



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Διαμόρφωση Ολοκληρωμένου Συστήματος Τμηματοποίησης Εστιακών Αλλοιώσεων Πολλαπλού Μυελώματος από Ολόσωμες Μαγνητικές Τομογραφίες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΣΤΥΛΙΑΝΟΥ ΚΑΤΣΗ

Επιβλέπων : Γεώργιος Ματσόπουλος
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπουσα : Ουρανία Πετροπούλου
ΕΔΙΠ Ε.Μ.Π.

Αθήνα, Οκτώβριος 2025



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Διαμόρφωση Ολοκληρωμένου Συστήματος Τμηματοποίησης Εστιακών Αλλοιώσεων Πολλαπλού Μυελώματος από Ολόσωμες Μαγνητικές Τομογραφίες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΣΤΥΛΙΑΝΟΥ ΚΑΤΣΗ

Επιβλέπων : Γεώργιος Ματσόπουλος
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπουσα : Ουρανία Πετροπούλου
ΕΔΙΠ Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 17 Οκτωβρίου 2025.

(Υπογραφή)

.....
Γ. Ματσόπουλος
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

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

(Υπογραφή)

.....
Π. Τσανάκας
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2025



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Copyright © Στυλιανός Κατσης, 2025.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

(Υπογραφή)

.....
ΣΤΥΛΙΑΝΟΣ ΚΑΤΣΗΣ

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

Περίληψη

Το Πολλαπλό Μυέλωμα αποτελεί μια σύνθετη και ετερογενή αιματολογική κακοήθεια, η οποία διηθεί τον μυελό των οστών και εκδηλώνεται με τρεις διαφορετικές μορφές: διάχυτο, εστιακό ή ποικίλο πρότυπο. Η ετερογένεια της νόσου καθιστά την ακρίβεια της διάγνωσης και της αντιμετώπισης εξαιρετικά απαιτητική. Η πολυτροπική προσέγγιση μέσω Whole-Body MRI (WB-MRI), και ειδικότερα βάσει του πρωτοκόλλου MY-RADS, προσφέρει ένα ισχυρό εργαλείο για τη μη επεμβατική αξιολόγηση της έκτασης της νόσου. Ωστόσο, η ερμηνεία των MRI εξετάσεων εξακολουθεί να βασίζεται σε μεγάλο βαθμό στην οπτική εκτίμηση των εικόνων από ειδικούς, γεγονός που μπορεί να οδηγήσει σε ασυνέπειες, παραλείψεις ή ασάφειες στις διαγνώσεις, ειδικά σε περιπτώσεις με κακή ποιότητα εικόνας ή συνύπαρξη διάχυτων και εστιακών αλλοιώσεων. Στο πλαίσιο αυτό, η παρούσα εργασία επιχειρεί να καλύψει αυτό το κενό, προτείνοντας μια αυτοματοποιημένη και επεκτάσιμη μεθοδολογία με τη χρήση της Μηχανικής Μάθησης (ML) για την ανίχνευση εστιακών σπονδυλικών αλλοιώσεων σε ασθενείς με Πολλαπλό Μυέλωμα. Για το εγχείρημα αυτό γίνεται χρήση μίας κατηγορίας Νευρωνικών Δικτύων, η οποία, αν και καινοτόμα, δεν έχει υπάρξει αντικείμενο μεγάλης έρευνας, και ειδικά στον συγκεκριμένο τομέα. Συγκεκριμένα θα χρησιμοποιηθούν τα Νευρωνικά Δίκτυα Γράφων (Graph Neural Networks – GNNs), προκειμένου να εντοπιστούν με τη μέγιστη δυνατή ακρίβεια κακοήθειες της νόσου. Σύμφωνα με τη μέθοδο αυτή, ένα σύνολο από εικόνες κάθε ασθενή θα κατακερματιστεί σε γεωμετρικά συναφείς περιοχές, οι οποίες με τη σειρά τους θα αποτελέσουν τους κόμβους για τη δημιουργία ενός γράφου. Σκοπός της παρούσας διπλωματικής εργασίας είναι η εξερεύνηση μιας εναλλακτικής μεθόδου εντοπισμού αλλοιώσεων πέραν των καθιερωμένων, η οποία, εκτός από απλές συσχετίσεις μεταξύ χρωμάτων και σχημάτων, θα μπορεί να λαμβάνει υπόψη και γεωμετρικές και άλλες εξαρτήσεις. Η έρευνα αυτή είναι επεκτάσιμη και μπορεί να προσαρμοστεί περεταίρω με την εισαγωγή επιπλέον πληροφορίας στους γράφους και την κάλυψη περισσότερων τύπων αλλοιώσεων, αυξάνοντας έτσι την ακρίβεια στην αυτόματη σταδιοποίηση και παρακολούθηση της θεραπευτικής ανταπόκρισης σε ασθενείς με Πολλαπλό Μυέλωμα.

Λέξεις κλειδιά: Πολλαπλό Μυέλωμα (ΠΜ), Νευρωνικά Δίκτυα Γράφων (GNN), supervoxels, radiomics, αλλοιώσεις, κατακερματισμός, εξαγωγή/επιλογή χαρακτηριστικών

Abstract

Multiple Myeloma is a complex and heterogeneous hematologic malignancy that infiltrates the bone marrow and manifests in three distinct patterns: diffuse, focal, or mixed. The heterogeneity of the disease makes accurate diagnosis and management extremely challenging. The multiparametric approach using Whole-Body MRI (WB-MRI), and in particular based on the MY-RADS protocol, provides a powerful tool for the non-invasive assessment of disease extent. However, the interpretation of MRI examinations still relies heavily on the visual evaluation of images by experts, which may lead to inconsistencies, omissions, or ambiguities in diagnoses—especially in cases of poor image quality or the coexistence of diffuse and focal lesions. In this context, the present study aims to address this gap by proposing an automated and scalable methodology using Machine Learning (ML) for the detection of focal spinal lesions in patients with Multiple Myeloma. For this purpose, a class of Neural Networks is employed, which, although innovative, has not been the subject of extensive research—particularly in this specific domain. Specifically, Graph Neural Networks (GNNs) will be used to detect disease-related malignancies with the highest possible accuracy. According to this approach, a set of images from each patient will be segmented into geometrically coherent regions, which will in turn serve as the nodes for the construction of a graph. The objective of this thesis is to explore an alternative method for lesion detection beyond conventional approaches—one that, apart from simple correlations between colors and shapes, can also consider geometric and other dependencies. This research is extensible and can be further adapted by incorporating additional information into the graphs and encompassing more lesion types, thereby increasing the accuracy of automated staging and monitoring of therapeutic response in patients with Multiple Myeloma.

Keywords: Multiple Myeloma (MM), Graph Neural Networks (GNN), supervoxels, radiomics, lesions, segmentation, feature extraction/selection

Ευχαριστίες

Πρώτα απ' όλα θα ήθελα να ευχαριστήσω τους επιβλέποντες καθηγητές κ. Γεώργιο Ματσόπουλο και Δρ. Ουρανία Πετροπούλου που μου εμπιστεύτηκαν το συγκεκριμένο θέμα και ακόμα περισσότερο τον υποψήφιο διδάκτορα Γεώργιο Μπότη για τη βοήθεια και την καθοδήγησή που μου προσέφερε καθ' όλη τη διάρκεια της έρευνας και της συγγραφής της διπλωματικής μου εργασίας.

Θα ήθελα, επίσης, να ευχαριστήσω την οικογένεια μου, που με ενθάρρυνε με τον δικό της τρόπο να προσπαθώ για το καλύτερο και να θέτω υψηλούς στόχους. Τέλος ευχαριστώ θερμά τους φίλους μου και την κοπέλα μου για τη δύναμη, την υπομονή και το κουράγιο να ολοκληρώσω τη διπλωματική μου και στάθηκαν στο πλευρό μου καθ' όλη τη διάρκεια των σπουδών μου.

Αθήνα, Οκτώβριος 2025
Κατσής Στυλιανός

Πίνακας περιεχομένων

Εισαγωγή	14
1 Θεωρητικό Υπόβαθρο	15
1.1 Το Πολλαπλό Μυέλωμα	15
1.1.1 Γενικά στοιχεία	15
1.1.2 Ο ρόλος της απεικόνισης στην διάγνωση και την διαχείριση του ΠΜ	16
1.1.3 Οι απεικονιστικές μέθοδοι και οι ενδείξεις τους	16
1.1.4 Πρωτόκολλα WB-MRI στην διαχείριση του ΠΜ	17
1.2 Η χρήση MRI στη διάγνωση	17
1.2.1 Βασικές Αρχές MRI	17
1.2.1.1 Πυρηνικός Μαγνητικός Συντονισμός	18
1.2.2 Τύποι Ακολουθιών	19
1.2.2.1 Μορφολογικές Ακολουθίες	19
1.2.2.2 Ποσοτικές Ακολουθίες	21
1.3 Επεξεργασία Ιατρικών Εικόνων	23
1.3.1 Supervoxels και κατακερματισμός (SLIC)	23
1.3.2 Εξαγωγή και Επιλογή Χαρακτηριστικών	24
1.3.2.1 Εξαγωγή Χαρακτηριστικών	25
1.3.2.2 Επιλογή Χαρακτηριστικών	26
1.4 Γράφοι και Μοντέλα GNN στην Ιατρική Απεικόνιση	27
1.4.1 Γράφοι: ορισμοί και χαρακτηριστικά	27
1.4.2 Θεμελιώδεις αρχές των GNN	27
1.4.3 Θεμελιώδεις αρχιτεκτονικές GNN	28
1.4.3.1 Graph Convolutional Network (GCN)	28
1.4.3.2 Graph Attention Network (GAT)	28
1.4.3.3 Graph Sample and Aggregate (GraphSAGE)	29
1.5 Σχετικές Εργασίες	29
1.5.1 Συστήματα τμηματοποίησης αλλοιώσεων σε MRI ασθενών με Πολλαπλό Μυέλωμα	29
1.5.2 GNN και Ιατρική Απεικόνιση	30
2 Μεθοδολογία	31
2.1 Διαμόρφωση γράφων	31
2.1.1 Κατακερματισμός (Supervoxel Segmentation)	31
2.1.2 Εξαγωγή Χαρακτηριστικών	31
2.1.3 Επιλογή Χαρακτηριστικών	32
2.1.4 Κατασκευή Γράφων	32
2.2 Αρχιτεκτονικές δικτύων	33

2.2.1 Model Overview	33
2.2.2 GCN	34
2.2.3 GAN	34
2.2.4 GraphSAGE.....	34
2.3 Στρατηγική Εκπαίδευσης	35
2.3.1 Αλγόριθμος Εκπαίδευσης.....	35
2.3.2 Συνάρτηση Απώλειας.....	35
3 Πειράματα.....	36
3.1 Διαχείριση dataset.....	36
3.1.1 Επισκόπηση του συνόλου δεδομένων	36
3.1.2 Δημιουργία των ground truth ετικετών	37
3.1.3 Προεπεξεργασία των εικόνων	37
3.2 Λεπτομέρειες υλοποίησης.....	38
3.3 Αξιολόγηση	39
3.3.1 Ποσοτική αξιολόγηση	39
3.3.2 Ποιοτική αξιολόγηση	42
3.4 Πειράματα	42
3.4.1 Επιλογή Πλήθους Χαρακτηριστικών	42
3.4.2 Αξιολόγηση Μοντέλων	43
3.5 Στατιστική ανάλυση.....	44
4 Αποτελέσματα.....	44
4.1 Επιλογή Πλήθους Χαρακτηριστικών.....	44
4.2 Αξιολόγηση Μοντέλων	45
5 Συζήτηση	55
5.1 Ανάλυση Πειραμάτων και Στατιστικών Αποτελεσμάτων.....	55
5.2 Σχέση με βιβλιογραφία	56
5.3 Ποιοτική Αξιολόγηση Μοντέλων.....	57
5.4 Περιορισμοί.....	60
6 Συμπεράσματα	60
Βιβλιογραφία	62

Ευρετήριο Εικόνων

ΕΙΚΟΝΑ 1: ΑΠΕΙΚΟΝΙΣΗ ΤΗΣ ΒΑΣΙΚΗΣ ΑΡΧΗΣ ΛΕΙΤΟΥΡΓΙΑΣ ΤΗΣ ΜΑΓΝΗΤΙΚΗΣ ΤΟΜΟΓΡΑΦΙΑΣ (MRI)	18
ΕΙΚΟΝΑ 2: ΑΠΕΙΚΟΝΙΣΗ ΤΗΣ ΑΠΟΚΑΤΑΣΤΑΣΗΣ T_1 ΚΑΙ ΤΗΣ ΑΠΟΣΥΝΘΕΣΗΣ T_2 ΤΟΥ ΜΑΓΝΗΤΙΣΜΟΥ	19

ΕΙΚΟΝΑ 3: ΕΣΤΙΑΚΗ ΑΛΛΟΙΩΣΗ (ΑΣΠΡΟ ΒΕΛΟΣ) ΠΑΡΑΤΗΡΟΥΜΕΝΗ ΑΠΟ ΤΗΝ STIR (Α) ΚΑΙ ΤΗΝ T1W-TSE (Β) ΑΚΟΛΟΥΘΙΑ	21
ΕΙΚΟΝΑ 4: ΑΛΛΟΙΩΣΗ ΠΑΡΑΤΗΡΟΥΜΕΝΗ ΑΠΟ ΤΗΝ T1W DIXON IP (ΑΣΠΡΟ ΒΕΛΟΣ) (Α) ΚΑΙ ΤΗΝ DWI, B=900 S/MM ² (ΜΑΥΡΟ ΒΕΛΟΣ) (Β) ΑΚΟΛΟΥΘΙΑ.....	23
ΕΙΚΟΝΑ 5: ΜΙΑ ΕΦΑΡΜΟΓΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ SLIC. ΠΑΡΑΤΗΡΟΥΜΕ ΟΤΙ ΠΕΡΙΟΧΕΣ ΕΝΔΙΑΦΕΡΟΝΤΟΣ ΕΧΟΥΝ ΑΠΟΜΟΝΩΘΕΙ ΚΑΤΑΛΛΗΛΑ	24
ΕΙΚΟΝΑ 6: ΣΥΝΟΠΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΤΗΣ ΡΟΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΤΟΥ ΠΑΚΕΤΟΥ PYRADIOMICS ΤΗΣ ΡΥΘΜΟΝ.....	25
ΕΙΚΟΝΑ 7: ΑΠΕΙΚΟΝΙΣΗ ΤΗΣ ΒΑΣΙΚΗΣ ΑΡΧΗΣ ΤΩΝ GRAPH NEURAL NETWORKS, ΜΕ ΚΑΘΕ ΚΟΜΒΟ ΝΑ ΕΙΝΑΙ ΕΝΑΣ ΣΥΝΔΥΑΣΜΟΣ ΤΩΝ ΓΕΙΤΟΝΩΝ ΤΟΥ	28
ΕΙΚΟΝΑ 8: ΤΟ ΑΡΧΙΚΟ PIPELINE ΤΗΣ ΕΡΓΑΣΙΑΣ, ΤΟ ΟΠΟΙΟ ΞΕΚΙΝΑ ΜΕ ΤΗΝ ΕΙΣΑΓΩΓΗ ΤΩΝ ΕΙΚΟΝΩΝ, ΑΚΟΛΟΥΘΕΙ ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ, ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ, ΕΞΑΓΩΓΗ ΚΑΙ ΕΠΙΛΟΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΚΑΙ ΟΛΟΚΛΗΡΩΝΕΤΑΙ ΜΕ ΤΗΝ ΠΑΡΑΓΩΓΗ ΤΩΝ 5 ΔΙΑΦΟΡΕΤΙΚΩΝ ΤΥΠΩΝ ΓΡΑΦΩΝ	33
ΕΙΚΟΝΑ 9: Ο ΣΚΕΛΕΤΟΣ ΤΗΣ ΑΡΧΙΤΕΚΤΟΝΙΚΗΣ ΤΩΝ ΜΟΝΤΕΛΩΝ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ, ΟΠΟΥ Ν ΕΙΝΑΙ Ο ΑΡΙΘΜΟΣ ΤΩΝ ΚΟΜΒΩΝ ΤΟΥ ΓΡΑΦΟΥ, IN Ο ΑΡΙΘΜΟΣ ΤΩΝ ΔΙΑΣΤΑΣΕΩΝ ΚΑΘΕ ΚΟΜΒΟΥ, Η ΤΑ HIDDEN FEATURES ΚΑΙ L Ο ΑΡΙΘΜΟΣ ΤΩΝ LAYERS.....	35
ΕΙΚΟΝΑ 10: ΑΠΕΙΚΟΝΙΣΗ ΤΩΝ ΣΥΝΤΕΛΕΣΤΩΝ ΤΟΥ ΚΑΘΕ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΥ (ΤΑΞΙΝΟΜΗΜΕΝΑ ΚΑΤΑ ΑΠΟΛΥΤΗ ΤΙΜΗ) ΜΕΤΑ ΤΗΝ ΕΞΑΓΩΓΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΑΠΟ ΤΟ LASSO.....	44
ΕΙΚΟΝΑ 11: ΑΣΘΕΝΗΣ 5. ΟΙ ΕΣΤΙΑΚΕΣ ΑΛΛΟΙΩΣΕΙΣ ΕΙΝΑΙ ΣΗΜΕΙΩΜΕΝΕΣ ΜΕ ΚΟΚΚΙΝΟ ΠΕΡΙΘΩΡΙΟ ΚΑΙ ΤΟ GROUND TRUTH ΜΕ ΠΡΑΣΙΝΟ. ΑΡΙΣΤΕΡΑ ΦΑΙΝΕΤΑΙ Η T1 ΑΚΟΛΟΥΘΙΑ, ΕΝΩ ΔΕΞΙΑ ΦΑΙΝΕΤΑΙ Η T2.....	48
ΕΙΚΟΝΑ 12: ΑΣΘΕΝΗΣ 8. ΟΙ ΕΣΤΙΑΚΕΣ ΑΛΛΟΙΩΣΕΙΣ ΕΙΝΑΙ ΣΗΜΕΙΩΜΕΝΕΣ ΜΕ ΚΟΚΚΙΝΟ ΠΕΡΙΘΩΡΙΟ ΚΑΙ ΤΟ GROUND TRUTH ΜΕ ΠΡΑΣΙΝΟ. ΑΡΙΣΤΕΡΑ ΦΑΙΝΕΤΑΙ Η T1 ΑΚΟΛΟΥΘΙΑ, ΕΝΩ ΔΕΞΙΑ ΦΑΙΝΕΤΑΙ Η T2.....	49
ΕΙΚΟΝΑ 13: ΑΣΘΕΝΗΣ 11. ΟΙ ΕΣΤΙΑΚΕΣ ΑΛΛΟΙΩΣΕΙΣ ΕΙΝΑΙ ΣΗΜΕΙΩΜΕΝΕΣ ΜΕ ΚΟΚΚΙΝΟ ΠΕΡΙΘΩΡΙΟ ΚΑΙ ΤΟ GROUND TRUTH ΜΕ ΠΡΑΣΙΝΟ. ΑΡΙΣΤΕΡΑ ΦΑΙΝΕΤΑΙ Η T1 ΑΚΟΛΟΥΘΙΑ, ΕΝΩ ΔΕΞΙΑ ΦΑΙΝΕΤΑΙ Η T2.....	50
ΕΙΚΟΝΑ 14: ΑΣΘΕΝΗΣ 20. ΟΙ ΕΣΤΙΑΚΕΣ ΑΛΛΟΙΩΣΕΙΣ ΕΙΝΑΙ ΣΗΜΕΙΩΜΕΝΕΣ ΜΕ ΚΟΚΚΙΝΟ ΠΕΡΙΘΩΡΙΟ ΚΑΙ ΤΟ GROUND TRUTH ΜΕ ΠΡΑΣΙΝΟ. ΑΡΙΣΤΕΡΑ ΦΑΙΝΕΤΑΙ Η T1 ΑΚΟΛΟΥΘΙΑ, ΕΝΩ ΔΕΞΙΑ ΦΑΙΝΕΤΑΙ Η T2.....	51
ΕΙΚΟΝΑ 15: ΑΣΘΕΝΗΣ 22. ΟΙ ΕΣΤΙΑΚΕΣ ΑΛΛΟΙΩΣΕΙΣ ΕΙΝΑΙ ΣΗΜΕΙΩΜΕΝΕΣ ΜΕ ΚΟΚΚΙΝΟ ΠΕΡΙΘΩΡΙΟ ΚΑΙ ΤΟ GROUND TRUTH ΜΕ ΠΡΑΣΙΝΟ. ΑΡΙΣΤΕΡΑ ΦΑΙΝΕΤΑΙ Η T1 ΑΚΟΛΟΥΘΙΑ, ΕΝΩ ΔΕΞΙΑ ΦΑΙΝΕΤΑΙ Η T2.....	52
ΕΙΚΟΝΑ 16: ΑΣΘΕΝΗΣ 27. ΟΙ ΕΣΤΙΑΚΕΣ ΑΛΛΟΙΩΣΕΙΣ ΕΙΝΑΙ ΣΗΜΕΙΩΜΕΝΕΣ ΜΕ ΚΟΚΚΙΝΟ ΠΕΡΙΘΩΡΙΟ ΚΑΙ ΤΟ GROUND TRUTH ΜΕ ΠΡΑΣΙΝΟ. ΑΡΙΣΤΕΡΑ ΦΑΙΝΕΤΑΙ Η T1 ΑΚΟΛΟΥΘΙΑ, ΕΝΩ ΔΕΞΙΑ ΦΑΙΝΕΤΑΙ Η T2.....	53
ΕΙΚΟΝΑ 17: ΑΣΘΕΝΗΣ 28. ΟΙ ΕΣΤΙΑΚΕΣ ΑΛΛΟΙΩΣΕΙΣ ΕΙΝΑΙ ΣΗΜΕΙΩΜΕΝΕΣ ΜΕ ΚΟΚΚΙΝΟ ΠΕΡΙΘΩΡΙΟ ΚΑΙ ΤΟ GROUND TRUTH ΜΕ ΠΡΑΣΙΝΟ. ΑΡΙΣΤΕΡΑ ΦΑΙΝΕΤΑΙ Η T1 ΑΚΟΛΟΥΘΙΑ, ΕΝΩ ΔΕΞΙΑ ΦΑΙΝΕΤΑΙ Η T2.....	54
ΕΙΚΟΝΑ 18: ΣΥΓΚΡΙΣΗ ΑΣΘΕΝΗ ΜΕ ΔΥΟ ΔΙΑΦΟΡΕΤΙΚΑ ΜΟΝΤΕΛΑ. ΤΟ ΚΟΚΚΙΝΟ ΕΙΝΑΙ ΤΟ ΜΟΝΤΕΛΟ ΜΕ ΤΟ ΜΕΓΑΛΥΤΕΡΟ F1-SCORE (GCN – 64 x 5) ΚΑΙ ΤΟ ΠΡΑΣΙΝΟ ΕΙΝΑΙ ΤΟ ΜΟΝΤΕΛΟ ΜΕ ΤΟ ΜΕΓΑΛΥΤΕΡΟ LESION F1-SCORE (SAGE – 128 x 4). ΜΕ ΜΠΛΕ ΣΧΕΔΙΑΖΕΤΑΙ ΤΟ GROUND TRUTH.	56
ΕΙΚΟΝΑ 19: Η ΠΕΡΙΟΧΗ ΤΟΥ ΙΕΡΟΥ ΤΟΥ ΑΣΘΕΝΗ 20. ΑΡΙΣΤΕΡΑ ΦΑΙΝΕΤΑΙ Η T1 ΑΚΟΛΟΥΘΙΑ, ΕΝΩ ΔΕΞΙΑ ΦΑΙΝΕΤΑΙ Η T2.....	58
ΕΙΚΟΝΑ 20: ΨΕΥΔΩΣ ΘΕΤΙΚΗ ΠΕΡΙΟΧΗ ΣΤΟΝ ΑΣΘΕΝΗ 22. ΑΡΙΣΤΕΡΑ ΦΑΙΝΕΤΑΙ Η T1 ΑΚΟΛΟΥΘΙΑ, ΕΝΩ ΔΕΞΙΑ ΦΑΙΝΕΤΑΙ Η T2.....	58
ΕΙΚΟΝΑ 21: ΨΕΥΔΩΣ ΑΡΝΗΤΙΚΗ ΠΕΡΙΟΧΗ ΣΤΟΝ ΑΣΘΕΝΗ 22. ΑΡΙΣΤΕΡΑ ΦΑΙΝΕΤΑΙ Η T1 ΑΚΟΛΟΥΘΙΑ, ΕΝΩ ΔΕΞΙΑ ΦΑΙΝΕΤΑΙ Η T2.....	59

Ευρετήριο Πινάκων

ΠΙΝΑΚΑΣ 1. ΑΠΕΙΚΟΝΙΣΤΙΚΕΣ ΠΑΡΑΜΕΤΡΟΙ ΓΙΑ ΤΙΣ ΑΚΟΛΟΥΘΙΕΣ T1W-TSE ΚΑΙ STIR	36
ΠΙΝΑΚΑΣ 2. ΑΠΟΤΕΛΕΣΜΑΤΑ (MEAN ± SD) ΤΗΣ ΑΠΟΔΟΣΗΣ ΤΩΝ ΜΟΝΤΕΛΩΝ ΣΕ ΣΧΕΣΗ ΜΕ ΤΗΝ ΤΙΜΗ TOP K.....	45
ΠΙΝΑΚΑΣ 3. ΜΕΓΙΣΤΕΣ ΑΠΟΔΟΣΕΙΣ ΜΟΝΤΕΛΩΝ ΑΝΑ TOP K.....	45
ΠΙΝΑΚΑΣ 4. ΑΠΟΤΕΛΕΣΜΑΤΑ (MEAN ± SD) ΤΗΣ ΑΠΟΔΟΣΗΣ ΤΩΝ ΜΟΝΤΕΛΩΝ ΣΕ ΣΧΕΣΗ ΜΕ ΤΟΝ ΤΥΠΟ ΤΟΥ ΓΡΑΦΟΥ	46
ΠΙΝΑΚΑΣ 5. ΜΕΓΙΣΤΕΣ ΑΠΟΔΟΣΕΙΣ ΜΟΝΤΕΛΩΝ ΑΝΑ ΚΑΤΗΓΟΡΙΑ ΑΡΧΙΤΕΚΤΟΝΙΚΗΣ.....	47
ΠΙΝΑΚΑΣ 6. ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΒΕΛΤΙΣΤΗΣ ΑΡΧΙΤΕΚΤΟΝΙΚΗΣ ΜΕ ΔΙΑΦΟΡΕΤΙΚΑ ΥΠΟΣΥΝΟΛΑ ΑΣΘΕΝΩΝ	60

Εισαγωγή

Η Whole-Body μαγνητική τομογραφία (WB-MRI) αποτελεί μια ταχέως εξελισσόμενη, μη επεμβατική απεικονιστική τεχνική, η οποία επιτρέπει την ολική εκτίμηση του σώματος με υψηλή ανατομική ακρίβεια και εξαιρετική ανάλυση αντίθεσης, καθιστώντας την ιδανική για την παρακολούθηση νοσημάτων πολλαπλών συστημάτων χωρίς την ανάγκη ιονίζουσας ακτινοβολίας [1]. Η εισαγωγή ταχύτερων πρωτοκόλλων λήψης και η βελτίωση της ποιότητας εικόνας οδήγησαν στην ευρεία υιοθέτηση της WB-MRI στην ογκολογική απεικόνιση, με εφαρμογές τόσο σε διαγνωστικό όσο και σε προγνωστικό επίπεδο.

Στο πλαίσιο αιματολογικών νοσημάτων, η WB-MRI έχει καθιερωθεί ως βασικό εργαλείο απεικόνισης στο Πολλαπλό Μυέλωμα (ΠΜ), παρέχοντας σημαντικές πληροφορίες για τον εντοπισμό, τον χαρακτήρα και την έκταση των οστικών βλαβών, καθώς και για την παρακολούθηση της ανταπόκρισης στη θεραπεία. Ωστόσο, η ερμηνεία των δεδομένων WB-MRI παραμένει ιδιαίτερα απαιτητική λόγω του ετερογενούς προτύπου κατανομής των βλαβών και του μεγάλου όγκου πληροφορίας, που συχνά απαιτεί λεπτομερή καταγραφή και αξιολόγηση παραμέτρων όπως το μέγεθος, η ένταση σήματος, η υφή και η θέση κάθε αλλοίωσης [2].

Η ανάγκη για αντικειμενικότερη και ταχύτερη ανάλυση δεδομένων έχει οδηγήσει στην ένταξη τεχνικών τεχνητής νοημοσύνης (AI) και μηχανικής μάθησης (ML) σε πολλές απεικονιστικές εφαρμογές, συμπεριλαμβανομένης της τμηματοποίησης βλαβών. Τα τελευταία χρόνια, τα Νευρωνικά Δίκτυα Γράφων (Graph Neural Networks – GNNs) έχουν αναδειχθεί ως ισχυρό εργαλείο για τη μοντελοποίηση της χωρικής και δομικής συσχέτισης περιοχών ενδιαφέροντος στην ιατρική εικόνα. Παρότι οι GNNs έχουν εφαρμοστεί σε πλήθος δομών και οργάνων, η αξιοποίησή τους στο πολλαπλούν μυέλωμα παραμένει αχαρτογράφητη ερευνητική περιοχή [3].

