Institutional Review Board (IRB) of the University of Pennsylvania (UPenn, Philadelphia, USA). Quantitative evaluation was based on public and private clinical data from four retrospective cohorts (Spleen (3D-CT, $n=20/41$, Medical Segmentation Decathlon [75]); Breast tumor (2D-DCE-MRI, $n=50$, multimodality trial at UPenn; NIH P01CA85484); Lung nodules (2D-CT, $n=50/89$, The Cancer Imaging Archive [76]–[78]); Brain Glioblastoma (3D-MRI, $n=20/335$, BraTS'19 [3], [4], [32])). Cohort subsets were created, following random selection, to facilitate the exhaustive manual annotations described hereafter. The Brain (11 Males / 9 Females: Age=62.84/64.36, Range=44.82-77.48/39.64-77.09) and Breast (Female: Age=50.41, Range=32.68-71.97) datasets were acquired from 2006-2014 and 2002-2006, respectively. The Spleen (13 Males / 7 Females: Age=63.85/58, Range=40-81/48-68) and Lung (34 Males / 16 Female) were acquired from 2000-2013 and 2004-2011, respectively. Age information was not available for the Lung dataset. Ground truth segmentations were available for all datasets, except for Lung which were created by a fellowship trained, board certified, thoracic radiologist (S.K., 21 years of experience).

4.2. Proposed Segmentation Algorithm

The algorithm can segment N regions of interest (ROIs) at one time by initializing $N+1$ different labels, where the additional one accounts for the “background”. As a first step, the user briefly draws over the different ROIs using at least two distinct labels (Fig.7).

Table 1: Outline of the datasets used.

Cohort	Modalities	Number of cases
Spleen	3D-CT	20
Breast	2D-DCE-MRI: pre-contrast first post-contrast second post-contrast	50
Lung	2D-CT	50
Glioblastoma	3D-MRI: T1 T2 T1CE (T1Gd) FLAIR	20

For each pair of image and class labels, an AGD map [16] (Fig.8) is produced reflecting a composite of intensity and spatial distance from the drawings, such that voxels far away and/or with very different intensity have higher values (in the figure, such voxels appear brighter). Additionally, for the purpose of providing more spatial information, three “coordinate” maps are used, one for each dimension of the image, where the values range from 0 to the size of the image in that dimension.

An ensemble [41] of SVMs is trained on voxels that belong to the drawings and segments the remainder of the scan. Each training sample (i.e., voxel) is described by the following features: (i) intensity across all co-registered images, (ii) distance in all AGD maps, and (iii) value in all coordinate maps. Three SVM classification models (i.e., radial basis function (RBF), chi-squared, histogram intersection kernels) are trained and their hyperparameters are selected through cross-validation. Each voxel’s final prediction is obtained by fusing the three model predictions via majority voting and the RBF classifier is used to resolve ties.

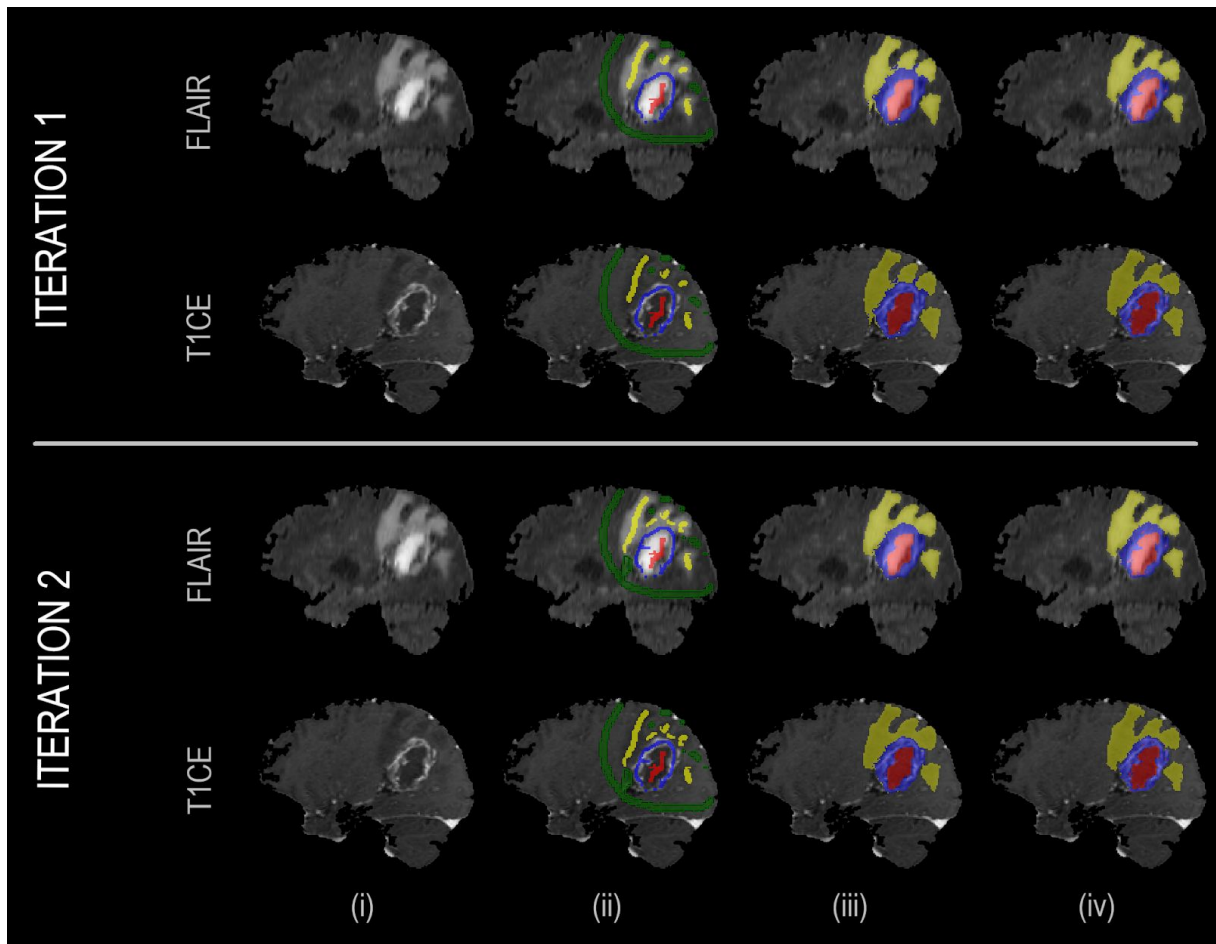


Figure 7: Example showcasing the result improving as a function of invested time. In the first iteration the user quickly draws over the different areas. In the second iteration, the user places few additional labels to correct representative misclassified areas, which are then used to retrain the machine learning model. From left to right: (i) Anatomical image (ii) User annotations (iii) result segmentation (iv) ground truth segmentation.

