



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Εύρεση Βιοδεικτών σε Εικόνες Μαγνητικού Τομογράφου για την Διάγνωση του Καρκίνου των Ωοθηκών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Τζανακάκη Πέτρου

Επιβλέπων: Γεώργιος Ματσόπουλος

Καθηγητής, Ε.Μ.Π

Αθήνα, Μάρτιος 2021



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Εύρεση Βιοδεικτών σε Εικόνες Μαγνητικού Τομογράφου για την Διάγνωση του Καρκίνου των Ωοθηκών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Τζανακάκη Πέτρου

Επιβλέπων: Γεώργιος Ματσόπουλος
Καθηγητής, Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 22α Μαρτίου 2021.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Γεώργιος Ματσόπουλος
Καθηγητής, Ε.Μ.Π

.....
Κωνσταντίνα Νικήτα
Καθηγήτρια, Ε.Μ.Π.

.....
Δημήτριος-Διονύσιος Κουτσούρης
Καθηγητής, Ε.Μ.Π

Αθήνα, Μάρτιος 2021



ΕΘΝΙΚΟ ΜΕΤΕΩΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

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

Πέτρος Τζανακάκης, 2021.

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

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

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

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

(Υπογραφή)

.....
Πέτρος Τζανακάκης

Περίληψη

Τα Συστήματα Υποστήριξης Κλινικών Αποφάσεων ή Clinical Decision Support Systems (CDSS) είναι ολοκληρωμένα συστήματα ανάλυσης και ανάπτυξης, σχεδιασμένα για να διευκολύνουν τη λήψη αποφάσεων που αφορά την υγειονομική περίθαλψη ασθενών. Συνδυάζουν τρέχουσες πληροφορίες των ασθενών με εκείνες προηγούμενων διαγνώσεων και θεραπειών. Η αποθήκευση τους γίνεται σε βάση δεδομένων με σκοπό να παρέχουν ανατροφοδότηση ή συστάσεις που θα βοηθήσουν στη διαδικασία λήψης αποφάσεων στο χώρο της περίθαλψης. Στην κατηγορία αυτή ανήκει και το ερευνητικό έργο Destiny, στόχος του οποίου είναι η ανάπτυξη ενός αυτόνομου συστήματος διάγνωσης του Καρκίνου των Ωοθηκών (ΚΩ) που θα συνδυάζει πληροφορίες από την εξαγωγή ραδιολογικών χαρακτηριστικών μέσω Μαγνητικής Τομογραφίας (MT) με τα γονιδιώματα και το οποίο θα επιτρέπει την αυτόματη κατηγοριοποίηση της επικινδυνότητας των περιοχών του όγκου των ωοθηκών σε τρεις κατηγορίες (μικρή, μέση, μεγάλη). Τα δεδομένα που χρησιμοποιήθηκαν στην εργασία, προήλθαν από ασθενείς του τμήματος Γυναικολογικής Ογκολογίας της Α' Μαιευτικής Γυναικολογικής Κλινικής Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών (ΕΚΠΑ), του Νοσοκομείου Αλεξάνδρα ενώ η συλλογή και κλινική επεξεργασία τους πραγματοποιήθηκε στο Α' Εργαστήριο Ακτινολογίας της Ιατρικής Σχολής του ΕΚΠΑ, το οποίο στεγάζεται στο Αρεταίειο Νοσοκομείο. Για τα πειράματα έγινε χρήση ενός συστήματος απεικόνισης μαγνητικού συντονισμού (ΑΜΣ), κλειστού τύπου με στατικό μαγνητικό πεδίο H_0 .

Στόχος της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη ενός εργαλείου που (α) θα εξάγει ραδιολογικά χαρακτηριστικά ως δυνητικούς βιοδείκτες (β) θα επιλέγει-ελαχιστοποιεί τα χαρακτηριστικά εκείνα που ταξινομούν βέλτιστα τα δεδομένα στις προαναφερθείσες τρεις κατηγορίες στην βάση αλγορίθμων επιλογής χαρακτηριστικών. Τα δεδομένα που χρησιμοποιούνται ως είσοδος στο εργαλείο έχουν προκύψει σε πρότερο στάδιο του έργου και συνίστατο στις προβολές φωτεινότητας των τριών περιοχών του όγκου αποκλειστικά. Καθότι το εν εξελίξει έργο δεν έχει ολοκληρωθεί προκύπτουν συνεχώς νέα δεδομένα, ωστόσο στη τρέχουσα διπλωματική εργασία έχουν χρησιμοποιηθεί δεδομένα έξι ασθενών. Η χρήση των αλγορίθμων επιλογής και ταξινόμησης χαρακτηριστικών έγινε συνδυαστικά καθότι η φύση της εργασίας είναι διερευνητική και αφορά την στατιστική, υπολογιστική επαλήθευση των ορίων των περιοχών των κλινικών πειραμάτων. Επιλέχθηκαν οι συνδυασμοί εκείνοι που παρέχουν τα πιο αποδοτικά και εύρωστα αποτελέσματα συσχέτισης των περιοχών με τους αντίστοιχους βιοδείκτες. Τέλος βάσει των συμπερασμάτων καθώς και των περιορισμών της φύσης των δεδομένων και της υπολογιστικής διαδικασίας έγιναν προτάσεις για μελλοντική έρευνα.

Λέξεις Κλειδιά

καρκίνος των ωοθηκών, ραδιολογικά χαρακτηριστικά, εύρεση βιοδεικτών, επιλογή/μείωση χαρακτηριστικών, μηχανική μάθηση, νευρωνικά δίκτυα, ταξινόμηση, MRI, pyradiomics, scikit-learn

Abstract

Clinical Decision Support Systems (CDSS) are integrated analysis and deployment systems designed to facilitate decision-making in patient health care. They combine information about the current patients with information about past diagnoses and treatments stored in a database to provide feedback or recommendations that will aid in decision-making process at the point of care. This category includes the research project Destiny, which aims to develop an autonomous system for the diagnosis of Ovarian Cancer (OC). Destiny combines information from the extraction of radiomic features by MRI with the genomes, allowing the automatic categorization of the risk of ovarian tumor areas into three categories (small, medium, high). The data used in this study came from patients of the Department of Gynecological Oncology of the 1st Obstetrics and Gynecology Clinic of the National and Kapodistrian University of Athens (EKPA), Alexandra Hospital. Simultaneously, their collection and clinical processing took place in the 1st Laboratory of Radiology, which is housed in Aretaio Hospital. A closed-type magnetic resonance imaging (AMS) imaging system with a static magnetic field H_0 was used for the experiments.

This diploma thesis aims to develop a tool that (a) will extract radiomic features as potential biomarkers (b) will select-minimize those features that optimally classify the data into the above three categories based on feature selection algorithms. This input data was obtained at an earlier stage of the project and consisted of the brightness projections of the three areas of the volume exclusively. As the ongoing project is not completed, new data are continually emerging; however, in the current diploma thesis, the data comes from six patients. s, the nature of this project, is exploratory; a combined formula of feature selection and classification algorithms were used for the statistical and computational verification of the boundaries of the areas of clinical experiments. Those combinations that provide the most efficient and robust results which associate the regions with the respective biomarkers were selected. Finally, suggestions were made for future research based on the conclusions as well as the limitations of the nature of the data and the computational process in total.

Keywords

tumor of the ovary, radiomics, bio-markers discovery, feature selection/reduction, machine learning, neural networks, classification, MRI, pyradiomics, scikit-learn

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή Γεώργιο Ματσόπουλο για την επίβλεψη αυτής της διπλωματικής εργασίας και για την ευκαιρία που μου έδωσε να ασχοληθώ με ένα πραγματικά ενδιαφέρον θέμα της υπολογιστικής βιοϊατρικής.

Επίσης θα ήθελα να ευχαριστήσω ιδιαίτερα τον διδάκτορα Θοδωρή Οικονομόπουλο για την καθοδήγησή του καθ' όλη την διάρκεια της εκπόνησης της εργασίας και την εξαιρετική συνεργασία που είχαμε και τις καίριες συμβουλές που μου έδωσε.

Τέλος θα ήθελα να ευχαριστήσω την οικογένεια μου, τους γονείς μου και τον αδερφό μου, τους φίλους μου για την πολύμορφη και συχνά ετερόκλητη επίδραση τους πάνω μου, η οποία και συνέβαλε στην διαμόρφωση και εξέλιξη μου έως σήμερα.

Αθήνα, Μάρτιος 2021

Πέτρος Τζανακάκης

Περιεχόμενα

Περίληψη	1
Abstract	3
Ευχαριστίες	5
1 Εισαγωγή	19
1.1 Σκοπός Διπλωματικής Εργασίας	20
1.2 Διάρθρωση της Διπλωματικής Εργασίας	21
2 Ιατρο/Βιολογικά Δεδομένα του Καρκίνου των Ωοθηκών	23
2.1 Κλινικά και Γενετικά Χαρακτηριστικά του Καρκίνου των Ωοθηκών	23
2.2 Σταδιοποίηση-Εξέλιξη της Νόσου	24
2.3 Τεχνικές Διάγνωσης του Καρκίνου Ωοθηκών	25
2.4 Καταγραφή Τεχνικών/Ευρημάτων από τη Γενετική Ανάλυση του Καρκίνου των Ωοθηκών	26
2.5 Περιπτώσεις Χρήσης και Κριτήρια Επιλογής του Δείγματος - Χρηστών	27
3 Εξαγωγή Ποσοτικών Πληροφοριών από Ιατρικές Εικόνες (Radiomics)	31
3.1 Ραδιολογικά Χαρακτηριστικά ως Δυνητικοί Βιοδείκτες	31
3.1.1 Χαρακτηριστικά σχήματος (Shape features)	33
3.1.2 Χαρακτηριστικά βάσει ιστογράμματος ή πρώτης τάξης χαρακτηριστικά (Histogram-based features or first-order statistics)	34
3.1.3 Χαρακτηριστικά υφής ή στατιστικά δεύτερης τάξης (Textural features or second-order statistics)	35
3.1.4 Χαρακτηριστικά στατιστικών υψηλότερης τάξης (Higher-order statistics features)	46
3.2 Ραδιολογικά Χαρακτηριστικά - Τεχνικές Βαθιάς Μάθησης	46
3.3 Παραδοχή	48
4 Μηχανική Μάθηση	49
4.1 Εισαγωγή	49
4.1.1 Επιβλεπόμενη Μάθηση	49
4.1.2 Μη Επιβλεπόμενη Μάθηση	50
4.1.3 Ενισχυτική Μάθηση	50
4.2 Ταξινομητές Μηχανικής Μάθησης	50

4.2.1	Γραμμικά Μοντέλα - Linear Models	51
4.2.2	Μοντέλα Συνόλων - Ensemble models	51
4.2.3	Μηχανές Διανυσμάτων Υποστήριξης - (Support Vector Machines (SVM))	54
4.2.4	Πολυεπίπεδα Νευρωνικά Δίκτυα (Multilayer Perceptrons (MLP))	56
4.2.5	Δένδρα Αποφάσεων (Decision Trees)	58
5	Πειραματική Διαδικασία και Μεθοδολογία Επεξεργασίας και Ανάλυσης Δεδομένων	61
5.1	Σχεδίαση Κλινικού Πρωτοκόλλου	61
5.1.1	Κριτήρια Εισαγωγής και Εξαίρεσης Ασθενών	61
5.1.2	Περιγραφή Διαδικασίας Κλινικού Χειρισμού Ασθενών	62
5.1.3	Πρωτόκολλο Μαγνητικής Τομογραφίας (MT)	63
5.1.4	Διαδικασία Επισήμανσης Δειγμάτων	63
5.1.5	Παθολογοανατομική Ανάλυση Δειγμάτων	64
5.2	Πειραματική Διαδικασία	64
5.2.1	Συλλεχθέντα Δεδομένα	64
5.2.2	Περιγραφή Υλικού Συλλογής Δεδομένων	65
5.2.3	Λειτουργικές Προδιαγραφές Λογισμικού για τα Δεδομένα Εισόδου	67
5.3	Προεπεξεργασία Δεδομένων Εισόδου	67
5.3.1	Ευθυγράμμιση Τρισδιάστατων Ιατρικών Δεδομένων - Registration	68
5.3.2	Κατάτμηση Τρισδιάστατων Ιατρικών Δεδομένων - Segmentation	71
5.3.3	Ανακατασκευή, Σύντηξη και Απεικόνιση Τρισδιάστατων Ιατρικών Δεδομένων - Reconstruction, Subtraction, Visualization	72
5.4	Εξαγωγή Χαρακτηριστικών - Feature Extraction	72
5.4.1	Κλάσεις Χαρακτηριστικών	73
5.4.2	Κλάσεις Φίλτρων	74
5.4.3	Επιλογή Χαρακτηριστικών και Ταξινόμηση	75
6	Αποτελέσματα	87
6.1	Απόδοση Ταξινομητών - Ασθενής 6	87
6.2	Απόδοση Ταξινομητών - Ασθενής 12	89
6.3	Απόδοση Ταξινομητών - Ασθενής 16	92
6.4	Απόδοση Ταξινομητών - Ασθενής 18	94
6.5	Απόδοση Ταξινομητών - Ασθενής 22	97
6.6	Απόδοση Ταξινομητών - Ασθενής 23	99
6.7	Απόδοση Ταξινομητών - Ασθενής 27	102
6.8	Απόδοση Ταξινομητών - Ασθενής 28	104
6.9	Απόδοση Ταξινομητών - Ασθενής 29	107
6.10	Απόδοση Ταξινομητών - Ασθενής 30	109
6.11	Στατιστική Αξιολόγηση - Τυπική Απόκλιση	112

7 Συζήτηση	115
7.1 Συμπεράσματα	115
7.1.1 Προεπεξεργασία Δεδομένων	115
7.1.2 Αλγόριθμοι Επιλογής Χαρακτηριστικών	116
7.1.3 Ταξινομητές	116
7.2 Περιορισμοί	117
7.2.1 Φύση των δεδομένων	117
7.2.2 Υπολογιστική διαδικασία	117
7.3 Μελλοντικές Προεκτάσεις	118
Βιβλιογραφία	129

Κατάλογος Σχημάτων

2.1	Ταξινόμηση κατά Fίγο για Καρκίνο Ωοθηκών	25
2.2	Αναλυτική λίστα των γονιδίων που θα αναλυθούν μέσω ΝΓΣ για την ανίχνευση γαμετικών μεταλλάξεων.	27
2.3	Αναλυτική λίστα των γονιδίων που θα αναλυθούν μέσω ΝΓΣ για την ανίχνευση ενδο-ογκικών μεταλλάξεων.	27
4.1	Μοντέλα Συνόλων	52
4.2	Υπερεπίπεδα διαχωρισμού. Πλήθος υπερεπιπέδων διαχωρισμού με διαφορετικό λειτουργικό περιθώριο το καθένα.	55
4.3	Υπερεπίπεδα διαχωρισμού. Αναπαράσταση μεγιστοποίησης της γεωμετρικής απόστασης μεταξύ εγγύτερων σημείων των κλάσεων και υπερεπιπέδου.	56
4.4	Δομή ενός ταξινομητή MLP.	57
5.1	Σύστημα Απεικόνισης Μαγνητικού Συντονισμού το οποίο χρησιμοποιήθηκε για τη συλλογή των δεδομένων του έργου.	65
5.2	Διαδικασία κλιμακωτής ευθυγράμμισης απεικονιστικών δεδομένων βάσει αρχικού συμπαγούς μετασχηματισμού και δευτερεύοντος ελαστικού μετασχηματισμού.	70
5.3	Τρισδιάστατες απεικονίσεις από ανακατασκευή δεδομένων μαγνητικής τομογραφίας, μέσω τεχνικών απόδοσης όγκου.	73
5.4	2-D Ορθογώνια Κυμάτια.	75
6.1	scikit-feature . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	89
6.2	scikit-feature . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	92
6.3	scikit-feature . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	94
6.4	scikit-feature . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	97
6.5	scikit-feature . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	99
6.6	scikit-feature . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	102
6.7	scikit-feature . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	104