Η παρούσα μελέτη προτείνει μια μέθοδο αυτόματης ανάλυσης της ολόσωμης μαγνητικής τομογραφίας σπονδυλικής στήλης σε ασθενείς με Πολλαπλό Μυέλωμα, με σκοπό την αυτόματη τμηματοποίηση και ταξινόμηση των εστιακών αλλοιώσεων στον οστικό μυελό με τη χρήση GNNs [4]. Η μεθοδολογία ενσωματώνει σύγχρονες τεχνικές επεξεργασίας ιατρικών εικόνων, εξαγωγής radiomic χαρακτηριστικών και νευρωνικών δικτύων γράφων, στοχεύοντας στην αποδοτική ανίχνευση και ταξινόμηση σπονδυλικών αλλοιώσεων.

Η συμβολή της παρούσας εργασίας έγκειται στην πρωτοτυπία της εφαρμογής των GNNs στη συγκεκριμένη παθολογία, καθώς μέχρι σήμερα δεν έχουν αναφερθεί μελέτες που να εφαρμόζουν γραφική μοντελοποίηση στη WB-MRI για τον εντοπισμό αλλοιώσεων σε ασθενείς με πολλαπλό μυέλωμα. Η μελέτη φιλοδοξεί να συμβάλει τόσο στη βελτίωση της ακρίβειας των μεθόδων ανάλυσης όσο και στη δημιουργία θεμελίων για κλινικά εργαλεία βασισμένα σε τεχνητή νοημοσύνη, τα οποία θα υποστηρίζουν την έγκαιρη και εξατομικευμένη διάγνωση [7], [8], [9]. Ο κώδικας για την παρούσα εργασία είναι διαθέσιμος εδώ: <https://github.com/stelioskt/MultipleMyelomaDetector>

1 Θεωρητικό Υπόβαθρο

1.1 Το Πολλαπλό Μυέλωμα

1.1.1 Γενικά στοιχεία

Το Πολλαπλό Μυέλωμα (ΠΜ) χαρακτηρίζεται από την νεοπλασματική υπερπλασία μονής κυτταρικής σειράς πλασματοκυτάρων που παράγουν συγκεκριμένη μονοκλωνική ανοσοσφαιρίνη οδηγώντας σε τεράστιο αριθμό αντιγράφων της ανοσοσφαιρίνης αυτής. Συνήθως η ανοσοσφαιρίνη είναι τύπου IgG ή IgA. Η μέση ηλικία διάγνωσης κυμαίνεται στο εύρος των 65-75 ετών και η νόσος είναι συχνότερη στους άρρενες. Ο επιπολασμός του ΠΜ είναι μεγαλύτερος στην μαύρη φυλή. Παρόλο που η αιτιολογία του ΠΜ δεν είναι σαφώς περιγεγραμμένη μέχρι σήμερα, εν τούτοις παράγοντες κινδύνου όπως ο αυξημένος δείκτης μάζας σώματος και η έκθεση σε συγκεκριμένα χημικά φαίνεται πως συσχετίζονται με την εκδήλωσή του. Στην πορεία της νόσου, τα στοιχεία του μυελού αντικαθίστανται από κακοήγη πλασματοκύτταρα με αποτέλεσμα την εμφάνιση αναιμίας, λευκοπενίας και θρομβοκυτταροπενίας [10]. Τα υπόλοιπα σημεία της νόσου συνίστανται στην υπερασβαστιαμία, την νεφρική ανεπάρκεια και τις οστεολυτικές αλλοιώσεις.

Κλινικά χαρακτηριστικά

Κάθε σημείο της νόσου σχετίζεται με το εκάστοτε κλινικό χαρακτηριστικό. Οι οστεολυτικές αλλοιώσεις οδηγούν σε οξύ οστικό πόνο ιδιαίτερα στην οσφυϊκή μοίρα της σπονδυλικής στήλης, το στήθος και την κάτω γνάθο, σε παθολογικά κατάγματα, και την καθ' ύψος μείωση των ασθενών λόγω της κατάρρευσης των σπονδύλων [11]. Η διήθηση του οστικού μυελού από τα πλασματοκύτταρα οδηγεί σε αναιμία. Η κυκλοφορία των μονοκλωνικών ανοσοσφαιρινών στο αίμα οδηγεί στην εμφάνιση του “μυελωματικού” νεφρού η οποία επιτείνει περεταίρω την αναιμία [12]. Παράλληλα, η λευκοπενία οδηγεί σε επαναλαμβανόμενες λοιμώξεις, ιδιαιτέρως του αναπνευστικού και ουροποιητικού συστήματος λόγω της έλλειψης φυσιολογικών ανοσοσφαιρινών. Η εμφάνιση πλασματοκυτωμάτων ή οστικών θραυσμάτων μπορεί να οδηγήσει σε συμπίεση του νωτιαίου μυελού και την εμφάνιση των αντίστοιχων κινητικών/αισθητικών επιπλοκών.

Διάγνωση

Συνολικά, η διάγνωση τίθεται τόσο από βιοχημικές όσο και από απεικονιστικές παραμέτρους:

- Ηλεκτροφόρηση πρωτεϊνών ορού και ούρων ± ανοσοηλεκτροφόρηση (SPEP/UPEP + IFE): αποκαλύπτουν την παρουσία μονοκλωνικής πρωτεΐνης (παραπρωτεΐνης).
- Δοκιμασία ελεύθερων ελαφρών αλυσίδων (sFLC): μετράει κ και λ ελεύθερες αλυσίδες· η αναλογία κ/λ είναι παθολογικά ασύμμετρη σε μονοκλωνική νόσο.
- Απεικονίσεις (LDCT, WB-MRI):
 1. LDCT: ανάδειξη μιας ή περισσότερων οστεολυτικών αλλοιώσεων
 2. WB-MRI: ανάδειξη εστιακών αλλοιώσεων διαμέτρου μεγαλύτερης των 5 χιλ.
- Βιοψία μυελού των οστών: >10% κλωνικών πλασματοκυτάρων υποστηρίζει τη διάγνωση πολλαπλού μυελώματος.
- Υπερασβαστιαμία: ανερχόμενα επίπεδα Ca^{2+} λόγω οστικής καταστροφής.
- Αυξημένη ολική πρωτεΐνη ορού και αυξημένο κλάσμα γ-σφαιρινών (υπεργαμμασφαιριναιμία) από την παρουσία παραπρωτεΐνης.
- Επιβάρυνση νεφρικής λειτουργίας (αυξημένη κρεατινίνη): συχνά λόγω απόθεσης/τοξικότητας των ελεύθερων αλυσίδων (cast nephropathy).

- Κυτταροπενίες: αναιμία, λευκοπενία και θρομβοπενία (συχνά από διήθηση μυελού ή καταστολή της παραγωγής).
- Επίχρισμα περιφερικού αίματος: νορμοκυτταρική αναιμία με σχηματισμό «rouleaux» (στοίχιση ερυθρών).
- Μικροσκόπηση ούρων: παρουσία μυελωματικών καστών (large/waxy casts).

Η παρουσία τουλάχιστον 10% πλασματοκυττάρων στην βιοψία σε συνδυασμό με την ύπαρξη μονοκλωνικής πρωτεΐνης στον ορό ή τα ούρα ή την παρουσία οστεολύσεων καθορίζει την διάγνωση.

Θεραπεία

Οι ασθενείς με ΠΜ που παρουσιάζουν βλάβες σε όργανα-στόχους (end-organ damage) χρήζουν θεραπειάς. Το προτιμώμενο θεραπευτικό σχήμα περιλαμβάνει υψηλής δόσης χημειοθεραπεία ως επαγωγική αγωγή, ακολουθούμενη από αυτόλογη μεταμόσχευση αιμοποιητικών κυττάρων (auto-HCT) [13]. Σε ασθενείς με μία ή περισσότερες οστικές αλλοιώσεις συνιστάται η χορήγηση διφωσφονικών για τη μείωση των οστικών συμβαμάτων.

Πρόγνωση

Το ΠΜ έχει μεταβλητή πρόγνωση, η οποία εξαρτάται από πολλούς παράγοντες της νόσου και τις νοσηρότητες του ασθενούς (κλινικά, εργαστηριακά και μοριακά χαρακτηριστικά). Η πενταετής επιβίωση υπολογίζεται περίπου στο 50%, με σημαντική διακύμανση ανάλογα με το προφίλ κάθε ασθενούς.

1.1.2 Ο ρόλος της απεικόνισης στην διάγνωση και την διαχείριση του ΠΜ

Οι απεικονιστικές τεχνικές στο ΠΜ κατέχουν σημαντικό ρόλο στην διαχείριση της νόσου. Παρότι δεν αντικαθιστούν τις βιοχημικές εξετάσεις, εν τούτοις λειτουργούν συνεργικά. Ορισμένες από τις εφαρμογές τους είναι οι εξής:

- Ανίχνευση αλλοιώσεων σκελετικού συστήματος και οστικού μυελού για την διάγνωση και την σταδιοποίηση.
- Εύρεση επιπλοκών που χρήζουν άμεσης δράσης (παθολογικά κατάγματα, συμπίεση νωτιαίου μυελού).
- Καθοδήγηση της βιοψίας και της τοπικής θεραπείας (ακτινοθεραπεία/χειρουργικές επεμβάσεις).
- Χαρακτηρισμό της απόκρισης στην θεραπεία και υποτροπών.

1.1.3 Οι απεικονιστικές μέθοδοι και οι ενδείξεις τους

- Ολόσωμη μαγνητική τομογραφία (WB-MRI):
 1. Η εξέταση εκλογής για τον έλεγχο του οστικού μυελού και τον διαχωρισμό μεταξύ των προτύπων εστιακού (focal), διάχυτου (diffuse) και ποικίλου προτύπου (variegated).
 2. Χρήση τόσο στον αρχικό έλεγχο του μυελώματος, όσο και στην απόκριση στην θεραπεία αλλά και για την εκτίμηση της ελάχιστης υπολειπόμενης νόσου [14].
- Ολόσωμη αξονική τομογραφία χαμηλής δόσης (LDCT):
Ταχεία εξέταση, χρήσιμη για την ανίχνευση οστελυτικών αλλοιώσεων και για την ανίχνευση του ρίσκου κατάγματος.
- FDG PET/CT:
Αναδεικνύει την μεταβολική δραστηριότητα. Χρήσιμη για την ανίχνευση παραμυελικών αλλοιώσεων, για την επιλογή στόχων προς βιοψία και την εκτίμηση της απόκρισης στην θεραπεία.

1.1.4 Πρωτόκολλα WB-MRI στην διαχείριση του ΠΜ

Το πρωτόκολλο MY-RADS (Myeloma Response Assessment and Diagnosis System) παρέχει ένα κοινώς αποδεκτό πλαίσιο για την τυποποίηση των ακολουθιών και των παραμέτρων απεικόνισης στη διάγνωση και παρακολούθηση του πολλαπλού μυελώματος, καθώς και για τη συστηματική κατάταξη της απεικονιστικής απόκρισης στη θεραπεία.

Το πρωτόκολλο συνίσταται από τις κάτωθι ακολουθίες:

1. T1 weighted Turbo or Fast Spin Echo (T1W-TSE)
2. T2 weighted Short Tau Inverse Recovery (STIR)
3. Diffusion-Weighted Imaging (DWI) με χάρτες δείκτη φαινομενικής διάχυσης (ADC)
4. T1 weighted Dixon
5. T2 weighted Dixon

Οι ακολουθίες T1W-TSE και STIR χαρακτηρίζονται ως Μορφολογικές Ακολουθίες και παρουσιάζουν την μέγιστη χωρική διακριτική ικανότητα σε σχέση με τις υπόλοιπες [15], [16]. Οι χάρτες ADC σε συνδυασμό με τις T1/T2 weighted Dixon ονομάζονται Ποσοτικές Ακολουθίες και παρέχουν ποσοτικές πληροφορίες όπως τον δείκτη φαινομενικής διάχυσης και τον δείκτη κλάσματος λίπους αντίστοιχα αλλά η χωρική τους διακριτική ικανότητα, υπό τις παραμέτρους που χρησιμοποιούνται στο εν λόγω πρωτόκολλο είναι σημαντικά μικρότερη συγκριτικά με τις T1W-TSE και STIR. Σε επόμενη ενότητα. Θα αναλυθεί κάθε ακολουθία ξεχωριστά, προκειμένου να γίνουν αντιληπτές οι διαφορές τους και η χρησιμότητά της κάθε μίας για τον εντοπισμό περιοχών που σχετίζονται με το Πολλαπλό Μυέλωμα.

1.2 Η χρήση MRI στη διάγνωση

Η Μαγνητική Τομογραφία (Magnetic Resonance Imaging – MRI) αποτελεί μία από τις σημαντικότερες μεθόδους απεικόνισης στη σύγχρονη ιατρική, με ιδιαίτερη εφαρμογή στην απεικόνιση των μαλακών ιστών και του μυελού των οστών. Σε αντίθεση με άλλες τεχνικές όπως η Αξονική Τομογραφία (CT) και η Ακτινογραφία, η MRI δεν χρησιμοποιεί ιονίζουσα ακτινοβολία, γεγονός που την καθιστά ασφαλέστερη για επαναλαμβανόμενες εξετάσεις. Η συμβολή της είναι καθοριστική στη διάγνωση, παρακολούθηση και αντιμετώπιση παθολογιών, μεταξύ των οποίων και το Πολλαπλό Μυέλωμα [18].

Η Μαγνητική Τομογραφία αποτελεί ίσως την πιο περίπλοκη ιατρική μέθοδο απεικόνισης και η πολυπλοκότητά της έγκειται στη φυσική αρχή λειτουργίας της, στις παραμέτρους σήματος που καθορίζουν τον τύπο της αντίθεσης, καθώς και στις εξειδικευμένες ακολουθίες που χρησιμοποιούνται ανάλογα με την κλινική υποψία. Στο πλαίσιο της παρούσας εργασίας, η MRI δεν αντιμετωπίζεται μόνο ως εργαλείο απεικόνισης, αλλά και ως πηγή πολυδιάστατης πληροφορίας για την εξαγωγή ραδιομικών (radiomic) χαρακτηριστικών, την κατασκευή γραφημάτων και την ανάπτυξη συστημάτων μηχανικής μάθησης [19].

1.2.1 Βασικές Αρχές MRI

Η MRI βασίζεται στο φαινόμενο του πυρηνικού μαγνητικού συντονισμού (NMR), και συγκεκριμένα στην αλληλεπίδραση των πυρήνων υδρογόνου με ένα εξωτερικό μαγνητικό πεδίο και ένα παλμό ραδιοσυχνότητας.

1.2.1.1 Πυρηνικός Μαγνητικός Συντονισμός

Οι πυρήνες υδρογόνου (πρωτόνια), που είναι άφθονοι στο ανθρώπινο σώμα λόγω της παρουσίας νερού και λίπους, έχουν μαγνητική ροπή [20]. Όταν τοποθετούνται σε ένα ισχυρό στατικό μαγνητικό πεδίο B_0 , οι ροπές αυτές τείνουν να ευθυγραμμιστούν με το πεδίο, δημιουργώντας μια συνολική μαγνήτιση $\vec{M}$.

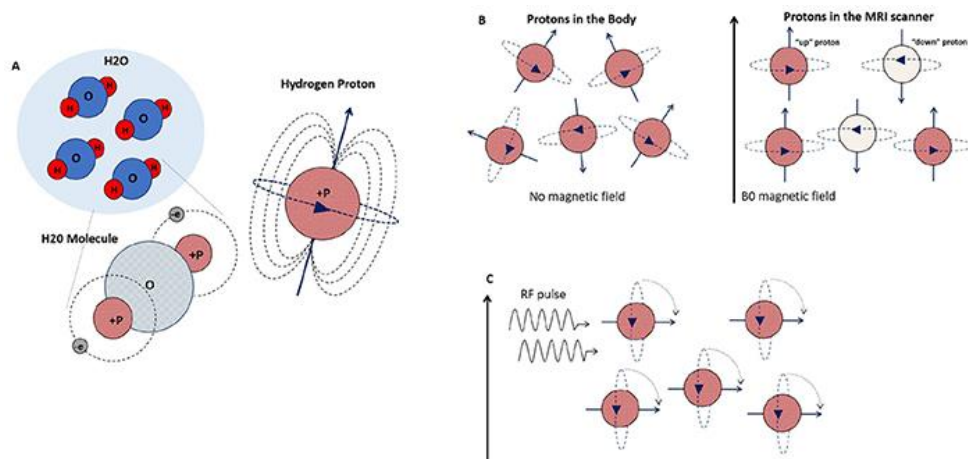
Με την εφαρμογή ενός παλμού ραδιοσυχνότητας (RF pulse) στη συχνότητα Larmor:

$$\omega_L = \gamma B_0 \quad (\text{Eq. 1})$$

όπου:

- ω_L είναι η συχνότητα Larmor
- γ είναι η γυρομαγνητική αναλογία
- B_0 είναι το στατικό μαγνητικό πεδίο

το διάνυσμα μαγνήτισης μετατοπίζεται από τον άξονα z στο εγκάρσιο επίπεδο.



Εικόνα 1: Απεικόνιση της βασικής αρχής λειτουργίας της μαγνητικής τομογραφίας (MRI)

1.2.1.2 Χρόνος Χαλάρωσης T1 (Spin-Lattice Relaxation)

Ο T1 αναφέρεται στον χρόνο που απαιτείται για να επανέρθει η διαμήκης μαγνήτιση M_z κατά μήκος του άξονα του B_0 στο 63% της αρχικής τιμής της. Εξαρτάται από τη μεταφορά ενέργειας των διεγερμένων πρωτονίων στο περιβάλλον τους (lattice). Μαθηματικά περιγράφεται ως:

$$M_z(t) = M_0(1 - e^{-t/T_1}) \quad (\text{Eq. 2})$$

Οι ιστοί με υψηλή περιεκτικότητα σε λίπος έχουν μικρό T1, άρα ανακτούν γρήγορα τη μαγνήτιση και εμφανίζονται υπερέντονοι στις T1-weighted εικόνες. Αντίθετα το νερό έχει μεγάλο T1, άρα εμφανίζεται υποέντονο σε T1W εικόνες.

Όσον αφορά τα πλεονεκτήματα της συγκεκριμένης ακολουθίας, χαρακτηρίζεται από υψηλή χωρική ανάλυση και οι ανατομικές περιοχές διαχωρίζονται ικανοποιητικά, δηλαδή είναι η καλύτερη εικόνα για μια αρχική εκτίμηση. Από την άλλη, η ακολουθία αυτή δεν είναι ευαίσθητη σε παθολογίες με υψηλή σε νερό περιεκτικότητα.

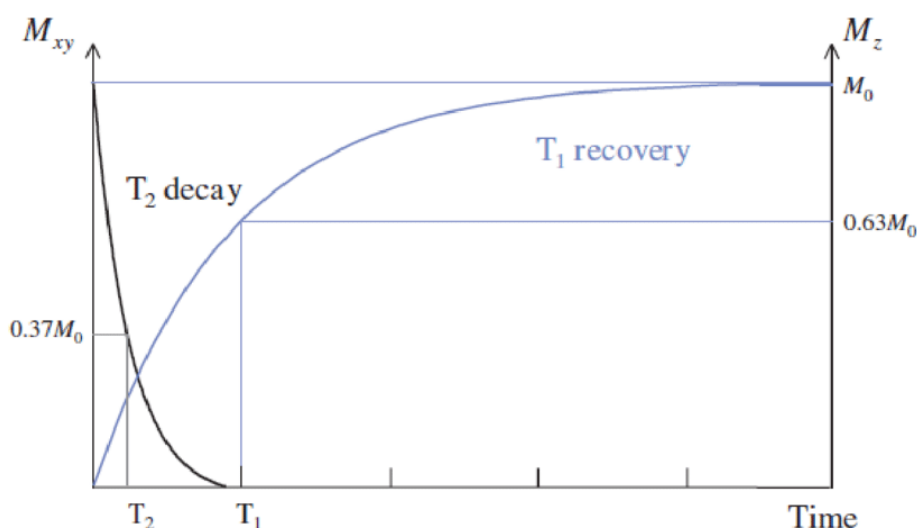
1.2.1.3 Χρόνος Χαλάρωσης T2 (Spin-Spin Relaxation)

Ο T2 αντιστοιχεί στον χρόνο που απαιτείται ώστε το μέτρο της εγκάρσιας μαγνήτισης M_{xy} να μειωθεί στο 37% της αρχικής του τιμής λόγω αποσυντονισμού μεταξύ γειτονικών πυρήνων που αλληλοεπιδρούν μεταξύ τους. Περιγράφεται από:

$$M_{xy}(t) = M_{xy}(0) \cdot e^{-t/T_2} \quad (\text{Eq. 3})$$

Οι ιστοί με υψηλή περιεκτικότητα σε νερό (οίδημα, φλεγμονή, κακοήθειες) έχουν μεγάλο T2, άρα παραμένουν φωτεινοί για μεγαλύτερο διάστημα και εμφανίζονται υπερέντοντοι στις T2-weighted εικόνες [21]. Αντίθετα, οι ιστοί με λίπος ή κολλαγόνο έχουν μικρό T2, το οποίο τους οδηγεί να εμφανίζονται υποέντοντοι στις T2W.

Σε αντίθεση με την προηγούμενη ακολουθία, οι T2W εικόνες είναι εξαιρετικά ευαίσθητες σε παθολογίες και χρησιμοποιούνται για ανάλογες διαγνώσεις. Παρ' όλα αυτά, η μειωμένη χωρική ανάλυση και ο θόρυβος που τις χαρακτηρίζει δυσχεραίνουν την ακριβή διάγνωση.



Εικόνα 2: Απεικόνιση της αποκατάστασης T_1 και της αποσύνθεσης T_2 του μαγνητισμού

1.2.2 Τύποι Ακολουθιών

Πέρα από τις βασικές ακολουθίες T1 και T2, η ανάγκη για πιο εξειδικευμένες διαγνώσεις απαιτεί πιο εξειδικευμένες τεχνικές, όπως αυτές που αναφέρονται στην Ενότητα 1.1.4. Οι ακολουθίες αυτές χωρίζονται στις **Μορφολογικές Ακολουθίες** και τις **Ποσοτικές Ακολουθίες** και στοχεύουν στη βελτιστοποίηση της αντίθεσης, στην καταστολή ανεπιθύμητων σημάτων και στη μείωση του χρόνου σάρωσης, προσφέροντας έτσι σημαντική διαγνωστική αξία σε παθολογίες όπως το Πολλαπλό Μυέλωμα.

1.2.2.1 Μορφολογικές Ακολουθίες

- T1W-Turbo Spin Echo (TSE)

Η TSE είναι μία τεχνική επιτάχυνσης της παραδοσιακής spin-echo ακολουθίας, επιτρέποντας την απόκτηση πολλαπλών echoes μετά από κάθε παλμό διέγερσης. Αυτό μειώνει σημαντικά τον συνολικό

χρόνο σάρωσης, καθιστώντας την ιδανική για εφαρμογές όπου απαιτείται υψηλή ποιότητα εικόνας, καθώς και για την αρχική αξιολόγηση του οστικού μυελού [22], [23].

Σε κάθε κύκλο, εφαρμόζεται ένας παλμός 90° και ακολουθείται από μια σειρά από παλμούς 180° που δημιουργούν echoes σε διαφορετικά χρονικά σημεία (echo train). Ο λιπώδης μυελός εμφανίζει υψηλό σήμα σε αυτήν την ακολουθία, ενώ ο ερυθρός μυελός, ο οποίος έχει μικρότερο ποσοστό λίπους και μεγαλύτερη κυτταροβρίθεια, εμφανίζει χαμηλότερο σήμα. Η ένταση του σήματος που παρατηρείται εξαρτάται άμεσα από τη σύσταση σε λιπώδη και ερυθρό μυελό [17]. Ο παθολογικός οστικός μυελός τυπικά εμφανίζει ακόμη χαμηλότερο σήμα σε σχέση με τον λιπώδη, λόγω της αυξημένης κυτταροβρίθειας, της μεγαλύτερης περιεκτικότητας σε νερό και της μειωμένης ποσότητας λίπους κατά όγκο [17].

Στην T1W-TSE, οι εστιακές αλλοιώσεις, συνήθως, παρουσιάζονται ως καλώς περιγεγραμμένες περιοχές χαμηλού σήματος σε ένα υπόστρωμα φυσιολογικού μυελού υψηλού σήματος (Εικόνα 3).

- T2W-Short Tau Inversion Recovery (STIR)

Τα πρωτόνια του λίπους έχουν μεγαλύτερο χρόνο αποδιέγερσης T2 σε σχέση με τα πρωτόνια του ύδατος. Για τον λόγο αυτό, η ένταση του σήματος του λίπους είναι χαμηλότερη από αυτήν των δομών πλούσιων σε νερό, όπως οι αλλοιώσεις του πολλαπλού μυελώματος. Στις αντίστοιχες ακολουθίες, η ένταση σήματος του μυελού λίπους είναι παρόμοια ή ελαφρώς χαμηλότερη από αυτήν του υποδόριου λίπους, ενώ ο ερυθρός μυελός εμφανίζει ενδιάμεση ένταση, χαμηλότερη από αυτήν των αλλοιώσεων. Η διαφοροποίηση του ερυθρού από τον λιπώδη μυελό είναι επομένως πιο δύσκολη σε σχέση με τις T1W ακολουθίες.

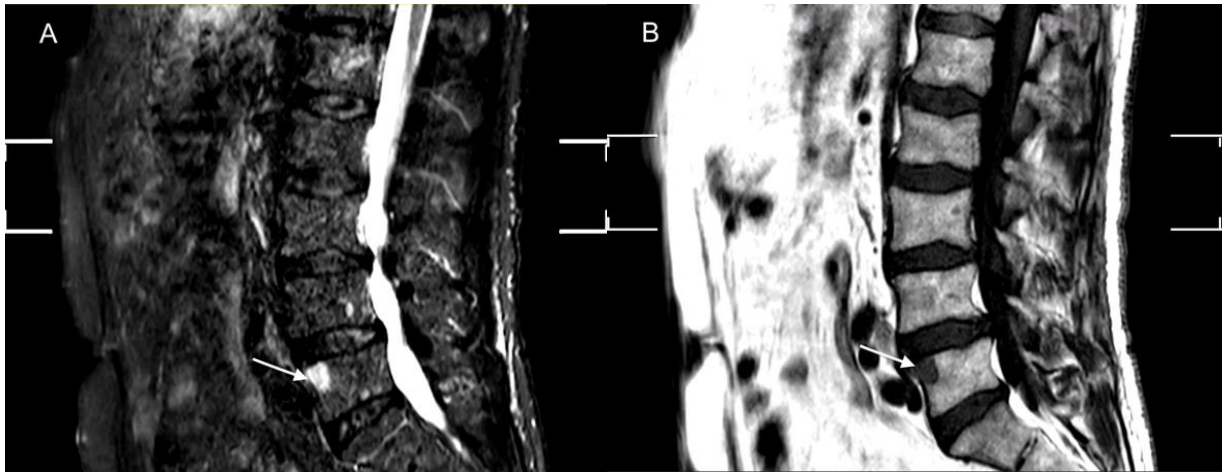
Εάν δεν εφαρμοστούν τεχνικές καταστολής του σήματος του λίπους, η ένταση του τελευταίου παραμένει υψηλή, δυσχεραίνοντας την αξιολόγηση του μυελού: τόσο ο λιπώδης μυελός όσο και αλλοιώσεις με υψηλή περιεκτικότητα σε νερό θα εμφανίζονται με ισχυρό σήμα [24]. Η καταστολή του λίπους είναι συνεπώς απαραίτητη για την απεικόνιση των αλλοιώσεων. Οι δύο κύριες τεχνικές που χρησιμοποιούνται είναι:

1. **Χημικώς επιλεκτική καταστολή**, όπου εφαρμόζεται ραδιοπαλμός που εξασθενεί επιλεκτικά το σήμα των πρωτονίων του λίπους. Όσο μικρότερη είναι η διαφορά της χημικής μετατόπισης μεταξύ ύδατος και λίπους (η οποία εξαρτάται από το κύριο μαγνητικό πεδίο του τομογράφου), τόσο δυσκολότερη είναι η διαφοροποίησή τους. Η επιτυχημένη εφαρμογή αυτής της τεχνικής οδηγεί σε:

- a. Χαμηλότερη ένταση σήματος του λιπώδους μυελού σε σχέση με τους παρακείμενους μύες,
- b. Ενδιάμεση ένταση σήματος για τον ερυθρό μυελό, περίπου ίση με τους μύες,
- c. Υψηλότερη ένταση σήματος για τις αλλοιώσεις σε σχέση με τους μύες.

2. **Τεχνική Short Tau Inversion Recovery (STIR)**, όπου εφαρμόζεται ραδιοπαλμός 180° για την αναστροφή της μαγνήτισης όλων των ιστών. Μετά το χρόνο αναστροφής (T1, inversion time), εφαρμόζεται ραδιοπαλμός 90° για την παραγωγή σήματος από όλους τους ιστούς εκτός του λίπους. Καθώς οι ιστοί χωρίς λίπος επανακτούν την εγκάρσια μαγνήτιση ταχύτερα από τους λιπώδεις ιστούς, ο λόγος σήματος προς θόρυβο (SNR) είναι υψηλότερος, με αποτέλεσμα οι ιστοί με μακρούς χρόνους αποδιέγερσης T1 και T2 να εμφανίζονται με έντονο σήμα [25].