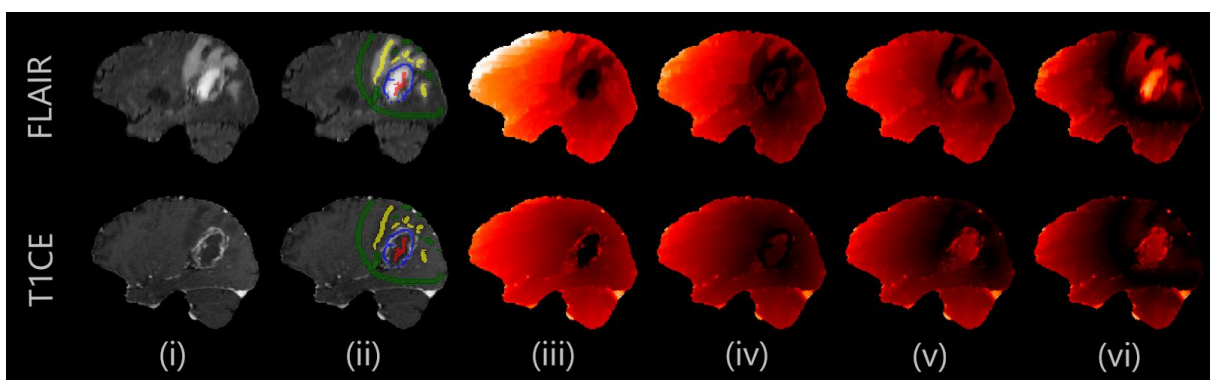


Figure 8: Example of AGD maps. Darker values indicate lower adaptive geodesic distance from the user drawings. In glioblastomas, NE and ET boundaries are clearer in T1CE, while the boundary between ED and background is clearer in FLAIR. From left to right: (i) Anatomical image (ii) User annotations (iii) AGD map for NE (Non-enhancing tumor core) annotation (iv) AGD map for ET (Enhancing tumor core) (v) AGD map for ED (peritumoral edema) (vi) AGD map for background.

4.3. Experimental Design

4.3.1. Evaluation protocol

To quantitatively evaluate the method, eight experts were included, two for each cohort. Each expert was asked to segment every scan four times, thereby producing two manual and two IML-assisted annotations, in addition to the extensively defined ground truth segmentations. The experts were given brief instructions for the method and were asked to note average time for their segmentations. To have a fair assessment of inter-rater consistency for glioblastoma segmentation, the experts were instructed to perform the manual segmentation of the various tumor sub-regions (enhancing tumor (ET), non-enhancing tumor (NE), and peritumoral edematous/infiltrated tissue (ED) [3], [4]) in 1 hour or less.

4.3.2. Experiment 1: Overall performance evaluation

Initially, the spatial overlap agreement of each approach was evaluated relative to the ground truth by utilizing the Dice Similarity Coefficient (DSC) as a metric to select one IML-assisted and one manual segmentation from each rater. For glioblastoma, only the whole tumor (WT) area was used for these selections. The DSCs of the two sets were statistically compared. Additionally, volumes calculated for IML and manual segmentations were quantitatively compared with the ground truth. Lastly, the Pearson's Correlation Coefficient [66] was estimated for each of the paired segmentations separately, i.e., (i) IML vs ground truth and (ii) manual vs ground truth. The average active drawing time was compared for each cohort between IML-assisted and manual segmentation.

4.3.3. Experiment 2: Intra-rater segmentation consistency

The DSCs within the two IML-assisted and the two manual segmentations of each rater were calculated (i.e., $DSC_{IML1/IML2}$ and $DSC_{Manual1/Manual2}$) for each case. The DSCs were statistically compared between the IML-assisted and manual segmentations. Furthermore, the existence

of significant differences between the DSC of the manual and IML-assisted segmentations relative to ground truth was also examined for each rater separately.

4.3.4. Experiment 3 - Inter-rater segmentation consistency

The best segmentations of each rater were selected with the same selection criteria as experiment 1. The DSCs within the best IML-assisted segmentations across raters, and within their best manual annotations, were calculated for each case, and their significant differences were evaluated.

4.3.5. Statistical Analysis

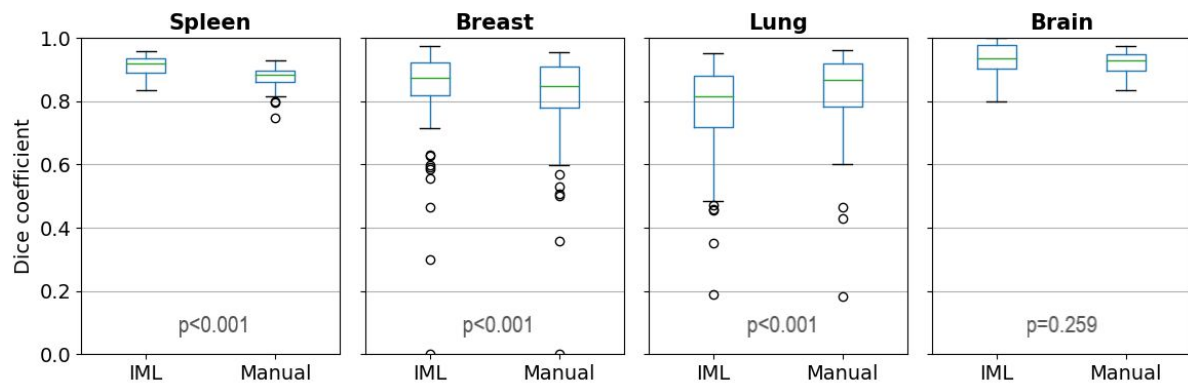
Paired Wilcoxon-signed rank non-parametric statistical tests [79] were used for statistical comparisons (assuming a type I error rate of 0.05), because the samples were paired and tended not to follow a Gaussian distribution.

5. Results

In this section, the results of the experimental validation are presented for 20 spleen, 50 breast tumor, 50 lung tumor, and 20 glioblastoma cases. For each patient, each expert produced two manual and two IML-assisted segmentations, which were additional to the extensively defined ground truth segmentations.