- 6.8 **scikit-feature**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες). 107
- 6.9 **scikit-feature**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες). 109
- 6.10 **scikit-feature**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες). 112

Κατάλογος Εικόνων

1.1	Γενική δομή Συστήματος Υποστήριξης Κλινικών Αποφάσεων Destiny.	20
6.1	scikit-feature - 100 επιλεγέντα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	87
6.2	scikit-learn. Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).	88
6.3	colin-mico. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	88
6.4	scikit-feature - 200 επιλεγέντα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	89
6.5	scikit-feature - 100 επιλεγέντα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	90
6.6	scikit-learn. Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).	90
6.7	colin-mico. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	91
6.8	scikit-feature - 200 επιλεγέντα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	91
6.9	scikit-feature - 100 επιλεγέντα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	92
6.10	scikit-learn. Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).	93
6.11	colin-mico. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	93
6.12	scikit-feature - 200 επιλεγέντα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	94
6.13	scikit-feature - 100 επιλεγέντα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	95

6.31	colin-mico . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	106
6.32	scikit-feature - 200 επιλεγέντα χαρακτηριστικά . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	106
6.33	scikit-feature - 100 επιλεγέντα χαρακτηριστικά . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	107
6.34	scikit-learn . Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).	108
6.35	colin-mico . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	108
6.36	scikit-feature - 200 επιλεγέντα χαρακτηριστικά . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	109
6.37	scikit-feature - 100 επιλεγέντα χαρακτηριστικά . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	110
6.38	scikit-learn . Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).	110
6.39	colin-mico . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	111
6.40	scikit-feature - 200 επιλεγέντα χαρακτηριστικά . Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).	111

Κατάλογος Πινάκων

5.1	Πρωτόκολλο ακολουθιών παλμών Μαγνητικής Τομογραφίας για προεγχειρητική εκτίμηση ασθενών με Καρκίνο Ωοθηκών (DESTINY)	82
5.2	Πρωτόκολλο ακολουθιών παλμών Μαγνητικής Τομογραφίας μεταθεραπευτικής παρακολούθησης ασθενών με Καρκίνο Ωοθηκών (DESTINY)	83
5.3	Πρωτόκολλο ακολουθιών παλμών Μαγνητικής Τομογραφίας εκτίμησης άνω κοιλίας για τον εντοπισμό μεταστάσεων εκτός πυέλου, ασθενών με Καρκίνο Ωοθηκών (DESTINY)	84
5.4	Χαρακτηριστικά: Πρώτης τάξης, Σχήματος, GLCM, GLSZM.	85
5.5	Χαρακτηριστικά: GLRLM, NGTDM, GLDM.	86
6.1	scikit-feature - Επιβλεπόμενη Μάθηση. Τυπική Απόκλιση (Standard Deviation) προγνωστικής ακρίβειας της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) της κάθε ασθενούς (σε σειρές).	112
6.2	colin-mico - Επιβλεπόμενη Μάθηση. Τυπική Απόκλιση (Standard Deviation) προγνωστικής ακρίβειας της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) της κάθε ασθενούς (σε σειρές).	113
6.3	scikit-learn - Επιβλεπόμενη Μάθηση. Τυπική Απόκλιση (Standard Deviation) προγνωστικής ακρίβειας της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) της κάθε ασθενούς (σε σειρές).	113
6.4	scikit-feature - Μη Επιβλεπόμενη Μάθηση. Τυπική Απόκλιση (Standard Deviation) προγνωστικής ακρίβειας της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) της κάθε ασθενούς (σε σειρές).	113

Κεφάλαιο 1

Εισαγωγή

Τα Συστήματα Υποστήριξης Κλινικών Αποφάσεων χρησιμοποιούνται ευρέως σε πολλές εφαρμογές της ιατρικής με σκοπό την υποβοήθηση της διαγνωστικής διαδικασίας, όπως αυτή πραγματοποιείται κατά περίπτωση από το εξειδικευμένο ιατρικό προσωπικό. Με βάση επιλεγμένα χαρακτηριστικά των εκάστοτε δεδομένων (απεικονιστικών και μη) και κατάλληλης διαδικασίας εκπαίδευσης τα συστήματα αυτά είναι σε θέση να κατηγοριοποιούν με καλή ακρίβεια τα δείγματα εισόδου σε δύο ή περισσότερες κατηγορίες [1]. Ένα τυπικό πλήρες Σύστημα Υποστήριξης Κλινικών Αποφάσεων αποτελείται από τρία υποσυστήματα σε σειριακή διάταξη :

- Τη μονάδα Εξαγωγής Χαρακτηριστικών - Feature Extraction
- Τη μονάδα Επιλογής Χαρακτηριστικών - Feature Selection
- Τη μονάδα Κατηγοριοποίησης - Classification

Η μονάδα Εξαγωγής Χαρακτηριστικών δημιουργεί το αρχικό διάνυσμα χαρακτηριστικών εισόδου το οποίο περιγράφει τα δείγματα εισόδου (patterns). Η μονάδα Επιλογής Χαρακτηριστικών δημιουργεί το τελικό διάνυσμα χαρακτηριστικών, απομακρύνοντας από το αρχικό τα περιττά χαρακτηριστικά. Το διάνυσμα αυτό χρησιμοποιείται ως είσοδος στη μονάδα κατηγοριοποίησης. Η μονάδα Κατηγοριοποίησης πραγματοποιεί τον διαχωρισμό των δεδομένων εισόδου στις κατηγορίες εξόδου.

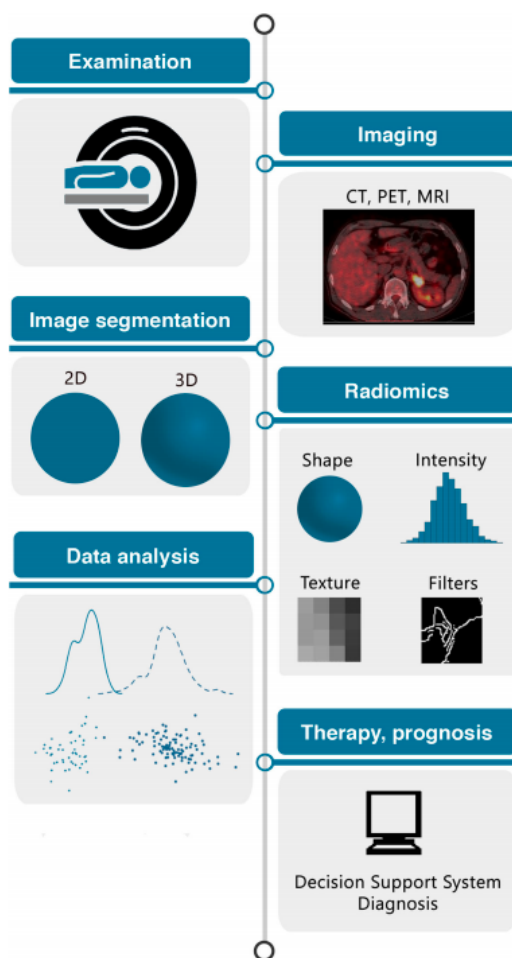
Οι πιο κρίσιμες παράμετροι για τον ποιοτικό σχεδιασμό ενός Συστήματος Υποστήριξης Κλινικών Αποφάσεων είναι το αν είναι εκ των προτέρων γνωστές ή όχι οι κατηγορίες εξόδου των δειγμάτων, το πλήθος των κατηγοριών εξόδου, το πλήθος των διαθέσιμων δειγμάτων εισόδου, το πλήθος των χαρακτηριστικών εισόδου, η ποιότητα των χαρακτηριστικών εισόδου και εκπροσώπησης κάθε κατηγορίας εξόδου στο διαθέσιμο δείγμα [2].

Στην περίπτωση του έργου οι κατηγορίες θα είναι εκ των προτέρων γνωστές καθώς καθορίζονται από την ανάλυση εξειδικευμένου ιατρικού προσωπικού. Ο ακριβής αριθμός των κατηγοριών εξόδου δεν είναι γνωστός σε αυτή τη φάση του έργου και εξαρτάται από τη διαίρεση σε στάδια της νόσου που θα προκύψει κατά την έρευνα. Ωστόσο, στην ελάχιστη περίπτωση είναι δύο : Υγιής και Ασθενής.

Αντίστοιχα, το πλήθος των διαθέσιμων δειγμάτων και το ποσοστό εκπροσώπησης κάθε κατηγορίας εξόδου στο διαθέσιμο δείγμα δεν είναι γνωστά, και εξαρτώνται από το αποτέλεσμα

της έρευνας που θα πραγματοποιηθεί σε προηγούμενα στάδια του έργου. Τα χαρακτηριστικά εισόδου αποτελούνται από μετρήσιμες ποσότητες οι οποίες εξάγονται, άμεσα ή έμμεσα, από τα υπό επεξεργασία δεδομένα (τριδιάστατα απεικονιστικά δεδομένα στην περίπτωση του έργου). Στην περίπτωση της επεξεργασίας ιατρικών εικόνων, η εξαγωγή των χαρακτηριστικών αυτών γίνεται δυνατή μέσω της εφαρμογής αλγορίθμων ευθυγράμμισης, σύντηξης και τμηματοποίησης, έτσι ώστε να απομονωθούν οι επιθυμητές περιοχές ενδιαφέροντος και να αναλυθούν κατάλληλα.

Η γενική δομή του προτεινόμενου συστήματος περιγράφεται από το διάγραμμα της Εικόνας 1.1



Εικόνα 1.1: Γενική δομή Συστήματος Υποστήριξης Κλινικών Αποφάσεων Destiny.

1.1 Σκοπός Διπλωματικής Εργασίας

Στα πλαίσια της παρούσας διπλωματικής εργασίας αναπτύχθηκε ένα καινοτόμο Σύστημα Υποστήριξης Κλινικών Αποφάσεων το οποίο περιλαμβάνει και τα τρία προαναφερθέντα υποσυστήματα. Στην υλοποίηση έγινε χρήση βιβλιοθηκών και εργαλείων τα οποία χρησιμοποιούνται κατά κόρον από την επιστημονική κοινότητα και αποτελούν αιχμή των σύγχρονων ερευνητικών προσεγγίσεων. Δοκιμάσαμε πληθώρα μοντέλων και αλγορίθμων ιδίως στις μο-

νάδες επιλογής χαρακτηριστικών και κατηγοριοποίησης, ενώ διερευνήθηκαν διεξοδικά παράμετροι των μοντέλων του συστήματος και επιλέχθηκαν εκείνες οι οποίες βελτιστοποιούν την απόδοση του.

1.2 Διάρθρωση της Διπλωματικής Εργασίας

Η εργασία αυτή είναι οργανωμένη σε επτά κεφάλαια: Στο Κεφάλαιο 2, δίνεται το ιατρο-/βιολογικό θεωρητικό υπόβαθρο των δεδομένων που σχετίζονται με τον καρκίνο των ωοθηκών. Αρχικά παρουσιάζονται τα κλινικά και γενετικά χαρακτηριστικά της νόσου και εν συνεχεία η εξέλιξη της καθώς και κλινικές τεχνικές διάγνωσης. Στο Κεφάλαιο 3, δίδονται οι βασικές θεωρητικές αρχές της ραδιωματικής πάνω στις οποίες βασίζεται και η υπολογιστική ανάπτυξη του έργου. Στο Κεφάλαιο 4, γίνεται μία επισκόπηση της θεωρίας της μηχανικής μάθησης, επικεντρώνοντας στην λειτουργία των μοντέλων τα οποία χρησιμοποιήθηκαν στην παρούσα εργασία. Στο Κεφάλαιο 5, παρουσιάζεται η πειραματική διαδικασία που ακολουθήθηκε συνολικά στο έργο και αναλύεται η σχεδίαση του συστήματος υποστήριξης αποφάσεων, δηλαδή η περιγραφή των υποσυστημάτων και των εφαρμογών του. Η περιγραφή της υλοποίησης του συστήματος, αφορά την ανάλυση των βασικών αλγορίθμων καθώς και λεπτομέρειες σχετικά με τις πλατφόρμες και τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν. Στο Κεφάλαιο 6, παρουσιάζεται ο έλεγχος της λειτουργίας του συστήματος με βάση τα αποτελέσματα μιας σειράς υπολογιστικών πειραμάτων. Τέλος στο Κεφάλαιο 7, διατυπώνονται τα συμπεράσματα τα οποία προέκυψαν από τις παραπάνω προσομοιώσεις, περιορισμοί τόσο της φύσης των δεδομένων όσο και της υπολογιστικής διαδικασίας, καθώς και μελλοντικές επεκτάσεις.

Κεφάλαιο 2

ΙΑΤΡΟ/ΒΙΟΛΟΓΙΚΑ ΔΕΔΟΜΕΝΑ ΤΟΥ ΚΑΡΚΙΝΟΥ ΤΩΝ ΩΟΘΗΚΩΝ

Ο καρκίνος των ωοθηκών είναι ένας κακοήθης όγκος που προσβάλλει τη μία ή και τις δύο ωοθήκες. Υπάρχουν δύο τύποι: (α) ο επιθηλιακός καρκίνος των ωοθηκών, που είναι ο πιο συχνός και οφείλεται στη μη φυσιολογική ανάπτυξη και αναπαραγωγή των επιθηλιακών κυττάρων που βρίσκονται στην επιφάνεια της ωοθήκης και (β) ο μη επιθηλιακός καρκίνος των ωοθηκών, που μπορεί να αναπτυχθεί από τα γεννητικά αναπαραγωγικά κύτταρα ή από τα στρωματικά κύτταρα του συνδετικού ιστού, γύρω από την ωοθήκη [3].

2.1 Κλινικά και Γενετικά Χαρακτηριστικά του Καρκίνου των Ωοθηκών

Ο καρκίνος των ωοθηκών (ΚΩ) κατατάσσεται στην 7^η θέση ως προς το σύνολο των καρκίνων που διαγιγνώσκονται στις γυναίκες [4] και αποτελεί τον τέταρτο πιο θανατηφόρο καρκίνο στον γυναικείο πληθυσμό [5]. Ο κίνδυνος να αναπτύξει καρκίνο μια γυναίκα κατά τη διάρκεια της ζωής της (estimated lifetime risk) κυμαίνεται από 1 στις 54 έως 1 στις 70, ενώ η πιθανότητα να πεθάνει από τη νόσο είναι 1 στις 100 [3]. Οι ασθενείς τυπικά προσέρχονται σε προχωρημένο στάδιο με τον 5-ετη σχετικό δείκτη επιβίωσης να κυμαίνεται μεταξύ του 30%-40% σε διεθνές επίπεδο, με ελάχιστη βελτίωση (2%-4%) τις τελευταίες δεκαετίες [6], κάτι το οποίο αποτελεί σημαντικό κοινωνικό-οικονομικό πρόβλημα.