Η STIR καταστέλλει το σήμα του λίπους και τονίζει το σήμα οιδηματωδών περιοχών ή περιοχών με υψηλή περιεκτικότητα σε ύδωρ. Οι εστιακές αλλοιώσεις εμφανίζονται ως περιοχές υψηλού σήματος σε αυτές τις ακολουθίες σε ένα υπόστρωμα χαμηλού σήματος (Εικόνα 3.B)



Εικόνα 3: Εστιακή αλλοίωση (άσπρο βέλος) παρατηρούμενη από την STIR (A) και την T1W-TSE (B) ακολουθία

1.2.2.2 Ποσοτικές Ακολουθίες

- T1W/T2W Dixon

Οι ακολουθίες DIXON αποκτώνται κυρίως με τεχνικές GRE, αν και τα τελευταία χρόνια εφαρμόζονται και μέσω spin-echo μεθόδων. Η βασική αρχή τους στηρίζεται στην καταγραφή εικόνων σύμφωνης φάσης (in-phase, IP) και αντίθετης φάσης (opposed-phase, OP) [26]. Η ένταση σήματος (signal intensity, SI) στις εικόνες IP προέρχεται από το άθροισμα του σήματος των πρωτονίων του ύδατος και του λίπους στον απεικονιζόμενο ιστό:

$$IP = \text{Water} + \text{Fat} \quad (\text{Eq. 4})$$

Αντίστοιχα, στις εικόνες OP η SI αντιστοιχεί στη διαφορά ανάμεσα στο σήμα των πρωτονίων του ύδατος και σε εκείνο του λίπους:

$$OP = \text{Water} - \text{Fat} \quad (\text{Eq. 5})$$

Από τους παραπάνω συνδυασμούς προκύπτουν δύο επιπλέον κανάλια, το κανάλι του ύδατος (Water only image, WO) και το κανάλι του λίπους (Fat only image, FO), τα οποία περιγράφονται από τις εξισώσεις:

$$\text{Water} = \frac{IP+OP}{2}, \quad \text{Fat} = \frac{IP-OP}{2} \quad (\text{Eq. 6,7})$$

Οι αλλοιώσεις που υποκαθιστούν τον φυσιολογικό μυελό εμφανίζουν μηδενική ένταση σήματος στο κανάλι FO. Επιπλέον, οι ακολουθίες DIXON εμφανίζουν υψηλότερο λόγο αντίθεσης προς θόρυβο (CNR) σε σύγκριση με τις T1W ακολουθίες, γεγονός που καθιστά τις αλλοιώσεις πιο ευδιάκριτες και μειώνει την πιθανότητα ψευδώς αρνητικών αποτελεσμάτων [27]. Στην Εικόνα 4.A παρατηρείται ο φαινότυπος μιας εστιακής αλλοίωσης στην T1W DIXON ακολουθία, στο κανάλι IP, όπου παρατηρείται χαμηλότερη ένταση σήματος στην αλλοίωση συγκριτικά με τον παρακείμενο μυελό.

- Diffusion Weighted Imaging

Η αντίθεση στις ακολουθίες διάχυσης μπορεί να αξιολογηθεί τόσο ποιοτικά όσο και ποσοτικά.

1. **Ποιοτική ερμηνεία:** Στις ακολουθίες διάχυσης που αποκτώνται με τεχνικές SSFP (υποκατηγορία των GRE ακολουθιών), η διάκριση μεταξύ καλοηθών και κακοήθων αλλοιώσεων βασίζεται κυρίως στην οπτική εκτίμηση των εικόνων [28].

2. **Ποσοτική ερμηνεία:** Επιτυγχάνεται με τη λήψη DWI εικόνων χρησιμοποιώντας ≥ 2 τιμές βάρους διάχυσης (b-values) και τον υπολογισμό των χαρτών φαινομενικού δείκτη διάχυσης (ADC maps), σύμφωνα με τη σχέση:

$$ADC = -\frac{1}{b_2 - b_1} \ln \left(\frac{S(b_2)}{S(b_1)} \right) \quad (\text{Eq. 8})$$

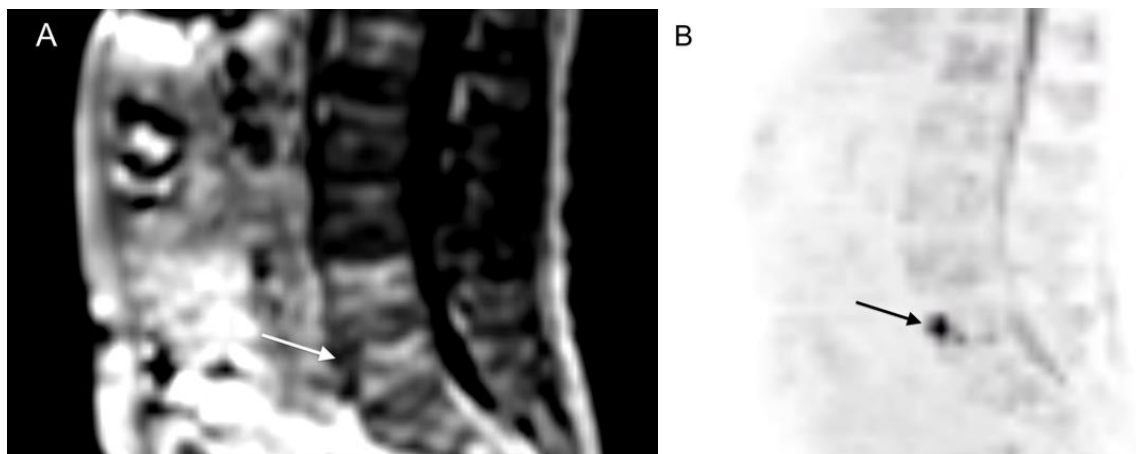
Χαρακτηριστικά τιμών ADC:

1. Οι τιμές του λιπώδους μυελού είναι χαμηλότερες σε σύγκριση με εκείνες του ερυθρού.
2. Οι κύριοι παράγοντες που επηρεάζουν το ADC είναι η περιεκτικότητα του μυελού σε λίπος καθώς και η δομή των δοκίδων του σπογγώδους οστού.
3. Παθολογικές διεργασίες που αντικαθιστούν τον φυσιολογικό μυελό εμφανίζονται συνήθως με αυξημένες τιμές ADC. Ωστόσο, το φαινόμενο αυτό περιορίζεται στον μυελό, καθώς σε άλλα όργανα όπως το ήπαρ ή ο εγκέφαλος οι παθολογικές αλλοιώσεις εμφανίζονται με μειωμένο ADC.
4. Φυσιολογικές τιμές ADC: $0.2 - 0.5 \times 10^{-3} \text{ mm}^2/\text{s}$
5. Παθολογικές τιμές ADC: $0.7 - 1.2 \times 10^{-3} \text{ mm}^2/\text{s}$
6. Στο διάχυτο πρότυπο του ΠΜ μπορεί να παρατηρηθεί επικάλυψη με τον φυσιολογικό μυελό. Η ζώνη επικάλυψης εντοπίζεται στο εύρος $0.5 - 0.7 \times 10^{-3} \text{ mm}^2/\text{s}$

Παρακάτω φαίνεται μια συνοπτική ταξινόμηση του μυελού με βάση τις τιμές ADC (κατά αύξουσα σειρά):

Λιπώδης μυελός < Ερυθρός μυελός < Διάχυτα παθολογικά πρότυπα < Εστιακά παθολογικά πρότυπα < Οξεία καλοήγη κατάγματα.

Οι εστιακές αλλοιώσεις του ΠΜ στην DWI αναδεικνύονται ως περιοχές περιορισμού διάχυσης, ήτοι με αυξημένο σήμα σε σχέση με τον παρακείμενο μυελό. Η φωτεινότητα των εικόνων αυτών συνήθως αναστρέφεται προκειμένου να διευκολύνεται η ανάγνωσή τους. Στην εικόνα 4.B παρουσιάζεται ο φαινότυπος μιας εστιακής αλλοίωσης σε μια DWI, $b = 900 \text{ s/mm}^2$ ακολουθία που έχει υποστεί αναστροφή φωτεινότητας.



Εικόνα 4: Αλλοίωση παρατηρούμενη από την T1W DIXON IP (άσπρο βέλος) (A) και την DWI, $b=900 \text{ s/mm}^2$ (μαύρο βέλος) (B) ακολουθία

1.3 Επεξεργασία Ιατρικών Εικόνων

Η προεπεξεργασία εικόνων αποτελεί το πρώτο στάδιο όσον αφορά την αναγνώριση εικόνας και παράλληλα ένα θεμελιώδες βήμα, καθώς η ενίσχυση της ευκρίνειας και η εξαγωγή των σωστών βασικών χαρακτηριστικών επηρεάζει σε μεγάλο βαθμό την απόδοση ολόκληρου του υπό κατασκευή συστήματος [29].

1.3.1 Supervoxels και κατακερματισμός (SLIC)

Το σημείο αυτό αφορά την κατάτμηση των εικόνων σε μικρές, συνεκτικές περιοχές, οι οποίες μοιράζονται κοινά χαρακτηριστικά και είναι σχετικά ομοιογενείς. Η επιλογή της μέθοδο του υπερ-κατακερματισμού (supervoxel segmentation) έναντι μιας απλής κατάτμησης των εικόνων σε χωρία ίδιου μεγέθους (πχ $25 \times 25 \times 25$) επιτρέπει:

- Τη διατήρηση τοπικών μορφολογικών χαρακτηριστικών και στατιστικών
- Την ενίσχυση της σταθερότητας των εξαγόμενων χαρακτηριστικών
- Τη διευκόλυνση μετέπειτα σταδίων να εξάγουν πληροφορία

Ο αλγόριθμος SLIC (Simple Linear Supervoxel Segmentation) αποτελεί μια παραλλαγή του διαδεδομένου **k-means**, ειδικά προσαρμοσμένου για τον κατακερματισμό εικόνων, και βασίζεται στη διάταξη πυρήνων (kernels) στον χώρο με τρόπο τέτοιο, ώστε κάθε πυρήνας να δημιουργεί ομοιόμορφα χωρία και να αναδιατάσσεται βάσει της τοπικής πληροφορίας [30]. Η βασική αρχή του SLIC είναι η ομαδοποίηση γειτονικών pixels (στη δική μας περίπτωση voxels) που παρουσιάζουν ομοιόμορφα ακτινολογικά χαρακτηριστικά και χωρική εγγύτητα, με τη χρήση μιας τροποποιημένης απόστασης που συνδυάζει τιμή έντασης και γεωμετρικής απόστασης.

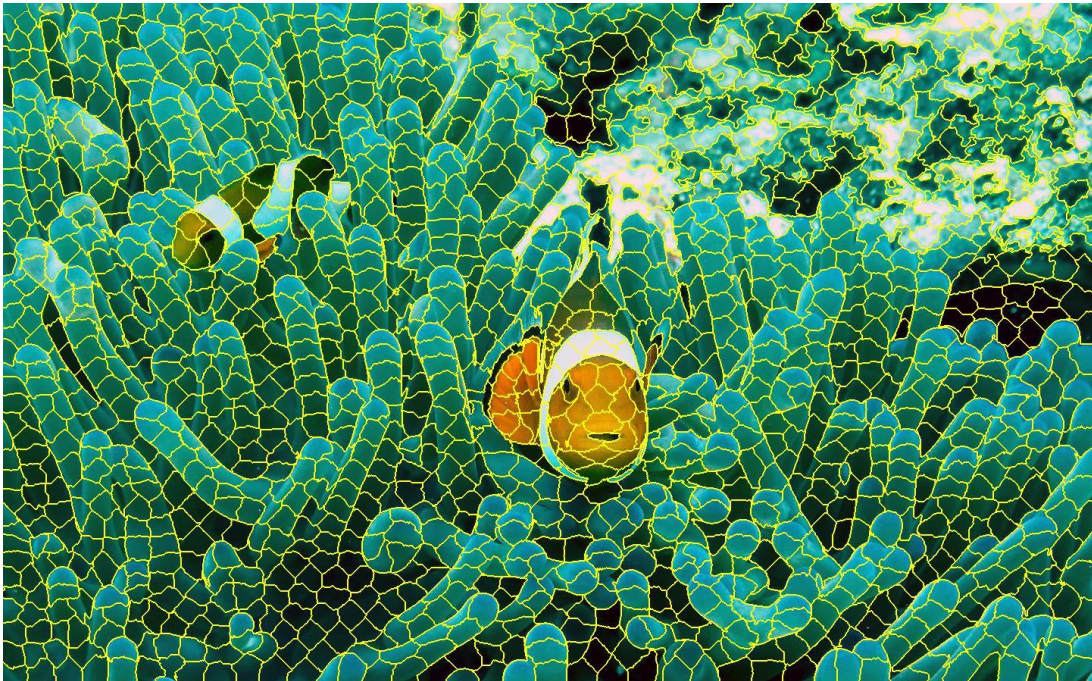
Ο αλγόριθμος λειτουργεί ως εξής:

1. Επιλέγεται ένας αρχικός αριθμός "cluster centers" (ανάλογα με το επιθυμητό πλήθος supervoxels), οι οποίοι τοποθετούνται ομοιόμορφα στο 3D πλέγμα της εικόνας.
2. Για κάθε voxel στην περιοχή επιρροής κάθε κέντρου (τοπική αναζήτηση), υπολογίζεται η απόσταση **D** που ορίζεται ως:

$$D = \sqrt{\left(\frac{d_c}{n_c}\right)^2 + \left(\frac{d_s}{n_s}\right)^2} \quad (\text{Eq. 9})$$

Όπου:

- d_c : διαφορά στην ένταση (color/contrast distance)
 - d_s : ευκλείδεια απόσταση στον 3D χώρο
 - n_c, n_s : παράγοντες κανονικοποίησης (intensity scale και spatial scale)
3. Ανατίθεται κάθε voxel στο πλησιέστερο κέντρο cluster βάσει του D.
 4. Ενημερώνονται τα κέντρα βάσει του μέσου όρου των voxels κάθε cluster.
 5. Επαναλαμβάνεται η διαδικασία έως ότου υπάρξει σύγκλιση ή συμπληρωθεί προκαθορισμένος αριθμός επαναλήψεων.



Εικόνα 5: Μια εφαρμογή του αλγορίθμου SLIC. Παρατηρούμε ότι περιοχές ενδιαφέροντος έχουν απομονωθεί κατάλληλα

Το πλεονέκτημα του SLIC έγκειται στην υψηλή ταχύτητα εκτέλεσης (λόγω τοπικής αναζήτησης) και της δυνατότητας ελέγχου του πόσο συμπαγή θα είναι τα παραγόμενα χωρία μέσω της μεταβλητής compactness. Η τιμή αυτή καθορίζει το βάρος της χωρικής απόστασης σε σχέση με τη διαφορά έντασης: Υψηλές τιμές δημιουργού πιο συμπαγή supervoxels, ενώ χαμηλές τιμές επιτρέπουν μεγαλύτερη παραμόρφωση [31], [32].

1.3.2 Εξαγωγή και Επιλογή Χαρακτηριστικών

Η εξαγωγή και επιλογή χαρακτηριστικών αποτελεί ένα κρίσιμο στάδιο στην ανάλυση ιατρικών εικόνων, ιδίως σε εφαρμογές που περιλαμβάνουν μηχανική μάθηση ή βαθιά μάθηση. Σκοπός της διαδικασίας είναι η μετατροπή των αρχικών απεικονιστικών δεδομένων σε ένα σύνολο αριθμητικών περιγραφών (features), που αποτυπώνουν τη μορφολογία, την υφή και την κατανομή των εντάσεων στις υπό μελέτη περιοχές. Η ποιότητα και η συνάφεια των χαρακτηριστικών αυτών επηρεάζουν καθοριστικά την απόδοση των ταξινομητών και των επιβλεπόμενων μοντέλων [33].

1.3.2.1 Εξαγωγή Χαρακτηριστικών

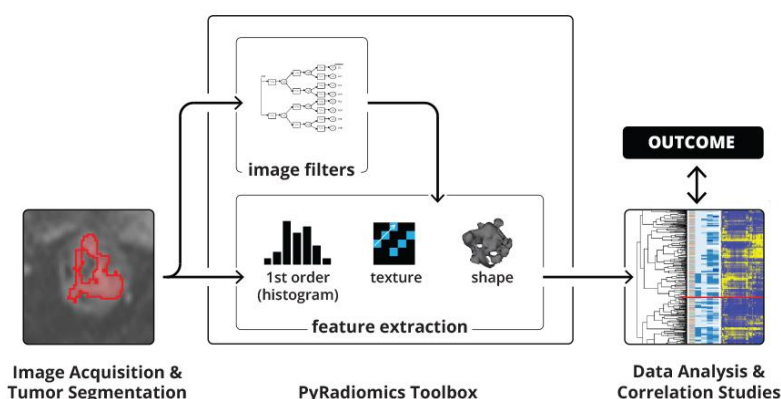
Η διαδικασία εξαγωγής χαρακτηριστικών μπορεί να βασιστεί είτε σε προκαθορισμένες μετρικές (handcrafted features) είτε σε αυτόματη μάθηση αναπαραστάσεων (representation learning) [34]. Η παρούσα θεώρηση εστιάζει σε προκαθορισμένες, στατιστικές μετρικές, οι οποίες είναι ιδιαίτερα χρήσιμες σε ρυθμίσεις μικρού δείγματος ή σε περιπτώσεις όπου η επεξηγησιμότητα είναι επιθυμητή.

Radiomic Χαρακτηριστικά

Τα radiomics είναι μία ευρέως διαδεδομένη τεχνική που περιλαμβάνει την εξαγωγή πληθώρας στατιστικών χαρακτηριστικών από περιοχές ενδιαφέροντος (ROIs) σε απεικονιστικά δεδομένα [35]. Τα χαρακτηριστικά αυτά ομαδοποιούνται σε τρεις βασικές κατηγορίες:

- **Χαρακτηριστικά πρώτης τάξης** (first-order features): Αφορούν την κατανομή της έντασης εντός της περιοχής, όπως μέση τιμή, διακύμανση, skewness, kurtosis, entropy, energy κ.ά.
- **Χαρακτηριστικά σχήματος** (shape features): Περιλαμβάνουν μετρικές όπως όγκος, επιφάνεια, λόγος επιφάνειας/όγκου, compactness, sphericity κ.ά.
- **Χαρακτηριστικά υφής** (texture features): Υπολογίζονται με βάση μήτρες συν-εμφάνισης εντάσεων όπως GLCM (Gray Level Co-occurrence Matrix), GLRLM (Gray Level Run Length Matrix), GLSZM, NGTDM κ.ά. και αποτυπώνουν τη συσχέτιση, ομοιομορφία και ανομοιομορφία των τιμών εντός του όγκου.

Η τυποποίηση των radiomics έχει ενσωματωθεί σε ανοιχτού κώδικα εργαλεία όπως το **PyRadiomics** [36], το οποίο εφαρμόζει σταθερές μεθοδολογίες σε εικόνες προεπεξεργασμένες με επαναδειγματοληψία (resampling) και κανονικοποίηση εντάσεων.



Εικόνα 6: Συνοπτική παρουσίαση της ροής πληροφορίας του πακέτου PyRadiomics της python

Intensity Percentiles

Μία εναλλακτική απλούστερη προσέγγιση περιλαμβάνει την περιγραφή της κατανομής έντασης μέσω επιλεγμένων εκατοστημορίων (percentiles), όπως το 20ό, 40ό, 60ό, 80ό και 100ό [37]. Οι μετρικές αυτές προσφέρουν συνοπτική αλλά χρήσιμη πληροφόρηση για την ένταση των voxels εντός κάθε υποπεριοχής (π.χ. supervoxel), με ιδιαίτερη χρησιμότητα σε περιπτώσεις με λιγότερο πολύπλοκες αναπαραστάσεις ή περιορισμένους υπολογιστικούς πόρους [38].

1.3.2.2 Επιλογή Χαρακτηριστικών

Η επιλογή χαρακτηριστικών αποτελεί ένα κρίσιμο στάδιο στη διαδικασία ανάλυσης δεδομένων, ιδιαίτερα σε προβλήματα υψηλής διαστασιμότητας όπως η ιατρική απεικόνιση. Σκοπός της είναι να εντοπίσει και να διατηρήσει εκείνα τα χαρακτηριστικά που προσφέρουν τη μεγαλύτερη διακριτική ικανότητα μεταξύ των τάξεων ενδιαφέροντος, απορρίπτοντας τα λιγότερο χρήσιμα ή θορυβώδη. Αυτό συμβάλλει στη βελτίωση της γενίκευσης των μοντέλων, στη μείωση της υπολογιστικής πολυπλοκότητας και στην αποτροπή της υπερεκπαίδευσης.

Για να γίνει εφικτή αυτή η επιλογή, χρησιμοποιούνται τεχνικές τακτικής κανονικοποίησης κατά την εκπαίδευση γραμμικών μοντέλων. Οι δύο βασικότερες μορφές κανονικοποίησης είναι οι:

1. L1 κανονικοποίηση, η οποία ορίζεται ως:

$$|w|_1 = \sum_j |w_j| \quad (\text{Eq. 10})$$

και έχει την ιδιότητα να προάγει αραιές λύσεις, δηλαδή να μηδενίζει εντελώς πολλούς από τους συντελεστές βαρών w_j . Αυτό σημαίνει ότι το μοντέλο επιλέγει αυτόματα ποια χαρακτηριστικά θεωρεί σημαντικά, απορρίπτοντας τα υπόλοιπα. Η L1 κανονικοποίηση χρησιμοποιείται χαρακτηριστικά στο LASSO [39].

2. L2 κανονικοποίηση, η οποία ορίζεται ως:

$$|w|_2^2 = \sum_j w_j^2 \quad (\text{Eq. 11})$$

και δεν μηδενίζει τα βάρη, αλλά τα εξαναγκάζει να παραμείνουν μικρά. Έτσι, αποφεύγονται φαινόμενα υπερπροσαρμογής, ειδικά όταν τα χαρακτηριστικά είναι πολυσυσχετισμένα. Η L2 κανονικοποίηση εφαρμόζεται στο Ridge Regression [40].

Στην πράξη, οι δύο αυτές τεχνικές μπορούν να συνδυαστούν μέσω της **Elastic Net κανονικοποίησης**, η οποία εισάγει έναν σταθμισμένο συνδυασμό των L1 και L2 ποινών στην αντικειμενική συνάρτηση εκπαίδευσης:

$$\mathcal{L}(w) = \mathcal{L}_{\log}(w) + \lambda[\alpha|w|_1 + (1 - \alpha)|w|_2^2] \quad (\text{Eq. 12})$$

όπου:

- $\mathcal{L}_{\log}(w)$ είναι η λογιστική απώλεια (cross-entropy),
- λ είναι ο συντελεστής κανονικοποίησης,
- $\alpha \in [0,1]$ ελέγχει την ισορροπία μεταξύ L1 και L2.

Η χρήση του Elastic Net [41] προσφέρει τα πλεονεκτήματα και των δύο τεχνικών: επιτρέπει την επιλογή χαρακτηριστικών μέσω της L1 συνιστώσας, ενώ παράλληλα διατηρεί τη σταθερότητα της λύσης και διαχειρίζεται την πολυσυγγραφοποίηση μέσω της L2. Έχει αποδειχθεί ιδιαίτερα αποτελεσματικό σε περιβάλλοντα με μεγάλο αριθμό συσχετισμένων χαρακτηριστικών, όπως συμβαίνει σε radiomic δεδομένα από ιατρικές εικόνες.

1.4 Γράφοι και Μοντέλα GNN στην Ιατρική Απεικόνιση

Η αναπαράσταση και ανάλυση ιατρικών δεδομένων πραγματοποιούνταν μέχρι πρότινος σχεδόν αποκλειστικά με τη χρήση CNN. Η αναπαράσταση των δεδομένων υπό τη μορφή γράφων έχει αποκτήσει αυξανόμενο ενδιαφέρον τα τελευταία χρόνια, λόγω της δυνατότητας των γράφων να αποτυπώνουν με φυσικό και εκφραστικό τρόπο χωρικές και σημασιολογικές σχέσεις. Η αυξανόμενη χρήση γράφων οδήγησε, λοιπόν, και στην δημιουργία **Συνελκτικών Δικτύων Γράφων (Graph Neural Networks – GNN)**, τα οποία είναι συμβατά με τους γράφους και μπορούν βέλτιστα να μεταφέρουν πληροφορίες μεταξύ γειτονικών κόμβων [45].

1.4.1 Γράφοι: ορισμοί και χαρακτηριστικά

Ένας μη κατευθυνόμενος γράφος ορίζεται ως $G = (V, E)$, όπου:

- V : σύνολο κόμβων (στην περίπτωσή μας, supervoxels)
- E : σύνολο ακμών που συνδέουν ζεύγη κόμβων (γειτονικές περιοχές)

Κάθε κόμβος συνοδεύεται από ένα διάνυσμα χαρακτηριστικών $x_i \in \mathbb{R}^F$, όπου F είναι το πλήθος των γνωρισμάτων που περιγράφουν τον κόμβο (π.χ. radiomic χαρακτηριστικά). Οι ακμές είτε είναι δυαδικές (1 για ύπαρξη σύνδεσης, 0 για μη σύνδεση) είτε φέρουν βάρος που περιγράφει τη δύναμη ή την ποιότητα της σχέσης μεταξύ κόμβων.

Η χρήση γράφων επιτρέπει:

- Τη διατήρηση της δομής του ιστού σε επίπεδο σπονδύλου,
- Την αναγνώριση από το μοντέλο δομών, οργάνων, χωρίων,
- Την εφαρμογή εξειδικευμένων μοντέλων μάθησης που ενσωματώνουν δομικές πληροφορίες.

1.4.2 Θεμελιώδεις αρχές των GNN

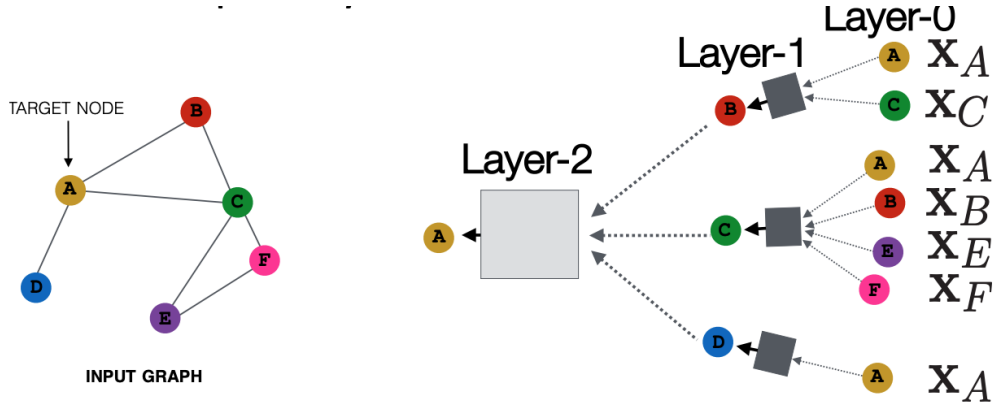
Τα Νευρωνικά Δίκτυα Γράφων - GNN είναι μοντέλα βαθιάς μάθησης σχεδιασμένα για δομές δεδομένων με γραφική μορφή. Σε αντίθεση με τα συμβατικά CNN που λειτουργούν πάνω σε πλέγματα (εικόνες), τα GNN λειτουργούν απευθείας πάνω σε γράφους, εκτελώντας "message passing" — δηλαδή, κάθε κόμβος επικαιροποιεί την αναπαράστασή του λαμβάνοντας υπόψη τις πληροφορίες από τους γείτονές του [46]. Η γενική μορφή κάθε layer GNN έχει τη μορφή:

$$h_i^{(l+1)} = \sigma \left(\text{AGGREGATE}^{(l)} \left(\{h_j^{(l)} : j \in \mathcal{N}(i)\} \right) \right) \quad (\text{Eq. 13})$$

Όπου:

- $h_i^{(l)}$: τα χαρακτηριστικά του κόμβου i στο επίπεδο l
- $\mathcal{N}(i)$: το σύνολο γειτόνων του κόμβου i
- σ : μη γραμμική συνάρτηση ενεργοποίησης

Η βασική ιδέα είναι η εκμάθηση αντιπροσωπευτικών ενσωματώσεων (embeddings) για κάθε κόμβο, λαμβάνοντας υπόψη τόσο τις ιδιότητές του όσο και τη θέση του στον γράφο.