Performance

(i) Whole segmented area



(ii) Individual labels

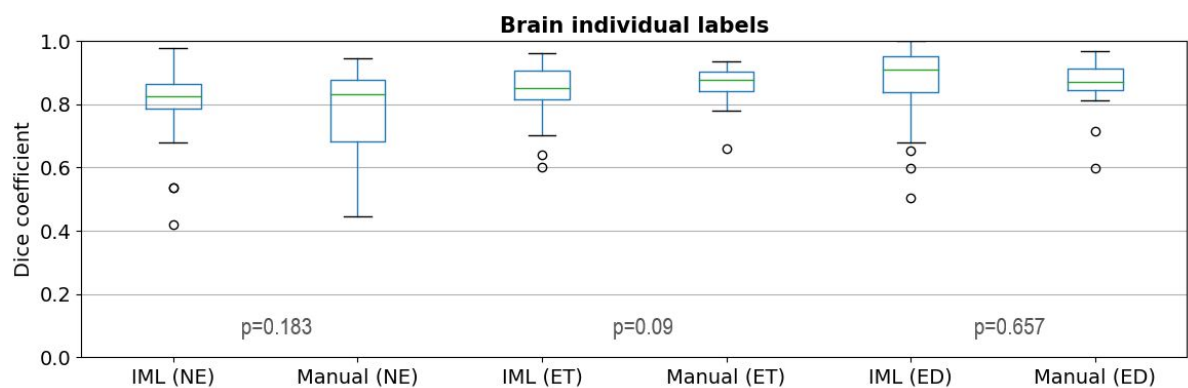
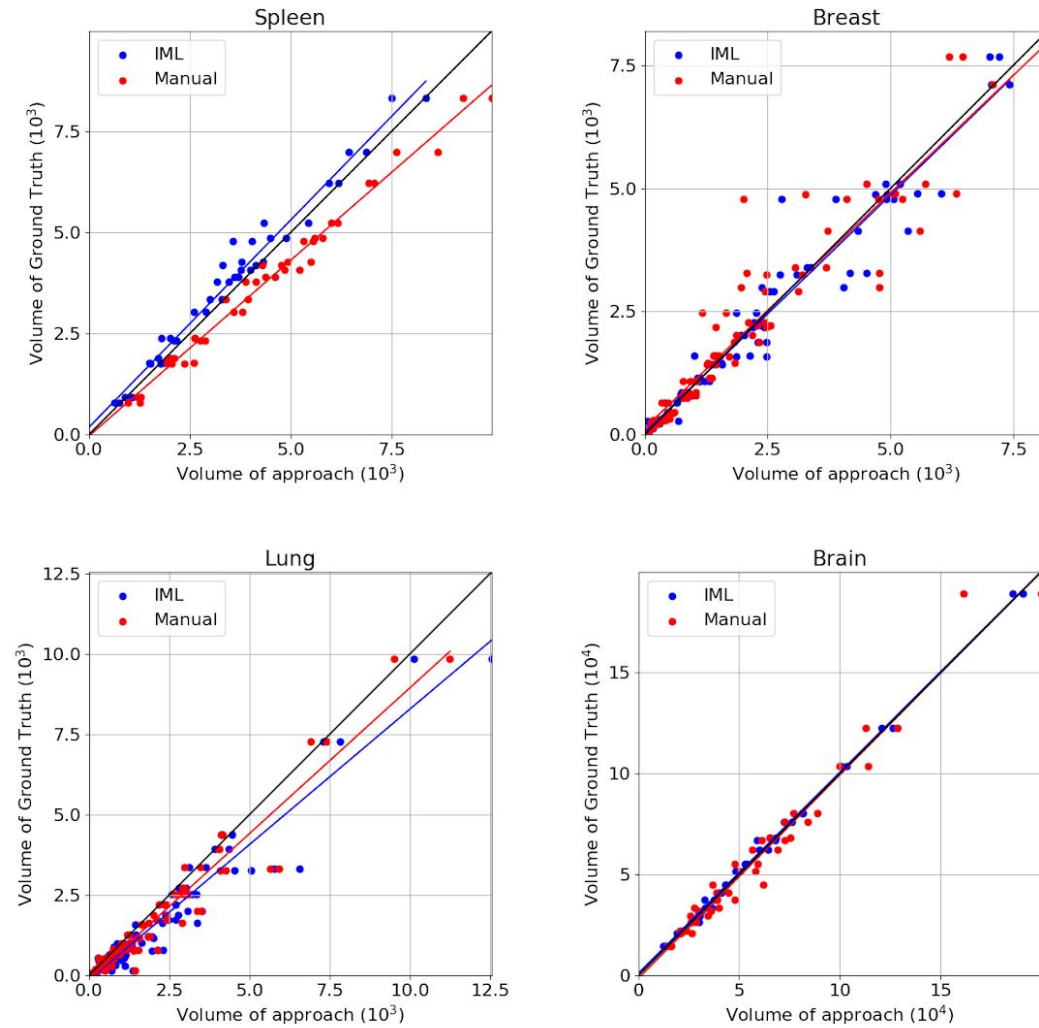


Figure 9: Dice coefficient, compared to ground truth, of: (i) all individual labels representing different areas of the structure counted as one, (ii) the individual areas of glioblastomas.

Analysis of segmentation volume

(i) Whole segmented area



(ii) Individual labels

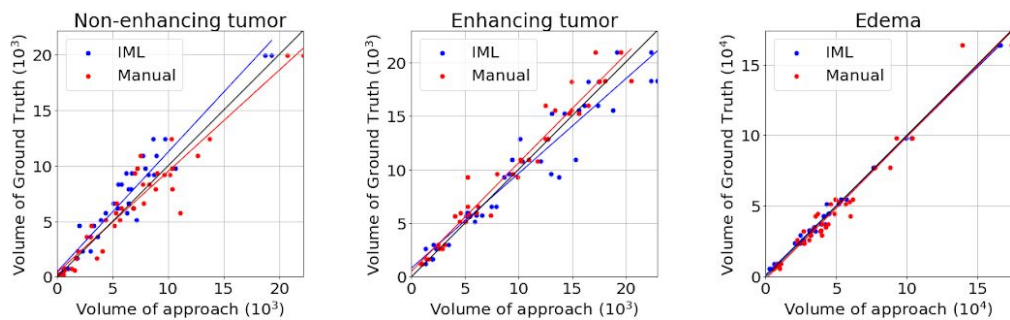


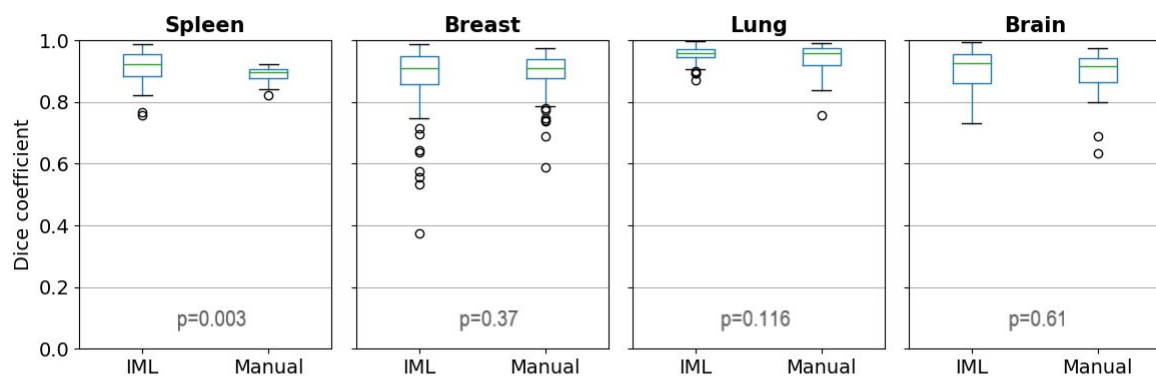
Figure 10: Scatterplots in which blue points are the pairs of volume of IML method and volume of ground truth and red are the pairs of manual segmentation volume and ground truth. The black line represents the ground truth's volume. The plots belong to (i) different cohorts where all individual labels, representing different areas of the structure, are counted as one, (ii) the sub-regions of glioblastomas.