Ο καρκίνος των ωοθηκών θεωρείται ένας «σιωπηρός» καρκίνος. Ο επιθηλιακός καρκίνος των ωοθηκών παρουσιάζεται με ευρύ φάσμα μη ειδικών συμπτωμάτων. Στα αρχικά στάδια στερείται πρακτικά κλινικής συμπτωματολογίας. Κατά την πρόοδο της νόσου τα συμπτώματα συνήθως συνίστανται σε πυελικό άλγος, μετεωρισμό/διάταση της κοιλίας, πιεστικά φαινόμενα από την ουροδόχο κύστη (π.χ. έπειξη για ούρηση) και το παχύ έντερο (δυσκοιλιότητα), διάρροια, κολπική αιμορραγία, δυσπεψία, ανορεξία, γαστροοισοφαγική παλινδρόμηση, δύσπνοια, εύκολη κόπωση, κακουχία και απώλεια βάρους. Παρανεοπλασματικά σύνδρομα μπορεί να οδηγήσουν σε διάφορες κλινικές εκδηλώσεις, όπως για παράδειγμα θρόμβωση των κάτω άκρων. Ευρήματα στην κλινική εξέταση είναι ασυνήθη στα αρχικά στάδια της νόσου. Ασθενείς σε προχωρημένο στάδιο μπορεί να παρουσιάσουν ψηλαφητή πυελική ή ενδοκοιλιακή μάζα, βατραχοειδή κοιλία (λόγω ασκίτη), πλευριτική συλλογή, ή σημεία ειλεού λόγω εντερικής απόφραξης. Η παρουσία οιδήματος στα κάτω άκρα, συνήθως απότοκη

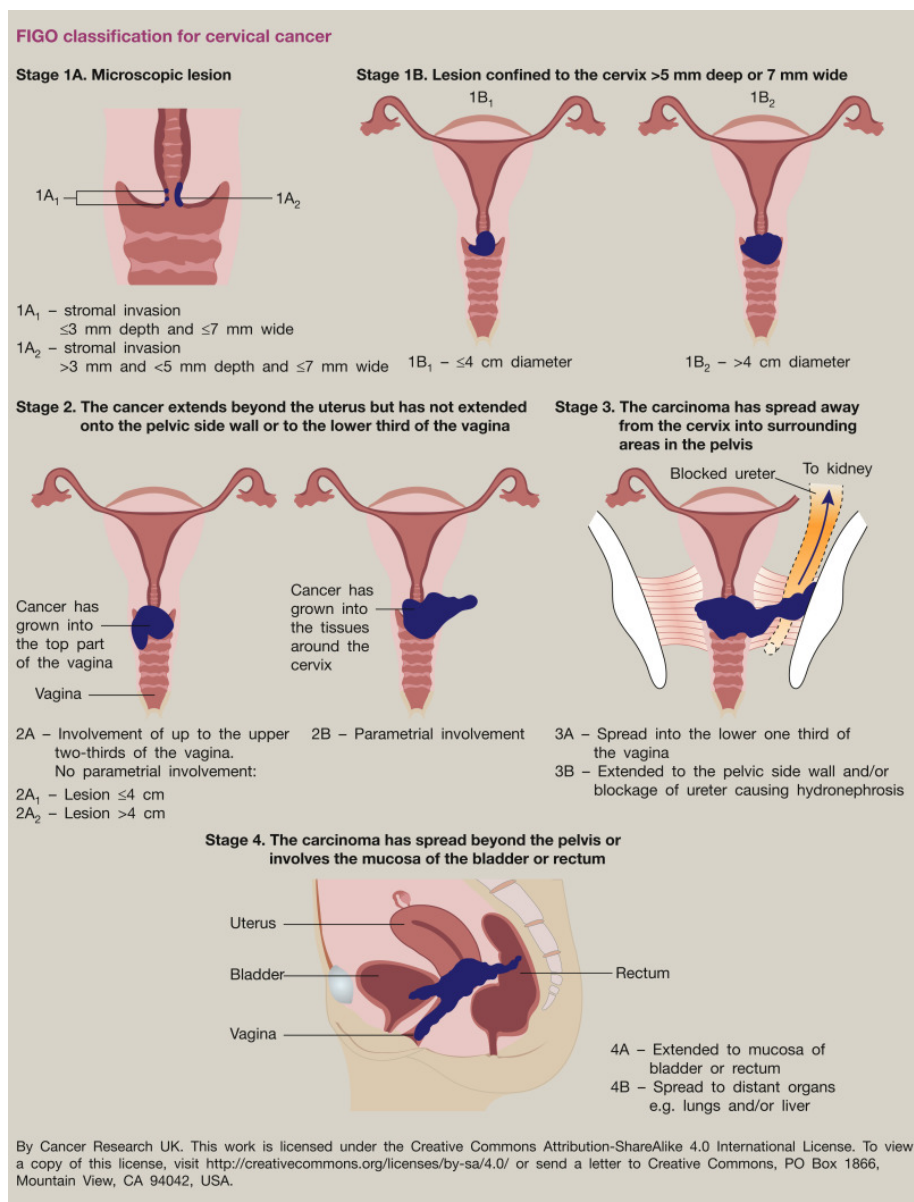
φλεβοθρόμβωσης, δεν είναι σπάνια. Επίσης, μπορεί να παρουσιαστούν ψηλαφητές λεμφαδενικές διογκώσεις στις βουβωνικές, μασχαλιαίες ή υπερκλείδιες χώρες [4]. Η συντριπτική πλειοψηφία των κακοήθων νεοπλασμάτων της ωοθήκης (70%) ανήκουν στην κατηγορία του ορώδους ΚΩ υψηλού βαθμού κακοήθειας (high-grade serous ovarian carcinoma - (HG-SOC)), ο οποίος ανήκει στους επιθηλιακού τύπου καρκίνους που, στην πλειονότητα των περιπτώσεων, ξεκινούν από το επιθήλιο των κροσσών των σαλπίνγων (serous tubal intraepithelial carcinoma theory - (STIC)) [7]. Τυπικά, η θεραπευτική αντιμετώπιση συνίσταται σε κυτταρομειωτική χειρουργική, ακολουθούμενη από χημειοθεραπεία με πλατίνα και ταξάνη. Ωστόσο, το ποσοστό υποτροπής παραμένει υψηλό (70%) εντός 3 ετών από την αρχική διάγνωση. Η σημαντική μοριακή ετερογένεια του αρχικού όγκου θεωρείται πλέον η πιο πιθανή εξήγηση για την αντίσταση στη θεραπεία [8].

Στους επιθηλιακούς όγκους των ωοθηκών εντοπίζεται ένα ευρύ φάσμα γενετικών βλαβών, που αφορούν ως επί το πλείστον γονίδια, τα πρωτεϊνικά παράγωγα των οποίων μετέχουν στη διαδικασία επιδιόρθωσης του DNA μέσω ομόλογου ανασυνδυασμού. Τα γονίδια BRCA1 και BRCA2 διαδραματίζουν κεντρικό ρόλο στη διαδικασία αυτή της επιδιόρθωσης του DNA, ενώ οι κληρονομικές μεταλλάξεις της γαμετικής σειράς στα γονίδια αυτά προδιαθέτουν στον καρκίνο των ωοθηκών. Σύμφωνα με τα μέχρι τώρα δεδομένα και ανάλογα με τον τρόπο επιλογής των ασθενών, υπολογίζεται ότι το 10 – 25% αυτών φέρουν γαμετικές παθογόνους παραλλαγές στα γονίδια BRCA1 και BRCA2, ενώ ένα επιπλέον 5 – 10% φέρει σωματικές μεταλλάξεις στα γονίδια αυτά. Ιδιαίτερο ενδιαφέρον παρουσιάζει το γεγονός ότι οι όγκοι που αναπτύσσονται σε ασθενείς που φέρουν τόσο γαμετικές, όσο και επίκτητες (σωματικές) μεταλλάξεις στα γονίδια BRCA1 και BRCA2 και ενδεχομένως και σε άλλα γονίδια, επιδεικνύουν ιδιαίτερη ευαισθησία σε φάρμακα που ονομάζονται αναστολείς της πολυμεράσης της πολυαδενοφωσφορικής ριβόζης (polyADP-ribose polymerase - (PARP)), δεδομένης της απώλειας λειτουργίας της επιδιόρθωσης του DNA μέσω ομόλογου ανασυνδυασμού [9, 10].

Με βάση τα παραπάνω, η διάγνωση του γονιδιακού προφίλ των ωοθηκικών όγκων έχει κεφαλαιώδη σημασία για την κλινική διαχείριση της νόσου, καθώς οι ασθενείς με μεταλλάξεις στα γονίδια του ομόλογου ανασυνδυασμού, για παράδειγμα, μπορούν να ωφεληθούν από νεότερες στοχευμένες θεραπείες, όπως οι αναστολείς PARP και να επιδείξουν καλύτερη ανταπόκριση και επιβίωση [11].

2.2 Σταδιοποίηση-Εξέλιξη της Νόσου

Η σταδιοποίηση κατά FIGO (Federation International of Gynecology and Obstetrics) παραμένει ο ισχυρότερος προγνωστικός δείκτης της πορείας της νόσου [12]. Αν και η σταδιοποίηση είναι χειρουργική, η προεγχειρητική εκτίμηση με τις νεότερες απεικονιστικές μεθόδους είναι σημαντική καθώς χαρτογραφεί την έκταση της νόσου και καθοδηγεί την χειρουργική προσπέλαση. Η λήψη βιοψίας για ιστολογική επιβεβαίωση της νόσου είναι απαραίτητη. Σε περιπτώσεις που η χειρουργική επέμβαση δεν είναι δυνατή (π.χ. ασθενείς αυξημένου περιεγχειρητικού κινδύνου), θα πρέπει να λαμβάνονται βιοψίες υπό απεικονιστική καθοδήγηση ή λαπαροσκοπικά. Τα στάδια κατά FIGO σε ασθενείς με υψηλόβαθμο επιθηλιακό καρκίνο ωοθηκών παρουσιάζονται αδρά (Σχ.2.1). Το ίδιο σύστημα σταδιοποίησης εφαρμόζεται και στις περιπτώσεις πρωτοπαθούς καρκίνου σάλπιγγας και πρωτοπαθούς



Σχήμα 2.1: Ταξινόμηση κατά FIGO για Καρκίνο Ωοθηκών

καρκίνου του περιτοναίου.

2.3 Τεχνικές Διάγνωσης του Καρκίνου Ωοθηκών

Εκτός της κλινικής εξέτασης, ως μέθοδος ρουτίνας για την διάγνωση του καρκίνου των ωοθηκών χρησιμοποιείται η μέτρηση του καρκινικού βιοδείκτη CA 125 . Ωστόσο, η αξία του CA 125 στην ανίχνευση του πρώιμου διηθητικού καρκίνου των ωοθηκών αμφισβητείται καθώς αυξάνει μόνο στο 50% των ασθενών με νόσο σταδίου I κατά FIGO. Σε ασθενείς με προχωρημένη νόσο, το CA 125 είναι αυξημένο στο 85% περίπου των ασθενών. Επίσης, ο συγκεκριμένος δείκτης δεν είναι ειδικός για τον καρκίνο των ωοθηκών και μπορεί να αυξηθεί και σε άλλους καρκίνους (π.χ. μαστού, πνεύμονα, παχέος εντέρου, παγκρέατος) αλλά και σε καλοήθεις νόσους (π.χ. ενδομητρίωση, ρήξη ωχρού σωματίου ή πυελική φλεγμονή).

Το υπερηχοτομογράφημα κοιλίας είναι συνήθως η πρώτη απεικονιστική μέθοδος, η οποία συστήνεται σε γυναίκες με υποψία καρκίνου ωοθηκών. Το διακολπικό υπερηχοτομογράφημα προσφέρει εξαιρετική απεικόνιση των ωοθηκών και βοηθά στη διαφορική διάγνωση καλοήθων από κακοήθεις νόσους. Η ευαισθησία του υπερηχοτομογραφήματος στη διάγνωση είναι πολύ υψηλή (88-100%), ωστόσο ποικίλει αρκετά η ειδικότητα (39-87%) Στην περίπτωση κακοήθειας, το υπερηχοτομογράφημα έχει σημαντικούς τεχνικούς περιορισμούς στην ανάδειξη της έκτασης και, κατ' επέκταση, στη σταδιοποίηση της νόσου. Οι περιορισμοί του υπερηχοτομογραφήματος αφορούν κυρίως στο σωματότυπο της εξεταζόμενης (παχυσαρκία), την παρουσία εντερικού περισταλτισμού και στην εμπειρία του εξεταστή.

Επί του παρόντος, απεικονιστική εξέταση εκλογής για την διερεύνηση της έκτασης της νόσου και τον χειρουργικό σχεδιασμό είναι η Υπολογιστική Τομογραφία (ΥΤ). Η ΥΤ είναι ευρέως διαθέσιμη και επιπλέον προσφέρει καλύτερη χωρική διακριτική ικανότητα από το υπερηχοτομογράφημα, κάτι που βοηθάει στην ανίχνευση των περιτοναϊκών μεταστάσεων οι οποίες είναι πολύ συχνές κατά τη διάγνωση της νόσου. Η διαγνωστική ακρίβεια της μεθόδου ποικίλει από 70-90% για όλα τα στάδια της νόσου [13, 14]. Αρχικά η Μαγνητική Τομογραφία (ΜΤ) χρησιμοποιήθηκε συμπληρωματικά της Υπολογιστικής Τομογραφίας, κυρίως λόγω περιορισμών που αφορούσαν στο κόστος και τη χρονική διάρκεια της εξέτασης, σε αλλοιώσεις που η τελευταία δεν μπορούσε να χαρακτηρίσει. Ωστόσο, τα τελευταία χρόνια, ιδιαίτερα λόγω της εισαγωγής λειτουργικών τεχνικών (ακολουθία μοριακής διάχυσης, Diffusion Weighted Imaging-DWI) έχει ολοένα και πιο ουσιαστικό ρόλο στην διάγνωση, τη σταδιοποίηση και την παρακολούθηση των ασθενών με καρκίνο ωοθηκών. Η ΜΤ έχει εξίσου υψηλή ευαισθησία στο χαρακτηρισμό των ωοθηκικών βλαβών (96.6%) και πολύ μεγαλύτερη ειδικότητα (39.5% έναντι 83.7%) συγκριτικά με το υπερηχοτομογράφημα. Η συνολική διαγνωστική ακρίβεια της μεθόδου ποικίλει από 78-88% και είναι παρόμοια με εκείνη της Υπολογιστικής Τομογραφίας [14, 15], γεγονός που την καθιστά αξιόπιστη μέθοδο για την εκτίμηση του ωοθηκικού καρκίνου.

2.4 Καταγραφή Τεχνικών/Ευρημάτων από τη Γενετική Ανάλυση του Καρκίνου των Ωοθηκών

Η γονιδιακή ανάλυση θα πραγματοποιηθεί με τη μέθοδο μαζικής παράλληλης αλληλούχισης (Next Generation Sequencing - (NGS)), τόσο για την ανάλυση των δειγμάτων γενωμικού DNA, όσο και του DNA που θα εξαχθεί από τους όγκους, στοχεύοντας στην μαζική ανάλυση πολλών γονιδίων συγχρόνως. Πιο συγκεκριμένα, στο ειδικό γονιδιακό πάνελ που αφορά την ανάλυση των δειγμάτων γενωμικού DNA θα αναλυθούν 94 γονίδια, μεταλλάξεις των οποίων προδιαθέτουν στον καρκίνο. Αναλυτικά τα γονίδια αυτά συνοψίζονται στην (Σχ.2.2).

Στην περίπτωση των ωοθηκικών όγκων, θα επιστρατευτεί ένα εκτεταμένο πάνελ γονιδίων, μέσω του οποίου δύναται να ταυτοποιηθούν μεταλλάξεις σε τουλάχιστον 50 γονίδια, τα οποία παίζουν σημαντικό ρόλο στη διάγνωση και πρόγνωση, αλλά προσφέρουν και σημαντικές πληροφορίες για το θεραπευτικό σχεδιασμό στη κλινική διαχείριση των ασθενών. Στην (Σχ.4.1) απεικονίζονται τα γονίδια προς ανάλυση.

Συνοπτικά, για τη διαδικασία της επόμενης γενιάς αλληλούχισης, αρχικά απομονώνεται

AIP	BUB1B	DDB2	EXT2	FANCL	MEN1	PALB2	RB1	SLX4	WRN
ALK	CDC73	DIGER1	EZH2	FANCM	MET	PHOX2B	RECQL4	SMAD4	WT1
APC	CDH1	DIS3L2	FANCA	FH	MLH1	PMS1	RET	SMARCB1	XPA
ATM	CDK4	EGFR	FANCB	FLCN	MSH2	PMS2	RHBDP2	STK11	XPC
BAP1	CDKN1C	EPCAM	FANCC	GATA2	MSH6	PRF1	RUNX1	SUFU	
BLM	CDKN2A	ERCC2	FANCD2	GPC3	MUTYH	PRKAR1A	SBDS	TMEM127	
BMPR1A	CEBPA	ERCC3	FANCE	HN1A	NBN	PTCH1	SDHAF2	TP53	
BRCA1	CEP57	ERCC4	FANCF	HRAS	NF1	PTEN	SDHB	TSC1	
BRCA2	CHEK2	ERCC5	FANCG	KIT	NF2	RAD51C	SDHC	TSC2	
BRIP1	CYLD	EXT1	FANCI	MAX	NSD1	RAD51D	SDHD	VHL	

Σχήμα 2.2: Αναλυτική λίστα των γονιδίων που θα αναλυθούν μέσω ΝΓΣ για την ανίχνευση γαμετικών μεταλλάξεων.