Εικόνα 7: Απεικόνιση της βασικής αρχής των Graph Neural Networks, με κάθε κόμβο να είναι ένας συνδυασμός των γειτόνων του

1.4.3 Θεμελιώδεις αρχιτεκτονικές GNN

Μερικές από τις πιο διαδεδομένες αρχιτεκτονικές GNN που έχουν χρησιμοποιηθεί σε έρευνες και μελέτες, ενώ παράλληλα έχουν υλοποιηθεί και από γνωστές βιβλιοθήκες της Python, είναι οι ακόλουθες:

1.4.3.1 Graph Convolutional Network (GCN)

Το GCN βασίζεται σε έναν φιλτραρισμένο μηχανισμό συνελκτικών πράξεων επάνω στη δομή του γράφου. Σε κάθε layer, τα χαρακτηριστικά κάθε κόμβου επικαιροποιούνται βάσει του μέσου όρου των χαρακτηριστικών των γειτόνων του [42]. Το GCN μοντελοποιεί καλά ομαλές συναρτήσεις επάνω στον γράφο, αλλά μπορεί να αντιμετωπίσει δυσκολίες σε περιοχές με υψηλή ετερογένεια πληροφορίας. Ο μαθηματικός τύπος ενός GCN layer είναι:

$$H^{(l+1)} = \sigma \left(D^{-1/2} \tilde{A} D^{-1/2} H^{(l)} W^{(l)} \right) \quad (\text{Eq. 14})$$

Όπου:

- $\tilde{A} = A + I$: ο πίνακας γειτνίασης με προσθήκη ταυτότητας (self-loops)
- $\tilde{D}$: ο διαγώνιος πίνακας βαθμού των κόμβων
- $W^{(l)}$: βάρη προς εκμάθηση για το layer l
- σ : μη γραμμική συνάρτηση (π.χ. ReLU)

1.4.3.2 Graph Attention Network (GAT)

Το GAT εισάγει μηχανισμούς προσοχής (attention), επιτρέποντας στο μοντέλο να μαθαίνει διαφορετικό "βάρος" για κάθε γείτονα ενός κόμβου κατά την ενημέρωση [44]. Έτσι, ενισχύεται η σημασία σημαντικών κόμβων στη γειτονιά και περιορίζεται η επίδραση άσχετων περιοχών. Είναι ιδιαίτερα χρήσιμο όταν ο γράφος περιέχει κόμβους με ποικίλες σημασιολογικές ιδιότητες. Ο μαθηματικός τύπος είναι:

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)} \right), \quad \alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a^T [W h_i || W h_j]))}{\sum_{k \in \mathcal{N}(i)} \exp(\text{LeakyReLU}(a^T [W h_i || W h_k]))} \quad (\text{Eq. 15, 16})$$

Όπου:

- $\alpha_{ij}^{(l)}$: ο συντελεστής προσοχής μεταξύ κόμβων i και j :

- $\mathbf{a}$: διάνυσμα attention (προς εκμάθηση)
- $\parallel$: συνεκκόλληση (concatenation)

1.4.3.3 Graph Sample and Aggregate (GraphSAGE)

Το GraphSAGE υλοποιεί ένα πιο ευέλικτο πλαίσιο "δειγματοληψίας και συνάθροισης" (sampling & aggregation), επιτρέποντας τη μάθηση σε γράφους μεταβλητού μεγέθους και τη δυνατότητα generalization σε μη-γνωστά υποδίκτυα [43]. Η επιλογή συνάρτησης συνάθροισης (mean, LSTM, pooling) επηρεάζει την τελική απόδοση του μοντέλου. Γενικός τύπος:

$$h_i^{(l+1)} = \sigma \left(W^{(l)} \cdot \text{AGGREGATE} \left(\{h_i^{(l)}\} \cup \{h_j^{(l)} : j \in \mathcal{N}(i)\} \right) \right) \quad (\text{Eq. 17})$$

Η συναρτησιακή μορφή της AGGREGATE μπορεί να είναι μέση τιμή (mean), μέγιστο (max pooling), ή ακόμα και LSTM-based συνάθροιση.

1.5 Σχετικές Εργασίες

Η μελέτη του Πολλαπλού Μυελώματος (ΠΜ) και η προσπάθεια ανάπτυξης αυτοματοποιημένων συστημάτων υποβοήθησης της διάγνωσης, παρ' ότι δεν είχε προσελκύσει μεγάλο μέρος της επιστημονικής κοινότητας μέχρι πριν μία δεκαετία, έχει αποτελέσει πεδίο με σημαντική πρόοδο τα τελευταία χρόνια. Η ραγδαία εξέλιξη των τεχνικών ιατρικής απεικόνισης, σε συνδυασμό με την πρόοδο της τεχνητής νοημοσύνης, έχει επιτρέψει την ανάδυση προηγμένων μεθόδων για την ανάλυση εικόνων MRI. Ειδικότερα, η εφαρμογή Graph Neural Networks (GNNs) σε δομημένα ιατρικά δεδομένα ανοίγει νέους δρόμους για την κατανόηση της παθοφυσιολογίας και την υποστήριξη της διάγνωσης και πρόγνωσης. Στην παρούσα ενότητα παρουσιάζονται σχετικές εργασίες τόσο για το ΠΜ όσο και για τη χρήση GNNs στον τομέα της ιατρικής απεικόνισης.

1.5.1 Συστήματα τμηματοποίησης αλλοιώσεων σε MRI ασθενών με Πολλαπλό Μυέλωμα

Οι μελέτες που αφορούν την αυτόματη τμηματοποίηση των αλλοιώσεων του πολλαπλού μυελώματος (ΠΜ) μέσω συστημάτων τεχνητής νοημοσύνης (TN) παραμένουν ακόμα σε πολύ αρχικό στάδιο. Μια πρόωμη προσέγγιση [47], χρησιμοποίησε χαρακτηριστικά εικόνων T1W και T2W για τη διαμόρφωση ενός ταξινομητή Random Forest σε επίπεδο ογκοστοιχείου (voxel), ο οποίος ταξινομούσε τα voxel σε δύο κατηγορίες: εστιακής αλλοίωσης ή φυσιολογικού μυελού. Τα δεδομένα εκπαίδευσης προέρχονταν από χειροκίνητα εξαγμένες περιοχές εντός των εστιακών αλλοιώσεων. Συγκεκριμένα, εφαρμόστηκαν δύο παραλλαγές: στην πρώτη, ο ταξινομητής εκπαιδεύτηκε με την ακριβή τιμή της φωτεινότητας, ενώ στη δεύτερη χρησιμοποιήθηκαν χαρακτηριστικά της φωτεινότητας (μέγιστη, ελάχιστη, μέση και διάμεσος τιμή) σε ένα παράθυρο 5×5 ογκοστοιχείων. Η απόδοση στο σύνολο δοκιμής ήταν περίπου 95% ακρίβεια.

Μεταγενέστερη μελέτη [49] υιοθέτησε δίκτυο αρχιτεκτονικής κωδικοποιητή–αποκωδικοποιητή για την τμηματοποίηση διαφορετικών τύπων αλλοιώσεων (εστιακών και διάχυτων), χρησιμοποιώντας δεδομένα από 45 ασθενείς με T1W και STIR ακολουθίες. Το δείγμα εμπλουτίστηκε με τεχνικές επαύξησης (augmentation). Παρά το γεγονός ότι η μελέτη δεν εισήγαγε κάποια καινοτόμο μέθοδο στην τμηματοποίηση, ανέδειξε ότι τα συνελκτικά δίκτυα δύνανται να τμηματοποιήσουν επιτυχώς αλλοιώσεις ΠΜ. Παράλληλα, σε παρόμοια εργασία [48], παρουσιάστηκε πως τα συνελκτικά δίκτυα είναι ικανά στην τμηματοποίηση του μυελού και στην επακόλουθη εξαγωγή ραδιομικών

χαρακτηριστικών, αν και δεν πραγματοποιήθηκε περαιτέρω ανάλυση για την εξαγωγή και ποσοτικοποίηση των αλλοιώσεων, πέραν της επίδειξης αυτής της δυνατότητας.

1.5.2 GNN και Ιατρική Απεικόνιση

Μία από τις μελέτες που επηρέασαν καθοριστικά τον σχεδιασμό της παρούσας εργασίας είναι η [42]. Σε αυτήν, οι συγγραφείς εισάγουν ένα υβριδικό σύστημα που συνδυάζει deep features από CNN με μια GCN-βασισμένη αρχιτεκτονική για ημι-εποπτευόμενη ταξινόμηση εικόνων βυθού του οφθαλμού.

Τα σημεία ενδιαφέροντος εντοπίζονται με βάση τον αλγόριθμο SURF, τα οποία στη συνέχεια θα αποτελέσουν τους κόμβους ενός γράφου. Κάθε ROI θα περιέχει ένα σύνολο χαρακτηριστικών (embeddings), με βάση τα οποία θα δημιουργηθούν και οι ακμές των γράφων. Πιο συγκεκριμένα, θα επιλεγούν οι k -κοντινότεροι γείτονες με κριτήριο την απόσταση των features στο latent space. Η αρχιτεκτονική περιλαμβάνει δύο επίπεδα **GCN**, τα οποία εφαρμόζονται πάνω στο γράφο για να ενσωματώσουν πληροφορία από γειτονικές περιοχές. Σαν τελικό στάδιο, η έξοδος του GCN και του CNN ενσωματώνονται μέσω μιας διαδικασίας ενοποίησης (concatenation) και ενός τελικού fully connected layer. Ο συνδυασμός GNN για την εκμετάλλευση της τοπολογικής πληροφορίας αποδείχτηκε καθοριστική στην έρευνα αυτή, καθώς και στην παρούσα εργασία.

Μια ακόμα έρευνα που ασχολείται με GNN και brain tumor segmentation [43] προτείνει ένα υβριδικό μοντέλο που συνδυάζει GNNs και CNNs για τη βελτίωση της ακρίβειας στην τμηματοποίηση εγκεφαλικών όγκων από πολυπαραμετρικές MRI. Η διαδικασία ξεκινά με την εφαρμογή του αλγορίθμου SLIC για την εξαγωγή υπερ-ογκιδίων (supervoxels), τα οποία χρησιμοποιούνται ως κόμβοι σε έναν γράφο με χαρακτηριστικά βασισμένα σε quintiles έντασης από τις διάφορες MRI ακολουθίες. Οι ακμές καθορίζονται με βάση τη χωρική γειτνίαση. Πάνω στον γράφο εφαρμόζεται ένα **GraphSAGE** δίκτυο, το οποίο αποδίδει πιθανότητες κλάσης σε κάθε κόμβο, προσφέροντας μια πρώτη εκτίμηση της κατανομής του όγκου. Οι προβλέψεις του GNN προβάλλονται πίσω στον αρχικό 3D χώρο της εικόνας και χρησιμοποιούνται για την περικοπή της περιοχής ενδιαφέροντος, η οποία εν συνεχεία προωθείται μαζί με τα αρχικά δεδομένα και τα logits σε ένα απλό CNN δύο επιπέδων. Η πληροφορία του CNN ενισχύει την ακρίβεια στα όρια των βλαβών, με βελτίωση που αγγίζει το 2% στο Dice score, σύμφωνα με το BraTS 2021 dataset.

Τέλος, μια ακόμα ενδιαφέρουσα εργασία που σχετίζεται με πολυκατηγορική τμηματοποίηση εγκεφαλικών όγκων [38] προτείνει ένα πλήρως αυτοματοποιημένο pipeline, το οποίο βασίζεται σε Graph Attention Networks (**GAT**). Αρχικά, οι MRI εικόνες προεπεξεργάζονται με histogram matching και z-score normalization, ενώ εφαρμόζεται thresholding στο Flair channel για εξαγωγή περιοχών ενδιαφέροντος. Η υπερτμηματοποίηση γίνεται με τον αλγόριθμο SLIC και κάθε superpixel αντιστοιχεί σε κόμβο γράφου. Τα χαρακτηριστικά των κόμβων περιλαμβάνουν 5 στατιστικά ($\{10, 25, 50, 75, 90\}^{\text{th}}$ percentiles) από κάθε modality (T1, T1ce, T2, Flair), με συνολικά 20 διαστάσεις. Οι ακμές του γράφου κατασκευάζονται με βάση τη γειτνίαση και παραμένουν στατικές κατά την εκπαίδευση. Το GAT αποτελείται από 8 διαδοχικά attention layers με LeakyReLU και ένα τελικό στάδιο με softmax κανονικοποίηση για τα attention weights, ώστε να εξάγει προβλέψεις για κάθε superpixel, μια για κάθε label. Οι έξοδοι προβάλλονται πίσω στην εικόνα για την τελική χαρτογράφηση των κατηγοριών tumor core, enhancing tumor και whole tumor. Το pipeline αξιολογήθηκε στο BraTS 2021 dataset με βελτιωμένες επιδόσεις σε σχέση με baseline CNN-only μοντέλα.

2 Μεθοδολογία

2.1 Διαμόρφωση γράφων

Η αναπαράσταση κάθε ασθενούς υπό τη μορφή γράφου αποτέλεσε τον θεμέλιο λίθο της παρούσας μελέτης. Η διαδικασία κατασκευής των γράφων περιλαμβάνει τη δημιουργία supervoxels, την εξαγωγή χαρακτηριστικών, επιλογή χαρακτηριστικών και τον ορισμό των γράφων με βάση τις προηγούμενες επιλογές, εξετάζοντας πέντε διαφορετικές αναπαραστάσεις.

2.1.1 Κατακερματισμός (Supervoxel Segmentation)

Ο κατακερματισμός πραγματοποιήθηκε με τον αλγόριθμο SLIC στον όγκο εντός των σπονδύλων, μετά την περικοπή της εικόνας στο σχήμα της κάθε μάσκα. Στο πλαίσιο της μελέτης διερευνήθηκαν δύο διαφορετικές στρατηγικές:

- Σταθερός αριθμός supervoxels per patient (~1000 supervoxels): Ενιαία κατανομή ανεξαρτήτως ανατομικού όγκου, κατάλληλη για πιο συνεπή αναπαράσταση across patients.
- Καθορισμένος όγκος κάθε supervoxel (~150 mm³/supervoxel): Η παράμετρος `n_segments` προσαρμόζεται με βάση ένα target supervoxel size, ώστε να προσαρμόζεται δυναμικά σε περιοχές ή γενικότερα σε ασθενείς με μεγαλύτερες ανατομικές περιοχές.

Η δεύτερη στρατηγική στηρίχθηκε σε κλινικά ευρήματα αναφορικά με το ελάχιστο μέγεθος των παθολογικών εστιών στο μυελό των οστών. Σύμφωνα με τη βιβλιογραφία [50], [51], η μικρότερη αναγνωρίσιμη εστιακή βλάβη έχει ελάχιστη διάμετρο περίπου 5 mm, γεγονός που αντιστοιχεί σε όγκο περίπου 60 voxels, αλλά αντισταθμίζοντας την ευκρίνεια και την υπολογιστική πολυπλοκότητα, επιλέχθηκε ως αντιπροσωπευτικό μέγεθος supervoxel τα 150 voxels.

Και στις δύο περιπτώσεις χρησιμοποιήθηκε compactness ίσο με 0.1, ενώ οι τεχνικές συγκρίθηκαν ως προς την επιτεύξιμη ακρίβεια τμηματοποίησης (**Achievable Segmentation Accuracy**), υπολογισμένη μέσω του Dice score μεταξύ των ground-truth βλαβών και των supervoxels που τις περιείχαν. Η εν λόγω μετρική εκτιμά το μέγιστο Dice που θα μπορούσε θεωρητικά να επιτευχθεί από ένα τέλειο μοντέλο, το οποίο απλώς ταξινομεί κάθε supervoxel ως παθολογικό ή όχι, με βάση την αλληλοεπικάλυσή του με την πραγματική μάσκα της βλάβης (κάθε supervoxel χαρακτηριζόταν ως παθολογικό όταν τουλάχιστον το 10% του όγκου του συνέπιπτε με την ground truth μάσκα της βλάβης). Τα αποτελέσματα έδειξαν ότι η δεύτερη προσέγγιση (150 voxels per supervoxel) παρήγαγε σαφώς υψηλότερες τιμές Dice, υποδεικνύοντας καλύτερη ανατομική αντιστοίχιση των υπερτμημάτων με τις περιοχές ενδιαφέροντος [52].

2.1.2 Εξαγωγή Χαρακτηριστικών

Αφότου εφαρμόστηκε ο κατακερματισμός και παράχθηκαν τα επιθυμητά supervoxels, παράχθηκαν δύο διαφορετικά σύνολα χαρακτηριστικών για κάθε ένα από αυτά:

- Radiomics-based χαρακτηριστικά: Περιλαμβάνουν στατιστικά πρώτης τάξης (mean, std, skewness, kurtosis), καθώς και δείκτες υφής (GLCM, GLRLM, GLSZM, NGTDM), τόσο από τις T1 όσο και από τις T2 εικόνες.
- Κατανομή εντάσεων (Quintile-based): Για κάθε supervoxel, υπολογίστηκαν τα 20ό, 40ό, 60ό, 80ό και 100ό εκατοστημόριο (percentiles) των εντάσεων, προσφέροντας μια συνοπτική

περιγραφή της κατανομής. Η μέθοδος αυτή, παρ' ότι είναι απλούστερη, χρησιμοποιείται από πολλές παραπλήσιες εργασίες και θεωρήθηκε άξια περαιτέρω διερεύνησης.

Αμφότερες οι μεθοδολογίες εξετάστηκαν ανεξάρτητα ή σε συνδυασμό, και τα εξαγόμενα χαρακτηριστικά αποθηκεύονται για κάθε supervoxel, με αποτέλεσμα να προκύψουν 5 παραλλαγές γράφων:

1. 1000 Supervoxels – Radiomic χαρακτηριστικά
2. 1000 Supervoxels – Intensity Quintiles
3. 150 mm³ Supervoxels – Radiomic χαρακτηριστικά
4. 150 mm³ Supervoxels – Intensity Quintiles
5. 150 mm³ Supervoxels – Radiomic + Intensity Quintiles (συνδυαστικό)

Αυτή η επιλογή των γράφων ήταν στοχευμένη, ώστε να εξεταστούν όλες οι πτυχές και τα επίπεδα πολυπλοκότητας και να συμπεράνουμε εάν μεγαλύτερη πολυπλοκότητα συνεπάγεται μεγαλύτερη επίδοση του εκάστοτε μοντέλου.

2.1.3 Επιλογή Χαρακτηριστικών

Δεδομένων των supervoxels και των χαρακτηριστικών καθενός από αυτά, πραγματοποιήθηκε feature selection σε global επίπεδο, αφού πρώτα συγκεντρώθηκαν όλα τα διανύσματα χαρακτηριστικών από όλους τους ασθενείς. Ο στόχος ήταν η μείωση των διαστάσεων, με γνώμονα την αύξηση της γενίκευσης και την ενίσχυση της διαχωρισιμότητας μεταξύ των υγιών και παθολογικών supervoxels. Η διαδικασία ακολουθεί τρία διαδοχικά βήματα:

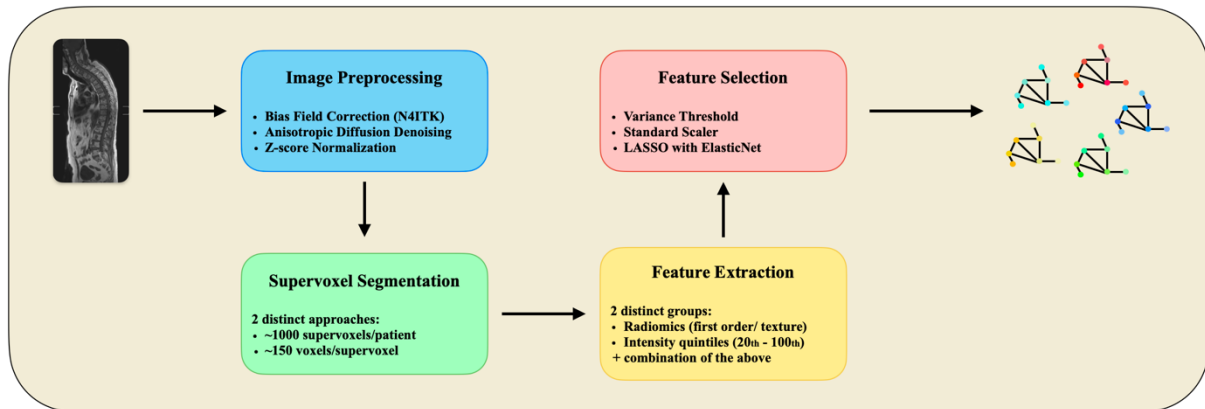
1. Variance Thresholding: Αρχικά, απομακρύνθηκαν όλα τα χαρακτηριστικά με ελάχιστη διακύμανση (κατώφλι: $1e-8$), διατηρώντας μόνο όσα παρουσιάζουν έστω και μικρές διαφοροποιήσεις μεταξύ των δειγμάτων. Επιλέχθηκε τέτοια κλίμακα, διότι μια τιμή της τάξης του $1e-3$ θα απέκλειε πιθανώς σημαντικές διαφοροποιήσεις (για dataset με τη θετική κλάση να καταλαμβάνει το 4%, η τιμή $4e-3$ περιγράφει σημαντική πληροφορία).
2. Κανονικοποίηση (Scaling): Τα διατηρημένα χαρακτηριστικά τυποποιήθηκαν με StandardScaler, έτσι ώστε να έχουν μέση τιμή 0 και τυπική απόκλιση 1, προαπαιτούμενο για L1-βασισμένες μεθόδους.
3. LASSO με ElasticNet Logistic Regression: Η επιλογή βασίστηκε σε λογιστική παλινδρόμηση με κανονικοποίηση Elastic-Net, προκειμένου να επιλεγθούν τα χαρακτηριστικά εκείνα, τα οποία έχουν τη μεγαλύτερη συμβολή στον καθορισμό του αποτελέσματος (επιλογή των χαρακτηριστικών με τους υψηλότερους συντελεστές).

2.1.4 Κατασκευή Γράφων

Οι τελικοί γράφοι δημιουργήθηκαν ορίζοντας τα supervoxels του κάθε ασθενή ως κόμβους, με τα χαρακτηριστικά του καθενός να προκύπτουν από το feature extraction, και ακμές μεταξύ των κόμβων που προκύπτουν από τη συνδεσιμότητα των supervoxels με βάση τον 3D κανόνα του 26-γείτονα [53], [54]. Σύμφωνα με τον κανόνα αυτό, δυο supervoxels θεωρούνται γειτονικά εάν βρίσκονται σε απόσταση 1 κατά μήκος οποιουδήποτε από τους τρεις άξονες (x, y, z), συμπεριλαμβανομένων και των διαγώνιων. Η τελική αναπαράσταση γράφου ορίστηκε ως αντικείμενο τύπου `torch_geometric.data.Data`, με τα εξής πεδία:

- x : Ο πίνακας χαρακτηριστικών κόμβων ($N \times F$), όπου N ο αριθμός των supervoxels και F τα επιλεγμένα χαρακτηριστικά.
- y : Ο πίνακας ετικετών ($N \times 1$), που δηλώνει την παρουσία ή απουσία βλάβης στο αντίστοιχο supervoxel.
- $edge_index$: Δείκτης ακμών του γράφου ($2 \times E$), ο οποίος κωδικοποιεί τη συνδεσιμότητα μεταξύ κόμβων με βάση 3D 26-γειτονία.
- pos : Οι χωρικές συντεταγμένες των κέντρων βάρους των supervoxels ($N \times 3$), χρήσιμες για positional encoding.
- sv_ids : Τα μοναδικά IDs των supervoxels, χρήσιμα για ανακατασκευή και debugging.

Μια συνολική ανασκόπηση της παραπάνω διαδικασίας φαίνεται και στην Εικόνα 8, με την έξοδο να είναι 5 διαφορετικά είδη γράφων:



Εικόνα 8: Το αρχικό pipeline της εργασίας, το οποίο ξεκινά με την εισαγωγή των εικόνων, ακολουθεί προεπεξεργασία, κατακερματισμός, εξαγωγή και επιλογή χαρακτηριστικών και ολοκληρώνεται με την παραγωγή των 5 διαφορετικών τύπων γράφων

2.2 Αρχιτεκτονικές δικτύων

Στο παρόν έργο υλοποιήθηκαν και αξιολογήθηκαν τρεις διαφορετικές αρχιτεκτονικές γραφικών νευρωνικών δικτύων (GNN): **Graph Convolutional Network (GCN)**, **Graph Attention Network (GAT)** και **GraphSAGE**. Παρά τις διαφορές στον τρόπο επεξεργασίας της δομής του γράφου, όλες οι αρχιτεκτονικές μοιράζονται ένα κοινό πλαίσιο σχεδίασης που διασφαλίζει την ομοιομορφία στη σύγκριση των επιδόσεών τους (Εικόνα 9).

2.2.1 Model Overview

Η βασική δομή όλων των μοντέλων περιλαμβάνει τα ακόλουθα στάδια:

- **Input Linear Layer**: Τα χαρακτηριστικά εισόδου $\mathbf{In}$ κάθε κόμβου προβάλλονται σε κοινό χώρο $\mathbf{H}$ διαστάσεων μέσω ενός πλήρως συνδεδεμένου επιπέδου.
- **GNN Layers**: Ακολουθούν πολλαπλά επίπεδα συνελιζων γραφήματος. Κάθε επίπεδο περιλαμβάνει:
 - GNN ConvLayer (GCN/ GATConv/ SAGEConv): $\mathbf{L} \rightarrow \mathbf{L}$,
 - GraphNorm normalization,
 - GELU activation function,

- dropout ($p=0.5$)
- Residual Connections: Εφαρμόζονται σε κάθε επίπεδο με συντελεστή κλίμακας που μαθαίνεται, για σταθερότερη εκπαίδευση.
- Edge Dropout: Κατά την εκπαίδευση εφαρμόζεται τυχαία αφαίρεση ακμών ($p = 0.2$), ώστε να μειώνεται η εξάρτηση από συγκεκριμένες τοπικές συνδέσεις.
- Jumping Knowledge (JK): Τα embeddings όλων των επιπέδων συνενώνονται (concatenation) ώστε να αξιοποιείται πληροφορία από διαφορετικά βάθη του γράφου. Η έξοδος θα έχει διαστάσεις $H \times L$, λόγω της συνένωσης
- Output Head: Η τελική αναπαράσταση περνά από ένα output head, το οποίο με τη σειρά του περιλαμβάνει:
 - Linear Layer ($H \times L \rightarrow H$)
 - GELU activation function
 - Dropout ($p=0.5$)
 - Linear Layer ($H \rightarrow 2$).

Η αρχιτεκτονική αυτή παράγει στην έξοδο δύο logits ανά κόμβο (supervoxel), αντιστοιχίζοντας την πιθανότητα απουσίας ή μη βλάβης.

2.2.2 GCN

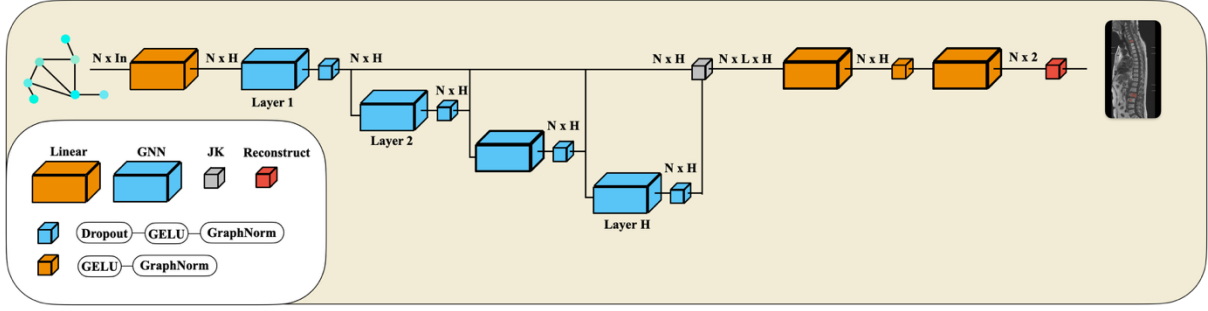
Το μοντέλο GCN αξιοποιεί τον τελεστή **GCNConv** του PyTorch Geometric, ο οποίος συνδιάζει την πληροφορία του κάθε κόμβου με τους γειτονικούς μέσα από μία διαδικασία συμμετρικής κανονικοποίησης του πίνακα γειτνίασης και σε απλή μέση αναγωγή (mean aggregation) των χαρακτηριστικών των γειτονικών κόμβων. Σε σχέση με τα άλλα μοντέλα, τα GCN αποτελούν την πιο απλή αρχιτεκτονική, χωρίς μεγάλο flexibility ή κάποιον learnable μηχανισμό [55].

2.2.3 GAT

Το GAT διαφοροποιείται από το GCN εισάγοντας έναν multi-head attention μηχανισμό μέσω του τελεστή **GATv2Conv**. Κάθε κόμβος υπολογίζει διαφορετικά βάρη προσοχής για κάθε γείτονά του, επιτρέποντας στο μοντέλο να εστιάζει σε πιο σχετικούς κόμβους. Η έξοδος κάθε επιπέδου ήταν αποτέλεσμα συνένωσης (enation) από 4 heads, οι οποίοι εκπαιδεύονταν ανεξάρτητα. Ο μηχανισμός αυτός καθιστά το GAT περισσότερο εκφραστικό, ειδικά σε περιβάλλοντα με υψηλή δομική ποικιλομορφία. Ωστόσο, έχει αυξημένη υπολογιστική πολυπλοκότητα, καθώς κάθε layer περιλαμβάνει περισσότερες παραμέτρους και πράξεις. Τέλος, η συγκεκριμένη αρχιτεκτονική παρέχει τη δυνατότητα εισαγωγής ενός dropout μηχανισμού εντός του GAT (attn_dropout = 0.2), ο οποίος αποτρέπει το overfitting των attentional weights [56].

2.2.4 GraphSAGE

Τέλος, το GraphSAGE χρησιμοποιεί τον τελεστή **SAGEConv** με στρατηγική max pooling aggregation, μια θεμελιώδη διαφοροποίηση από το GCN και GAT. Αντί να εφαρμόσει normalized sum ή attention, το SAGE συγκεντρώνει τα χαρακτηριστικά των γειτόνων μέσω ενός μηχανισμού pooling (π.χ., max ή mean), και ακολούθως προβάλλει τα συγκεντρωμένα χαρακτηριστικά σε νέο χώρο. Το συγκεκριμένο μοντέλο χρησιμοποίησε "max" ως aggr strategy και ενεργοποίησε την επιλογή project=True ώστε η έξοδος να ανήκει στον ίδιο χώρο με το hidden. Η αρχιτεκτονική αυτή έχει το προνόμιο της ανθεκτικότητας σε τοπικό θόρυβο και είναι ικανή να λειτουργεί σε γράφους με δυναμικό αριθμό γειτόνων [57].