5.1. Experiment 1: Overall Performance Evaluation

In the first experiment, the performance of the proposed method was evaluated (Table 3, Fig.9). For glioblastomas, manual and IML-assisted segmentations yielded similar pairs of DSCs both for WT and individual sub-regions, thereby indicating no significant difference between them, whereas the converse was true for other cohorts. The method achieved higher DSC on average for spleen and breast tumors, but lower for lung nodules when compared with the manual segmentations. However, the method was substantially faster than manual annotation in all cohorts (Table 3), by **53.1%** on average.

Intra-rater consistency

(i) Whole segmented area



(ii) Individual labels

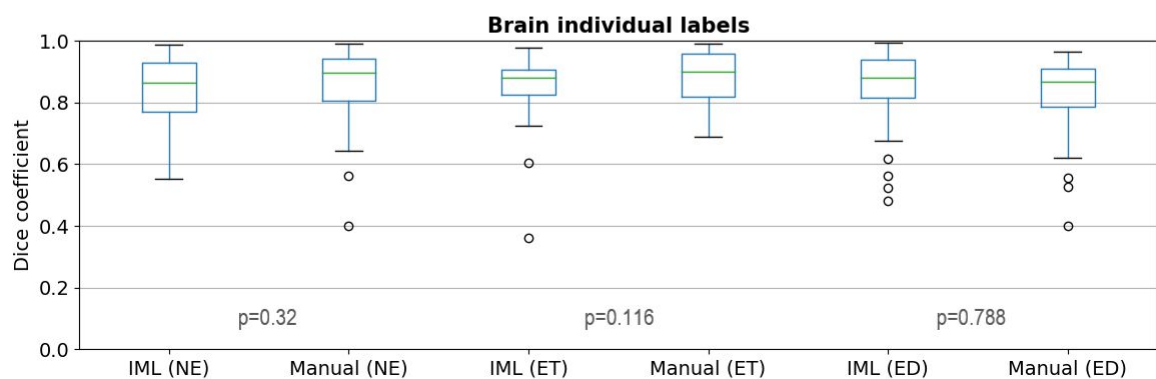


Figure 11: Dice coefficient, for intra-rater consistency, between the first and second round of the raters of: (i) all individual labels representing different areas of the structure counted as one, (ii) the individual areas of glioblastomas.

An analysis of the volume of ground truth, manual and IML-assisted segmentations (Fig.10, Table 3 “Corr. Coef.” column) shows that errors made by the method were mostly systematic. This is more evident in spleen images, where IML-assisted and manual segmentations revealed systematic under- and over-segmentation, respectively. The method made some non-systematic errors in lung nodules and the ET glioblastoma sub-region, but these areas were also more erroneous in manual segmentations. Notably, ET is regarded as the most challenging area of glioblastoma, because it frequently has unclear and smooth boundaries [3].

Table 2: Results (p-values) of a paired Wilcoxon test for each rater, comparing the dice coefficient results of the different approaches relative to ground. $p < 0.05$ indicates a significant difference.

	Rater 1		Rater 2	
Label	IML	Manual	IML	Manual
Spleen				
-	0.9405	0.3507	0.1454	0.433
Breast				
-	0.2425	0.5116	0.5921	0.1358
Lung				
-	0.0422	<0.0001	0.1358	<0.0001
Brain				
WT	0.156	0.0001	0.8813	0.0008
NE	0.0522	0.0859	0.9405	0.0001
ET	0.1913	0.3317	0.0522	0.0001
ED	0.4781	0.0001	0.6274	0.0008

5.2. Experiment 2: Intra-rater segmentation consistency

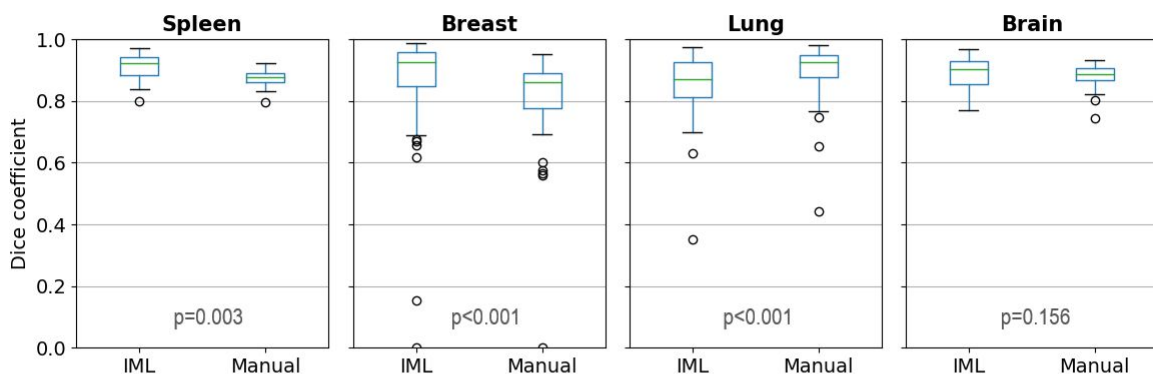
The second experiment attempts to quantify intra-rater consistency, comparing the two cycles of segmentations of each rater, separately for IML and manual (Table 3, Fig.11). No

significant difference was found between manual and IML-assisted segmentations for any of the cohorts, except spleen where segmentations using the method had higher mean overlap.

Additional analysis of DSC relative to ground truth (Table 2) found a significant difference in only one of the raters when using the IML method, while revealing a significant difference in 4/8 raters for manual annotations. Furthermore, there was no significant difference when using the IML method for any of the two raters for individual sub-regions of glioblastoma. The same tests for manual annotations revealed a significant difference in all sub-regions, except ET in one of the raters.