ATM	TP53	MSH6
BAP1	TSC1	NBN
BRCA1	TSC2	NOTCH2
BRCA2	ARID1A	NOTCH3
CDKN2A	ATR	PALB2
FBXW7	ATRX	PMS2
MSH2	CDK12	POLE
NF1	CDKN1B	RAD50
NF2	CDKN2B	RAD51
NOTCH1	CHEK1	RAD51B
PIK3R1	CREBBP	RAD51C
PTCH1	FANCA	RAD51D
PTEN	FANCD2	RNF43
RB1	FANCI	SETD2
SMARCB1	MLH1	SLX4
STK11	MRE11A	SMARCA4

Σχήμα 2.3: Αναλυτική λίστα των γονιδίων που θα αναλυθούν μέσω ΝΓΣ για την ανίχνευση ενδο-ογκικών μεταλλάξεων.

γενωμικό ή καρκινικό DNA και θρυμματίζεται σε μικρότερα τμήματα ενζυμικά με στόχο την δημιουργία βιβλιοθηκών με ζευγαρωμένα άκρα. Όσον αφορά τη διαχείριση των μονιμοποιημένων ιστών, αρχικά θα αξιολογείται το ποσοστό των καρκινικών κυττάρων, καθώς και το ποσοστό νέκρωσης. Το εξαγόμενο DNA πολλαπλασιάζεται επιλεκτικά με βάση τις αλληλουχίες-στόχους που έχουν επιλεγεί, οι οποίες συμπεριλαμβάνουν κατά το ελάχιστο, την πλήρη κωδικοποιούσα αλληλουχία των γονιδίων που έχουν επιλεγεί. Οι γονιδιωματικές βιβλιοθήκες που δημιουργούνται επεξεργάζονται στον αναλυτή επόμενης γενεάς Illumina MiSeq, ενώ η βιοπληροφορική ανάλυση πραγματοποιείται με δύο ανεξάρτητα λογισμικά, το εμπορικά διαθέσιμο, αλλά και το ειδικά ανεπτυγμένο in-house λογισμικό. Το βάθος ανάγνωσης (coverage) είναι κατά το ελάχιστο 100x για > 99% των εξονίων που αναλύονται και ο μέσος όρος ανάγνωσης εξονίων ανά δείγμα είναι > 500x. Κατά την αξιολόγηση και για λόγους ακριβείας το κατώφλι συχνότητας παραλλαγμένου αλληλομόρφου (variant allele frequency) τέθηκε στο 5%. Όλα τα ευρήματα που θα προκύψουν, θα ταυτοποιηθούν με ανεξάρτητη μέθοδο (αλληλούχιση κατά Sanger, MLPA, Real-Time PCR).

2.5 Περιπτώσεις Χρήσης και Κριτήρια Επιλογής του Δείγματος – Χρηστών

Όπως τονίστηκε και στις παραπάνω ενότητες, ο καρκίνος ωθηκών είναι μια γονιδιακά ποικιλόμορφη νόσος και αυτή η γονιδιακή ετερογένεια έχει συσχετιστεί με την αντίσταση στη θεραπεία με πλατίνα [7, 16]. Επιπρόσθετα, ετερογένεια στο γονιδίωμα απεικονίζεται και μεταξύ του πρωτοπαθούς όγκου και των περιτοναϊκών μεταστατικών εναποθέσεων. Τα τελευταία χρόνια γίνεται προσπάθεια συσχέτισης των απεικονιστικών χαρακτηριστικών (ποιοτικών ή ποσοτικών) του πρωτοπαθούς όγκου σε ασθενείς με καρκίνο ωθηκών με την πρόγνωση και την ανταπόκριση αυτών των ασθενών στη θεραπεία.

Σε μελέτη που πραγματοποιήθηκε από τους Sala και συνεργάτες, σε ασθενείς με καρκίνο ωοθηκών που υπεβλήθησαν σε ΜΤ, φαίνεται ότι υπάρχει φαινοτυπική ετερογένεια μεταξύ του πρωτοπαθούς ωοθηκικού όγκου και της μεταστατικής περιτοναϊκής νόσου κατά τη διάγνωση, κάτι το οποίο για πρώτη φορά διαπιστώνεται με την απεικόνιση. Συγκεκριμένα σε δείγμα 22 ασθενών, οι συγγραφείς βρήκαν στατιστικά σημαντικές διαφορές στις τιμές του φαινομενικού συντελεστής διάχυσης (ADC) μεταξύ του πρωτοπαθούς όγκου, του επιπλοϊκού «κέικ» και των περιτοναϊκών εναποθέσεων, ευρήματα που υποδηλώνουν ότι το προφίλ διαχυσιμότητας (diffusivity) μπορεί να εξαρτάται από τη εντόπιση της νόσου [17].

Επίσης, συγκεκριμένα απεικονιστικά (ποιοτικά) χαρακτηριστικά και πρότυπα ανάπτυξης του όγκου στην Υπολογιστική Τομογραφία φαίνεται να συσχετίζονται με συγκεκριμένα βιολογικά χαρακτηριστικά του καρκίνου των ωοθηκών και κατ'έπекταση με την πρόγνωση αυτών των ασθενών. Σε μελέτη που πραγματοποιήθηκε από τους Vargas και συνεργάτες η διάχυτη περιτοναϊκή νόσος και ειδικότερα η δεσμοπλαστική αντίδραση στο μεσεντέριο (tethering) συσχετίστηκε με τον μεσεγχυματογενή υπότυπο CLOVAR, ο οποίος εμφανίζει και τη μεγαλύτερη αντίσταση στη θεραπεία με πλατίνα (63%) [18]. Πρόσφατα, οι Nougaret και συνεργάτες συσχέτισαν το διάχυτο πρότυπο περιτοναϊκής ανάπτυξης του καρκίνου των ωοθηκών, την παρουσία νόσου στο μεσεντέριο και την παρουσία παθολογικών πρόσθιων διαφραγματικών λεμφαδένων με την απουσία μετάλλαξης BRCA, η οποία αποτελεί πω-χό προγνωστικό δείκτη ανταπόκρισης των ασθενών στη θεραπεία. Σε αντίθεση, το οζώδες πρότυπο περιτοναϊκής καρκινωμάτωσης και η παρουσία νόσου στον ηπατογαστρικό σύνδεσμο φαίνεται να συσχετίζεται με την παρουσία μετάλλαξης BRCA και πιθανά με λιγότερο επιθετική συμπεριφορά της νόσου [19]. Ωστόσο, ανάλυση των ποιοτικών απεικονιστικών χαρακτηριστικών και ειδικότερα αυτών της Υπολογιστικής Τομογραφίας, χαρακτηρίζεται από χαμηλή συμφωνία μεταξύ των διαφόρων εξεταστών (low interobserver reproducibility), κάτι που δυνητικά περιορίζει τη χρήση τους.

Τα τελευταία χρόνια υπάρχει προσπάθεια να συσχετιστούν ποσοτικά χαρακτηριστικά υφής του όγκου (texture), με την πρόγνωση των ασθενών με καρκίνο ωοθηκών, ανεξάρτητα από το γονιδιακό CLOVAR προφίλ. Έτσι, ασθενείς με λιγότερες διαφορές (inter-site dissimilarities) στα χαρακτηριστικά υφής εντός του ίδιου όγκου (inter-site similarity entropy, similarity level cluster shade, inter-site similarity level cluster prominence and variance) φαίνεται να έχουν καλύτερη πρόγνωση ανεξάρτητα από το γονιδιακό CLOVAR προφίλ τους. Αντίθετα ασθενείς με περισσότερες διαφορές (inter-site dissimilarities) στα χαρακτηριστικά υφής εντός του ίδιου όγκου έχουν χειρότερη πρόγνωση, ανεξάρτητα από το γονιδιακό CLOVAR προφίλ [20].

Επίσης, συγκεκριμένα ραδιολογικά χαρακτηριστικά (radiomics) που προέκυψαν από την Υπολογιστική Τομογραφία και σχετίζονται με το μέγεθος της ωοθηκικής μάζας (compactness1 below median), την τυχαιότητα (randomness, GrayLevelCooccurrenceMatrix250/0-1InformationMeasureCorr2 below the median) και την ανομοιογένεια (GrayLevelCooccurrenceMatrix25/-333-1InverseVariance above the median) συσχετίστηκαν με μεγαλύτερο κίνδυνο υπολειπόμενης νόσου κατά το χειρουργείο. Στην ίδια μελέτη προτείνεται ένα συνδυαστικό μοντέλο κλινικών χαρακτηριστικών και χαρακτηριστικών υφής το οποίο φαίνεται να έχει καλύτερα αποτελέσματα από ένα αμιγώς κλινικό μοντέλο για να προβλέψει την πρόοδο της νόσου εντός 12 μηνών [21].

Τα ως άνω αποτελέσματα ενισχύουν την υπόθεση ότι τα επόμενα χρόνια τα χαρακτηριστικά υφής (texture analysis) του όγκου, τα οποία είναι πιο αντικειμενικά από τα μορφολογικά (ποιοτικά) χαρακτηριστικά του όγκου, μπορεί να χρησιμοποιηθούν ως βιοδείκτες για την επιλογή της καταλληλότερης θεραπείας των ασθενών με καρκίνο ωοθηκών. Προϋπόθεση είναι να λυθούν θέματα που αφορούν την συστηματικοποιημένη (standardized) και ομοιόμορφη (uniform) μέτρηση αυτών των παραμέτρων, τον μεγάλο όγκο δεδομένων που προκύπτουν από τις μετρήσεις και ειδικότερα για τον καρκίνο των ωοθηκών, λόγω του μεγάλου φορτίου νόσου, τη χρονοβόρα διαδικασία χειροκίνητης (manual) κατάτμησης (segmentation) των δεδομένων που αφορούν στο σύνολο του όγκου. Η ανάπτυξη αυτοματοποιημένων συστημάτων μετα-επεξεργασίας των δεδομένων, θα βοηθήσει στην εξαγωγή της μέγιστης χρήσιμης πληροφορίας από τις εικόνες και ενδεχομένως να βοηθήσει ουσιαστικά στην αναγνώριση αυτών των όγκων που θα ανταποκριθούν στη θεραπεία. Επί του παρόντος, ένα τέτοιο πρόγραμμα έχει εφαρμοστεί πιλοτικά μόνο στον καρκίνο του ορθού [22].

Κεφάλαιο 3

Εξαγωγή Ποσοτικών Πληροφοριών από Ιατρικές Εικόνες (Radiomics)

Τα ραδιολογικά χαρακτηριστικά στοχεύουν στην εξαγωγή ποσοτικών παραμέτρων από δεδομένα ιατρικής απεικόνισης που αποκτήθηκαν με συστηματικό τρόπο, επιτρέποντας έτσι την ανάλυση των δεδομένων αυτών με χαμηλό κόστος. Τα περισσότερα από τα χαρακτηριστικά αυτά αντιπροσωπεύουν την ετερογενετικότητα της υποκείμενης εικόνας (όγκος). Στις κλασικές ραδιολογικές προσεγγίσεις που βασίζονται στα χαρακτηριστικά, τα ραδιολογικά χαρακτηριστικά που εξάγονται είναι προκαθορισμένα και υπολογίζονται από μια εικόνα που έχει τμηματοποιηθεί χειροκίνητα ή ημι-αυτοματοποιημένα. Αντίθετα, στα ραδιολογικά χαρακτηριστικά που βασίζονται σε τεχνικές βαθιάς μάθησης ακολουθείται διαφορετική προσέγγιση για την εξαγωγή των ποσοτικών παραμέτρων. Συγκεκριμένα, τα εν λόγω χαρακτηριστικά δεν είναι προκαθορισμένα, αλλά αναγνωρίζονται και παράγονται με την εφαρμογή υπολογιστικών μοντέλων στα υποκείμενα δεδομένα. Επιπλέον, η τμηματοποίηση των εικόνων δεν είναι απαραίτητη, ωστόσο, συχνά η χρήση τμηματοποιημένης εικόνας βελτιώνει την απόδοση του μοντέλου. Αν και η αξία των περισσότερων ραδιολογικών μοντέλων στο κλινικό περιβάλλον δεν είναι αποδεδειγμένη, κατά την διαδικασία εξαγωγής χαρακτηριστικών εφαρμόζονται πλήρεις ή ημι-αυτοματοποιημένες μέθοδοι (προηγμένη στατιστική, μηχανική μάθηση) που μπορούν ως εκ τούτου να οδηγήσουν σε εύρωστα, αναπαράξιμα και αξιόπιστα αποτελέσματα σε σύγκριση με την εξαρτώμενη από τον κλινικό ιατρό ερμηνεία των δεδομένων απεικόνισης [23].

3.1 Ραδιολογικά Χαρακτηριστικά ως Δυνητικοί Βιοδείκτες

Για τον υπολογισμό των ραδιολογικών χαρακτηριστικά, είναι απαραίτητη η χειροκίνητη ή ημι-αυτόματη τμηματοποίηση της περιοχής ενδιαφέροντος (ROI) ή του όγκου ενδιαφέροντος (VOI). Συνήθως, για την ραδιολογική ανάλυση χρησιμοποιείται η βελτίωση της αντίθεσης του τμήματος αντίθεσης του όγκου στη μαγνητική τομογραφία. Ωστόσο, συμπεριλαμβανομένων των πληροφοριών που περιέχονται στη ζώνη διήθησης της βλάβης, λαμβάνοντας επίσης υπόψη τυχόν ανωμαλίες του σήματος στις σταθμισμένες T2 εικόνες ή στην αντίστροφη ανάκτηση εξασθενημένου υγρού (FLAIR), οι εικόνες MRI παρέχουν μια πιο ρεαλιστική απεικόνιση ολόκληρου του όγκου. Παράλληλα, επιτρέπουν την διεξαγωγή ραδιολογικής ανάλυσης σε μεγαλύτερα τμήματα του όγκου, κωδικοποιώντας ενδεχομένως περισσότερες

πληροφορίες, που σε οδηγούν σε καλύτερα επίπεδα διάγνωσης. Παρα τον μικρό αριθμό μελετών που χρησιμοποιούν εικόνες PET για ραδιολογική ανάλυση, η συνδυασμένη ανάλυση PET και MRI εικόνων φαίνεται πως κωδικοποιεί περισσότερες διαγνωστικές πληροφορίες [24] και μπορεί να αποκτήσει κλινική σημασία. Μέθοδοι από το πεδίο της μηχανικής μάθησης όπως η ανάλυση χαρακτηριστικών υφής ή η χρήση τεχνικών βαθιάς μάθησης είναι ήδη διαθέσιμες και βρίσκονται υπό διερεύνηση για υιοθέτηση ρουτινών με κλινικές εφαρμογές [25].

Όπως αναφέρθηκε παραπάνω, τα ραδιολογικά χαρακτηριστικά συνίστανται στην εξαγωγή ποσοτικών χαρακτηριστικών από ιατρικές εικόνες. Κατά συνέπεια, προκειμένου να επιτύχουμε υψηλή αναπαραξιμότητα και καθολικότητα των αποτελεσμάτων, οι εικόνες πρέπει να υποβληθούν σε ορισμένα βήματα προεπεξεργασίας πριν την εξαγωγή των χαρακτηριστικών. Η παρεμβολή (interpolation) σε διάταξη ισοτροπικών ογκοστοιχείων (voxel spacing) είναι ένα από τα πρώτα βήματα προεπεξεργασίας των εικόνων, που επιτρέπει καλύτερη σύγκριση των ετερογενών, απεικονιστικών δεδομένων των οποίων η διαχείριση γίνεται δυνητικά από πολλούς οργανισμούς/ιδρύματα. Επιπλέον, ο υπολογισμός των ραδιολογικών χαρακτηριστικών, ιδίως των χαρακτηριστικών υφής, απαιτεί περιστροφικά αμετάβλητα ογκοστοιχεία. Η χρήση παρεμβολής καλύπτει την απαίτηση αυτή. Οι εικόνες μπορούν είτε να υπερδειγματοληπτηθούν όπου καθώς εισάγονται τεχνητές πληροφορίες μπορεί να αυξηθεί ο θόρυβος της εικόνας, είτε να υποδειγματοληπτηθούν που μπορεί να οδηγήσει σε απώλεια πληροφοριών. Κατά συνέπεια, προς το παρόν δεν υπάρχει σαφής σύσταση για οποιαδήποτε από τις δύο διαδικασίες [26].