Εικόνα 9: Ο σκελετός της αρχιτεκτονικής των μοντέλων που χρησιμοποιήθηκαν, όπου N είναι ο αριθμός των κόμβων του γράφου, In ο αριθμός των διαστάσεων κάθε κόμβου, H τα hidden features και L ο αριθμός των layers.

2.3 Στρατηγική Εκπαίδευσης

2.3.1 Αλγόριθμος Εκπαίδευσης

Η εκπαίδευση των μοντέλων μπορεί να διαχωριστεί σε δύο βασικά στάδια. Αρχικά, αφού χωριστεί το dataset σε train και test sets (20% του συνολικού dataset), εφαρμόζεται 5-fold cross-validation πάνω στο σύνολο εκπαίδευσης. Στόχος της διαδικασίας είναι να προσδιοριστεί το βέλτιστο κατώφλι ταξινόμησης, πάνω από το οποίο ένα supervoxel χαρακτηρίζεται ως βλάβη (lesion), και κάτω από το οποίο χαρακτηρίζεται ως μη-βλάβη (non-lesion). Το βέλτιστο κατώφλι προκύπτει από την εκπαίδευση του μοντέλου για 100 εποχές σε κάθε fold και επιλέγεται αυτό που βελτιστοποιεί το F1 score πάνω στα validation δεδομένα κάθε fold. Για την εκπαίδευση αυτή εφαρμόστηκε ReduceLROnPlateau scheduler και early stopping μετά από 20 εποχές μη αύξησης του F1 [60], [61].

Αφού προσδιοριστεί το κατώφλι, το επιλεγμένο μοντέλο επανεκπαιδεύεται πάνω στο σύνολο train + validation (δηλαδή σε όλους τους ασθενείς πλην του test set), για 200 εποχές. Η εκπαίδευση χρησιμοποιεί early stopping σε περίπτωση μηδενικού F1 για 70 συνεχόμενες εποχές, και χρησιμοποιεί CosineAnnealingLR scheduler. Τέλος, η επίδοση του μοντέλου αξιολογείται με την εξαγωγή μετρικών πάνω στο test set. Η διαδικασία αυτή εξασφαλίζει τόσο βέλτιστη επιλογή των supervoxels που αντιστοιχούν σε βλάβες όσο και την εκπαίδευση του μοντέλου σε όσο το δυνατόν περισσότερα δεδομένα.

2.3.2 Συνάρτηση Απώλειας

Για την επίλυση του προβλήματος δυαδικής ταξινόμησης των κόμβων των γράφων (supervoxels) σε «βλάβη» και «μη βλάβη», χρησιμοποιήθηκε η συνάρτηση **CrossEntropyLoss** [58], [59] του PyTorch. Η επιλογή αυτή καθιστά εφικτή τη χρήση δύο εξόδων (logits) από το μοντέλο, μία για κάθε κλάση, χωρίς την ανάγκη εφαρμογής sigmoid στη τελική στρώση. Η CrossEntropyLoss συνδυάζει εσωτερικά τη λογιστική συνάρτηση ενεργοποίησης Softmax με την αρνητική log-likelihood, και ενδείκνυται για πολυταξινόμηση με one-hot ετικέτες.

Η απώλεια υπολογίζεται ως εξής για κάθε κόμβο:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \log \left(\frac{e^{y_{i,c_i}}}{\sum_j e^{y_{i,j}}} \right) \quad (\text{Eq. 18})$$

όπου:

- $\hat{y}_{i,j}$ είναι το logit της κλάσης j για τον κόμβο i ,
- $c_i \in \{0,1\}$ είναι η ground truth κλάση του κόμβου i ,
- N ο συνολικός αριθμός κόμβων του batch.

Δεδομένης της **ανισορροπίας των ετικετών** (πολύ λιγότεροι κόμβοι αντιστοιχούν σε βλάβες), εφαρμόστηκε weighting ανά κλάση μέσω της παραμέτρου weight της CrossEntropyLoss, δίνοντας αυξημένο βάρος στην θετική κλάση. Αυτό επιτρέπει στο μοντέλο να μην παραβλέπει την κλάση των lesions, η οποία στην περίπτωσή μας είναι πολύ πιο σπάνια από την υγιή κλάση.

3 Πειράματα

3.1 Διαχείριση dataset

3.1.1 Επισκόπηση του συνόλου δεδομένων

Τα δεδομένα της παρούσας μελέτης συλλέχθηκαν αναδρομικά από εξετάσεις Whole-Body MRI (WBMRI) ασθενών με Πολλαπλό Μυέλωμα [62]. Για τους σκοπούς της παρούσας εργασίας επιλέχθηκαν δύο βασικές ακολουθίες: η οβελιαία T1-weighted turbo spin-echo (T1W-TSE) και η Short Tau Inversion Recovery (STIR), καθώς η τελευταία θεωρείται η καταλληλότερη μορφολογική ακολουθία για την εξαγωγή ραδιομικών χαρακτηριστικών.

Αποκλείστηκαν από το σύνολο δεδομένων όλες οι εξετάσεις που περιείχαν τεχνικά σφάλματα ή αλλοιώσεις, όπως αυτές που προκαλούνται από σπονδυλοδεσία ή θερμικό θόρυβο [63]. Το τελικό σύνολο δεδομένων που χρησιμοποιήθηκε για την ανάπτυξη των μοντέλων αποτελείται από **37 εξετάσεις**.

Οι απεικονιστικές παράμετροι των χρησιμοποιούμενων ακολουθιών συνοψίζονται στον **Πίνακα 1**:

Πίνακας 1. Απεικονιστικές παράμετροι για τις ακολουθίες T1W-TSE και STIR

Παράμετρος	T1W-TSE	T2W-STIR
Επίπεδο λήψης	Sagittal	Sagittal
Πεδίο θέασης (mm)	763	764
Μήτρα	296 × 240	280 × 203
Πάχος τομής (mm)	4	4
Κενό μεταξύ τομών (mm)	0.4	0.4
Αριθμός τομών	20	20
TR (ms)	440	4043
TE (ms)	16	70
TSE Factor	7	28
TI (ms)	-	170
Τεχνική Καταστολής Λίπους	-	STIR

Voxel size (mm)	4.4 × 0.4 × 0.4	4.4 × 0.7 × 0.7
Μέση Ανάλυση Εικόνας	718 × 1919 × 19	412 × 1111 × 19

Όλες οι εξετάσεις MRI πραγματοποιήθηκαν στην Α' Ακτινολογική Κλινική του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών στο Νοσοκομείο «Αρεταίειο», χρησιμοποιώντας τον ίδιο μαγνητικό τομογράφο **1.5T Achieva MRI** (Philips Healthcare, Best, Ολλανδία). Η μελέτη εγκρίθηκε από την Επιτροπή Ηθικής και Δεοντολογίας του νοσοκομείου.

3.1.2 Δημιουργία των ground truth ετικετών

Όλες οι ακολουθίες που καθορίζονται από το πρωτόκολλο **MY-RADS** χρησιμοποιήθηκαν για τη δημιουργία των ground truth annotations, σύμφωνα με τα κριτήρια του MY-RADS για τις εστιακές βλάβες, ενώ η T1W-TSE χρησιμοποιήθηκε ειδικά για τη διαδικασία της σημείωσης των βλαβών.

Η ανατομική περιοχή του μυελού των οστών (με εξαίρεση το φλοιώδες οστό) απομονώθηκε και σημειώθηκε χειροκίνητα σε κάθε τομή (slice-by-slice), με βάση τις οβελιαίες εικόνες Whole-Body T1W-TSE. Η χειροκίνητη τμηματοποίηση όλων των εστιακών βλαβών του πολλαπλού μυελώματος εντός του προκαθορισμένου χώρου μυελού πραγματοποιήθηκε από τρεις ειδικευόμενους ακτινολόγους, οι οποίοι έλαβαν ειδικές οδηγίες και εκπαίδευση για τη συγκεκριμένη εργασία από δύο εξειδικευμένους ακτινολόγους με εμπειρία 31 και 14 ετών, αντίστοιχα, στην ερμηνεία μαγνητικών τομογραφιών του μυελού των οστών.

Ο χώρος του μυελού των οστών ορίστηκε να περιλαμβάνει:

- τα σπονδυλικά σώματα των αυχενικών σπονδύλων **C2–C7**,
- των θωρακικών σπονδύλων **T1–T12**,
- των οσφυϊκών σπονδύλων **L1–L5**,
- καθώς και των **ιεροκοκκυγικών** σπονδύλων.

Ο άτλας (C1) εξαιρέθηκε από τις μάσκες. Οι αρχικές σημειώσεις επανεξετάστηκαν μετά την ολοκλήρωσή τους από τους δύο έμπειρους ακτινολόγους και αναθεωρήθηκαν όπου κρίθηκε απαραίτητο.

3.1.3 Προεπεξεργασία των εικόνων

Οι εικόνες του συνόλου δεδομένων υποβλήθηκαν στα ακόλουθα βήματα προεπεξεργασίας. Η διαδικασία ξεκινά με την επαναδειγματοληψία (resampling) όλων των εικόνων σε ιστροπική ανάλυση $1 \times 1 \times 1 \text{ mm}^3$, προκειμένου να διασφαλιστεί κοινή κλίμακα σε όλους τους άξονες. Κεντρικό σημείο αναφοράς είναι η εικόνα T1, η οποία ορίζεται ως πρότυπο γεωμετρίας. Όλες οι υπόλοιπες εικόνες, δηλαδή η T2 ακολουθία και οι μάσκες της σπονδυλικής στήλης και των βλαβών, ευθυγραμμίζονται χωρικά με την εικόνα T1 μέσω της συνάρτησης Resample της βιβλιοθήκης SimpleITK, διατηρώντας το ίδιο μέγεθος voxel και τις ίδιες γεωμετρικές παραμέτρους (origin, direction, spacing). Για τις συνεχείς εικόνες (T2) εφαρμόζεται γραμμική παρεμβολή (linear interpolation), ενώ για τις μάσκες χρησιμοποιείται πλησιέστερος γείτονας (nearest neighbor) ώστε να μην προκύψουν ενδιάμεσες τιμές που θα αλλοίωναν τη δυαδική φύση τους.

Αμέσως μετά εφαρμόζεται **διόρθωση πεδίου ανομοιογένειας** (Bias Field Correction) με τον αλγόριθμο **N4ITK** [64], [65], με στόχο την αντιστάθμιση των μεταβολών έντασης που οφείλονται σε

παραμορφώσεις του μαγνητικού πεδίου ή ανομοιογένειες στην RF διέγερση. Η διαδικασία αυτή βασίζεται σε μάσκα ιστού που εξάγεται με **Otsu thresholding**, επιτρέποντας στο bias field να εκτιμηθεί μόνο εντός του σχετικού σημασιολογικού χώρου (δηλ. εντός της περιοχής του σώματος και του μυελού).

Στο επόμενο στάδιο πραγματοποιείται **εξομάλυνση εικόνας** με τη μέθοδο ανισοτροπικής διάχυσης (anisotropic diffusion filtering) [66], ώστε να μειωθεί ο θόρυβος υψηλής συχνότητας ενώ διατηρούνται τα ανατομικά όρια, όπως αυτά μεταξύ φλοιού και μυελού των σπονδύλων. Ακολουθεί **κανονικοποίηση της έντασης** με χρήση z-score normalization, η οποία εφαρμόζεται τοπικά εντός της μάσκας. Οι τιμές έντασης εντός της περιοχής ενδιαφέροντος κανονικοποιούνται ως:

$$x_{norm} = \frac{x - \mu}{\sigma} \quad (\text{Eq. 19})$$

όπου μ και σ είναι η μέση τιμή και η τυπική απόκλιση των εντάσεων των voxels εντός της μάσκας. Με αυτόν τον τρόπο, εξαλείφεται η επίδραση των απόλυτων εντάσεων και διευκολύνεται η γενίκευση των ραδιομικών χαρακτηριστικών μεταξύ ασθενών.

Τέλος, γίνεται **ανάλυση συνδεδεμένων περιοχών** (connected components) στη δυαδική μάσκα της σπονδυλικής στήλης, με στόχο την απομόνωση και ανασήμανση (relabeling) των σπονδύλων [67]. Κάθε περιοχή-υποψήφιος σπόνδυλος φιλτράρεται βάσει όγκου, αποκλείοντας πολύ μικρές ή πολύ μεγάλες περιοχές που πιθανώς αντιστοιχούν σε θόρυβο ή στο παρασκήνιο [68]. Οι αποδεκτές συνιστώσες ταξινομούνται κατά μήκος του άξονα Y (sagittal scan), και κάθε μία λαμβάνει νέο μοναδικό αναγνωριστικό βάσει της θέσης της, διασφαλίζοντας έτσι αρίθμηση σπονδύλων και αντιστοίχισή τους με τις πραγματικές ονομασίες τους (L1, T7 etc)

Μετά την ολοκλήρωση της προεπεξεργασίας, το σύνολο δεδομένων χωρίστηκε σε δύο υποσύνολα: εκπαίδευσης (80%) και δοκιμής (20%), αποδίδοντας 8 ασθενείς για επαλήθευση των αποτελεσμάτων. Στις υπόλοιπες περιπτώσεις εφαρμόστηκε 5-fold Cross Validation, προκειμένου να αξιοποιηθεί καλύτερα το ελλιπές σύνολο δεδομένων.

3.2 Λεπτομέρειες υλοποίησης

Η υλοποίηση του πειραματικού pipeline πραγματοποιήθηκε εξ ολοκλήρου σε Python, με χρήση της βιβλιοθήκης PyTorch (v2.7.1) [69] για την εκπαίδευση των GNN μοντέλων και της PyTorch Geometric (v2.6.1) [70] για τη διαχείριση των γράφων.

Η εκπαίδευση πραγματοποιήθηκε σε υπολογιστικό σύστημα με:

- **OS:** Ubuntu 22.04
- **GPU:** NVIDIA RTX 4080 (16 GB VRAM)
- **CPU:** Intel Core i5-12400F
- **RAM:** 64 GB
- **CUDA version:** έκδοση 12.6 [71]

Στα πειράματα χρησιμοποιείται weighted CrossEntropyLoss, AdamW optimizer με αρχικό learning rate $1e-3$ και weight decay $1e-4$. Για την αποφυγή overfitting και την ενίσχυση του ντετερμινισμού,

ορίζεται σταθερό seed (42) σε όλα τα υποσυστήματα (NumPy, Python, PyTorch) και ενεργοποιείται deterministic mode της PyTorch. Επίσης γίνεται χρήση gradient clipping με μέγιστο norm 5.0

Όλα τα ενδιάμεσα αποτελέσματα, logs και checkpoints αποθηκεύονται σε δομημένους φακέλους ανά πείραμα, ώστε να διασφαλίζεται η ευκολία παρακολούθησης, αξιολόγησης και αναπαραγωγής των αποτελεσμάτων.

3.3 Αξιολόγηση

3.3.1 Ποσοτική αξιολόγηση

Η αξιολόγηση της μεθοδολογίας πραγματοποιείται σε τρία διακριτά επίπεδα, αντισταθμίζοντας διαφορετικές πτυχές του προβλήματος: (α) **voxel-wise** ταξινόμηση [72], [73] (σε επίπεδο supervoxels/voxels), (β) **overlap/boundary-wise** αποτίμηση μασκών τμηματοποίησης [74], [75] (ογκομετρική επικάλυψη και γεωμετρική εγγύτητα), και (γ) **lesion-wise** αξιολόγηση σε επίπεδο μεμονωμένων βλαβών (instances). Όλες οι αποστάσεις υπολογίζονται σε χιλιοστά (mm), με χρήση του *image spacing*. Πριν από τον υπολογισμό των μετρικών, οι μάσκες αναφοράς ευθυγραμμίστηκαν στη γεωμετρία των μασκών πρόβλεψης (ίδιο μέγεθος, origin, direction) μέσω nearest-neighbor επαναδειγματοληψίας, ώστε να διασφαλίζεται ισομορφία ετικετών.

A. Voxel-wise Μετρικές (Ταξινόμηση)

Για ένα κατώφλι $\tau \in [0,1]$, παράγονται δυαδικές προβλέψεις και υπολογίζονται:

$$Precision = \frac{TP}{TP+FP} \quad (\text{Eq. 20})$$

$$Recall = \frac{TP}{TP+FN} \quad (\text{Eq. 21})$$

$$Specificity = \frac{TN}{TN+FP} \quad (\text{Eq. 22})$$

όπου:

- **TP** (True Positives): αριθμός των voxels που ανήκουν σε βλάβες και προβλέφθηκαν σωστά ως βλάβες.
- **FP** (False Positives): αριθμός των voxels που προβλέφθηκαν ως βλάβες, ενώ στην πραγματικότητα δεν είναι.
- **TN** (True Negatives): αριθμός των voxels που ανήκουν σε φυσιολογικό ιστό και προβλέφθηκαν σωστά ως φυσιολογικά.

Από τα παραπάνω προκύπτει και:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (\text{Eq. 23})$$

Επιπλέον, αναφέρουμε δύο επιπλέον μετρικές ανεξάρτητες κατωφλίου, την AUROC (εμβαδό κάτω από ROC) και την AUPRC (εμβαδό κάτω από διάγραμμα Precision–Recall). Ως κύριοι δείκτες προτείνονται οι F1 και AUPRC. Η μετρική Accuracy δεν χρησιμοποιείται ως πρωτεύων δείκτης

λόγω της γνωστής μεροληψίας της σε σπάνιες θετικές κλάσεις, κάτι που δεν ισχύει για τις F1 και AUPRC, οι οποίες συνδυάζουν πληροφορία τόσο από Precision όσο και από Recall.

B. Boundary-wise Metrics

- **Dice Similarity Coefficient (DSC):** Μετρά την ογκομετρική επικάλυψη μεταξύ πρόβλεψης P και αλήθειας T:

$$DSC = \frac{2|T \cap P|}{|T| + |P|} = \frac{2TP}{2TP + FP + FN}, \quad DSC \in [0,1] \quad (\text{Eq. 24})$$

Ο DSC χρησιμοποιείται ως δευτερεύουσα μετρική, καθώς οι μάσκες πρέπει να είναι λεπτομερώς σχεδιασμένες, κάτι που δεν ισχύει κατ' ανάγκη στην περίπτωση μας.

- **95th Percentile Hausdorff Distance (HD95):** Η HD95 ποσοτικοποιεί την «τυπικά χειρότερη» απόκλιση μεταξύ των συνόρων, απομειώνοντας την επίδραση μεμονωμένων ακραίων outliers. Έστω ∂T , ∂P τα σύνορα και

$$d(x, S) = \min_{s \in S} |x - s|_2 \quad (\text{Eq. 25})$$

Τότε

$$HD95 = \text{percentile}_{95} (\{d(t, \partial P): t \in \partial T\} \cup \{d(p, \partial T): p \in \partial P\}) \quad (\text{mm}) \quad (\text{Eq. 26})$$

με μικρότερες τιμές να υποδηλώνουν καλύτερη εγγύτητα συνόρων. Η HD95 συνιστάται ως βασική οριακή μετρική, σε συνδυασμό με άλλες.

- **Average Symmetric Surface Distance (ASSD):** Ο μέσος συμμετρικός όρος των αποστάσεων συνόρων ορίζεται ως:

$$ASSD(T, P) = \frac{1}{|\partial T| + |\partial P|} (\sum_{t \in \partial T} d(t, \partial P) + \sum_{p \in \partial P} d(p, \partial T)) \quad (\text{mm}) \quad (\text{Eq. 27})$$

παρέχοντας συνολική εκτίμηση της μέσης οριακής απόκλισης (μικρότερο είναι καλύτερο).

- **Surface Dice @ τ mm (sDSC@ τ):** Ο Surface Dice με ανοχή τ ορίζεται ως ο μέσος όρος των κλασμάτων σημείων επιφανείας των δύο масκών που βρίσκονται εντός τ mm από το αντίστοιχο σύνορο:

$$sDSC_{\tau} = \frac{1}{2} \left(\frac{|\{p \in \partial P: d(p, \partial T) \leq \tau\}|}{|\partial P|} + \frac{|\{t \in \partial T: d(t, \partial P) \leq \tau\}|}{|\partial T|} \right), \quad sDSC_{\tau} \in [0,1] \quad (\text{Eq. 28})$$

Η μετρική αυτή (τυπικά με $\tau=2$ mm) είναι ανθεκτική σε μικρές χωρικές μετατοπίσεις και έχει άμεση κλινική ερμηνεία.

C. Lesion-wise Metrics(Instances)

Δεδομένου ότι η εφαρμογή επικεντρώνεται στην ανίχνευση/τμηματοποίηση μεμονωμένων βλαβών, αξιολογούμε σε επίπεδο συνδεδετών συνιστωσών (3D instances). Τα instances εξάγονται από τις μάσκες T και P, υπολογίζεται ο πίνακας IoU μεταξύ όλων των ζευγών, και εκτελείται 1-

προς-1 αντιστοίχιση μέσω αλγορίθμου Hungarian που μεγιστοποιεί το συνολικό IoU. Ένα ζεύγος θεωρείται TP αν $\text{IoU} \geq \theta$ (στην πράξη $\theta=0.2$, με αναλύσεις ευαισθησίας σε $\theta \in \{0.1, 0.2\}$).

- **Lesion Precision / Recall / F1:** Οι δείκτες αυτοί απαντούν στο ερώτημα «ανιχνεύθηκαν οι βλάβες;» και είναι λιγότερο ευαίσθητοι σε μικρο-αποκλίσεις περιγράμματος του GT σε σχέση με τις voxel-wise overlap μετρικές. Συνιστώνται ως κύριοι δείκτες σε επίπεδο βλάβης, με παράθεση επιπλέον ανάλυσης ανά κατηγορία μεγέθους βλάβης (μικρές/μεσαίες/μεγάλες).

$$\text{Precision}_l = \frac{TP}{\#Pred} \quad (\text{Eq. 29})$$

$$\text{Recall}_l = \frac{TP}{\#GT} \quad (\text{Eq. 30})$$

όπου:

- **TP (True Positives):** αριθμός σωστά εντοπισμένων βλαβών
- **#Pred:** συνολικός αριθμός προβλεπόμενων βλαβών
- **#GT (Ground Truth lesions):** συνολικός αριθμός πραγματικών βλαβών

Από τα παραπάνω προκύπτει:

$$F1_l = \frac{2 * \text{Precision}_l * \text{Recall}_l}{\text{Precision}_l + \text{Recall}_l} \quad (\text{Eq. 31})$$

- **Panoptic Quality (PQ):** Η Panoptic Quality [76], [77] αποδομείται σε Detection Quality (DQ) και Segmentation Quality (SQ):

$$DQ = \frac{TP}{TP + 0.5FP + 0.5FN} \quad (\text{Eq. 32})$$

$$SQ = \frac{1}{TP} \sum_{(g,p) \in M} \text{IoU}(g, p) \quad (\text{Eq. 33})$$

$$PQ = DQ \times SQ \quad (\text{Eq. 34})$$

όπου M το σύνολο των αντιστοιχισμένων ζευγών με $\text{IoU} \geq \theta$. Η PQ παρέχει συνεκτική περίληψη της ικανότητας ανίχνευσης (DQ) και της ποιότητας σχήματος (SQ).

- **Contact-Threshold Score (CTS):** Για να ληφθεί υπόψη ο θόρυβος/η ατέλεια του GT, χρησιμοποιείται ένας custom δείκτης, ο οποίος είναι ειδικά προσαρμοσμένος στο τρέχον πρόβλημα:

$$CTS = \frac{TP_\theta}{TP_\theta + \alpha U_\theta + \beta FN + \gamma FP} \quad (\text{Eq. 35})$$

όπου TP_θ τα ζεύγη με $\text{IoU} \geq \theta$, U_θ τα «near-miss» ζεύγη με $0 < \text{IoU} < \theta$, ενώ $\alpha, \beta, \gamma > 0$ είναι βάρη (εν προκειμένω $\alpha=0.5, \beta=\gamma=1$). Ο δείκτης CTS επιβραβεύει τις «ορθές» ανιχνεύσεις και τιμωρεί ηπιότερα τις σχεδόν-σωστές αντιστοιχίσεις, προσφέροντας μία σύντομη, ερμηνεύσιμη περίληψη απόδοσης υπό ατελές GT.

Στο πλαίσιο της παρούσας εργασίας, όπου οι ετικέτες αναφοράς (ground truth) παρουσιάζουν αβεβαιότητες και μικρές αποκλίσεις στη χωρική τους ακρίβεια, η αξιολόγηση βασίστηκε σε μετρικές ανθεκτικές σε θόρυβο και μερική ασυμφωνία περιγραμμάτων, οι οποίες αποτυπώνουν πιο αξιόπιστα την πραγματική απόδοση του μοντέλου. Οι πιο σημαντικές από τις παραπάνω μετρικές αποτυπώνονται συνοπτικά παρακάτω:

- **Voxel-wise F1 (↑):** Παρέχει συνολική εκτίμηση της τοπικής ταξινομητικής ακρίβειας σε επίπεδο voxel, αντανakλώντας τη σωστή διάκριση μεταξύ υγιούς και παθολογικού ιστού.
- **Lesion F1-score (↑):** Εκτιμά τη συνολική ικανότητα του μοντέλου να εντοπίζει σωστά τις βλάβες, ισορροπώντας ανάμεσα σε Precision και Recall. Υψηλές τιμές υποδεικνύουν επιτυχή ανίχνευση χωρίς υπερβολικά ψευδώς θετικές προβλέψεις.
- **Lesion Precision (↑):** Αντιπροσωπεύει το ποσοστό των εντοπισμένων βλαβών που είναι πράγματι υπαρκτές. Υψηλή τιμή σημαίνει ότι το μοντέλο αποφεύγει ψευδείς εντοπισμούς και εμφανίζει υψηλή ειδικότητα.
- **Lesion Recall (↑):** Μετρά το ποσοστό των πραγματικών βλαβών που εντοπίζονται από το μοντέλο. Υψηλή τιμή δηλώνει ικανότητα ανίχνευσης των περισσότερων υπαρκτών βλαβών, περιορίζοντας τις παραλείψεις.
- **Contact-Threshold Score (CTS) (↑):** Ενσωματώνει τιμωρίες για ψευδώς θετικές, ψευδώς αρνητικές και μερικώς επικαλυπτόμενες («near-miss») περιπτώσεις, αποδίδοντας πιο δίκαιη εκτίμηση της επίδοσης σε ατελές ground truth.
- **AUPRC (↑):** Αποτιμά τη συνολική ταξινομητική συμπεριφορά του μοντέλου σε όλα τα κατώφλια, λαμβάνοντας υπόψη την ανισορροπία κλάσεων. Υψηλή τιμή δείχνει καλή διακριτική ικανότητα μεταξύ βλάβης και φυσιολογικού ιστού.

3.3.2 Ποιοτική αξιολόγηση

Τα αποτελέσματα της τμηματοποίησης αξιολογήθηκαν ποιοτικά μέσω οπτικής σύγκρισης των προβλέψεων του μοντέλου με τις πραγματικές σημειώσεις (ground truth). Οι αρχικές εικόνες, οι μάσκες αναφοράς και οι προβλέψεις τοποθετήθηκαν πλευρικά, ώστε να εκτιμηθεί η ακρίβεια και η ευθυγράμμιση του μοντέλου στην ανίχνευση και τμηματοποίηση των εστιακών βλαβών. Η ανάλυση αυτή προσέφερε ποιοτικές ενδείξεις για τη σταθερότητα και την κλινική αξιοπιστία του μοντέλου.

3.4 Πειράματα

3.4.1 Επιλογή Πλήθους Χαρακτηριστικών

Στο πρώτο πείραμα διερευνήθηκε η επίδραση του αριθμού χαρακτηριστικών που διατηρούνται μετά από διαδικασία επιλογής μέσω γραμμικού μοντέλου LASSO στην απόδοση των GNN. Στόχος ήταν να προσδιοριστεί το ιδανικό πλήθος χαρακτηριστικών (top-k) που βελτιστοποιεί τη διαχωριστική ικανότητα των μοντέλων ως προς τον εντοπισμό παθολογικών supervoxels.

Η επιλογή των πιο σημαντικών χαρακτηριστικών πραγματοποιήθηκε μέσω λογιστικής παλινδρόμησης με κανονικοποίηση ElasticNet, η οποία συνδυάζει L1 (LASSO) και L2 (Ridge) κανονικοποίηση. Η μέθοδος εφαρμόστηκε μέσω της LogisticRegressionCV της βιβλιοθήκης scikit-learn, με τις ακόλουθες παραμέτρους:

- **Solver:** saga
- **Penalty:** elasticnet
- **l1_ratios:** [0.15, 0.5, 0.85]
- **Cs:** 20 τιμές σε λογαριθμική κλίμακα από $1e-6$ έως $1e2$
- **Συνάρτηση κόστους:** average_precision
- **Cross-validation:** Stratified 5-Fold CV με class_weight="balanced"
- **Μέγιστος αριθμός επαναλήψεων:** 8000

Αφού ολοκληρώθηκε η παραπάνω διαδικασία, επιλέχθηκαν 6 διαφορετικές τιμές της παραμέτρου top_k με βάση τον αριθμό των χαρακτηριστικών που παρέμεναν. Δεδομένου ότι το πείραμα πραγματοποιήθηκε με τους γράφους που προέκυψαν από τμηματοποίηση με ~150 voxels/supervoxel σε radiomic χαρακτηριστικά (βλ. Ενότητα 2.1), ο αριθμός χαρακτηριστικών είναι 176. Οι τιμές που επιλέχθηκαν είναι: **16, 32, 64, 128, 88 και 176**

Για κάθε τιμή του παραμέτρου top_k, εκπαιδεύτηκαν όλοι οι δυνατοί συνδυασμοί μοντέλου (GCN, GAT, GraphSAGE), διαστάσεων ενδιάμεσων χαρακτηριστικών (hidden $\in \{64, 128\}$) και αριθμού επιπέδων (L $\in \{4, 5\}$), ώστε να αξιολογηθεί η γενικευσιμότητα κάθε διαμόρφωσης ως προς το πλήθος επιλεγμένων χαρακτηριστικών.