Inter-rater consistency

(i) Whole segmented area



(ii) Individual labels

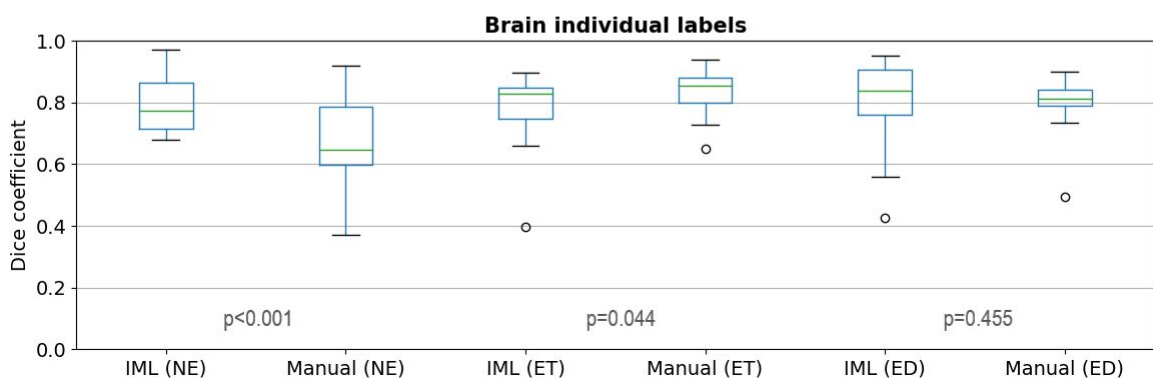


Figure 12: Dice coefficient, for inter-rater consistency, between segmentations of different raters, for: (i) all individual labels representing different areas of the structure counted as one, (ii) the individual areas of glioblastomas.

5.3. Experiment 3: Inter-rater segmentation consistency

In the last experiment, inter-rater consistency of IML and manual segmentations was calculated and compared (Table 3, Fig.12). There was a significant difference for spleen, breast, lung, and the NE and ET glioblastoma sub-regions. From those, the method achieved a higher overlap for spleen, breast, and the NE glioblastoma sub-region. Conversely, manual segmentations had higher agreement for lung and the ET region.

Table 3: Results for all three experiments. Experiment 1: Performance is calculated as dice coefficient relative to ground truth. p-values are a result of paired comparisons between the highest scoring IML-assisted and manual segmentations for each rater and each case. Correlation coefficient is calculated between the resultant and ground truth volumes. Experiment 2: Values indicate overlap between the first and second cycle of each rater. $p < 0.05$ indicates a significant difference between the results of IML and manual segmentations. Experiment 3: Values indicate overlap between segmentations of different raters. $p < 0.05$ indicates a significant difference in the inter-rater results for the respective cohort.

Label	Experiment 1						Experiment 2			Experiment 3			
	Mean Overlap			Time		Corr. Coef.		Intra-rater mean			Inter-rater mean		
	IML	Manual	p	IML	Manual	IML	Manual	IML	Manual	p	IML	Manual	p
Spleen													
-	0.91	0.87	$<10^{-3}$	66s	100s	0.99	0.99	0.91	0.89	0.003	0.91	0.87	0.003
Breast													
-	0.84	0.82	$<10^{-3}$	19s	70s	0.98	0.95	0.88	0.9	0.37	0.86	0.81	$<10^{-3}$
Lung													
-	0.78	0.83	$<10^{-3}$	93s	125s	0.96	0.97	0.96	0.95	0.116	0.85	0.89	$<10^{-3}$
Brain													
WT	0.94	0.92	0.259	21m	60m	1	0.98	0.91	0.89	0.61	0.89	0.88	0.156
NE	0.81	0.79	0.183			0.97	0.95	0.85	0.86	0.32	0.79	0.67	$<10^{-3}$
ET	0.85	0.87	0.09			0.96	0.98	0.85	0.88	0.116	0.79	0.84	0.044
ED	0.88	0.87	0.657			1	0.98	0.85	0.83	0.788	0.81	0.8	0.455

6. Discussion

In this study, a general-purpose, easy-to-use, and fast IML-based segmentation method was presented that can be applied in a multitude of research applications without requiring any adaptations to different domains or training of users. The method takes as input co-registered images and quick user drawings, to create AGD maps and train an ensemble of SVMs, used for segmenting the whole scan. The performance of the method was evaluated on solid structures across different cohorts, image modalities, and anatomical sites.

The method utilizes the power of ML; however, it mitigates one of its known weaknesses, i.e., the need for extensive training and lack of reproducibility on new datasets. By virtue of being trained interactively, segmentation models are optimal for the specific individual's scans. Additional benefits include the ability of the method to be parallelized and low hardware requirements. The disadvantage of this approach is that it is not fully automated.

Quantitative evaluation showed great promise for the applicability of the method in various structures relevant to medical research. Accuracy and inter-rater agreement were comparable to manual segmentation, while intra-rater agreement was high, indicating that the method is stable. Volumetric errors were mostly systematic, indicating that results can be improved through further iterations or volumetric operations like shrinking/expanding. The method was also shown to be fast and not require excessive interaction.

The presented method can be utilized in a vast array of applications, because fully automatic segmentation has not received widespread clinical adoption yet. Training of automatic methods requires large datasets that are difficult to create and obtain. Additionally, deep learning methods are often inaccessible to clinicians, because they require specialized hardware and don't have a standardized distribution method yet. Furthermore, automatic

methods target specific problems, which means that it would take years, expensive acquisition trials, and vast inter-institutional cooperation to build robust models for every structure of the body. Some automatic approaches also face the problem of domain adaptation, meaning that the trained models have trouble segmenting images acquired from machines with different intensity and noise profiles than the ones they were trained on. As a result, interactive segmentation is currently highly relevant and will probably continue to be, to some extent, in the future. Lastly, medical image segmentation is complicated, because label assignment often relies on advanced medical knowledge and patient information, especially for cancerous structures. Consequently, automatic methods will probably never become synonymous with ground truth and there will be a need for quick expert corrections, potentially with interactive methods that use a segmentation produced automatically alongside input from a clinician to improve the result.

ITK-SNAP [14] also provides a method for interactive segmentation. However, it requires users to follow a more complex protocol to achieve multi-label segmentation. The user first provides quick drawings for the different ROIs and trains a model. Afterwards and separately for each class, the user must place seeds and evolve a contour. According to their evaluation on high-grade glioblastomas, also on the BraTS dataset, there is a lower mean agreement with the ground truth for ITK-SNAP in the regions this thesis also evaluated, particularly ET ($Dice_{IML}/Dice_{ITK-SNAP}=0.85/0.69$) and WT ($Dice_{IML}/Dice_{ITK-SNAP}=0.94/0.85$). Average user interaction time was also lower for the method presented here ($Time_{IML}/Time_{ITK-SNAP}=21\text{min}/27.8\text{min}$). 3D Slicer's "grow from seeds" effect follows a workflow similar to the one that was presented here, but it can only support one image as input.