Η διακριτοποίηση ή η ποσοτικοποίηση της έντασης της εικόνας έχει ιδιαίτερη σημασία ώστε η εξαγωγή χαρακτηριστικών να γίνει με κατανοητό τρόπο [26]. Δύο μέθοδοι διακριτοποίησης των εικόνων χρησιμοποιούνται ευρέως. Η πρώτη μέθοδος αφορά την διακριτοποίηση της έντασης της εικόνας σε έναν καθορισμένο αριθμό διαστημάτων (fixed number of bins), ο οποίος επιτρέπει την άμεση σύγκριση των τιμών των χαρακτηριστικών μεταξύ διαφορετικών ασθενών ενώ παράλληλα κανονικοποιεί εν μέρει την εικόνα. Η λειτουργία αυτή είναι ιδιαίτερα σημαντική για απεικονιστικές διαδικασίες όπως η δομική μαγνητική τομογραφία (structural MRI) που απαντώνται συνήθως σε αυθαίρετες μονάδες. Αξίζει να αναφερθεί πως δεν συσχετίζεται η προκύπτουσα ένταση με εκείνη της αρχικής εικόνας. Στην δεύτερη μέθοδο χρησιμοποιείται διάστημα σταθερού μεγέθους, όπου κάθε νέο διάστημα ανατίθεται σε διαφορετική τιμή έντασης με ορισμένο εύρος. Στην προκειμένη περίπτωση των δομικών μαγνητικών τομογραφιών η μέθοδος αυτή δεν συνιστάται. Ωστόσο, καθώς διατηρείται η σχέση με την αρχική κλίμακα της έντασης, το ορισμένο εύρος του διαστήματος διακριτοποίησης μπορεί να φανεί ιδιαίτερα χρήσιμο σε άλλες απεικονιστικές μεθόδους όπως το PET. Αναμφίβολα, η διακριτοποίηση της εικόνας έχει ουσιαστικό αντίκτυπο στα εξαγόμενα ραδιολογικά χαρακτηριστικά και ως εκ τούτου και στην αναπαραξιμότητα των αποτελεσμάτων [26]. Η κανονικοποίηση της έντασης της εικόνας εξασφαλίζει καλύτερη συγκρισιμότητα των αποτελεσμάτων μεταξύ διαφορετικών σαρωτών, πρωτοκόλλων και ασθενών. Οι συνήθεις χρησιμοποιούμενες διαδικασίες είναι η κανονικοποίηση white stripe [27] και η z-score κανονικοποίηση [28]. Άλλα τυπικά στάδια προεπεξεργασίας περιλαμβάνουν μεταξύ άλλων, τη χωρική εξομάλυνση (spatial smoothing), τη μείωση θορύβου (noise reduction), τη χωρική δειγματοληψία (spatial sampling) και τυχόν διορθώσεις της ανομοιογένειας του μαγνητικού

πεδίου της μαγνητικής τομογραφίας.

Μετά την προεπεξεργασία και τμηματοποίηση της εικόνας μπορούν να υπολογιστούν τα ραδιολογικά χαρακτηριστικά, τα περισσότερα εκ των οποίων αντικατοπτρίζουν την ετερογένεια του όγκου. Δεδομένου ότι τα εν λόγω χαρακτηριστικά είναι μαθηματικά προκαθορισμένα και βασισμένα σε τεράστιο αριθμό ελαφρώς τροποποιημένων μαθηματικών ορισμών, ένας μεγάλος αριθμός ραδιολογικών χαρακτηριστικών μπορούν να εξαχθούν από ιατρικές εικόνες. Τα χαρακτηριστικά αυτά διακρίνονται συνήθως σε 4 κύριες υποομάδες οι οποίες είναι:

3.1.1 Χαρακτηριστικά σχήματος (Shape features)

Ποσοτικοποιούν τις γεωμετρικές σχέσεις και ιδιότητες των τμηματοποιημένων αλλοιώσεων όπως η περίμετρος (perimeter), η επιφάνεια πλέγματος (mesh surface), μέγιστη διάμετρος (maximum diameter), σφαιρικότητα (sphericity) κ.α [29].

Θεωρούμε:

- N_p τον αριθμό των περιλαμβανομένων εικονοστοιχείων (pixels) της περιοχής ενδιαφέροντος (ROI),
- N_f τον αριθμό των γραμμών που καθορίζουν την περιφέρεια(περίμετρο) του πλέγματος,
- A την επιφάνεια του πλέγματος σε mm^2 ,
- P την περίμετρο του πλέγματος σε mm .

Με βάση τους άνωθεν συντελεστές υπολογίζονται τα εν λόγω χαρακτηριστικά. Ενδεικτικά, σε σύνολο 10, ορισμένα από τα οποία είναι:

1. Επιφάνεια Πλέγματος.

$$A = \sum_{i=1}^{N_f} A_i \quad (3.1)$$

Όπου το $O_i\alpha_i$ και το $O_i\beta_i$ είναι οι ακμές του $i^{ου}$ τριγώνου στο πλέγμα, όπως σχηματίζονται από τις κορυφές α_i , β_i της περιμέτρου και της αρχής των αξόνων O .

Για τον υπολογισμό της επιφάνειας, υπολογίζεται πρώτα η προσημασμένη επιφάνεια A_i κάθε τριγώνου στο πλέγμα. Το συνολικό εμβαδόν προκύπτει στη συνέχεια λαμβάνοντας το άθροισμα όλων των υπολογιζόμενων υποπεριοχών (3.1), όπου το πρόσημο εξασφαλίζει την ορθή συμπερίληψη της επιφάνειας, καθώς η περιοχή των τριγώνων με αρνητικό πρόσημο που βρίσκονται εκτός της περιοχής ενδιαφέροντος (ROI) ακυρώνει την πλεονασματική περιοχή των τριγώνων που βρίσκονται εν μέρει εντός και εν μέρει εκτός της περιοχής ενδιαφέροντος (ROI).

2. Περίμετρος.

$$P_i = \sqrt{(\alpha_i - \beta_i)^2} \quad (3.2)$$

$$P = \sum_{i=1}^{N_f} P_i \quad (3.3)$$

Όπου, το α_i και το β_i είναι κορυφές της γραμμής i^{th} της περιμέτρου του πλέγματος.

Για τον υπολογισμό της περιμέτρου, υπολογίζουμε πρώτα την περίμετρο A_i της κάθε γραμμής της περιφέρειας του πλέγματος (3.2). Η συνολική περίμετρος προκύπτει αθροίζοντας όλες τις υπολογισμένες υποπεριοχές (3.3) [30].

3. Μέγιστο μήκος άξονα.

$$\text{major axis} = 4 \sqrt{\hat{f}_{\text{major}}} \quad (3.4)$$

Το χαρακτηριστικό αυτό αποδίδει το μήκος του μεγαλύτερου άξονα που περικλείεται από την ελλειψοειδή περιοχή ενδιαφέροντος (ROI) και υπολογίζεται με βάση την μεγαλύτερη κύρια συνιστώσα \hat{f}_{major} .

Η ανάλυση κύριων συνιστωσών (principal component analysis - PCA) πραγματοποιείται κάνοντας χρήση των φυσικών συνιστωσών των κέντρων των pixels όπως ορίζονται από την περιοχή ενδιαφέροντος (ROI). Συνεπώς λαμβάνει υπόψη τις αποστάσεις αλλά όχι το σχήμα του πλέγματος.

3.1.2 Χαρακτηριστικά βάσει ιστογράμματος ή πρώτης τάξης χαρακτηριστικά (Histogram-based features or first-order statistics)

Η κατανομή των τιμών της έντασης της εικόνας εντός των τμηματοποιημένων αλλοιώσεων αντιπροσωπεύεται συνήθως από τα ιστογράμματα. Από τα ιστογράμματα, μπορούν να υπολογιστούν ποσοτικά χαρακτηριστικά που δεν λαμβάνουν υπόψη χωρικό προσανατολισμό ή την σχέση των ογκοστοιχείων όπως η μέση τιμή(mean), η ενέργεια(energy), η εντροπία(entropy), η λοξότητα(skewness) ή κύρτωση(kurtosis) [31].

Θεωρούμε:

- το σύνολο \mathbf{X} των N_p περιλαμβανομένων ογκοστοιχείων (voxels) της περιοχής ενδιαφέροντος (ROI),
- το πρώτης τάξης ιστόγραμμα $\mathbf{P}(i)$ του \mathbf{X} με N_g διακριτά επίπεδα έντασης,
- το κανονικοποιημένο πρώτης τάξης ιστόγραμμα $p(i)$ που ισούται με $\frac{\mathbf{P}(i)}{N_p}$.

Με βάση τους άνωθεν συντελεστές υπολογίζονται τα εν λόγω χαρακτηριστικά. Ενδεικτικά, σε σύνολο 19, ορισμένα από τα οποία είναι:

1. Μέση Τιμή.

$$\text{Mean} = \frac{1}{N_p} \sum_{i=1}^{N_p} \mathbf{X}(i) \quad (3.5)$$

Η μέση τιμή των γκρι επιπέδων της έντασης εντός της περιοχής ενδιαφέροντος (ROI) της εικόνας \mathbf{X} .

2. Ενέργεια.

$$\text{Energy} = \sum_{i=1}^{N_p} (\mathbf{X}(i) + c)^2 \quad (3.6)$$

Όπου η μεταβλητή c είναι αυθαίρετη και εξαρτάται από το voxelArrayShift το οποίο μετατοπίζει τις εντάσεις για να αποτρέψει τυχόν αρνητικές τιμές του \mathbf{X} . Αυτό εξασφαλίζει

ότι τα voxels με την χαμηλότερη τιμή του γκρι συνεισφέρουν ελάχιστα στην συνολική ενέργεια εν αντιθέσει με τα voxels που λαμβάνουν τιμές του γκρι εγγύτερα στο 0.

Η ενέργεια είναι το μέτρο του μεγέθους των τιμών των voxels σε μια εικόνα. Μεγαλύτερες τιμές συνεπάγονται σε μεγαλύτερο άθροισμα των τετραγώνων αυτών των τιμών.

3. Εντροπία.

$$Entropy = - \sum_{i=1}^{N_g} p(i) \log_2 (p(i) + \epsilon) \quad (3.7)$$

Εδώ, το ϵ είναι ένας αυθαίρετα μικρός θετικός αριθμός $\approx 2.2 \times 10^{-16}$.

Η εντροπία καθορίζει την αβεβαιότητα / τυχειότητα στις τιμές της εικόνας, υπολογίζει δηλαδή τον μέσο όρο των πληροφοριών που απαιτούνται για την κωδικοποίηση των τιμών της εικόνας.

4. Λοξότητα (Ασυμμετρία).

$$Skewness = \frac{\mu_3}{\sigma^3} = \frac{\frac{1}{N_p} \sum_{i=1}^{N_p} (\mathbf{X}(i) - \bar{X})^3}{\left(\sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (\mathbf{X}(i) - \bar{X})^2} \right)^3} \quad (3.8)$$

Όπου, το μ_3 είναι η 3^η κεντρική ροπή.

Η λοξότητα υπολογίζει την ασυμμετρία της κατανομής των τιμών της εικόνας \mathbf{X} γύρω από τη μέση τιμή. Θετική λοξότητα σημαίνει ότι η καμπύλη εκτείνεται περισσότερο προς τα δεξιά της μέσης τιμής ενώ αρνητική λοξότητα σημαίνει μεγαλύτερη έκταση της καμπύλης προς τα αριστερά της μέσης τιμής.

5. Κύρτωση.

$$Kurtosis = \frac{\mu_4}{\sigma^4} = \frac{\frac{1}{N_p} \sum_{i=1}^{N_p} (\mathbf{X}(i) - \bar{X})^4}{\left(\frac{1}{N_p} \sum_{i=1}^{N_p} (\mathbf{X}(i) - \bar{X})^2 \right)^2} \quad (3.9)$$

Όπου, το μ_4 είναι η 4^η κεντρική ροπή.

Η κύρτωση είναι μέτρο της οξύτητας της κορυφής της κατανομής τιμών της περιοχής ενδιαφέροντος (ROI) της εικόνας \mathbf{X} . Υψηλότερες τιμές κύρτωσης υποδηλώνουν ότι η μάζα της κατανομής συγκεντρώνεται γύρω από την ουρά και όχι γύρω από τη μέση τιμή. Χαμηλότερες τιμές κύρτωσης υποδηλώνουν το αντίστροφο, ότι δηλαδή η μάζα της κατανομής συγκεντρώνεται γύρω από μία κορυφή κοντά στην μέση τιμή.

3.1.3 Χαρακτηριστικά υφής ή στατιστικά δεύτερης τάξης (Textural features or second-order statistics)

Τα χαρακτηριστικά υφής αντιπροσωπεύουν τη στατιστική σχέση μεταξύ των επίπεδων της έντασης γειτονικών εικονοστοιχείων (pixels) ή ογκοστοιχείων (voxels) ή ομάδων pixels ή voxels εντός της τμηματοποιημένης βλάβης και συνεπώς, ποσοτικοποιούν την ετερογένεια της εικόνας. Τα χαρακτηριστικά υφής δεν εξάγονται απευθείας από τις εικόνες αλλά μέσω διαφόρων αναλυτικών πινάκων που ήδη κωδικοποιούν συγκεκριμένες χωρικές σχέσεις μεταξύ των pixels ή voxels στη τμηματοποιημένη βλάβη. Οι πίνακες που χρησιμοποιούνται

πιο συχνά για τον υπολογισμό των χαρακτηριστικών υφής είναι η μήτρα μήκους διαδρομής επιπέδων του γκρι (Gray Level Run Length Matrix - GLRLM), που κωδικοποιεί το εύρος των ομοιογενών διαδρομών για κάθε ένταση της εικόνας [30], η μήτρα γεινίασης της διαφοράς των επιπέδων του γκρι (Neighbouring Gray Tone Difference Matrix - NGTDM), που αντιστοιχεί στη διαφορά των επιπέδων έντασης μεταξύ ενός voxel και των γειτονικών του voxels στις τρεις διαστάσεις [32], η μήτρα συνεμφάνισης επιπέδων του γκρι (Gray Level Co-occurrence Matrix - GLCM) [33], που αντιπροσωπεύει τη συχνότητα εμφάνισης δύο επιπέδων έντασης σε γειτονικά pixel ή voxels σε μια συγκεκριμένη απόσταση κατά μήκος μιας σταθερής κατεύθυνσης, η μήτρα μεγέθους ζωνών των επιπέδων του γκρι (Gray Level Size Zone - GLSZM), η οποία ποσοτικοποιεί τον αριθμό των συνδεδεμένων voxels που έχουν κοινή ένταση επιπέδων του γκρι [34] και η μήτρα εξάρτησης των επιπέδων του γκρι (Gray Level Dependence Matrix - GLDM), η οποία ποσοτικοποιεί τον αριθμό των συνδεδεμένων voxels εντός μιας απόστασης δ τα οποία εξαρτώνται από το κεντρικό voxel [35].

- Η μήτρα συν-εμφάνισης επιπέδων του γκρι (Gray Level Co-occurrence Matrix - GLCM), μεγέθους $N_g \times N_g$, περιγράφει την από κοινού συνάρτηση πιθανότητας δεύτερης τάξης second-order joint probability function του τμήματος εκείνου της εικόνας όπως καθορίζεται από την μάσκα και ορίζεται ως $\mathbf{P}(i, j | \delta, \theta)$. Το $(i, j)^{\text{στο}}$ στοιχείο της μήτρας αναπαριστά τον αριθμό των φορών που ο συνδυασμός των επιπέδων i και j εμφανίζεται σε δύο εικονοστοιχεία στην εικόνα, που διαχωρίζονται με απόσταση δ εικονοστοιχείων σε κάθε κατεύθυνση κατά μήκος της γωνίας θ , ενώ N_g είναι ο αριθμός των διακριτών επιπέδων έντασης του γκρι.