3.4.2 Αξιολόγηση Μοντέλων

Μετά την επιλογή των βέλτιστων χαρακτηριστικών από το προηγούμενο πείραμα σχεδιάστηκε ένα δεύτερο πείραμα στο οποίο αξιολογήθηκαν οι επιδόσεις των μοντέλων σε διαφορετικές εκδοχές γράφων. Χρησιμοποιήθηκε το βέλτιστο υποσύνολο χαρακτηριστικών (δηλαδή η τιμή top_k που απέδωσε τα καλύτερα αποτελέσματα) και εξετάστηκαν οι περιπτώσεις που αναλύθηκαν στην Ενότητα 2.1.2.

Για κάθε εκδοχή γράφου, εκπαιδεύτηκαν όλες οι αρχιτεκτονικές (GCN, GAT, GraphSAGE) δοκιμάζοντας όλους τους συνδυασμούς υπερπαραμέτρων που προέκυψαν από τις τιμές hidden $\in \{64, 128\}$ και layers $\in \{4, 5\}$. Το συνολικό πειραματικό πλάνο εξασφάλισε συγκρίσιμα αποτελέσματα σε όλο το φάσμα των διαφορετικών γραφικών αναπαραστάσεων.

Ως test set επιλέχθηκαν οι εικόνες 5, 8, 11, 20, 22, 27 και 28, οι οποίες είχαν σχολιαστεί με ιδιαίτερη προσοχή από έμπειρη καθηγήτρια ακτινολογίας, εξασφαλίζοντας αξιόπιστη και ποιοτική ετικετοδότηση (ground truth) για τον υπολογισμό της γενίκευσης των μοντέλων σε νέους ασθενείς.

Για λόγους σαφούς οργάνωσης και τεκμηρίωσης των πειραμάτων, κάθε εκδοχή γράφου κωδικοποιήθηκε με συγκεκριμένη ονομασία φακέλου ως εξής:

- 1000 Supervoxels / Intensity Quintiles → Case 1
- 1000 Supervoxels / Radiomic χαρακτηριστικά → Case 2
- 150 mm³ Supervoxels / Intensity Quintiles → Case 3
- 150 mm³ Supervoxels / Radiomic χαρακτηριστικά → Case 4
- 150 mm³ Supervoxels –/Radiomic + Intensity Quintiles → Case 5

Η συγκεκριμένη κωδικοποίηση χρησιμοποιήθηκε τόσο κατά την αποθήκευση των πειραματικών αρχείων όσο και κατά την ανάλυση και οπτικοποίηση των αποτελεσμάτων.

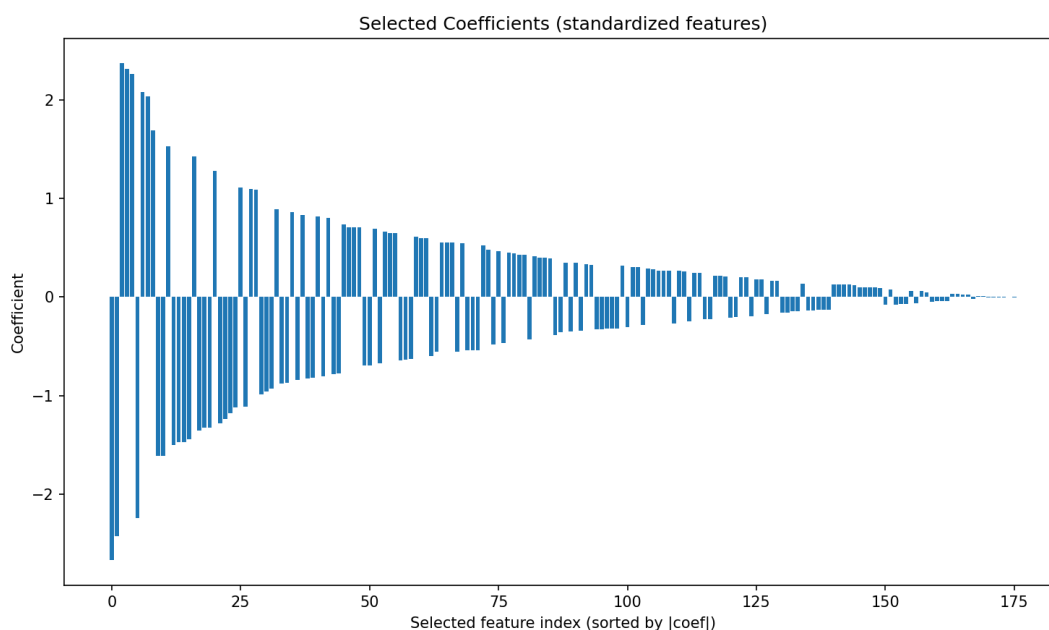
3.5 Στατιστική ανάλυση

Και για τη σύγκριση των μοντέλων, όπως περιγράφεται στην Ενότητα 3.4.2, πραγματοποιήθηκε στατιστική ανάλυση των αποτελεσμάτων με κύρια μετρική τη Lesion F1 (Eq. 31) για την αξιολόγηση της ακρίβειας της τμηματοποίησης. Υπολογίστηκαν περιγραφικά στατιστικά, όπως ο μέσος όρος και η τυπική απόκλιση, για την απόδοση κάθε μοντέλου. Η σύγκριση της απόδοσης των μοντέλων πραγματοποιήθηκε με ανάλυση διακύμανσης επαναλαμβανόμενων μετρήσεων (repeated measures ANOVA), με επίπεδο σημαντικότητας 0.05, υπό την προϋπόθεση ότι ικανοποιούνται οι απαιτήσεις εφαρμογής της. Σε περίπτωση που οι παραπάνω προϋποθέσεις δεν πληρούνται, η ανάλυση πραγματοποιήθηκε με τη μη παραμετρική δοκιμασία Friedman, δεδομένου ότι η ανάλυση έγινε σε μη ανεξάρτητα δεδομένα.

4 Αποτελέσματα

4.1 Επιλογή Πλήθους Χαρακτηριστικών

Στο πρώτο πείραμα μελετήθηκε η επίδραση του πλήθους χαρακτηριστικών k που διατηρούνται μετά την επιλογή μέσω Logistic Regression με ElasticNet (CV). Στο αρχικό σύνολο (116 991 δείγματα, 176 χαρακτηριστικά) το variance-thresholding δεν απέκλεισε χαρακτηριστικά, ενώ η βέλτιστη ρύθμιση του ElasticNet ($C=37.93$, $l1_ratio=0.15$) παρήγαγε 176 μη μηδενικούς συντελεστές (Εικόνα 10). Συνεπώς το top_k ορίστηκε με βάση την κατάταξη των coef σε απόλυτη τιμή. Αξιολογήθηκαν διαδοχικές τιμές ($k \in \{16, 32, 64, 88, 128, 176\}$) σε όλους τους συνδυασμούς αρχιτεκτονικής (GCN, GAT, GraphSAGE), διαστάσεων (64, 128) και βάθους (4, 5). Για την παρακάτω ανάλυση χρησιμοποιήθηκαν οι πιο περιγραφικές μετρικές, δηλαδή οι F1 (Eq. 23), AUPRC, lesion_F1 (Eq. 31), lesion_precision (Eq. 29), lesion_recall (Eq. 30) και CTS (Eq. 35).



Εικόνα 10: Απεικόνιση των συντελεστών του κάθε χαρακτηριστικού (ταξινομημένα κατά απόλυτη τιμή) μετά την εξαγωγή των αποτελεσμάτων από το LASSO

Σε μέσο όρο όλων των διαμορφώσεων, οι επιδόσεις βελτιώνονται μονοτονικά όσο αυξάνει το k . Η καλύτερη συνολικά απόδοση καταγράφηκε στο $k=176$ με $F1 = 0.572$ (SD = 0.193, 95% CI: 0.463–0.681), $AUPRC = 0.689$ (SD = 0.236, 95% CI: 0.556–0.823), $\text{lesion F1} = 0.219$ (SD = 0.089, 95% CI: 0.169–0.270), $\text{lesion precision} = 0.321$ (SD = 0.236, 95% CI: 0.187–0.454), $\text{lesion recall} = 0.573$ (SD = 0.187, 95% CI: 0.467–0.678) και $CTS = 0.195$ (SD = 0.108, 95% CI: 0.134–0.257). Στον αντίποδα, για ($k=16$) οι επιδόσεις ήταν οριακά πάνω από τυχαίο με μέσους όρους: $F1 = 0.027$, $AUPRC = 0.057$, $\text{lesion F1} = 0.021$ και $CTS = 0.014$.

Πίνακας 2. Αποτελέσματα (mean $\pm$ sd) της απόδοσης των μοντέλων σε σχέση με την τιμή top k

k	F1	AUPRC	lesion_F1	lesion_precision	lesion_recall	CTS
16	0.027 $\pm$ 0.027	0.057 $\pm$ 0.000	0.763 $\pm$ 0.237	0.104 $\pm$ 0.104	0.021 $\pm$ 0.021	0.014 $\pm$ 0.014
32	0.045 $\pm$ 0.027	0.076 $\pm$ 0.017	0.604 $\pm$ 0.237	0.174 $\pm$ 0.104	0.035 $\pm$ 0.021	0.023 $\pm$ 0.014
64	0.137 $\pm$ 0.083	0.283 $\pm$ 0.136	0.525 $\pm$ 0.241	0.289 $\pm$ 0.164	0.043 $\pm$ 0.025	0.026 $\pm$ 0.016
88	0.299 $\pm$ 0.124	0.462 $\pm$ 0.184	0.410 $\pm$ 0.227	0.440 $\pm$ 0.169	0.118 $\pm$ 0.056	0.088 $\pm$ 0.047
128	0.443 $\pm$ 0.150	0.625 $\pm$ 0.165	0.430 $\pm$ 0.192	0.498 $\pm$ 0.166	0.190 $\pm$ 0.068	0.167 $\pm$ 0.066
176	0.572 $\pm$ 0.099	0.689 $\pm$ 0.122	0.321 $\pm$ 0.125	0.573 $\pm$ 0.093	0.219 $\pm$ 0.048	0.195 $\pm$ 0.058

Σε επίπεδο μεμονωμένων διαμορφώσεων, το υψηλότερο F1 επιτεύχθηκε με αρχιτεκτονική GraphSAGE (128 $\times$ 4) για τιμή $k=176$: $F1 = 0.705$, $AUPRC = 0.802$, $\text{lesion precision} = 0.365$, $\text{lesion recall} = 0.597$, $\text{lesion F1} = 0.319$ και $CTS = 0.320$. Η μέγιστη AUPRC παρατηρήθηκε στο $k=88$ με GraphSAGE (64 $\times$ 5): $AUPRC = 0.808$ (με $F1 = 0.440$, $\text{lesion precision} = 0.129$, $\text{lesion recall} = 0.681$, $\text{lesion F1} = 0.204$, $CTS = 0.142$), υποδεικνύοντας ιδιαίτερα καλή κατάταξη και υψηλή ευαισθησία σε επίπεδο βλαβών, αλλά χαμηλότερη ισορροπία precision–recall στη δυαδική απόφαση.

Πίνακας 3. Μέγιστες αποδόσεις μοντέλων ανά top k

k	Αρχιτεκτονική	H x L	F1	AUPRC	Lesion F1	Lesion Precision	Lesion Recall	CTS
k16	GCN	128 x 4	0.108	0.057	0.084	0.050	0.416	0.054
k32	GCN	64 x 4	0.108	0.057	0.084	0.050	0.416	0.054
k64	GCN	64 x 5	0.382	0.696	0.126	0.076	0.659	0.082
k88	SAGE	64 x 4	0.612	0.798	0.286	0.211	0.638	0.268
k128	SAGE	64 x 4	0.668	0.792	0.324	0.388	0.597	0.320
k176	SAGE	64 x 5	0.689	0.787	0.329	0.374	0.597	0.332

4.2 Αξιολόγηση Μοντέλων

Μετά την πραγματοποίηση του προηγούμενου πειράματος, πραγματοποιήθηκε το δεύτερο πείραμα, στο οποίο αξιολογήθηκαν οι επιδόσεις των GNN μοντέλων σε διαφορετικές εκδοχές γράφων. Κάθε εκδοχή αντιστοιχούσε σε διαφορετικό πλήθος supervoxels και τύπο χαρακτηριστικών (Radiomic, Intensity ή συνδυασμό αυτών), όπως περιγράφηκε στην Ενότητα 2.1.2. Για κάθε περίπτωση εκπαιδεύτηκαν οι αρχιτεκτονικές GCN, GAT και GraphSAGE, σε όλους τους συνδυασμούς υπερπαραμέτρων ($\text{hidden} \in \{64, 128\}$, $\text{layers} \in \{4, 5\}$), με σταθερό test set. Όσον αφορά τα features

των γράφων επιλέχθηκε σε κάθε περίπτωση να μην αφαιρεθεί κανένα, καθώς αυτό υπέδειξε το προηγούμενο πείραμα.

Η αξιολόγηση πραγματοποιήθηκε και εδώ με βάση τις μετρικές F1 (Eq. 23), AUPRC, lesion_F1 (Eq. 31), lesion_precision (Eq. 29), lesion_recall (Eq. 30) και CTS (Eq. 35). Τα αποτελέσματα αναφέρονται για τις πέντε εκδοχές γράφων:

- **Case 1** (1000 supervoxels, Intensity)
- **Case 2** (1000 supervoxels, Radiomic)
- **Case 3** (150 mm³ supervoxels, Intensity)
- **Case 4** (150 mm³ supervoxels, Radiomic)
- **Case 5** (150 mm³ supervoxels, Radiomic + Intensity).

Σε μέσο όρο όλων των διαμορφώσεων, οι επιδόσεις παρουσίασαν σταθερή βελτίωση όσο αυξανόταν η πληροφoρία της γραφικής αναπαράστασης. Η Εκδοχή 5 εμφάνισε την υψηλότερη συνολική απόδοση, με lesion F1 = 0.417 ± 0.044 και CTS = 0.377 ± 0.054 , επιβεβαιώνοντας τη σημαντική βελτίωση τόσο στην ακρίβεια εντοπισμού όσο και στη συνολική συμφωνία των βλαβών. Η Εκδοχή 4 ακολούθησε με lesion F1 = 0.257 ± 0.051 και CTS = 0.228 ± 0.076 , παρουσιάζοντας ικανοποιητική, αν και μειωμένη, επίδοση σε σχέση με την εκδοχή 5. Αντίθετα, οι Εκδοχές 1, 2 και 3 κατέγραψαν πολύ χαμηλές τιμές lesion F1 (≤ 0.113) και CTS (≤ 0.073).

Πίνακας 4. Αποτελέσματα (mean $\pm$ sd) της απόδοσης των μοντέλων σε σχέση με τον τύπο του γράφου

Εκδοχή Γράφου	F1	AUPRC	Lesion F1	Lesion Precision	Lesion Recall	CTS
Case 1	0.471 $\pm$ 0.049	0.554 $\pm$ 0.138	0.113 $\pm$ 0.035	0.067 $\pm$ 0.023	0.657 $\pm$ 0.027	0.073 $\pm$ 0.028
Case 2	0.726 $\pm$ 0.039	0.796 $\pm$ 0.05	0.003 $\pm$ 0.002	0.002 $\pm$ 0.001	0.011 $\pm$ 0.008	0.002 $\pm$ 0.001
Case 3	0.442 $\pm$ 0.057	0.578 $\pm$ 0.174	0.004 $\pm$ 0.0	0.002 $\pm$ 0.0	0.014 $\pm$ 0.0	0.003 $\pm$ 0.0
Case 4	0.636 $\pm$ 0.055	0.786 $\pm$ 0.015	0.257 $\pm$ 0.051	0.27 $\pm$ 0.1	0.633 $\pm$ 0.056	0.228 $\pm$ 0.076
Case 5	0.703 $\pm$ 0.024	0.814 $\pm$ 0.019	0.417 $\pm$ 0.044	0.345 $\pm$ 0.046	0.793 $\pm$ 0.031	0.377 $\pm$ 0.054

Για την αξιολόγηση της ακρίβειας τμηματοποίησης και ικανότητας γενίκευσης των τριών μοντέλων (GCN, GAT και GraphSAGE) ως προς τη μετρική Lesion F1, πραγματοποιήθηκε στατιστική ανάλυση των αποτελεσμάτων. Ο έλεγχος κανονικότητας πραγματοποιήθηκε με το Shapiro–Wilk test, το οποίο έδειξε ότι τα δεδομένα όλων των μοντέλων απέκλιναν σημαντικά από την κανονική κατανομή (GCN: $p=0.0069$, GAT: $p=0.0015$, SAGE: $p=0.0058$). Καθώς οι προϋποθέσεις της κανονικότητας δεν ικανοποιούνταν, επιλέχθηκε η μη παραμετρική προσέγγιση. Το αποτέλεσμα του Friedman test ($p = 0.246$) έδειξε ότι δεν υπάρχουν στατιστικά σημαντικές διαφορές στην απόδοση των μοντέλων, επιβεβαιώνοντας ότι οι παρατηρούμενες διαφοροποιήσεις οφείλονται σε τυχαία μεταβλητότητα και όχι σε ουσιαστική υπεροχή κάποιας αρχιτεκτονικής.

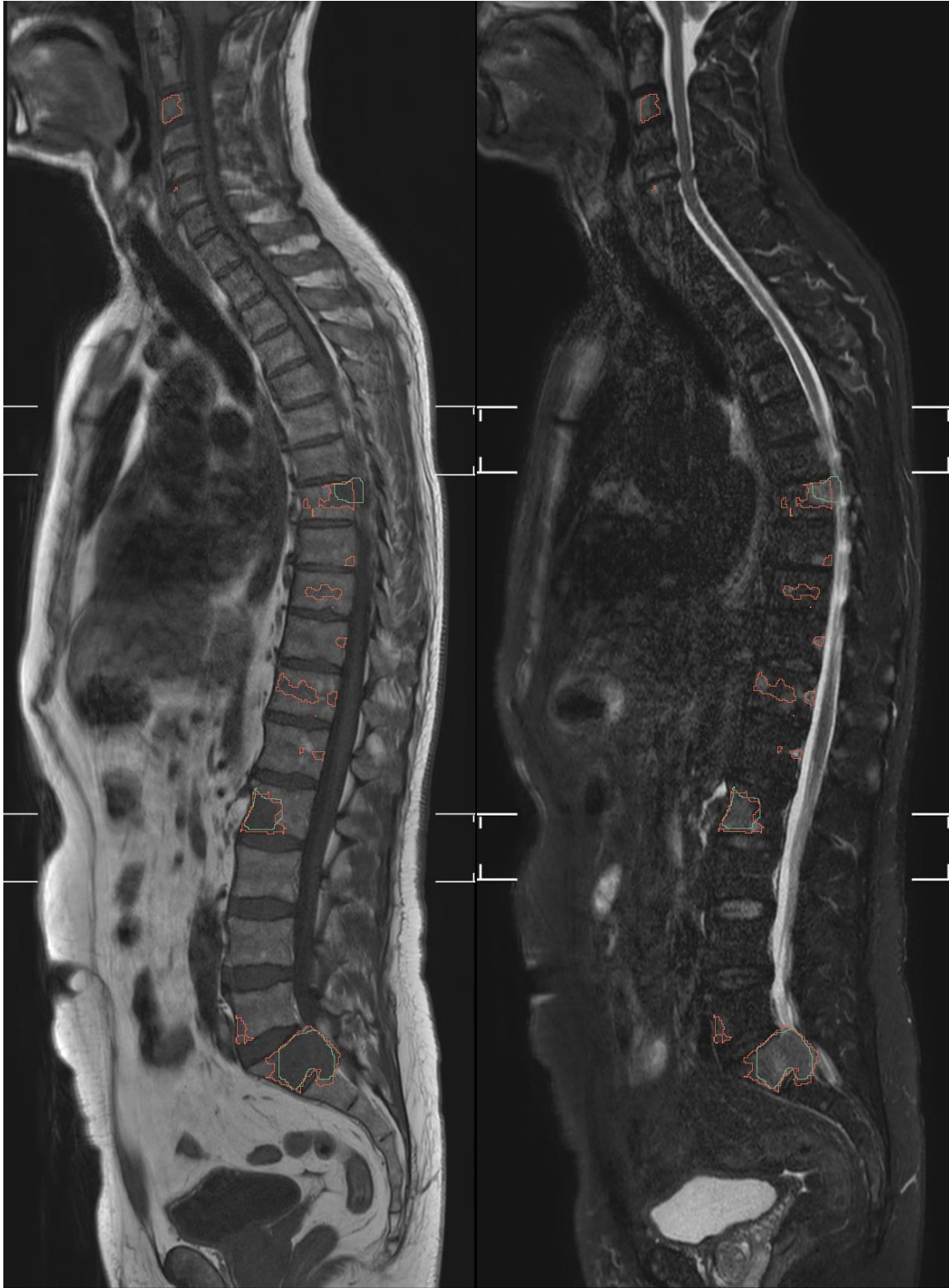
Σε επίπεδο μεμονωμένων διαμορφώσεων, η υψηλότερη απόδοση καταγράφηκε με το GraphSAGE (128×4) στην εκδοχή *Graphs*, παρουσιάζοντας F1 = 0.738, AUPRC = 0.808, lesion precision = 0.419, lesion recall = 0.826, lesion F1 = 0.488 και CTS = 0.420. Παρόμοια υψηλή επίδοση σημείωσε το GCN (128×4) στην ίδια εκδοχή, με F1 = 0.735, AUPRC = 0.796, lesion precision = 0.408, lesion recall = 0.806, lesion F1 = 0.468 και CTS = 0.405. Ελαφρώς χαμηλότερα αποτελέσματα παρατηρήθηκαν για το GAT (64×5), με F1 = 0.718, AUPRC = 0.834, lesion precision = 0.381, lesion recall = 0.778, lesion F1 = 0.451 και CTS = 0.414.

Πίνακας 5. Μέγιστες αποδόσεις μοντέλων ανά κατηγορία αρχιτεκτονικής

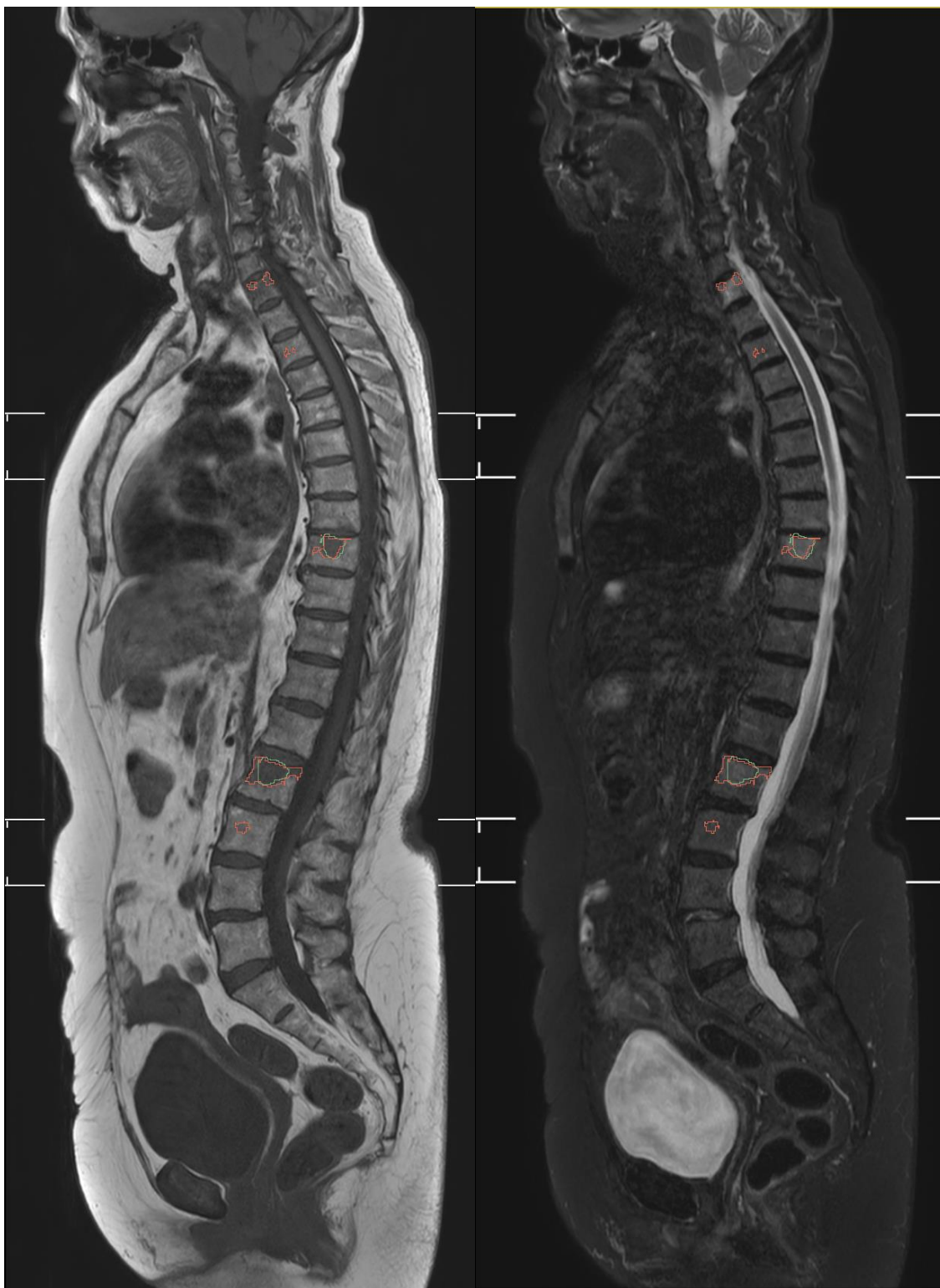
Αρχιτεκτονική	Graph Variant	H x L	F1	AUPRC	Lesion F1	Lesion Precision	Lesion Recall	CTS
GCN	Case 5	128 x 4	0.735	0.796	0.468	0.408	0.805	0.405
GAT	Case 5	64 x 5	0.718	0.834	0.451	0.380	0.778	0.413
SAGE	Case 5	128 x 4	0.738	0.808	0.488	0.419	0.826	0.420

Για τον περαιτέρω έλεγχο της σταθερότητας των μοντέλων για συγκεκριμένες περιπτώσεις ασθενών πραγματοποιήθηκε στατιστική σύγκριση των μοντέλων με την μέγιστη τιμή Lesion F1 καθενός ασθενή. Αρχικά, διαπιστώθηκε ότι τα ζεύγη GCN–GAT ($p = 0.0047$) και GCN–SAGE ($p = 0.0350$) δεν ακολουθούν κανονική κατανομή, ενώ το ζεύγος GAT–SAGE ($p = 0.995$) δεν εμφάνισε απόκλιση. Εφόσον οι προϋποθέσεις κανονικότητας δεν πληρούνται, η ανάλυση πραγματοποιήθηκε με τη μη παραμετρική δοκιμασία Friedman, κατάλληλη για δεδομένα επαναλαμβανόμενων μετρήσεων. Ο έλεγχος Friedman απέδωσε $p = 0.895$, γεγονός που υποδηλώνει την απουσία στατιστικά σημαντικών διαφορών μεταξύ των τριών μοντέλων.