Future research can improve this method on multiple fronts. Advanced ML techniques, such as semi-supervised learning, can potentially increase the accuracy and consistency of the results. Transfer learning could expand the range of tasks to non-solid structures, such as

brain lesions. If a specific task is targeted, pre-trained population-derived models, atlases, and specialized preprocessing techniques can potentially aid in producing better segmentations. Furthermore, a prospective dataset, especially one acquired under different acquisition settings, would lend further validity to the method.

The results showed that the method has accuracy and inter-rater consistency on par with manual segmentation across different solid anatomical structures and modalities. Additionally, the method showed high intra-rater consistency and minimized user interaction.

7. Conclusion

The purpose of this thesis was to compare the proposed method with manual segmentation in an experimental setting that included segmentation of spleens and brain, breast, and lung tumors. Various tests concluded that the method is comparable to manual segmentation in accuracy, intra- and inter-rater consistency in most situations, while even outperforming manual segmentation in some others. Additionally, the method was always faster, making it suitable for applications that require segmenting many cases.

8. References

- [1] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annu. Rev. Biomed. Eng.*, vol. 2, pp. 315–337, 2000, doi: 10.1146/annurev.bioeng.2.1.315.
- [2] N. Sharma and L. M. Aggarwal, "Automated medical image segmentation techniques," *J. Med. Phys.*, vol. 35, no. 1, pp. 3–14, Jan. 2010, doi: 10.4103/0971-6203.58777.
- [3] B. H. Menze *et al.*, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015, doi: 10.1109/tmi.2014.2377694.
- [4] S. Bakas *et al.*, "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," *Sci Data*, vol. 4, p. 170117, 2017, doi: 10.1038/sdata.2017.117.
- [5] S. Golemati and K. Nikita, Eds., *Cardiovascular Computing-Methodologies and Clinical Applications*. Springer, 2020.
- [6] S. Rathore *et al.*, "Radiomic signature of infiltration in peritumoral edema predicts subsequent recurrence in glioblastoma: implications for personalized radiotherapy planning," *J Med Imaging (Bellingham)*, vol. 5, no. 2, p. 021219, Apr. 2018, doi: 10.1117/1.Jmi.5.2.021219.
- [7] B. Sahiner *et al.*, *The effect of nodule segmentation on the accuracy of computerized lung nodule detection on CT scans: comparison on a data set annotated by multiple radiologists*, vol. 6514. SPIE, 2007, p. MI.
- [8] S. Golemati, J. Stoitsis, E. G. Sifakis, T. Balkizas, and K. S. Nikita, "Using the Hough transform to segment ultrasound images of longitudinal and transverse sections of the carotid artery," *Ultrasound Med. Biol.*, vol. 33, no. 12, pp. 1918–1932, Dec. 2007, doi: 10.1016/j.ultrasmedbio.2007.05.021.
- [9] A. S. Panayides *et al.*, "AI in Medical Imaging Informatics: Current Challenges and Future Directions," *IEEE Journal of Biomedical and Health Informatics*. pp. 1–1, 2020, doi: 10.1109/jbhi.2020.2991043.
- [10] J. H. Thrall *et al.*, "Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success," *J. Am. Coll. Radiol.*, vol. 15, no. 3, Part B, pp. 504–508, Mar. 2018, doi: 10.1016/j.jacr.2017.12.026.
- [11] P. A. Yushkevich *et al.*, "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006, doi: 10.1016/j.neuroimage.2006.01.015.
- [12] S. Bakas *et al.*, "GLISTRboost: Combining Multimodal MRI Segmentation, Registration, and Biophysical Tumor Growth Modeling with Gradient Boosting Machines for Glioma Segmentation," *Brainlesion*, vol. 9556, pp. 144–155, 2016, doi: 10.1007/978-3-319-30858-6_1.
- [13] K. Zeng *et al.*, "Segmentation of Gliomas in Pre-operative and Post-operative Multimodal Magnetic Resonance Imaging Volumes Based on a Hybrid Generative-Discriminative Framework," *Brainlesion*, vol. 10154, pp. 184–194, Oct. 2016, doi: 10.1007/978-3-319-55524-9_18.
- [14] P. A. Yushkevich *et al.*, "User-Guided Segmentation of Multi-modality Medical Imaging Datasets with ITK-SNAP," *Neuroinformatics*, vol. 17, no. 1, pp. 83–102, Jan. 2019, doi: 10.1007/s12021-018-9385-x.
- [15] A. Fedorov *et al.*, "3D Slicer as an image computing platform for the Quantitative Imaging Network," no. 1873–5894 (Electronic), 2012.
- [16] B. Gaonkar *et al.*, "Automated Tumor Volumetry Using Computer-Aided Image Segmentation," *Acad. Radiol.*, vol. 22, no. 5, pp. 653–661, 2015, doi: 10.1016/j.acra.2015.01.005.