Ως δισδιάστατο παράδειγμα, θεωρούμε το ακόλουθο πλέγμα \mathbf{I} που αντιπροσωπεύει μια εικόνα 5×5 , με 5 διακριτά επίπεδα του γκρι:

$$\mathbf{I} = \begin{bmatrix} 1 & 2 & 5 & 2 & 3 \\ 3 & 2 & 1 & 3 & 1 \\ 1 & 3 & 5 & 5 & 2 \\ 1 & 1 & 1 & 1 & 2 \\ 1 & 2 & 4 & 3 & 5 \end{bmatrix} \quad (3.10)$$

Για απόσταση $\delta = 1$ και γωνία $\theta = 0^\circ$ (οριζόντιο επίπεδο, δηλαδή ογκοστοιχεία (voxels) προς τα αριστερά και δεξιά από το κεντρικό ογκοστοιχείο (voxel), λαμβάνεται η ακόλουθη συμμετρική μήτρα:

$$\mathbf{P} = \begin{bmatrix} 6 & 4 & 3 & 0 & 0 \\ 4 & 0 & 2 & 1 & 3 \\ 3 & 2 & 0 & 1 & 2 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 3 & 2 & 0 & 2 \end{bmatrix} \quad (3.11)$$

Θεωρούμε:

- το ϵ ως μία αυθαίρετα μικρή θετική σταθερά ($\approx 2.2 \times 10^{-16}$),

- $\mathbf{P}(i,j)$ την μήτρα συν-εμφάνισης για αυθαίρετο δ και θ ,
- $p(i,j)$ την κανονικοποιημένη μήτρα συν-εμφάνισης που ισούται με $\frac{\mathbf{P}(i,j)}{\sum \mathbf{P}(i,j)}$,
- N_g τον αριθμό των διακριτών επιπέδων έντασης του γκρι της εικόνας,
- $p_x(i) = \sum_{j=1}^{N_g} P$ τις κατά γραμμή περιθωριακές πιθανότητες,
- $p_y(j) = \sum_{i=1}^{N_g} P$ τις κατά στήλη περιθωριακές πιθανότητες,
- μ_x την μέση τιμή της έντασης των επιπέδων του γκρι p_x που ισούται με $\mu_x = \sum_{i=1}^{N_g} p_x(i)i$,
- μ_y την μέση τιμή της έντασης των επιπέδων του γκρι p_y που ισούται με $\mu_y = \sum_{j=1}^{N_g} p_y(j)j$,
- σ_x την τυπική απόκλιση του p_x ,
- σ_y την τυπική απόκλιση του p_y ,
- $p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)$, όπου $i+j=k$, και $k=2,3,\dots,2N_g$,
- $p_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)$, όπου $|i-j|=k$, και $k=0,1,\dots,N_g-1$,
- $HX = -\sum_{i=1}^{N_g} p_x(i) \log_2(p_x(i) + \epsilon)$ την εντροπία του p_x ,
- $HY = -\sum_{j=1}^{N_g} p_y(j) \log_2(p_y(j) + \epsilon)$ την εντροπία του p_y ,
- $HXY = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log_2(p(i,j) + \epsilon)$ την εντροπία του $p(i,j)$,
- $HXY1 = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log_2(p_x(i)p_y(j) + \epsilon)$,
- $HXY2 = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_x(i)p_y(j) \log_2(p_x(i)p_y(j) + \epsilon)$.

Με βάση τους άνωθεν συντελεστές υπολογίζονται τα εν λόγω χαρακτηριστικά. Ενδεικτικά, σε σύνολο 24, ορισμένα από τα οποία είναι:

1. Αυτοσυσχέτιση (Autocorrelation).

$$autocorrelation = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)ij \quad (3.12)$$

Η αυτοσυσχέτιση αποτελεί μέτρο του μεγέθους της λεπιότητας και της τραχύτητας της υφής. Όσο υψηλότερη είναι η τιμή της τόσο περισσότερα ζεύγη με υψηλά επίπεδα του γκρι παρουσιάζει η υφή της εικόνας.

2. Τάση Δέσμης (Cluster Tendency).

$$cluster\ tendency = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i+j - \mu_x - \mu_y)^2 p(i,j) \quad (3.13)$$

Η τάση δέσμης αποτελεί μέτρο ομαδοποίησης ογκοστοιχείων (voxels) με παρόμοιες τιμές επιπέδων του γκρι.

3. Διαφορά Διακύμανσης (Difference Variance).

$$\text{difference variance} = \sum_{k=0}^{N_g-1} (k - DA)^2 p_{x-y}(k) \quad (3.14)$$

Η διαφορά διακύμανσης είναι ένα μέτρο της ετερογένειας που τοποθετεί υψηλότερα βάρη σε διαφορετικά ζεύγη επιπέδων έντασης που αποκλίνουν περισσότερο από την μέση τιμή.

4. Ομοιογένεια 1 (Homogeneity 1).

$$\text{homogeneity 1} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1 + |i - j|} \quad (3.15)$$

Η ομοιογένεια 1 είναι ένα μέτρο της ομοιότητας στις τιμές έντασης για γειτονικά ογκοστοιχεία (voxels). Είναι ένα μέτρο τοπικής ομοιογένειας που αυξάνεται με λιγότερη αντίθεση στο παράθυρο.

5. Αντίστροφη Διαφορά (Inverse Difference (ID)).

$$\text{inverse variance} = \sum_{k=1}^{N_g-1} \frac{p_{x-y}(k)}{k^2} \quad (3.16)$$

Η αντίστροφη διαφορά είναι ένα άλλο μέτρο της τοπικής ομοιογένειας μιας εικόνας. Με περισσότερο ομοιόμορφα επίπεδα του γκρι, ο παρονομαστής παραμένει χαμηλός, με αποτέλεσμα συνολική τιμή υψηλότερη.

6. Άθροισμα Εντροπίας (Sum Entropy).

$$\text{sum entropy} = \sum_{k=2}^{2N_g} p_{x+y}(k) \log_2 (p_{x+y}(k) + \epsilon) \quad (3.17)$$

Το άθροισμα εντροπίας είναι το άθροισμα γειτονικών τιμών διαφοράς της έντασης μίας εικόνας.

- Η μήτρα μήκους διαδρομής επιπέδων του γκρι (Gray Level Run Length Matrix - GLRLM) ποσοτικοποιεί τις διαδρομές των επιπέδων του γκρι, οι οποίες ορίζονται ως το μήκος (αριθμός) εικονοστοιχείων (pixels), συνεχόμενων εικονοστοιχείων, που έχουν την ίδια τιμή επιπέδου του γκρι. Σε μία μήτρα μήκους διαδρομής επιπέδων του γκρι $\mathbf{P}(i,j|\theta)$, το $(i,j)^{\text{στο}}$ στοιχείο περιγράφει τον αριθμό των διαδρομών με επίπεδο του γκρι i και μήκος j που συναντάται στην περιοχή ενδιαφέροντος (ROI) κατά μήκος της γωνίας θ .

Ως δισδιάστατο παράδειγμα, θεωρούμε το ακόλουθο πλέγμα \mathbf{I} που αντιπροσωπεύει μια

εικόνα 5x5, με 5 διακριτά επίπεδα του γκρι:

$$\mathbf{I} = \begin{bmatrix} 5 & 2 & 5 & 4 & 4 \\ 3 & 3 & 3 & 1 & 3 \\ 2 & 1 & 1 & 1 & 3 \\ 4 & 2 & 2 & 2 & 3 \\ 3 & 5 & 3 & 3 & 2 \end{bmatrix} \quad (3.18)$$

Η μήτρα μήκους διαδρομής επιπέδων του γκρι (Gray Level Run Length Matrix - GLRLM) για γωνία $\theta = 0$, όπου 0 μοίρες είναι ο οριζόντιος άξονας, προκύπτει:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 \\ 4 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (3.19)$$

Θεωρούμε:

- N_g τον αριθμό των διακριτών επιπέδων εντάσης του γκρι της εικόνας,
- N_r τον αριθμό των διακριτών μηκών των διαδρομών της εικόνας,
- N_p τον αριθμό των διακριτών ογκοστοιχείων (voxels) της εικόνας,
- $N_r(\theta)$ τον αριθμό των διαδρομών κατα μήκος των γωνιών θ της εικόνας, ο οποίος ισούται με $\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \mathbf{P}(i,j|\theta)$ και $1 \leq N_r(\theta) \leq N_p$
- $\mathbf{P}(i,j|\theta)$ την μήτρα μήκους διαδρομών για αυθαίρετη κατεύθυνση θ ,
- $p(i,j|\theta)$ την κανονικοποιημένη μήτρα μήκους διαδρομών, η οποία ορίζεται ως:

$$p(i,j|\theta) = \frac{\mathbf{P}(i,j|\theta)}{N_r(\theta)}$$

Με βάση τους άνωθεν συντελεστές υπολογίζονται τα εν λόγω χαρακτηριστικά. Ενδεικτικά, σε σύνολο 24, ορισμένα από τα οποία είναι:

1. Έμφαση Μικρού Μήκους (Short Run Emphasis (SRE)).

$$SRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{\mathbf{P}(i,j|\theta)}{j^2}}{N_r(\theta)} \quad (3.20)$$

Η έμφαση μικρού μήκους αποτελεί μέτρο της κατανομής των διαδρομών μικρού μήκους. Μεγαλύτερες τιμές είναι ενδεικτικές διαδρομών μικρότερου μήκους και πιο λεπτές υφές.

2. Μη Ομοιομορφία των Επιπέδων του Γκρι (Gray Level Non-Uniformity (GLN)).

$$GLN = \frac{\sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_r} \mathbf{P}(i,j|\theta) \right)^2}{N_r(\theta)} (i+j - \mu_x - \mu_y)^2 p(i,j) \quad (3.21)$$

Η μη ομοιομορφία των επιπέδων του γκρι μετρά την ομοιότητα των τιμών των επιπέδων του γκρι στην εικόνα, όπου μια χαμηλότερη τιμή μη ομοιομορφίας συσχετίζεται με μεγαλύτερη ομοιότητα στις τιμές έντασης.

3. Διακύμανσης Διαδρομής (Run Variance (RV)).

$$RV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\partial)(j - \mu)^2 \quad (3.22)$$

Η διακύμανσης διαδρομής είναι ένα μέτρο της διακύμανσης των διαδρομών ως προς το μήκος τους.

4. Έμφαση Μικρών Διαδρομών Χαμηλών Επιπέδων του Γκρι (Short Run Low Gray Level Emphasis (SRLGLE)).

$$SRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j|\partial)}{i^2 j^2}}{N_r(\partial)} \quad (3.23)$$

Η έμφαση μικρών διαδρομών χαμηλών επιπέδων του γκρι μετρά την συνδυασμένη κατανομή διαδρομών μικρότερου μήκους με χαμηλότερες τιμές επιπέδων του γκρι.

- Η μήτρα μεγέθους ζωνών των επιπέδων του γκρι (Gray Level Size Zone - GLSZM) ποσοτικοποιεί τις ζώνες επιπέδων του γκρι σε μία εικόνα. Μία ζώνη επιπέδου του γκρι ορίζεται ως ο αριθμός των συνδεδεμένων ογκοστοιχείων voxels που έχουν κοινή ένταση επιπέδου του γκρι. Ένα ογκοστοιχείο θεωρείται συνδεδεμένο αν η απόσταση είναι 1 σύμφωνα με την νόρμα απείρου (infinity norm). Σε μία μήτρα μεγέθους ζωνών των επιπέδων του γκρι $P(i,j)$ το $(i,j)^{στο}$ στοιχείο ισούται με τον αριθμό των ζωνών με i επίπεδο του γκρι και μέγεθος j που εμφανίζονται στην περιοχή ενδιαφέροντος (ROI). Σε αντίθεση με την μήτρα συν-εμφάνισης επιπέδων του γκρι και την μήτρα μήκους διαδρομής επιπέδων του γκρι, η μήτρα μεγέθους ζωνών των επιπέδων του γκρι είναι ανεξάρτητη της περιστροφής, επομένως απαιτείται ο υπολογισμός μόνο μίας μήτρας για όλες τις κατευθύνσεις της περιοχής ενδιαφέροντος.

Ως δισδιάστατο παράδειγμα, θεωρούμε το ακόλουθο πλέγμα \mathbf{I} που αντιπροσωπεύει μια εικόνα 5x5, με 5 διακριτά επίπεδα του γκρι:

$$\mathbf{I} = \begin{bmatrix} 5 & 2 & 5 & 4 & 4 \\ 3 & 3 & 3 & 1 & 3 \\ 2 & 1 & 1 & 1 & 3 \\ 4 & 2 & 2 & 2 & 3 \\ 3 & 5 & 3 & 3 & 2 \end{bmatrix} \quad (3.24)$$

Η παραγόμενη μήτρα μεγέθους ζωνών των επιπέδων του γκρι είναι:

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (3.25)$$

Θεωρούμε:

- N_g τον αριθμό των διακριτών τιμών έντασης της εικόνας,
- N_s τον αριθμό των διακριτών μεγεθών των ζωνών της εικόνας,
- N_z τον αριθμό των ζωνών εντός της περιοχής ενδιαφέροντος της εικόνας, που ισούται με $\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \mathbf{P}(i,j)$ και $1 \leq N_z \leq N_p$,
- $\mathbf{P}(i,j)$ την μήτρα μεγέθους ζωνών,
- $p(i,j)$ την κανονικοποιημένη μήτρα μεγέθους ζωνών, η οποία ορίζεται ως $p(i,j) = \frac{\mathbf{P}(i,j)}{N_z}$

Με βάση τους άνωθεν συντελεστές υπολογίζονται τα εν λόγω χαρακτηριστικά. Ενδεικτικά, σε σύνολο 16, ορισμένα από τα οποία είναι:

1. Έμφαση Μικρής Περιοχής (Small Area Emphasis (SAE)).

$$SAE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_s} \frac{\mathbf{P}(i,j)}{j^2}}{N_z} \quad (3.26)$$

Η έμφαση μικρής περιοχής αποτελεί ένα μέτρο της κατανομής των ζωνών μικρού μεγέθους. Μεγαλύτερες τιμές είναι ενδεικτικές για περισσότερες ζώνες μικρότερου μεγέθους και για πιο λεπτές υφές.

2. Μη Ομοιομορφία του Μεγέθους των Ζωνών (Size-Zone Non-Uniformity (SZN)).

$$SZN = \frac{\sum_{j=1}^{N_s} \left(\sum_{i=1}^{N_g} \mathbf{P}(i,j) \right)^2}{N_z} \quad (3.27)$$

Η μη ομοιομορφία του μεγέθους των ζωνών μετρά τη μεταβλητότητα του μεγέθους των όγκων ζώνης στην εικόνα. Χαμηλότερη τιμή δείχνει περισσότερη ομοιογένεια στα μεγέθη των όγκων ζώνης.

3. Διακύμανσης Ζώνης (Zone Variance (ZV)).

$$ZV = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j)(j - \mu)^2 \quad (3.28)$$

$$\text{Όπου, } \mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_s} p(i,j)j.$$

Η διακύμανσης ζώνης μετρά την διακύμανση του μεγέθους των όγκων των ζωνών.

4. Ποσοστό Ζώνης (Zone Percentage (ZP)).

$$ZP = \frac{N_z}{N_p} \quad (3.29)$$

Το ποσοστό ζώνης μετρά την τραχύτητα της υφής λαμβάνοντας το λόγο του αριθμού των ζωνών προς τον αριθμού των ογκοστοιχείων στην περιοχή ενδιαφέροντος.