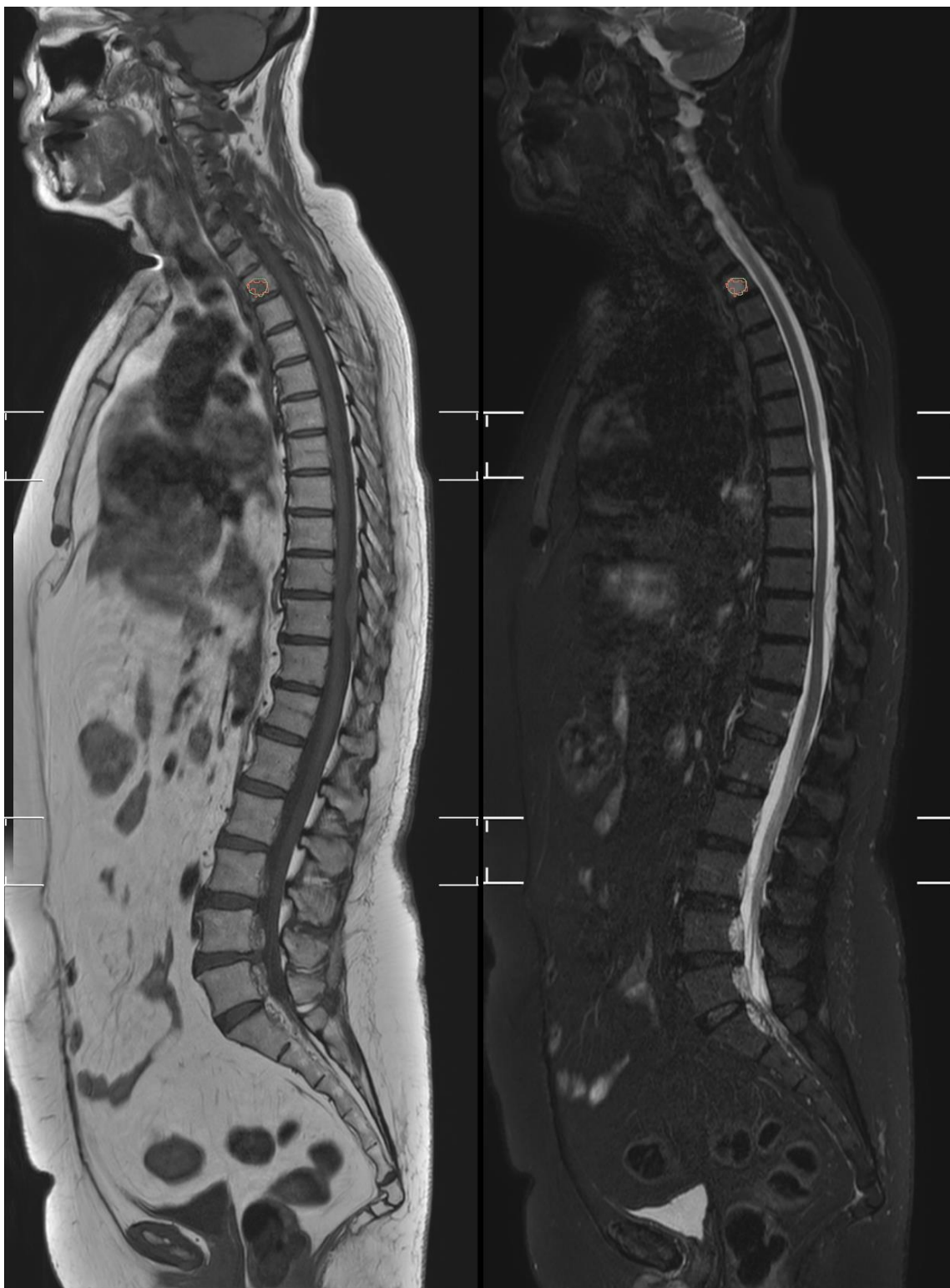
Με βάση το βέλτιστο μοντέλο που προέκυψε από την ανάλυση των αποτελεσμάτων του παραπάνω πίνακα, παρουσιάζονται στη συνέχεια τα αποτελέσματα του μοντέλου για τους ασθενείς του *test set* στις Εικόνες 11-17:



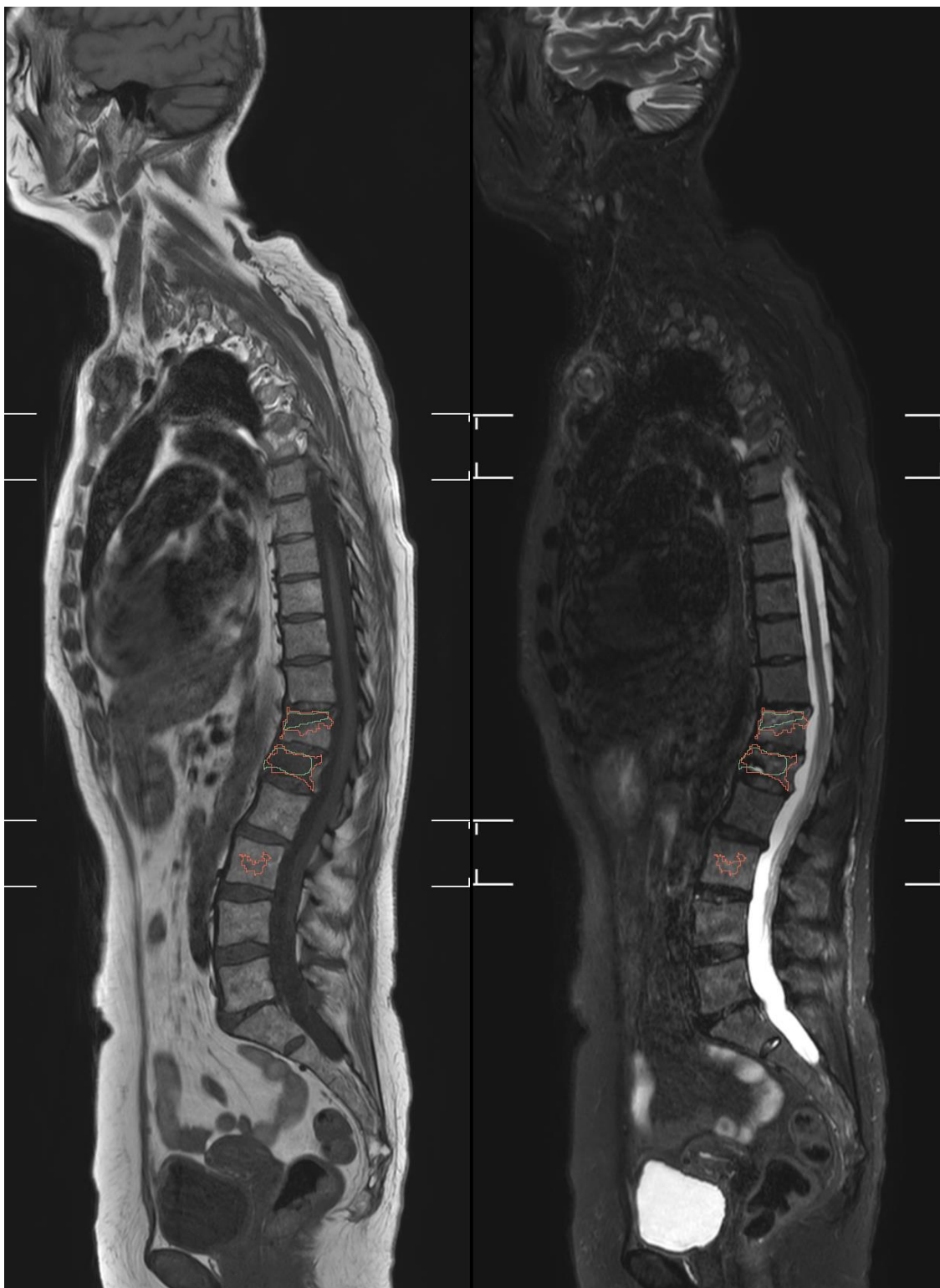
Εικόνα 11: Ασθενής 5. Οι εστιακές αλλοιώσεις είναι σημειωμένες με κόκκινο περιθώριο και το ground truth με πράσινο. Αριστερά φαίνεται η T1 ακολουθία, ενώ δεξιά φαίνεται η T2.



Εικόνα 12: Ασθενής 8. Οι εστιακές αλλοιώσεις είναι σημειωμένες με κόκκινο περιθώριο και το ground truth με πράσινο. Αριστερά φαίνεται η T1 ακολουθία, ενώ δεξιά φαίνεται η T2.



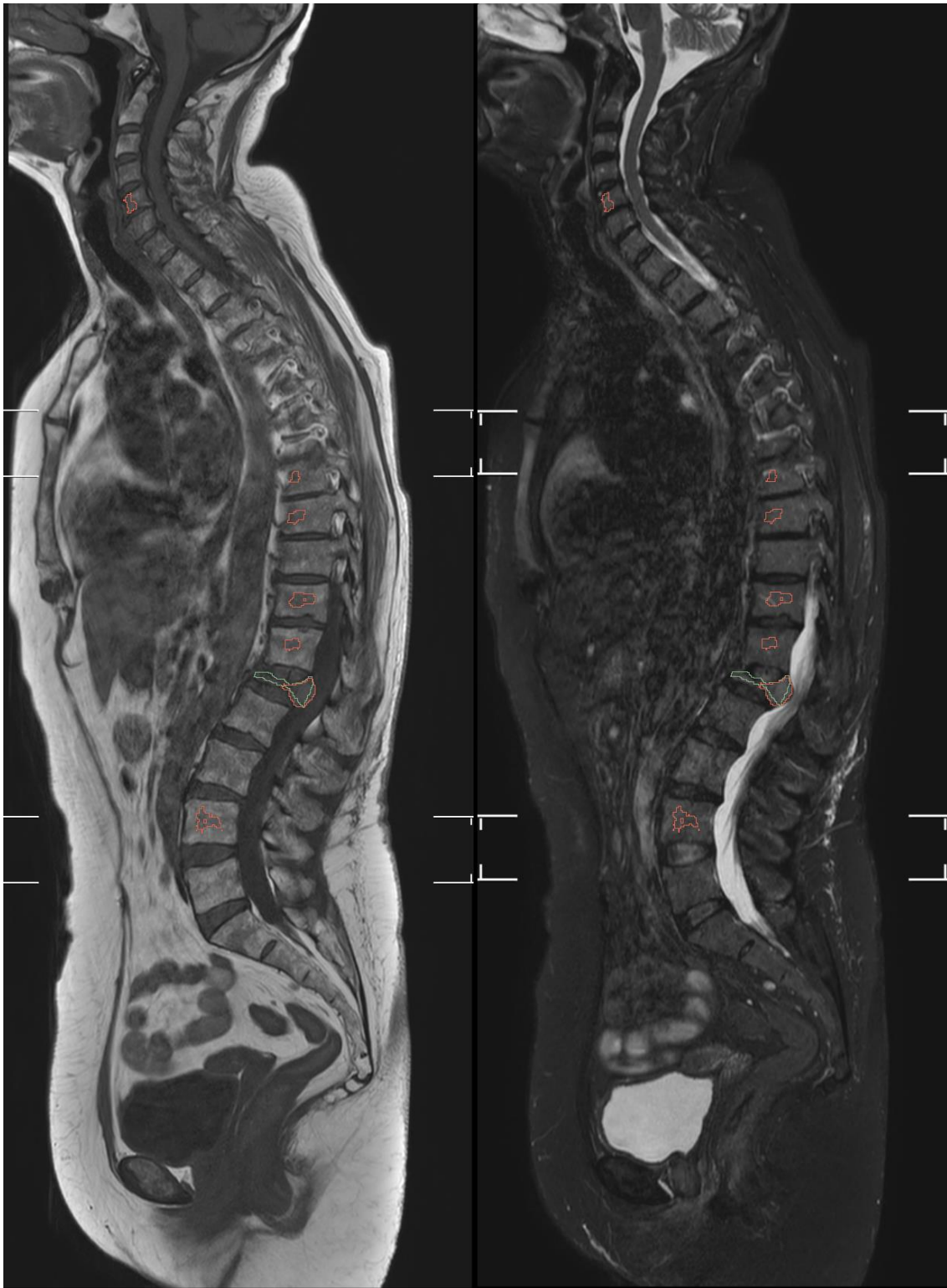
Εικόνα 13: Ασθενής 11. Οι εστιακές αλλοιώσεις είναι σημειωμένες με κόκκινο περιθώριο και το ground truth με πράσινο. Αριστερά φαίνεται η T1 ακολουθία, ενώ δεξιά φαίνεται η T2.



Εικόνα 14: Ασθενής 20. Οι εστιακές αλλοιώσεις είναι σημειωμένες με κόκκινο περιθώριο και το ground truth με πράσινο. Αριστερά φαίνεται η T1 ακολουθία, ενώ δεξιά φαίνεται η T2.



Εικόνα 15: Ασθενής 22. Οι εστιακές αλλοιώσεις είναι σημειωμένες με κόκκινο περιθώριο και το ground truth με πράσινο. Αριστερά φαίνεται η T1 ακολουθία, ενώ δεξιά φαίνεται η T2.



Εικόνα 16: Ασθενής 27. Οι εστιακές αλλοιώσεις είναι σημειωμένες με κόκκινο περιθώριο και το ground truth με πράσινο. Αριστερά φαίνεται η T1 ακολουθία, ενώ δεξιά φαίνεται η T2.



Εικόνα 17: Ασθενής 28. Οι εστιακές αλλοιώσεις είναι σημειωμένες με κόκκινο περιθώριο και το ground truth με πράσινο. Αριστερά φαίνεται η T1 ακολουθία, ενώ δεξιά φαίνεται η T2.

5 Συζήτηση

5.1 Ανάλυση Πειραμάτων και Στατιστικών Αποτελεσμάτων

Η πρώτη φάση των πειραμάτων έδειξε ξεκάθαρα ότι η μείωση του πλήθους χαρακτηριστικών μέσω επιλογής βάσει συντελεστών ElasticNet μπορεί μεν να περιορίσει τη διαστατικότητα, αλλά ενέχει κίνδυνο απώλειας κρίσιμης πληροφορίας. Καθώς το σύνολο δεδομένων περιλαμβάνει σύνθετα και σπάνια μοτίβα αλλοιώσεων, η πλήρης διατήρηση όλων των χαρακτηριστικών αποδείχθηκε καταλληλότερη, πιθανώς λόγω του υψηλού redundancy αλλά και της συμπληρωματικότητας μεταξύ radiomic και intensity παραμέτρων. Ειδικά στις περιπτώσεις με πολύ χαμηλό k , η απόδοση παρέμεινε στα όρια της τυχαιότητας, κάτι που υποδεικνύει ότι η εξαγωγή πληροφορίας για αλλοιώσεις απαιτεί πολυδιάστατη αναπαράσταση. Είναι, επίσης, προφανές ότι οι πρώτοι όροι, όπως φαίνεται και στην Εικόνα 10, έχουν μεγαλύτερη βαρύτητα από τους τελευταίους, το οποίο επιβεβαιώνεται και από το γεγονός ότι στις μικρότερες περιπτώσεις τα αποτελέσματα είναι σχεδόν μηδενικά.

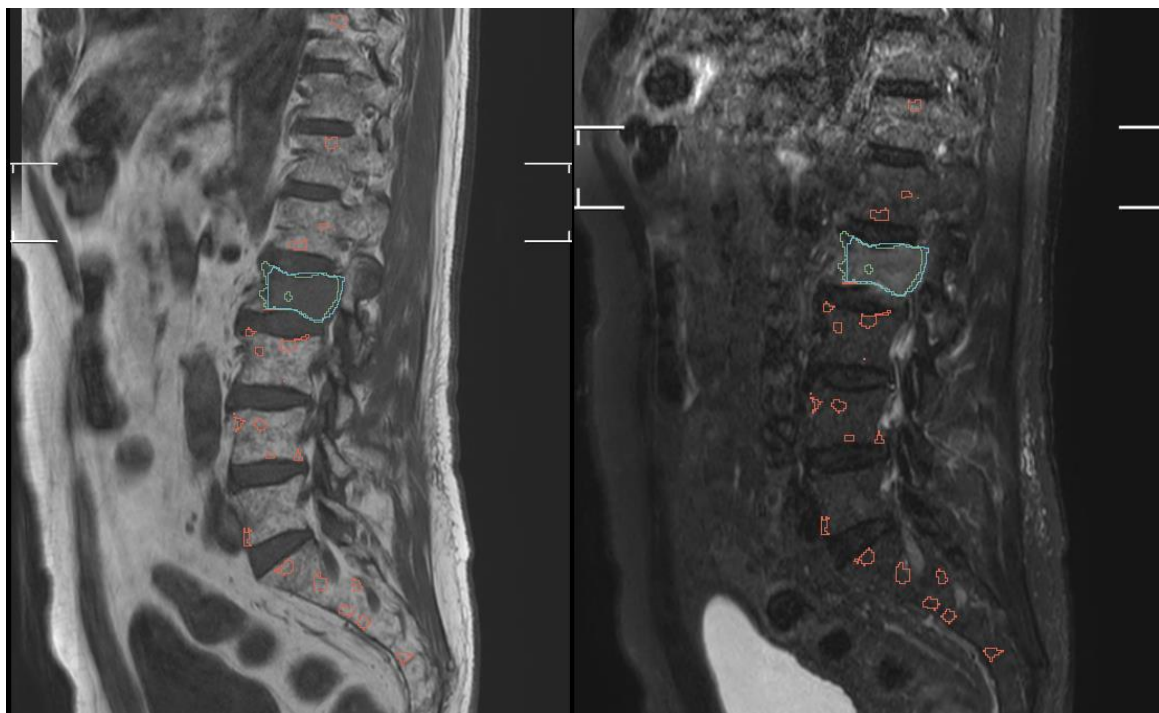
Η δεύτερη φάση ανέδειξε τη συνδυαστική εκδοχή γράφων (**Case 5**) ως τη βέλτιστη προσέγγιση, ιδιαίτερα σε ό,τι αφορά την ισορροπία μεταξύ ακρίβειας και ευαισθησίας. Η ενσωμάτωση τόσο radiomic όσο και intensity χαρακτηριστικών προσφέρει μια πολυτροπική αναπαράσταση των supervoxels, που ενισχύει την ικανότητα εντοπισμού. Αντίθετα, η απομονωμένη χρήση intensity features (Case 1 ή 3) απέτυχε να διαχωρίσει αποτελεσματικά τις παθολογικές περιοχές, λόγω έλλειψης πληροφορίας μορφολογικής πολυπλοκότητας.

Η επιλογή του αριθμού supervoxels (1000 vs 150 mm^3) επίσης διαδραμάτισε κρίσιμο ρόλο. Η σταθεροποίηση του μεγέθους (150 mm^3) αντί της απόλυτης ποσότητας οδήγησε σε ομοιογενή και αναπαραγώγιμη κατανομή, αποφεύγοντας την υπερσυνάθροιση σε πυκνές ανατομικές περιοχές (όπως οι σπονδυλικές πλάκες) και ενισχύοντας την εκπροσώπηση των μεταβλητών περιοχών. Έτσι, η απόδοση βελτιώθηκε αισθητά σε επίπεδο Lesion F1 και CTS. Αντίθετα, παρατηρούμε πως δεν ισχύει το ίδιο σε μετρικές όπως το F1-score, κάτι που ερμηνεύεται ως **αδυναμία των voxel-wise μετρικών να είναι βασικοί παράγοντες** στον προσδιορισμό της καλύτερης περίπτωσης και προτείνεται να είναι συμπληρωματικός. Όπως φαίνεται και στην Εικόνα 18, συγκρίνοντας το Case 5 με το μεγαλύτερο Lesion F1-score και το Case 2 με το μεγαλύτερο F1-score, παρατηρούμε πως στη δεύτερη περίπτωση οι εντοπισμένες αλλοιώσεις είναι σχεδόν τυχαίες.

Όσον αφορά τις αρχιτεκτονικές, οι διαφορές μεταξύ GCN, GAT και GraphSAGE δεν κρίθηκαν στατιστικά σημαντικές, γεγονός που ενισχύει την άποψη πως το είδος της γραφικής αναπαράστασης και των χαρακτηριστικών είναι ο σημαντικότερος παράγοντας απόδοσης. Οι παρατηρούμενες διαφοροποιήσεις στις μέσες τιμές πιθανώς αντικατοπτρίζουν τυχαία μεταβλητότητα που οφείλεται στη φύση των δεδομένων και στο περιορισμένο μέγεθος δείγματος ($n = 7$). Παρ' όλα αυτά, το GraphSAGE εμφάνισε σταθερά υψηλές επιδόσεις σε κάθε πείραμα (Πίνακας 4, 5), πιθανώς λόγω του μηχανισμού sampling και της ικανότητάς του να γενικεύει πέρα από τις τοπικές συνδέσεις, καθιστώντας το πιο ανθεκτικό σε περιπτώσεις noise ή σπανιότητας. Αντίθετα, το GAT, ενώ παρουσίασε ισχυρές επιδόσεις στο AUPRC, ενδέχεται να υπερεστιάζει σε λίγους κόμβους με υψηλή προσοχή, παραμελώντας την αναγκαία χωρική συνέπεια, κάτι κρίσιμο σε ογκολογικές εικόνες.

Η ποιότητα των επιδόσεων αποδεικνύει ότι η καλύτερη επιλογή για την τμηματοποίηση αλλοιώσεων σε Πολλαπλό Μυέλωμα είναι η χρήση ενός **συνδυαστικού γραφήματος (Radiomic + Intensity)** με **σταθερό όγκο supervoxels (150 mm^3)** και εφαρμογή σε **μοντέλα τύπου GraphSAGE** με επαρκές

βάθος και κρυφές διαστάσεις. Ο συνδυασμός αυτός εξισορροπεί τη λεπτομέρεια σε μικρές αλλοιώσεις, τη σταθερότητα σε noise, και τη γενίκευση σε νέους ασθενείς, προσφέροντας μια αξιόπιστη λύση για κλινική υποστήριξη.



Εικόνα 18: Σύγκριση ασθενή με δύο διαφορετικά μοντέλα. Το κόκκινο είναι το μοντέλο με το μεγαλύτερο F1-score (GCN – 64 x 5) και το πράσινο είναι το μοντέλο με το μεγαλύτερο Lesion F1-score (SAGE – 128 x 4). Με μπλέ σχεδιάζεται το ground truth.

5.2 Σχέση με βιβλιογραφία

Σε σύγκριση με προηγούμενες μελέτες στον χώρο της αυτόματης ανάλυσης WB-MRI για πολλαπλό μυέλωμα, η παρούσα εργασία επιτυγχάνει ουσιαστική βελτίωση τόσο σε επίπεδο αξιολόγησης όσο και σε μεθοδολογική πληρότητα. Σχετικές προσπάθειες της βιβλιογραφίας (βλ. Ενότητα 1.5.1) έχουν κατά καιρούς παρουσιάσει εμφανώς υψηλές επιδόσεις, χωρίς όμως να καθίσταται σαφές αν τα χρησιμοποιούμενα δεδομένα περιλάμβαναν ομοειδείς ή πανομοιότυπες κλινικές εικόνες μεταξύ εκπαίδευσης και δοκιμής [47]. Τέτοια προσέγγιση εγείρει ζητήματα overfitting και μειώνει τη ρεαλιστικότητα της εκτίμησης της πραγματικής απόδοσης σε άγνωστα, ετερογενή περιστατικά. Επιπλέον, η πλειοψηφία των προηγούμενων εργασιών βασίζεται σε ψηφιοποιημένες τομές ή regions-of-interest (ROIs) που αποκόπονται από το σύνολο του σώματος, υποβαθμίζοντας τη σημασία της ολιστικής αναπαράστασης [48].

5.3 Ποιοτική Αξιολόγηση Μοντέλων

Η ποιοτική αξιολόγηση των μοντέλων πραγματοποιήθηκε μέσω της οπτικοποίησης των προβλέψεων τους σε όλες περιπτώσεις του test set, όπως αυτά διαφαίνονται στις Εικόνες 11-17. Η σύγκριση αυτή επιτρέπει την ποιοτική αξιολόγηση της συμπεριφοράς των μοντέλων, πέρα από τις συνολικές μετρικές απόδοσης. Ειδικότερα, εστιάζεται:

- στην μορφολογική ακρίβεια των προβλεπόμενων περιοχών (π.χ. εάν οι βλάβες ανιχνεύονται πλήρως ή μερικώς),
- στην συμφωνία των μασκών και του ground truth με τις πραγματικές αλλοιώσεις,
- και στην ευαισθησία/ειδικότητα έναντι μικρών ή δυσδιάκριτων αλλοιώσεων.

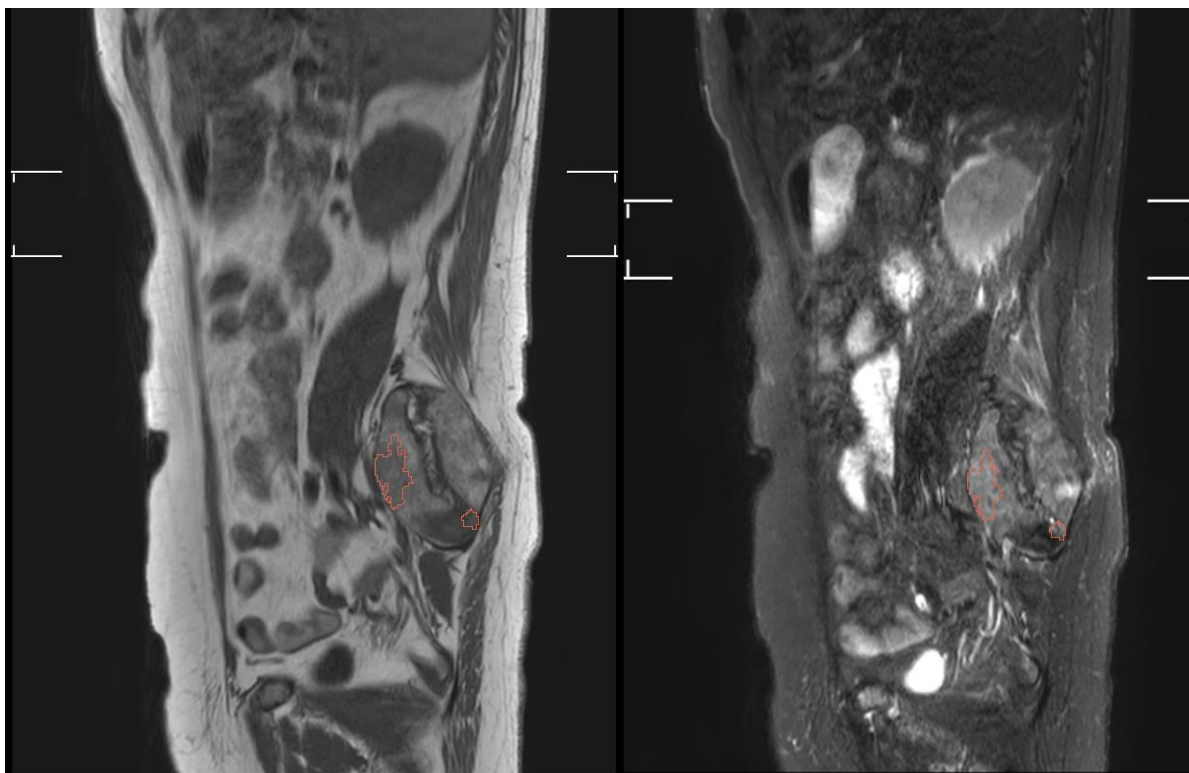
Με βάση αυτές τις συγκρίσεις, σχολιάζονται οι περιπτώσεις στις οποίες τα μοντέλα αποτυγχάνουν να συλλάβουν μικρού μεγέθους ή άτυπες βλάβες, καθώς και οι περιπτώσεις υπερεκτίμησης όπου εμφανίζονται ψευδώς θετικές περιοχές.

Ασθενής 5: Το μοντέλο επέδειξε υψηλή ευαισθησία, εντοπίζοντας διάχυτα κατανεμημένες περιοχές με παθολογικά χαρακτηριστικά. Σε αρκετές περιπτώσεις, παρατηρήθηκαν ευρήματα σε θέσεις μη επισημασμένες στο ground truth, πιθανώς λόγω της συνύπαρξης διάχυτων και εστιακών προτύπων Πολλαπλού Μυελώματος, ή λόγω παράλειψης από τον ειδικό ασήμαντων περιοχών κατά τη διαδικασία σήμανσης (Εικόνα 11).

Ασθενής 8: Η πρόβλεψη περιλάμβανε αλλοιώσεις εκτός των επισημασμένων περιοχών, χωρίς εμφανές οπτικό σφάλμα. Το γεγονός αυτό αναδεικνύει τη δυσκολία της κλινικής επιβεβαίωσης των ορίων μεταξύ υγιούς και παθολογικού ιστού σε περιπτώσεις ήπιας διάχυσης, και **θέτει υπό αμφισβήτηση την ακρίβεια του ground truth** (Εικόνα 12).

Ασθενής 11: Το μοντέλο ταυτίστηκε σχεδόν πλήρως με τις επισημασμένες βλάβες, γεγονός που καταδεικνύει την **ικανότητα πλήρους εντοπισμού καθαρών και εστιακών αλλοιώσεων**, ιδίως όταν η μορφολογία τους είναι σαφής και καλά οριοθετημένη (Εικόνα 13).

Ασθενής 20: Παρατηρήθηκαν τόσο επιτυχείς ανιχνεύσεις επιπλέον αλλοιώσεων (εκτός ground truth) όσο και περιορισμένος αριθμός ψευδώς θετικών (Εικόνα 14). Ενδιαφέρον παρουσιάζει η περιοχή του ιερού και συγκεκριμένα ο εντοπισμός τόσο ενός σημείου που κατά πάσα πιθανότητα αντιστοιχεί σε μυ, όσο και ενός σημείου πιθανώς αντιστοιχεί σε διάχυτη αλλοίωση, όπως φαίνεται και στην Εικόνα 19.