- [17] “Diagnostic imaging,” *World Health Organization*.
https://www.who.int/diagnostic_imaging/en/ (accessed Jul. 21, 2020).
- [18] E. Friedberg *et al.*, “Unifying the Silos of Subspecialized Radiology: The Essential Role of the General Radiologist,” *J. Am. Coll. Radiol.*, vol. 15, no. 8, pp. 1158–1163, Aug. 2018, doi: 10.1016/j.jacr.2018.05.016.
- [19] “Imaging and radiology,” *medlineplus.gov*.
<https://medlineplus.gov/ency/article/007451.htm> (accessed Jul. 20, 2020).
- [20] T. P. Murphy, “Introduction to clinical interventional radiology,” *Semin. Intervent. Radiol.*, vol. 22, no. 1, pp. 3–5, Mar. 2005, doi: 10.1055/s-2005-869569.
- [21] T. Kurki, N. Lundbom, and S. Valtonen, “Tissue characterisation of intracranial tumours: the value of magnetisation transfer and conventional MRI,” *Neuroradiology*, vol. 37, no. 7, pp. 515–521, Oct. 1995, doi: 10.1007/BF00593707.
- [22] P. Mildemberger, M. Eichelberg, and E. Martin, “Introduction to the DICOM standard,” *Eur. Radiol.*, vol. 12, no. 4, pp. 920–927, Apr. 2002, doi: 10.1007/s003300101100.
- [23] I. Wolf *et al.*, “The medical imaging interaction toolkit,” *Med. Image Anal.*, vol. 9, no. 6, pp. 594–604, Dec. 2005, doi: 10.1016/j.media.2005.04.005.
- [24] GCO, “All Cancers - GCO,” *International Agency for Research on Cancer*.
<https://gco.iarc.fr/today/data/factsheets/cancers/39-All-cancers-fact-sheet.pdf> (accessed Jul. 27, 2020).
- [25] P. Anand *et al.*, “Cancer is a preventable disease that requires major lifestyle changes,” *Pharm. Res.*, vol. 25, no. 9, pp. 2097–2116, Sep. 2008, doi: 10.1007/s11095-008-9661-9.
- [26] “Types of cancer,” *Cancer Research UK*.
<https://www.cancerresearchuk.org/what-is-cancer/how-cancer-starts/types-of-cancer> (accessed Jul. 27, 2020).
- [27] “What is Cancer?,” *cancer.net*.
<https://www.cancer.net/navigating-cancer-care/cancer-basics/what-cancer> (accessed Sep. 08, 2020).
- [28] A. Gupta and T. Dwivedi, “A Simplified Overview of World Health Organization Classification Update of Central Nervous System Tumors 2016,” *J. Neurosci. Rural Pract.*, vol. 8, no. 4, pp. 629–641, Oct. 2017, doi: 10.4103/jnrp.jnrp_168_17.
- [29] D. N. Louis *et al.*, “WHO classification and grading of tumours of the central nervous system,” *WHO Classification of Tumours of the Central Nervous System. 4th ed. , Revised. Lyon: International Agency for Research Centre*, pp. 12–13, 2016.
- [30] A. Brodbelt, D. Greenberg, T. Winters, M. Williams, S. Vernon, and V. P. Collins, “Glioblastoma in England: 2007–2011,” *Eur. J. Cancer*, vol. 51, no. 4, pp. 533–542, Mar. 2015, doi: 10.1016/j.ejca.2014.12.014.
- [31] L. Sun, S. Zhang, H. Chen, and L. Luo, “Brain Tumor Segmentation and Survival Prediction Using Multimodal MRI Scans With Deep Learning,” *Front. Neurosci.*, vol. 13, p. 810, Aug. 2019, doi: 10.3389/fnins.2019.00810.
- [32] S. Bakas *et al.*, “Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge,” *arXiv*, p. arXiv:1811.02629, 2018, [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2018arXiv181102629B>.
- [33] O. Eidel *et al.*, “Tumor Infiltration in Enhancing and Non-Enhancing Parts of Glioblastoma: A Correlation with Histopathology,” *PLoS One*, vol. 12, no. 1, p. e0169292, Jan. 2017, doi: 10.1371/journal.pone.0169292.
- [34] S. G. Armato 3rd, M. L. Giger, and H. MacMahon, “Automated detection of lung nodules in CT scans: preliminary results,” *Med. Phys.*, vol. 28, no. 8, pp. 1552–1561, Aug. 2001, doi: 10.1118/1.1387272.
- [35] J. Wei, Y. Hagihara, A. Shimizu, and H. Kobatake, “Optimal image feature set for

- detecting lung nodules on chest X-ray images,” in *CARS 2002 Computer Assisted Radiology and Surgery*, 2002, pp. 706–711, doi: 10.1007/978-3-642-56168-9_118.
- [36] M. Woźniak, D. Połap, G. Capizzi, G. L. Sciuto, L. Kośmider, and K. Frankiewicz, “Small lung nodules detection based on local variance analysis and probabilistic neural network,” *Comput. Methods Programs Biomed.*, vol. 161, pp. 173–180, Jul. 2018, doi: 10.1016/j.cmpb.2018.04.025.
- [37] C. A. Owens *et al.*, “Lung tumor segmentation methods: Impact on the uncertainty of radiomics features for non-small cell lung cancer,” *PLoS One*, vol. 13, no. 10, p. e0205003, Oct. 2018, doi: 10.1371/journal.pone.0205003.
- [38] C. E. DeSantis *et al.*, “Breast cancer statistics, 2019,” *CA Cancer J. Clin.*, vol. 69, no. 6, pp. 438–451, Nov. 2019, doi: 10.3322/caac.21583.
- [39] S. Park *et al.*, “Changes in Noninvasive Liver Fibrosis Indices and Spleen Size During Chemotherapy: Potential Markers for Oxaliplatin-Induced Sinusoidal Obstruction Syndrome,” *Medicine*, vol. 95, no. 2, p. e2454, Jan. 2016, doi: 10.1097/MD.0000000000002454.
- [40] F. A. Fasola and A. J. Adekanmi, “HAEMATOLOGICAL PROFILE AND BLOOD TRANSFUSION PATTERN OF PATIENTS WITH SICKLE CELL ANAEMIA VARY WITH SPLEEN SIZE,” *Ann Ib Postgrad Med*, vol. 17, no. 1, pp. 30–38, Jun. 2019, [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/31768154>.
- [41] L. Rokach, “Ensemble-based classifiers,” *Artificial Intelligence Review*, vol. 33, no. 1, pp. 1–39, Feb. 2010, doi: 10.1007/s10462-009-9124-7.
- [42] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27–21, 2011.
- [43] R. Filipovych and C. Davatzikos, “Semi-supervised pattern classification of medical images: Application to mild cognitive impairment (MCI),” *Neuroimage*, vol. 55, no. 3, pp. 1109–1119, Apr. 2011, doi: 10.1016/j.neuroimage.2010.12.066.
- [44] M. Ghafoorian *et al.*, “Transfer Learning for Domain Adaptation in MRI: Application in Brain Lesion Segmentation,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, 2017, pp. 516–524, doi: 10.1007/978-3-319-66179-7_59.
- [45] A. van Oopbroek, M. A. Ikram, M. W. Vernooij, and M. de Bruijne, “Transfer Learning Improves Supervised Image Segmentation Across Imaging Protocols,” *IEEE Trans. Med. Imaging*, vol. 34, no. 5, pp. 1018–1030, May 2015, doi: 10.1109/TMI.2014.2366792.
- [46] H.-C. Shin *et al.*, “Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.
- [47] N. Heller *et al.*, “The KiTS19 Challenge Data: 300 Kidney Tumor Cases with Clinical Context, CT Semantic Segmentations, and Surgical Outcomes,” *arXiv [q-bio.QM]*, Mar. 31, 2019.
- [48] Y. Sun *et al.*, “Multi-Site Infant Brain Segmentation Algorithms: The iSeg-2019 Challenge,” *arXiv [eess.IV]*, Jul. 04, 2020.
- [49] G. Litjens *et al.*, “Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge,” *Med. Image Anal.*, vol. 18, no. 2, pp. 359–373, Feb. 2014, doi: 10.1016/j.media.2013.12.002.
- [50] B. Boyer, C. Balleyguier, O. Granat, and C. Pharaboz, “CAD in questions/answers Review of the literature,” *Eur. J. Radiol.*, vol. 69, no. 1, pp. 24–33, Jan. 2009, doi: 10.1016/j.ejrad.2008.07.042.
- [51] J. Firth-Cozens, “Doctors, their wellbeing, and their stress,” *BMJ*, vol. 326, no. 7391, pp. 670–671, Mar. 2003, doi: 10.1136/bmj.326.7391.670.
- [52] S. G. Mougiakakou, S. Golemati, I. Gousias, A. N. Nicolaidis, and K. S. Nikita, “Computer-aided diagnosis of carotid atherosclerosis based on ultrasound image