- Η μήτρα γεινίασης της διαφοράς των επιπέδων του γκρι (Neighbouring Gray Tone Difference Matrix - NGTDM) ποσοτικοποιεί διαφορά μεταξύ μίας τιμής του γκρι και της μέση τιμής των τιμών του γκρι ενός ογκοστοιχείου και των γειτονικών του ογκοστοιχείων στις τρεις διαστάσεις εντός απόστασης δ . Το άθροισμα της απόλυτης διαφοράς στο i επίπεδο του γκρι αποθηκεύεται στην μήτρα.

Θεωρούμε \mathbf{X}_{gl} το σύνολο των τμηματοποιημένων ογκοστοιχείων και $x_{gl}(j_x, j_y, j_z) \in \mathbf{X}_{gl}$ το επίπεδο του γκρι ενός ογκοστοιχείου στην θέση (j_x, j_y, j_z) , τότε η μέση τιμή επιπέδου του γκρι της γειτονιάς είναι:

$$\begin{aligned} \bar{A}_i &= \bar{A}(j_x, j_y, j_z) \\ &= \frac{1}{W} \sum_{k_x=-\delta}^{\delta} \sum_{k_y=-\delta}^{\delta} \sum_{k_z=-\delta}^{\delta} x_{gl}(j_x + k_x, j_y + k_y, j_z + k_z), \end{aligned} \quad (3.30)$$

όπου $(k_x, k_y, k_z) \neq (0, 0, 0)$ και $x_{gl}(j_x + k_x, j_y + k_y, j_z + k_z) \in \mathbf{X}_{gl}$

Όπου, το W είναι ο αριθμός των ογκοστοιχείων της γειτονιάς που ανήκουν επίσης στο \mathbf{X}_{gl} .

Ως διδιάστατο παράδειγμα, θεωρούμε το ακόλουθο πλέγμα \mathbf{I} που αντιπροσωπεύει μια εικόνα 4x4, με 5 διακριτά επίπεδα του γκρι, χωρίς ογκοστοιχεία με επίπεδο του γκρι 4:

$$\mathbf{I} = \begin{bmatrix} 1 & 2 & 5 & 2 \\ 3 & 5 & 1 & 3 \\ 1 & 3 & 5 & 5 \\ 3 & 1 & 1 & 1 \end{bmatrix} \quad (3.31)$$

Προκύπτει η ακόλουθη μήτρα γεινίασης της διαφοράς των επιπέδων του γκρι:

i	n_i	p_i	s_i
1	6	0.375	13.35
2	2	0.125	2.00
3	4	0.25	2.63
4	0	0.00	0.00
5	4	0.25	10.075

(3.32)

Έξι εικονοστοιχεία έχουν επίπεδο του γκρι 1, οπότε:

$$s_1 = |1 - 10/3| + |1 - 30/8| + |1 - 15/5| + |1 - 13/5| + |1 - 15/5| + |1 - 11/3| = 13.35$$

Δύο εικονοστοιχεία έχουν επίπεδο του γκρι 2, οπότε:

$$s_2 = |2 - 15/5| + |2 - 15/5| = 2$$

Αντίστοιχα για τα επίπεδα του γκρι 3 και 5 ισχύει:

$$s_3 = |3 - 12/5| + |3 - 18/5| + |3 - 20/8| + |3 - 5/3| = 3.03$$

$$s_5 = |5 - 14/5| + |5 - 18/5| + |5 - 20/8| + |5 - 11/5| = 10.075$$

Θεωρούμε:

- n_i τον αριθμό των ογκοστοιχείων στο X_{gi} με επίπεδο του γκρι i ,
- $N_{v,p}$ τον συνολικό αριθμό των ογκοστοιχείων στο X_{gl} , ο οποίος ισούται με $\sum n_i$ (δηλαδή, τον αριθμό των ογκοστοιχείων με έγκυρη περιοχή τουλάχιστον ένα γε-ίτονα),
Ισχύει $N_{v,p} \leq N_p$, όπου N_p είναι ο συνολικός αριθμός των ογκοστοιχείων στην περιοχή ενδιαφέροντος.
- p_i τον αριθμό της πιθανότητας ενός επιπέδου του γκρι που ισούται με n_i/N_v
- $s_i = \begin{cases} \sum^{n_i} |i - \bar{A}_i| & \text{για } n_i \neq 0 \\ 0 & \text{για } n_i = 0 \end{cases}$ το άθροισμα των απολύτων διαφορών του i επιπέδου του γκρι,
- N_g τον αριθμό των διακριτών τιμών του γκρι της εικόνας,
- $N_{g,p}$ τον αριθμό των επιπέδων του γκρι όπου $p_i \neq 0$.

Με βάση τους άνωθεν συντελεστές υπολογίζονται τα εν λόγω χαρακτηριστικά. Ενδεικτικά, σε σύνολο 5, αυτά είναι:

1. Τραχύτητα (Coarseness).

$$Coarseness = \frac{1}{\sum_{i=1}^{N_g} p_i s_i} \quad (3.33)$$

Η τραχύτητα είναι ένα μέτρο της μέσης διαφοράς μεταξύ του κεντρικού ογκοστοιχείου και της γειτονιάς του και αποτελεί ένδειξη του χωρικού ρυθμού μεταβολής. Υψηλότερη τιμή της τραχύτητας δείχνει χαμηλότερο ρυθμό χωρικής μεταβολής και τοπικά πιο ομοιόμορφη υφή.

Το $\sum_{i=1}^{N_g} p_i s_i$ συγκλίνει πιθανώς στο 0 (σε περίπτωση εντελώς ομοιογενούς εικόνας).

Σε αυτήν την περίπτωση, επιστρέφεται μια αυθαίρετη τιμή: 10^6 .

2. Αντίθεση (Contrast).

$$Contrast = \left(\frac{1}{N_{g,p}(N_{g,p} - 1)} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_i p_j (i - j)^2 \right) \left(\frac{1}{N_{v,p}} \sum_{i=1}^{N_g} s_i \right), \text{ όπου } p_i \neq 0, p_j \neq 0 \quad (3.34)$$

Η αντίθεση είναι ένα μέτρο της αλλαγής της χωρικής έντασης, αλλά εξαρτάται επίσης από το συνολικό δυναμικό εύρος του επιπέδου του γκρι. Η αντίθεση είναι υψηλή όταν τόσο το δυναμικό εύρος όσο και ο ρυθμός χωρικής μεταβολής (spatial

change rate) είναι υψηλοί, δηλαδή μια εικόνα με μεγάλη εμβέλεια των επιπέδων του γκρι, με μεγάλες αλλαγές μεταξύ των ογκοστοιχείων και της γειτονιάς τους.

Σε περίπτωση εντελώς ομοιογενούς εικόνας, *spatialchangerate* που θα σήμαινε διαίρεση με 0, επιστρέφεται αυθαίρετα η τιμή 0.

3. Απασχολησιμότητα (Busyness).

$$Busyness = \frac{\sum_{i=1}^{N_g} p_i s_i}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |ip_i - jp_j|}, \text{ ωηερε } p_i \neq 0, p_j \neq 0 \quad (3.35)$$

Η απασχολησιμότητα μετρά την αλλαγή από ένα εικονοστοιχείο στο γείτονά του. Υψηλή τιμή απασχολησιμότητας δείχνει μια «απασχολημένη» εικόνα, δηλαδή μια εικόνα με γρήγορες αλλαγές έντασης μεταξύ του εικονοστοιχείου και της γειτονιάς του.

N.B. Εάν $N_{g,p} = 1$, τότε η *busyness* = $\frac{0}{0}$. Εάν συμβεί αυτό, επιστρέφεται 0, καθώς αφορά μια πλήρως ομοιογενή περιοχή.

4. Πολυπλοκότητα (Complexity).

$$Complexity = \frac{1}{N_{v,p}} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i-j| \frac{p_i s_i + p_j s_j}{p_i + p_j}, \text{ όπου } p_i \neq 0, p_j \neq 0 \quad (3.36)$$

Μια εικόνα θεωρείται περίπλοκη όταν υπάρχουν πολλά πρωτόγονα στοιχεία στην εικόνα, δηλαδή η εικόνα δεν είναι ομοιόμορφη και υπάρχουν πολλές γρήγορες αλλαγές στην ένταση του επιπέδου του γκρι.

5. Δύναμη (Strength).

$$Strength = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p_i + p_j)(i-j)^2}{\sum_{i=1}^{N_g} s_i}, \text{ όπου } p_i \neq 0, p_j \neq 0 \quad (3.37)$$

Η δύναμη είναι ένα μέτρο των πρωτόγονων στοιχείων μιας εικόνας. Η τιμή της είναι υψηλή όταν τα πρωτόγονα στοιχεία ορίζονται εύκολα και είναι ορατά, δηλαδή μια εικόνα με αργή αλλαγή στην ένταση αλλά πιο μεγάλες χονδροειδείς διαφορές στις εντάσεις του επιπέδου του γκρι.

Το άθροισμα $\sum_{i=1}^{N_g} s_i$ δυνητικά δίδει στο 0 (σε περίπτωση πλήρως ομοιογενούς εικόνας). Εάν συμβεί αυτό, επιστρέφεται 0.

- Η μήτρα εξάρτησης των επιπέδων του γκρι (Gray Level Dependence Matrix - GLDM) ποσοτικοποιεί τις εξαρτήσεις των επιπέδων του γκρι σε μια εικόνα. Η εξάρτηση των επιπέδων του γκρι ορίζεται ως ο αριθμός των συνδεδεμένων ογκοστοιχείων σε απόσταση δ που εξαρτώνται από το κεντρικό ογκοστοιχείο. Ένα γειτονικό ογκοστοιχείο με j επίπεδο του γκρι θεωρείται εξαρτώμενο από το κεντρικό ογκοστοιχείο με i επίπεδο του γκρι, αν $|i-j| \leq a$. Σε μία μήτρα εξάρτησης των επιπέδων του γκρι $\mathbf{P}(i,j)$ το $(i,j)^{sto}$ στοιχείο περιγράφει τον αριθμό των φορών ενός ογκοστοιχείου με επίπεδο του γκρι i με j εξαρτώμενα ογκοστοιχεία στη γειτονιά του που εμφανίζεται στην εικόνα.

Ως διδιάστατο παράδειγμα, θεωρούμε το ακόλουθο πλέγμα \mathbf{I} που αντιπροσωπεύει μια εικόνα 5×5 , με 5 διακριτά επίπεδα του γκρι:

$$\mathbf{I} = \begin{bmatrix} 5 & 2 & 5 & 4 & 4 \\ 3 & 3 & 3 & 1 & 3 \\ 2 & 1 & 1 & 1 & 3 \\ 4 & 2 & 2 & 2 & 3 \\ 3 & 5 & 3 & 3 & 2 \end{bmatrix} \quad (3.38)$$

Για $a = 0$ και $\delta = 1$ η μήτρα εξάρτησης των επιπέδων του γκρι γίνεται:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 2 & 1 \\ 1 & 2 & 3 & 0 \\ 1 & 4 & 4 & 0 \\ 1 & 2 & 0 & 0 \\ 3 & 0 & 0 & 0 \end{bmatrix} \quad (3.39)$$

Θεωρούμε:

- N_g τον αριθμό των διακριτών τιμών εντάσης της εικόνας,
- N_d τον αριθμό των διακριτών μεγεθών εξάρτησης της εικόνας,
- N_z τον αριθμό των ζωνών εξάρτησης της εικόνας, που ισούται με $\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \mathbf{P}(i,j)$,
- $\mathbf{P}(i,j)$ την μήτρα εξάρτησης,
- $p(i,j)$ την κανονικοποιημένη μήτρα εξάρτησης, η οποία ορίζεται ως $p(i,j) = \frac{\mathbf{P}(i,j)}{N_z}$

Με βάση τους άνωθεν συντελεστές υπολογίζονται τα εν λόγω χαρακτηριστικά. Ενδεικτικά, σε σύνολο 14, ορισμένα από τα οποία είναι:

1. Έμφαση Μεγάλης Εξάρτησης (Large Dependence Emphasis (LDE)).

$$LDE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \mathbf{P}(i,j)^2}{N_z} \quad (3.40)$$

Η έμφαση μεγάλης εξάρτησης αποτελεί ένα μέτρο της κατανομής των μεγάλων εξαρτήσεων. Μεγαλύτερη τιμή είναι ενδεικτική μεγαλύτερης εξάρτησης και πιο ομοιογενών υφών.

2. Μη Ομοιομορφίας Εξάρτησης (Dependence Non-Uniformity (DN)).

$$DN = \frac{\sum_{j=1}^{N_d} \left(\sum_{i=1}^{N_g} \mathbf{P}(i,j) \right)^2}{N_z} \quad (3.41)$$

Η μη ομοιομορφία εξάρτησης μετρά την ομοιότητα της εξάρτησης σε όλη την εικόνα. Χαμηλότερη τιμή δείχνει περισσότερη ομοιογένεια μεταξύ των εξαρτήσεων στην εικόνα.

3. Εντροπία Εξάρτησης (Dependence Entropy (DE)).

$$DependenceEntropy = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} p(i,j) \log_2(p(i,j) + \epsilon) \quad (3.42)$$

4. Έμφαση Μεγάλης Εξάρτησης Υψηλών Επιπέδων του Γκρι (Large Dependence High Gray Level Emphasis (LDHGLE)).

$$LDHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \mathbf{P}(i,j) i^2 j^2}{N_z} \quad (3.43)$$

Η έμφαση μεγάλης εξάρτησης υψηλών επιπέδων του γκρι μετρά την συνδυασμένη κατανομή της μεγάλης εξάρτησης των υψηλότερων τιμών των επιπέδων του γκρι.

3.1.4 Χαρακτηριστικά στατιστικών υψηλότερης τάξης (Higher-order statistics features)

Οι τρεις προηγούμενες υπο-ομάδες χαρακτηριστικών συνήθως υπολογίζονται στην επεξεργασμένη αρχική εικόνα χωρίς την προσθήκη φίλτρων. Τα χαρακτηριστικά στατιστικών υψηλότερης τάξης υπολογίζονται με την εφαρμογή συγκεκριμένων μαθηματικών μετασχηματισμών ή φίλτρων που δίνουν έμφαση σε ορισμένες πτυχές της τμηματοποιημένης αλλοίωσης όπως επαναλαμβανόμενα μοτίβα, άκρα, κλίσεις προσανατολισμένες βάσει ιστογράμματος ή τοπικά δυαδικά μοτίβα. Οι τυπικοί μαθηματικοί μετασχηματισμοί που χρησιμοποιούνται για την εξαγωγή στατιστικών στοιχείων υψηλότερης τάξης είναι οι μετασχηματισμοί κυματίου (Wavelet transforms) ή Fourier transforms, η μορφοκλασματική ανάλυση (fractal analysis), το συναρτησοειδές του Minkowski (Minkowski functionals) και ο Λαπλασιανός μετασχηματισμός με φίλτρο Γκάους (Laplacian transform of Gaussian-filtered images -Laplacian of Gaussian-) [36].

Συνοπτικά, με την διαδικασία αυτή μπορεί να υπολογιστεί ένας μεγάλος αριθμός χαρακτηριστικών από μία μόνο τμηματοποιημένη βλάβη. Πλέον, το πρόβλημα μετατρέπεται σε αναζήτηση των παραμέτρων εκείνων που σχετίζονται με το υπό διερεύνηση κλινικό πρόβλημα και διακρίνονται από τις αλυσιτελείς και περιπτώσεις. Η επιλογή χαρακτηριστικών έχει μεγάλη σημασία για τη δημιουργία ενός αξιόπιστου μοντέλου πρόγνωσης ή πρόβλεψης της φύσης της τμηματοποιημένης βλάβης που θα βασίζεται στα υπολογισθέντα χαρακτηριστικά, ειδικά εάν ο αριθμός των διαθέσιμων δεδομένων είναι περιορισμένος. Η διαδικασία επιλογής χαρακτηριστικών χρησιμοποιεί προηγμένες στατιστικές μεθόδους για τον προσδιορισμό ενός υποσυνόλου χαρακτηριστικών που θα δεν θα περιλαμβάνει τα περιττά, σταθερά, διπλότυπα, αλυσιτελή και υπερ-συσχετιζόμενα χαρακτηριστικά [21].