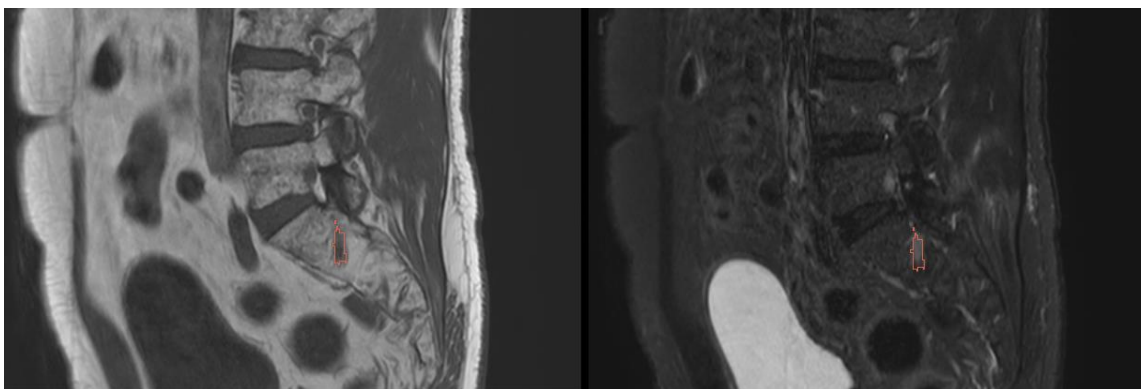


Εικόνα 19: Η περιοχή του ιερού του ασθενή 20. Αριστερά φαίνεται η T1 ακολουθία, ενώ δεξιά φαίνεται η T2.

Ασθενής 22: Η πρόβλεψη συνδυάζει επιτυχίες και αστοχίες. Παρότι εντοπίζονται όλες οι σημαντικές βλάβες, παράγεται μία επιπλέον πρόβλεψη και μία ψευδώς θετική (Εικόνα 15). Η χαμηλή αντίθεση των εικόνων και το μέγεθος των αλλοιώσεων ενδέχεται να συνέβαλαν σε αυτή την αστάθεια. Επισημαίνεται επίσης η δυσκολία εντοπισμού μικρών όγκων, όταν αυτοί δεν συνοδεύονται από έντονη διαμόρφωση σήματος. Οι παραπάνω διαπιστώσεις γίνονται πιο αντιληπτές στις Εικόνες 20 και 21.



Εικόνα 20: Ψευδώς θετική περιοχή στον ασθενή 22. Αριστερά φαίνεται η T1 ακολουθία, ενώ δεξιά φαίνεται η T2.



Εικόνα 21: Ψευδώς αρνητική περιοχή στον ασθενή 22. Αριστερά φαίνεται η T1 ακολουθία, ενώ δεξιά φαίνεται η T2.

Ασθενής 27: Η πρόβλεψη περιλάμβανε πλήθος περιοχών που δεν υπήρχαν στο ground truth, ωστόσο η οπτική επιθεώρηση υποδηλώνει ότι ενδέχεται να αντιστοιχούν σε πραγματικές αλλοιώσεις. Η περίπτωση αυτή εγείρει το ενδεχόμενο **υποεκπροσώπησης οριακών περιπτώσεων** στη φάση της σήμανσης, και αναδεικνύει τη δυναμική του μοντέλου να εντοπίζει λεπτές ή υποσυνείδητες μεταβολές του μυελού (Εικόνα 16).

Ασθενής 28: Το μοντέλο παρουσιάζει ιδιαίτερα υψηλή ευαισθησία, εντοπίζοντας το σύνολο σχεδόν των επισημασμένων περιοχών, αλλά συνοδεύεται από μέτρια ειδικότητα, με αρκετές ψευδώς θετικές περιοχές. Η συμπεριφορά αυτή αντικατοπτρίζει το γενικότερο μοτίβο απόδοσης του μοντέλου, το οποίο τείνει να προτιμά την υπερανίχνευση όταν η μορφολογία είναι ασαφής (Εικόνα 17).

Η οπτική αξιολόγηση των μασκών τμηματοποίησης κατέδειξε ότι το μοντέλο παρουσιάζει υψηλή ευαισθησία στην ανίχνευση εστιών, ιδίως σε περιπτώσεις ασθενών με διάσπαρτα ή ήπια πρότυπα προσβολής. Η ικανότητα του να εντοπίζει αλλοιώσεις που δεν έχουν καταγραφεί στο ground truth, όπως παρατηρήθηκε στους ασθενείς 5, 8 και 27, αναδεικνύει δύο κρίσιμες παραμέτρους:

- **Την ετερογένεια του πολλαπλού μυελώματος**, το οποίο συχνά εκδηλώνεται με πολυμορφικά και μη σαφώς οριοθετημένα πρότυπα προσβολής (π.χ. μεικτά διάχυτα-εστιακά).
- **Την ενδεχόμενη υποσήμανση του συνόλου εκπαίδευσης**, ιδίως όταν οι ιατροί δεν επισημαίνουν αλλοιώσεις σε περιοχές με ήπια ενδείξεις (π.χ. στο ιερό ή στα σπονδυλικά τόξα).

Αυτό επιβεβαιώνεται από το γεγονός ότι σε αρκετούς ασθενείς, το μοντέλο εντόπισε περιοχές πιθανής κακοήθειας που δεν είχαν σημειωθεί από ειδικούς, είτε λόγω της μορφολογικής ασάφειας είτε λόγω ελλιπούς συμφωνίας μεταξύ των ακολουθιών.

Από την άλλη πλευρά, η παραγωγή ψευδώς θετικών προβλέψεων (όπως στους ασθενείς 20, 22 και 28), περιορίστηκε κυρίως σε μικρές περιοχές στα όρια των σπονδύλων ή σε ανατομικές δομές χωρίς παθολογικό φορτίο (π.χ. ιερό οστό). Αυτό μπορεί να οφείλεται:

- στην **υψηλή ετερογένεια** της σύστασης του μυελού σε περιοχές με λιπώδη ή μεταβατικό μυελό, που ενδέχεται να προκαλούν "σύγχυση" στο σύστημα
- στο γεγονός ότι το ground truth δεν αποτυπώνει **απολύτως ακριβώς** τα περιγράμματα των βλαβών, καθώς η διαδικασία της ετικετοδότησης δεν ακολουθεί πάντα αυστηρή γεωμετρική συνέπεια ανάμεσα στις ακολουθίες.

Ιδιαίτερο ενδιαφέρον παρουσιάζει η απόλυτη συμφωνία του μοντέλου με τις ετικέτες στους ασθενείς 11 και 15. Η απόδοση αυτή αποδίδεται στην καθαρότητα του εστιακού προτύπου στις συγκεκριμένες περιπτώσεις, γεγονός που καθιστά τις βλάβες πιο "ευδιάκριτες" τόσο για τον ακτινολόγο όσο και για τον αλγόριθμο. Η παρούσα τεχνική μπορεί να επεκταθεί και στους υπόλοιπα πρότυπα του ΠΜ, όπως αυτό του διαχύτου όσο και του ποικίλου προτύπου, δεδομένης της ευπλαστότητας του SLIC αλγορίθμου όσον αφορά την παραμετροποίηση.

5.4 Περιορισμοί

Στην παρούσα μελέτη, εντοπίστηκαν ορισμένοι περιορισμοί που σχετίζονται με την ετικετοδότηση, την ποιότητα της απεικόνισης, αλλά και την ετερογένεια του φαινοτύπου του Πολλαπλού Μυελώματος. Οι περιορισμοί αυτοί μπορούν να συγκεντρωθούν στις παρακάτω κατηγορίες:

- πιθανά σφάλματα ή παραλείψεις κατά την ετικετοδότηση,
- χαμηλή ποιότητα των MRI εικόνων, ή
- συνδυασμός με άλλους τύπους αλλοιώσεων, όπως diffusion και variegated

Ταυτόχρονα και οι υπόλοιπες ακολουθίες του MY-RADS πρωτοκόλλου θα μπορούσαν να ενσωματωθούν στον παρόν αλγόριθμο προσφέροντας περεταίρω διαγνωστική πληροφορία η οποία ενδεχομένως να βελτιώσει την απόδοση των μοντέλων.

Για την αντιμετώπιση των παραπάνω περιορισμών πραγματοποιήθηκε συμπληρωματική ποσοτική ανάλυση σε τρεις ασθενείς που κρίθηκαν οι πιο αξιόπιστοι και αντιπροσωπευτικοί και συγκεκριμένα οι ασθενείς 11, 22 και 28. Τα αποτελέσματα αυτής της ανάλυσης επιβεβαιώνουν την υψηλή ακρίβεια και σταθερότητα του μοντέλου:

Πίνακας 6. Αποτελέσματα της βέλτιστης αρχιτεκτονικής με διαφορετικά υποσύνολα ασθενών

Ασθενείς	F1	AUPRC	Lesion F1	Lesion Precision	Lesion Recall	CTS
5, 8, 11, 20, 22, 27, 28	0.738	0.987	0.488	0.419	0.826	0.420
11, 22, 28	0.882	0.997	0.681	0.694	0.722	0.667

Όπως γίνεται εμφανές και από τον Πίνακα 6, τα αποτελέσματα βελτιώνονται σημαντικά μετά την επιλογή των πιο ποιοτικών και σωστά επεξεργασμένων εικόνων. Παρ' όλα αυτά, ένα πιο σωστά επιλεγμένο σετ δοκιμής και περεταίρω εμπλουτισμός των δεδομένων θα μπορούσε να αναδείξει καλύτερα αποτελέσματα και πιο γενικευμένες σχέσεις.

6 Συμπεράσματα

Συμπερασματικά, η παρούσα εργασία εισάγει μια καινοτόμο προσέγγιση για την αυτόματη ανάλυση της WB-MRI σπονδυλικής στήλης σε ασθενείς με Πολλαπλό Μυέλωμα, βασισμένη στη γραφική αναπαράσταση των δεδομένων μέσω supervoxels και την αξιοποίηση νευρωνικών δικτύων σε γράφους (GNNs). Η προτεινόμενη μεθοδολογία ενσωματώνει τεχνικές εξαγωγής ραδιομικών χαρακτηριστικών, επιλογής βέλτιστων χαρακτηριστικών και εκπαιδευσιμότητας με πολλαπλές αρχιτεκτονικές GNN (GCN, GAT, GraphSAGE), διερευνώντας την αλληλεπίδραση μεταξύ χωρικής ανάλυσης και μορφολογικής πληροφορίας.

Η ανάλυση των αποτελεσμάτων ανέδειξε τη σημασία της σωστής επιλογής χαρακτηριστικών και της δομής των γράφων για την επίδοση των μοντέλων, με τον συνδυασμό radiomic και intensity-based χαρακτηριστικών να προσφέρει αυξημένη ακρίβεια στην αναγνώριση παθολογικών supervoxels. Οι συγκριτικές αξιολογήσεις μεταξύ των αρχιτεκτονικών ανέδειξαν ισχυρές επιδόσεις, ιδίως για συγκεκριμένες παραμετροποιήσεις, ως προς τις μετρικές F1, AUPRC και Lesion-level F1, παρά τις

δυσκολίες του εγχειρήματος (ελλιπή δεδομένα, σφάλματα στη σχεδίαση, πολλαπλοί τύποι αλλοιώσεων).

Η εργασία αυτή αποτελεί το πρώτο, κατά τη γνώση μας, ολοκληρωμένο παράδειγμα εφαρμογής GNNs στην ανάλυση WB-MRI δεδομένων για το Πολλαπλό Μυέλωμα, με εστίαση στις εστιακές αλλοιώσεις Πολλαπλού Μυελώματος στον μυελό της σπονδυλικής στήλης. Η modular φύση του pipeline καθιστά δυνατή την προσαρμογή του σε άλλα ανατομικά σημεία ή τύπους αλλοιώσεων, υποστηρίζοντας μελλοντικές εφαρμογές στην αυτόματη εκτίμηση του φορτίου νόσου, τη σταδιοποίηση, και την παρακολούθηση απόκρισης σε θεραπεία.

Βιβλιογραφία

- [1] Tunariu N, Blackledge M, Messiou C, *et al.* What's new for clinical whole-body MRI (WB-MRI) in the 21st century. *Br J Radiol.* 2020;93(1115):20200562. <https://doi.org/10.1259/bjr.20200562>
- [2] Messiou C, Hillengass J, Delorme S, *et al.* Guidelines for acquisition, interpretation, and reporting of whole-body MRI in myeloma: Myeloma Response Assessment and Diagnosis System (MY-RADS). *Radiology.* 2019;291(1):5–13. <https://doi.org/10.1148/radiol.2019181949>
- [3] Fang AM, Gregg JR, Pettaway C, *et al.* Whole-body MRI for staging prostate cancer: a narrative review. *BJU Int.* 2024;133(1):1–10. <https://doi.org/10.1111/bju.16514>
- [4] Taylor SA, Mallett S, Beare S, *et al.* Diagnostic accuracy of whole-body MRI versus standard imaging pathways for metastatic disease in newly diagnosed colorectal cancer: the prospective Streamline C trial. *Lancet Gastroenterol Hepatol.* 2019;4(7):529–537. [https://doi.org/10.1016/S2468-1253\(19\)30056-1](https://doi.org/10.1016/S2468-1253(19)30056-1)
- [5] Taylor SA, Mallett S, Ball S, *et al.* Diagnostic accuracy of whole-body MRI versus standard imaging pathways for metastatic disease in newly diagnosed non-small-cell lung cancer: the prospective Streamline L trial. *Lancet Respir Med.* 2019;7(6):523–532. [https://doi.org/10.1016/S2213-2600\(19\)30090-6](https://doi.org/10.1016/S2213-2600(19)30090-6)
- [6] Ballinger ML, Best A, Mai PL, *et al.* Baseline surveillance in Li-Fraumeni syndrome using whole-body magnetic resonance imaging: a meta-analysis. *JAMA Oncol.* 2017;3(12):1634–1639. <https://doi.org/10.1001/jamaoncol.2017.1968>
- [7] Ahlawat S, Fayad LM, Khan MS, *et al.* Current whole-body MRI applications in the neurofibromatoses. *Neurology.* 2016;87(7 Suppl 1):S31–S39. <https://doi.org/10.1212/WNL.0000000000002929>
- [8] Gottumukkala RV, Gee MS, Hampilos PJ, Greer MLC. Current and emerging roles of whole-body MRI in evaluation of pediatric cancer patients. *RadioGraphics.* 2019;39(2):516–534. <https://doi.org/10.1148/rg.2019180130>
- [9] Peccatori FA, Codacci-Pisanelli G, Del Grande M, Scarfone G, Zugni F, Petralia G. Whole-body MRI for systemic staging of breast cancer in pregnant women. *Breast.* 2017;35:177–181. <https://doi.org/10.1016/j.breast.2017.07.014>
- [10] Elessawy SS, Abdelsalam EM, Abdel Razek E, Tharwat S. Whole-body MRI for full assessment and characterization of diffuse inflammatory myopathy. *Acta Radiol Open.* 2016;5(9):2058460116668216. <https://doi.org/10.1177/2058460116668216>
- [11] Rajkumar SV, Dimopoulos MA, Palumbo A, *et al.* International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol.* 2014;15(12):e538–e548. [https://doi.org/10.1016/S1470-2045\(14\)70442-5](https://doi.org/10.1016/S1470-2045(14)70442-5)

- [12] Moulopoulos LA, Dimopoulos MA, Alexanian R, Leeds NE, Libshitz HI. Multiple myeloma: MR patterns of response to treatment. *Radiology*. 1994;193(2):441–446. <https://doi.org/10.1148/radiology.193.2.7972760>
- [13] Moulopoulos LA, Koutoulidis V, Hillengass J, *et al.* Recommendations for acquisition, interpretation, and reporting of whole-body low-dose CT in patients with multiple myeloma and other plasma cell disorders: a report of the IMWG Bone Working Group. *Blood Cancer J*. 2018;8(10):95. <https://doi.org/10.1038/s41408-018-0124-1>
- [14] Shortt CP, Carty F, Murray JG. The role of whole-body imaging in the diagnosis, staging, and follow-up of multiple myeloma. *Semin Musculoskelet Radiol*. 2010;14(1):37–46. <https://doi.org/10.1055/s-0030-1248705>
- [15] Wennmann M, Klein A, Bauer F, *et al.* Combining deep learning and radiomics for automated, objective, comprehensive bone marrow characterization from whole-body MRI: a multicentric feasibility study. *Invest Radiol*. 2022;57(11):752–761. <https://doi.org/10.1097/RLI.0000000000000891>
- [16] Koutoulidis V, Fontara S, Terpos E, *et al.* Quantitative diffusion-weighted imaging of the bone marrow: an adjunct tool for the diagnosis of a diffuse MR imaging pattern in patients with multiple myeloma. *Radiology*. 2017;282(2):484–493. <https://doi.org/10.1148/radiol.2016160363>
- [17] Moulopoulos LA, Koutoulidis V. MRI of the normal bone marrow: pulse sequences. In: *Bone Marrow MRI*. Milano: Springer; 2015. p.7–23. https://doi.org/10.1007/978-88-470-5316-8_2
- [18] Plewes DB, Kucharczyk W. Physics of MRI: a primer. *J Magn Reson Imaging*. 2012;35(5):1008–1021. <https://doi.org/10.1002/jmri.23642>
- [19] Brown RW, Cheng Y-CN, Haacke EM, Thompson MR, Venkatesan R. Magnetic resonance imaging: principles and techniques. *Radiol Clin North Am*. 2015;53(1):1–19.
- [20] Haacke EM, Brown RW, Thompson MR, Venkatesan R. Basic principles of magnetic resonance imaging: an update. *NMR Biomed*. 1999;12(7):377–386. [https://doi.org/10.1002/\(SICI\)1099-1492\(199911\)12:7<377::AID-NBM593>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1099-1492(199911)12:7<377::AID-NBM593>3.0.CO;2-X)
- [21] Bernstein MA, King KF, Zhou XJ. Imaging principles: k-space, gradients, RF pulses, and sampling. In: *Handbook of MRI Pulse Sequences*. Amsterdam: Elsevier; 2004. p.29–83.
- [22] Mansfield P, Morris PG. Linear systems, Fourier transforms, point spread and resolution. In: *NMR Imaging in Biomedicine*. Oxford: Oxford University Press; 1993. p.51–76.
- [23] de Graaf RA. Principles of magnetic resonance. In: *High-Resolution NMR in the Solid State*. Hoboken, NJ: Wiley; 2007. p.1–30.
- [24] Odaibo SG. A quantum mechanical review of magnetic resonance imaging. *arXiv preprint*. 2012. <https://arxiv.org/abs/1210.0946>
- [25] Di Ieva A, Coppo L. Physics and mathematics of magnetic resonance imaging for life scientists. *World J Radiol*. 2015;3(1):17–42. <https://doi.org/10.4329/wjr.v3.i1.17>

- [26] Pohmann R, Speck O, Scheffler K. Signal-to-noise ratio and parallel imaging performance of phased-array coils: a simulation study. *Magn Reson Med*. 2016;75(3):951–958. <https://doi.org/10.1002/mrm.25679>
- [27] Jimeno MM, Vaughan JT, Geethanath S. Superconducting magnet designs and MRI accessibility: a review. *arXiv preprint*. 2022. <https://arxiv.org/abs/2201.XXXX>
- [28] Chilla GS, Tan CH, Xu C, Poh CL. Diffusion-weighted magnetic resonance imaging and its recent trend: a survey. *J Med Imaging Health Inform*. 2015;5(6):1229–1235. <https://doi.org/10.1166/jmihi.2015.1578>
- [29] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell*. 2012;34(11):2274–2282. <https://doi.org/10.1109/TPAMI.2012.120>
- [30] Lowekamp BC, Chen DT, Yaniv Z, Yoo TS. Scalable simple linear iterative clustering (SSLIC) using a generic and parallel approach. *arXiv preprint*. 2018. <https://arxiv.org/abs/1806.11141>
- [31] Irving B. maskSLIC: Regional superpixel generation with application to local pathology characterisation in medical images. *arXiv preprint*. 2016. <https://arxiv.org/abs/1606.09518>
- [32] Tian Z, Xiang Q, Zhu J, Liu R, Bai W. Superpixel-based segmentation for 3D prostate MR images. *Comput Math Methods Med*. 2015;2015:231–239. <https://doi.org/10.1155/2015/189715>
- [33] Jing J, Gao T, Zhang W, Gao Y, Sun C. Image feature information extraction for interest point detection: a review. *arXiv preprint*. 2021. <https://arxiv.org/abs/2106.07929>
- [34] Amraee S, Chinipardaz M, Charoosaei M. Analytical study of two feature extraction methods in comparison with deep learning methods for classification of small metal objects. *Vis Comput Ind Biomed Art*. 2022;5:13. <https://doi.org/10.1186/s42492-022-00111-6>
- [35] Zwanenburg A, Vallières M, Abdalah MA, *et al*. The Image Biomarker Standardisation Initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328–338. <https://doi.org/10.1148/radiol.2020191145>
- [36] van Griethuysen JJM, Fedorov A, Parmar C, *et al*. Computational radiomics system to decode the radiographic phenotype: the PyRadiomics library. *Cancer Res*. 2017;77(21):e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
- [37] Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *Radiology*. 2020;295(2):328–338. <https://doi.org/10.1148/radiol.2020191145>
- [38] Patel D, Patel D, Saxena R, Akilan T. Multi-class brain tumor segmentation using graph attention network. *arXiv preprint*. 2023. <https://doi.org/10.48550/arXiv.2302.05598>
- [39] Schmidt M, Fung G, Rosales R. Fast optimization methods for L1 regularization: a comparative study and two new approaches. In: *Lecture Notes in Computer Science*. Vol 4701. Berlin: Springer; 2007. p.286–297. https://doi.org/10.1007/978-3-540-74958-5_28

- [40] Cortes C, Mohri M, Rostamizadeh A. L2 regularization for learning kernels. *arXiv preprint*. 2012. <https://arxiv.org/abs/1205.2653>
- [41] Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005;67(2):301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
- [42] Duan S, Li Z, Li R, Gao S. Semi-supervised classification of fundus images combined with CNN and GCN. *Med Phys*. 2022;49(10):6560–6573. <https://doi.org/10.1002/mp.15979>
- [43] Saueressig C, Berkley A, Munbodh R, Singh R. A joint graph and image convolution network for automatic brain tumor segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2021*. Cham: Springer; 2021. p.324–334. https://doi.org/10.1007/978-3-031-08999-2_30
- [44] Shen Y, Zhou B, Xiong X, Gao R, Wang YG. How graph neural networks enhance convolutional neural networks towards mining the topological structures from histology. In: *Proceedings of the ICML Workshop on Computational Biology*. 2022. <https://doi.org/10.48550/arXiv.2206.12345>
- [45] Meng Y, Zhang H, Zhao Y, *et al*. Graph-based region and boundary aggregation for biomedical image segmentation. *IEEE Trans Med Imaging*. 2022;41(3):690–701. <https://doi.org/10.1109/TMI.2021.3123567>
- [46] Pérez-Cano J, Ma Y, Huang X, *et al*. Applications of graph convolutional networks in computer vision: a comprehensive review. *Comput Math Methods Med*. 2024;2024:1–21. <https://doi.org/10.1155/2024/9834217>
- [47] Fränzle A, Hillengass J, Bendl R. Spinal focal lesion detection in multiple myeloma using multimodal image features. In: *Medical Imaging 2015: Computer-Aided Diagnosis*. Proc SPIE. 2015;9414:94142Q. <https://doi.org/10.1117/12.2081990>
- [48] Wennmann M, Neher P, Stanczyk N, *et al*. Deep learning for automatic bone marrow apparent diffusion coefficient measurements from whole-body MRI in patients with multiple myeloma: a retrospective multicenter study. *Invest Radiol*. 2023;58(4):273–282. <https://doi.org/10.1097/RLI.0000000000000932>
- [49] Zhao X, Huang Y, Sun J, *et al*. Multiple Myeloma Segmentation Network (MMNet): an encoder–decoder deep learning approach for spinal MR lesion segmentation in multiple myeloma patients. *Sci Rep*. 2024;14(1):11532. <https://doi.org/10.1038/s41598-024-11532-x>
- [50] Dutoit JC, Hillengass J, Delorme S, *et al*. MRI in multiple myeloma: a pictorial review of diagnostic applications. *Insights Imaging*. 2016;7(3):327–345. <https://doi.org/10.1007/s13244-016-0509-y>
- [51] Ormond Filho AG, Martin J, *et al*. Whole-body imaging of multiple myeloma: diagnostic criteria and imaging recommendations. *RadioGraphics*. 2019;39(4):1077–1097. <https://doi.org/10.1148/rg.2019180155>
- [52] Liu M, Tuzel O, Ramalingam S, Chellappa R. Entropy rate superpixel segmentation. *IEEE Conf Comput Vis Pattern Recognit (CVPR)*. 2011;2097–2104. <https://doi.org/10.1109/CVPR.2011.5995618>
- [53] Wirjadi O. Survey of 3D image segmentation methods. *ITWM Report 123*. Kaiserslautern: Fraunhofer Institute for Industrial Mathematics; 2007.

- [54] Carvalho LE, Rodrigues P, Santos M. 3D segmentation algorithms for computerized tomographic imaging: a review. *J Digit Imaging*. 2018;31(4):523–539. <https://doi.org/10.1007/s10278-018-0060-8>
- [55] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *Int Conf Learn Represent (ICLR)*. 2017. <https://arxiv.org/abs/1609.02907>
- [56] Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. *Int Conf Learn Represent (ICLR)*. 2018. <https://arxiv.org/abs/1710.10903>
- [57] Hamilton WL, Ying Z, Leskovec J. Inductive representation learning on large graphs. *Adv Neural Inf Process Syst (NeurIPS)*. 2017;30:1024–1034. <https://papers.nips.cc/paper/2017/hash/5dd9db5e033da9c6fb5ba83c7a7e9-Abstract.html>
- [58] Pedregosa F, Varoquaux G, Gramfort A, *et al*. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.
- [59] Phan T, Yamamoto K. Resolving class imbalance in object detection with weighted cross-entropy losses. *arXiv preprint*. 2020. <https://doi.org/10.48550/arXiv.2006.01413>
- [60] Niculescu-Mizil A, Caruana R. Predicting good probability estimates from binary classifiers in high-imbalance scenarios. In: *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*. 2005;625–632. <https://doi.org/10.1145/1102351.1102430>
- [61] Al-Kababji A, Bensaali F, Dakua SP. Scheduling techniques for liver segmentation: ReduceLROnPlateau vs OneCycleLR. *arXiv preprint*. 2022. <https://doi.org/10.48550/arXiv.2202.06373>
- [62] Hillengass J, Usmani S, Rajkumar SV, *et al*. International Myeloma Working Group consensus recommendations on imaging in monoclonal plasma cell disorders—2019 update. *Lancet Oncol*. 2019;20(6):e302–e312. [https://doi.org/10.1016/S1470-2045\(19\)30309-2](https://doi.org/10.1016/S1470-2045(19)30309-2)
- [63] Rasche L, Angtuaco E, Alpe T, *et al*. Modern imaging of multiple myeloma: utility in diagnosis and treatment. *Blood Cancer J*. 2016;6(7):e428. <https://doi.org/10.1038/bcj.2016.61>
- [64] Tustison NJ, Avants BB, Cook PA, *et al*. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29(6):1310–1320. <https://doi.org/10.1109/TMI.2010.2046908>
- [65] Hossain MI, Amin MZ, Anyimadu DT, Suleiman TA. Comparative study of probabilistic atlas and deep learning approaches for automatic brain tissue segmentation from MRI using N4 bias field correction and anisotropic diffusion pre-processing techniques. *arXiv preprint*. 2024. <https://doi.org/10.48550/arXiv.2411.05456>
- [66] Gerig G, Kübler O, Kikinis R, Jolesz FA. Nonlinear anisotropic filtering of MRI data. *IEEE Trans Med Imaging*. 1992;11(2):221–232. <https://doi.org/10.1109/42.141646>
- [67] Sekuboyina A, Bayat A, Rempfler M, *et al*. VerSe: a vertebrae labelling and segmentation benchmark. *Med Image Anal*. 2021;73:102166. <https://doi.org/10.1016/j.media.2021.102166>
- [68] Lessmann N, van Ginneken B, de Jong PA, Išgum I. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. *Med Image Anal*. 2019;53:142–155. <https://doi.org/10.1016/j.media.2019.02.005>

- [69] Paszke A, Gross S, Massa F, *et al.* PyTorch: an imperative style, high-performance deep learning library. In: *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*. 2019. <https://papers.nips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html>
- [70] Fey M, Lenssen JE. Fast graph representation learning with PyTorch Geometric. *ICLR Workshop on Representation Learning on Graphs and Manifolds*. 2019. <https://arxiv.org/abs/1903.02428>
- [71] Nickolls J, Buck I, Garland M, Skadron K. Scalable parallel programming with CUDA. *Queue*. 2008;6(2):40–53. <https://doi.org/10.1145/1365490.1365500>
- [72] Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. 2015;15:29. <https://doi.org/10.1186/s12880-015-0068-x>
- [73] Müller D, Soto-Rey I, Kramer F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res Notes*. 2022;15(1):96. <https://doi.org/10.1186/s13104-022-06096-y>
- [74] Reinke A, Tizabi MD, Eisenmann M, *et al.* Common pitfalls and recommendations for the evaluation of segmentation algorithms. *Med Image Anal*. 2021;67:101876. <https://doi.org/10.1016/j.media.2020.101871>
- [75] Maier-Hein L, Reinke A, Kozubek M, *et al.* Metrics reloaded: recommendations for image analysis validation. *arXiv preprint*. 2022. <https://arxiv.org/abs/2206.01653>
- [76] Kirillov A, He K, Girshick R, Dollár P. Panoptic segmentation. *IEEE Conf Comput Vis Pattern Recognit (CVPR)*. 2019:9404–9413. <https://doi.org/10.1109/CVPR.2019.00963>
- [77] de Geus D, Meletis P, Wen X, *et al.* Panoptic quality: a principled metric for panoptic segmentation. *arXiv preprint*. 2021. <https://arxiv.org/abs/2109.03814>