- statistics, Laws' texture and neural networks," *Ultrasound in Medicine & Biology*, vol. 33, no. 1. pp. 26–36, 2007, doi: 10.1016/j.ultrasmedbio.2006.07.032.
- [53] M. Gletsos, S. G. Mougiakakou, G. K. Matsopoulos, K. S. Nikita, A. S. Nikita, and D. Kelekis, "A computer-aided diagnostic system to characterize CT focal liver lesions: design and optimization of a neural network classifier," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 3, pp. 153–162, Sep. 2003, doi: 10.1109/titb.2003.813793.
- [54] J. Stoitsis, I. Valavanis, S. G. Mougiakakou, S. Golemati, A. Nikita, and K. S. Nikita, "Computer aided diagnosis based on medical image processing and artificial intelligence methods," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 569, no. 2. pp. 591–595, 2006, doi: 10.1016/j.nima.2006.08.134.
- [55] K. S. Nikita, "Atherosclerosis: the evolving role of vascular image analysis," *Comput. Med. Imaging Graph.*, vol. 37, no. 1, pp. 1–3, Jan. 2013, doi: 10.1016/j.compmedimag.2012.12.001.
- [56] J. Yanase and E. Triantaphyllou, "A systematic survey of computer-aided diagnosis in medicine: Past and present developments," *Expert Syst. Appl.*, vol. 138, p. 112821, Dec. 2019, doi: 10.1016/j.eswa.2019.112821.
- [57] K. Nikita, D. Koutsouris, and S. Pavlopoulos, *Ιατρικά απεικονιστικά συστήματα*. ΤΖΙΟΛΑΣ, 2004.
- [58] N. N. Tsiaparas, S. Golemati, I. Andreadis, J. S. Stoitsis, I. Valavanis, and K. S. Nikita, "Comparison of multiresolution features for texture classification of carotid atherosclerosis from B-mode ultrasound," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 1, pp. 130–137, Jan. 2011, doi: 10.1109/TITB.2010.2091511.
- [59] S. G. Mougiakakou, I. Valavanis, K. S. Nikita, A. Nikita, and D. Kelekis, "Characterization of CT liver lesions based on texture features and a multiple neural network classification scheme," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*, 2003, vol. 2, pp. 1287–1290 Vol.2, doi: 10.1109/IEMBS.2003.1279504.
- [60] P. Asvestas, G. K. Matsopoulos, and K. S. Nikita, "A Power Differentiation Method of Fractal Dimension Estimation for 2-D Signals," *J. Vis. Commun. Image Represent.*, vol. 9, no. 4, pp. 392–400, Dec. 1998, doi: 10.1006/jvci.1998.0394.
- [61] J. J. M. van Griethuysen *et al.*, "Computational Radiomics System to Decode the Radiographic Phenotype," *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, Nov. 2017, doi: 10.1158/0008-5472.CAN-17-0339.
- [62] S. E. Chatzistergos, I. Andreadis, and K. S. Nikita, "Identification of architectural distortions in mammograms using local binary patterns and radial lengths through an exhaustive evaluation framework," *Expert Syst.*, vol. 35, no. 4, p. e12281, Aug. 2018, doi: 10.1111/exsy.12281.
- [63] V. Bolón-Canedo, N. Sánchez-Marroño, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowl. Inf. Syst.*, vol. 34, no. 3, pp. 483–519, 2013, [Online]. Available: https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1007/s10115-012-0487-8&casa_token=8S2e0NubWboAAAAA:Flb3uuxUkyfq-1gl8RrzXJs9FnHuMs1Q2lumEQHQXJ6-BQBiEhCpK8tWm2fPBHQd8CrSLh4KwY9tDofH.
- [64] B. Gaonkar, L. Shu, G. Hermosillo, and Y. Zhan, "Adaptive geodesic transform for segmentation of vertebrae on CT images," in *Medical Imaging 2014: Computer-Aided Diagnosis*, Mar. 2014, vol. 9035, p. 903516, doi: 10.1117/12.2043527.
- [65] K. H. Zou *et al.*, "Statistical validation of image segmentation quality based on a spatial overlap index," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, Feb. 2004, doi: 10.1016/s1076-6332(03)00671-8.

- [66] S. M. Stigler, "Francis Galton's Account of the Invention of Correlation," *Stat. Sci.*, vol. 4, no. 2, pp. 73–79, 1989.
- [67] C. Davatzikos *et al.*, "Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome," *J Med Imaging (Bellingham)*, vol. 5, no. 1, p. 011018, Jan. 2018, doi: 10.1117/1.Jmi.5.1.011018.
- [68] M. McCormick, X. Liu, J. Jomier, C. Marion, and L. Ibanez, "ITK: enabling reproducible research and open science," *Front. Neuroinform.*, vol. 8, p. 13, Feb. 2014, doi: 10.3389/fninf.2014.00013.
- [69] T. S. Yoo *et al.*, "Engineering and algorithm design for an image processing Api: a technical report on ITK--the Insight Toolkit," *Stud. Health Technol. Inform.*, vol. 85, pp. 586–592, 2002, [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/15458157>.
- [70] BRADSKI and G, "The OpenCV library," *Dr Dobb's J. Software Tools*, vol. 25, pp. 120–125, 2000, Accessed: Jun. 16, 2020. [Online]. Available: <https://ci.nii.ac.jp/naid/10028167478/>.
- [71] W. McKinney and Others, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, 2010, vol. 445, pp. 51–56, [Online]. Available: <http://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf>.
- [72] T. E. Oliphant, *A guide to NumPy*, vol. 1. Trelgol Publishing USA, 2006.
- [73] P. Virtanen *et al.*, "Author Correction: SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nat. Methods*, vol. 17, no. 3, pp. 352–352, Mar. 2020, doi: 10.1038/s41592-020-0772-5.
- [74] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, May 2007, doi: 10.1109/MCSE.2007.55.
- [75] A. L. Simpson *et al.*, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," *arXiv*, p. arXiv:1902.09063, 2019, [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2019arXiv190209063S>.
- [76] H. J. Aerts *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nat. Commun.*, vol. 5, p. 4006, 2014, doi: 10.1038/ncomms5006.
- [77] H. J. W. L. Aerts *et al.*, "Data From NSCLC-Radiomics [Data set]," *The Cancer Imaging Archive.*, 2019.
- [78] K. Clark *et al.*, "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *J. Digit. Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013, doi: 10.1007/s10278-013-9622-7.
- [79] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945, doi: 10.2307/3001968.