3.2 Ραδιολογικά Χαρακτηριστικά - Τεχνικές Βαθιάς Μάθησης

Η βαθιά μάθηση (deep learning) περιγράφει γενικά τη διαδικασία εφαρμογής αρχιτεκτονικών βαθέων νευρωνικών δικτύων και την χρήση συγκεκριμένων τύπων αλγορίθμων μηχανικής μάθησης για την επίλυση προβλημάτων. Αρχικά, η βάση των προηγμένων αρχι-

τεκτονικών αυτών είναι τα τεχνητά νευρωνικά δίκτυα (artificial neural networks), τα οποία είναι εμπνευσμένα από την αρχή λειτουργίας που διέπει το οπτικό σύστημα του ανθρώπου. Προσθέτοντας λοιπόν, επιπλέον στρώματα κρυφών νευρώνων πέρα από τα στρώματα εισόδου και εξόδου των τεχνητών νευρωνικών δικτύων προστέθηκε ένα νέο επίπεδο πολυπλοκότητας που επιτρέπει την βαθιά μάθηση. Τα ραδιολογικά χαρακτηριστικά στη βάση τεχνικών βαθιάς μάθησης είναι το αποτέλεσμα αυτόματης αναγνώρισης και εξαγωγής χαρακτηριστικών υψηλών διαστάσεων από τις εικόνες εισόδου. Οι χρησιμοποιούμενες εικόνες έχουν υποστεί κλιμάκωση και αφαίρεση ποικίλων επιπέδων, με αποτέλεσμα την δημιουργία ιδιαίτερα χρήσιμων μοντέλων για την αναγνώριση προτύπων ή την ταξινόμηση μη γραμμικών δεδομένων υψηλών διαστάσεων [37]. Η χρησιμότητα των μεθόδων βαθιάς μάθησης για την αυτοματοποιημένη αναγνώριση και τμηματοποίηση καρκινικών αλλοιώσεων στην θεραπευτική ακτινολογία έχει δημοσιευθεί από αρκετές μελέτες [38].

Η ροή εργασίας της εν λόγω μεθόδου διαφέρει από την αντίστοιχη που περιγράφηκε παραπάνω και αφορά τα ραδιολογικά χαρακτηριστικά στην βάση των χαρακτηριστικών. Στην προκειμένη περίπτωση, αντί των μαθηματικά προκαθορισμένων χαρακτηριστικών, γίνεται χρήση νευρωνικών δικτύων με διάφορες αρχιτεκτονικές, όπως τα συνελκτικά νευρωνικά δίκτυα (convolutional neural networks - CNNs) ή οι αυτόματοι κωδικοποιητές που χρησιμοποιούνται για την παραγωγή και την αναγνώριση των σημαντικότερων χαρακτηριστικών από τα δεδομένα εισόδου. Συγκεκριμένα, τα δίκτυα αυτόματης κωδικοποίησης, τα οποία αποτελούν μια ειδική εκδοχή μη επιβλεπόμενων συνελκτικών νευρωνικών δικτύων, στοχεύουν στη συμπίεση του περιεχομένου της εικόνας και στη απεικόνιση του σε ένα σχετικά βραχύ αλλά αντιπροσωπευτικό διάνυσμα χαρακτηριστικών [39]. Σε γενικές γραμμές, η μέθοδος χρησιμοποιεί ένα αλληλοδιαδοχικό σύστημα νευρωνικών δικτύων μονής στρώσης (cascaded system of single-layer neural networks) που εκπαιδεύονται να εντοπίζουν συσχετιζόμενες δομές που οδηγούν στην ταξινόμηση των δεδομένων απεικόνισης [40]. Έπειτα εφαρμόζονται περαιτέρω συνδυασμοί από τα εξαχθέντα διανύσματα χαρακτηριστικών για να δημιουργήσουν νέα χαρακτηριστικά με υψηλότερο επίπεδο αφαίρεσης. Στο τελικό στάδιο, τα αναγνωρισμένα χαρακτηριστικά μπορούν να χρησιμοποιηθούν από το ίδιο το νευρικό δίκτυο, ή εγκαταλείποντας το δίκτυο να υποβληθούν στη διαδικασία δημιουργίας μοντέλων παρόμοια με εκείνη της πρώτης κατηγορίας ραδιολογικών χαρακτηριστικών. Οι προσεγγίσεις αυτές αφορούν την χρήση συμβατικών μεθόδων μηχανικής μάθησης για την ταξινόμηση των χαρακτηριστικών, όπως δέντρα αποφάσεων (decision trees), μοντέλα παλινδρόμησης (regression models) ή μηχανές διανυσμάτων υποστήριξης (support vector machines) κ.α. Αξίζει να σημειωθεί, ότι ενώ η πρώτη κατηγορία ραδιολογικών χαρακτηριστικών (feature-based features) απαιτεί πάντα ως είσοδο τμηματοποιημένες εικόνες, στην περίπτωση της βαθιάς μάθησης, ειδικά των CNNs, η χρήση μη-τμηματοποιημένων εικόνων είναι δυνατή.

Δεδομένου ότι τα χαρακτηριστικά παράγονται και εξάγονται άμεσα από τα υποκείμενα δεδομένα καθώς και ένα υποσύνολο των πιο αποδοτικών εξάγεται αυτόματα, η ανάγκη για επακόλουθη επιλογή χαρακτηριστικών είναι σπάνια. Ωστόσο, προκειμένου να μειωθεί ο κίνδυνος υπερβολική προσαρμογής (overfitting risk) στα δεδομένα εκπαίδευσης, γίνεται χρήση τεχνικών κανονικοποίησης και αποβολής βαρών σύνδεσης που προέκυψαν κατά την εκπαίδευση. Ένας βασικός περιορισμός των τεχνικών αυτών είναι η υψηλή συσχέτιση μεταξύ των χαρακτηριστικών και των δεδομένων εισόδου, καθώς τα χαρακτηριστικά παράγονται

από τα δεδομένα αυτά. Επομένως, σε αντίθεση με την προηγούμενη κατηγορία ραδιολογικών χαρακτηριστικών, απαιτούνται μεγάλα σύνολα δεδομένων για τον προσδιορισμό ενός βέλτιστου και εύρωστου υποσυνόλου χαρακτηριστικών. Ωστόσο, ο περιορισμός μπορεί να ξεπεραστεί εν μέρει, αν χρησιμοποιήσουμε μια τεχνική μηχανικής μάθησης που ονομάζεται μάθηση με μεταφορά (transfer learning). Στη μάθηση με μεταφορά, χρησιμοποιείται ένα νευρωνικό δίκτυο που έχει ήδη προ-εκπαιδευτεί για διαφορετικό, αλλά στενά συνδεδεμένο σκοπό [41]. Συνεπώς, τόσο ο όγκος των δεδομένων που απαιτούνται για τον προσδιορισμό ενός βέλτιστου υποσυνόλου χαρακτηριστικών όσο και η υπολογιστική απαίτηση είναι μειωμένα.

3.3 Παραδοχή

Στα πλαίσια της παρούσας διπλωματικής μελετήθηκαν τα χαρακτηριστικά που εξήχθησαν από εικόνες καρκίνου των ωοθηκών. Μετά από ενδελεχή αναζήτηση στην πρόσφατη αλλά και παλαιότερη βιβλιογραφία η συντριπτική πλειονότητα των μελετών εστιάζει στην εξαγωγή ραδιολογικών χαρακτηριστικών στην βάση χαρακτηριστικών (feature-based features). Για τον λόγο αυτό, στα πλαίσια μιας πιο μεθοδικής προσέγγισης, καθ ότι η φύση της εργασίας είναι αμιγώς διερευνητική και δεν βασίζεται σε ήδη υπάρχουσα άμεσα συνδεδεμένη εργασία προτιμήθηκε η μονομερής εξαγωγή χαρακτηριστικών, αυτών δηλαδή της πρώτης κατηγορίας. Συγκεκριμένα χρησιμοποιήθηκαν εργαλεία ελεύθερου λογισμικού (π.χ pyradiomics) με έγκυρη και τεκμηριωμένη λειτουργία ενώ η αναζήτηση των χαρακτηριστικών υπήρξε εξαντλητική με περισσότερα από 1200 χαρακτηριστικά να προκύπτουν για κάθε εικόνα εισόδου. Για λόγους πληρότητας και αντικειμενικότητας παρατίθενται και οι δύο κατηγορίες εξαγωγής χαρακτηριστικών με την δεύτερη να αποτελεί πιθανό μελλοντικό ερευνητικό αντικείμενο.

Μηχανική Μάθηση

4.1 Εισαγωγή

Η Μηχανική Μάθηση (Machine Learning) αποτελεί υποκλάδο της Τεχνητής Νοημοσύνης (Artificial Intelligence), που ασχολείται με την κατασκευή μοντέλων που υλοποιούν συγκεκριμένους αλγόριθμους και χρησιμοποιούν πειραματικά δεδομένα με σκοπό την εξαγωγή χρήσιμων προβλέψεων ή συμπερασμάτων. Σκοπός των μοντέλων αυτά να αποκτήσουν την ικανότητα να βελτιώνουν την απόδοσή τους όταν λαμβάνουν επιπλέον δεδομένα χωρίς τον εκ νέου προγραμματισμό τους. Ο Mitchell [42] έχει δώσει τον εξής ορισμό για την μηχανική μάθηση :

« Ένα πρόγραμμα υπολογιστή λέμε ότι μαθαίνει από την εμπειρία E ως προς κάποια κλάση εργασιών T και μέτρο απόδοσης P , αν η απόδοσή του σε εργασίες από το T , όπως μετρείται από το P , βελτιώνεται μέσω της εμπειρίας E .»

Η διαδικασία εκπαίδευσης ξεκινά με την τροφοδότηση του μοντέλου με δεδομένα, συνήθως αριθμητικά από τα οποία δημιουργούνται διανύσματα (Διανύσματα Χαρακτηριστικών - Feature Vectors) με τα οποία τροφοδοτούνται κατάλληλοι αλγόριθμοι. Στο τέλος της διαδικασίας χρησιμοποιούνται μετρικές για τον υπολογισμό της αξιοπιστίας και της ευρωστίας του κάθε μοντέλου.

Οι αλγόριθμοι μηχανικής μάθησης χωρίζονται σε τρεις μεγάλες κατηγορίες :

- Επιβλεπόμενη Μάθηση - Supervised Learning,
- Μη Επιβλεπόμενη Μάθηση - Unsupervised Learning,
- και Ενισχυτική Μάθηση - Reinforcement Learning.

4.1.1 Επιβλεπόμενη Μάθηση

Η επιβλεπόμενη μάθηση (Supervised learning) περιλαμβάνει κάθε εργασία στην οποία ο αλγόριθμος έχει πρόσβαση σε τιμές εισόδου και εξόδου. Οι τιμές εισόδου ορίζονται ως οι εξωτερικές πληροφορίες που επιτρέπεται να χρησιμοποιεί ο αλγόριθμος, όπως τιμές χαρακτηριστικών και μετα-δεδομένα, ενώ οι τιμές εξόδου, είναι οι συγκεκριμένες ετικέτες του χαρακτηριστικού κλάσης. Αυτό σημαίνει ότι η δομή των δεδομένων είναι ήδη γνωστή και ο στόχος αυτών των προγραμμάτων είναι η εκχώρηση νέων δεδομένων στις σωστές τάξεις

ή κατηγορίες. Στην παρούσα διπλωματική χρησιμοποιήθηκαν αλγόριθμοι εποπτευόμενης μάθησης [43]. Μερικά παραδείγματα αλγορίθμων εποπτευόμενης μάθησης:

- Γραμμικά Μοντέλα - Linear Models,
- Μοντέλα Συνόλων - Ensemble models,
- Μηχανές Διανυσμάτων Υποστήριξης - Support Vector Machines (SVM),
- Πολυεπίπεδα Νευρωνικά Δίκτυα - Multilayer Perceptrons (MLP),
- Δένδρα Αποφάσεων - Decision Trees.

4.1.2 Μη Επιβλεπόμενη Μάθηση

Η μη επιβλεπόμενη μάθηση περιλαμβάνει όλες τις διαδικασίες που δεν έχουν πρόσβαση σε τιμές εξόδου και επομένως προσπαθούν να βρουν δομές εντός των δεδομένων δημιουργώντας τάξεις μόνοι τους [43]. Μερικά παραδείγματα αλγορίθμων μη εποπτευόμενης μάθησης: k-means (Clustering) και Νευρωνικά Δίκτυα (Neural Networks)

4.1.3 Ενισχυτική Μάθηση

Η ενισχυτική μάθηση διαφέρει από τις προηγούμενες κατηγορίες, καθώς εστιάζει στη λήψη βέλτιστων αποφάσεων. Τα μοντέλα ενισχυτικής μάθησης περιέχουν έναν πράκτορα (agent), ο οποίος αλληλεπιδρά με ένα περιβάλλον (environment). Ο πράκτορας λαμβάνει κάποια παρατήρηση (observation) από το περιβάλλον, εκτελεί μια ενέργεια και, στη συνέχεια, λαμβάνει μια ανταμοιβή (reward) από το περιβάλλον. Αυτή η διαδικασία συνεχίζεται επαναληπτικά. Η συμπεριφορά του πράκτορα διέπεται από μια πολιτική (policy), η οποία είναι μια συνάρτηση που αντιστοιχίζει τις παρατηρήσεις του περιβάλλοντος σε ενέργειες. Ο στόχος του αλγορίθμου ενίσχυσης της μάθησης είναι να παράγει μια καλή πολιτική [44].

4.2 Ταξινομητές Μηχανικής Μάθησης

Η ταξινόμηση στη μηχανική μάθηση και στη στατιστική είναι μια επιβλεπόμενη μαθησιακή προσέγγιση στην οποία ο αλγόριθμος μαθαίνει από τα δεδομένα που του δίνονται και κάνει νέες παρατηρήσεις ή ταξινομήσεις. Προβλέπει την κλάση ή κατηγορία για μια μεταβλητή εισόδου από τα στοιχεία δεδομένων και χρησιμοποιείται καλύτερα όταν η έξοδος έχει πεπερασμένες και διακριτές τιμές. Στόχος της διαδικασίας είναι να προσδιοριστεί σε ποια κατηγορία θα ανήκουν τα νέα δεδομένα. Ο ταξινομητής, σε αυτήν την περίπτωση, χρειάζεται εκπαιδευτικά δεδομένα για να κατανοήσει πώς οι δεδομένες μεταβλητές εισόδου σχετίζονται με την τάξη. Στα πλαίσια της διπλωματικής εργασίας χρειάστηκε να χρησιμοποιηθούν αλγόριθμοι ταξινόμησης προκειμένου να κατηγοριοποιηθούν οι καρκινικές βλάβες(όγκοι) που απεικονίζονται στις εικόνες μαγνητικού τομογράφου σε τρεις κατηγορίες:

- Περιοχή Χαμηλής Επικινδυνότητας 0,
- Περιοχή Μέσης Επικινδυνότητας 1,
- Περιοχή Υψηλής Επικινδυνότητας 2.

4.2.1 Γραμμικά Μοντέλα - Linear Models

Τα ακόλουθα είναι ένα σύνολο μεθόδων που προορίζονται για παλινδρόμηση (regression) στις οποίες η τιμή - στόχος αναμένεται να είναι γραμμικός συνδυασμός των χαρακτηριστικών. Στη μαθηματική σημειογραφία, εάν \hat{y} είναι η προβλεπόμενη τιμή :

$$\hat{y}(w, x) = w_0 + w_1 x_1 + \dots + w_p x_p \quad (4.1)$$

Για τους σκοπούς της ταξινόμησης χρησιμοποιήσουμε γενικευμένα γραμμικά μοντέλα, όπως η Λογιστική Παλινδρόμηση (Logistic regression).

Λογιστική Παλινδρόμηση (Logistic regression)

Η Λογιστική Παλινδρόμηση είναι μια στατιστική μέθοδος που μπορεί να χρησιμοποιηθεί για ταξινόμηση δυο ή περισσότερων κλάσεων. Ο ταξινομητής αυτός είναι ένα γραμμικό μοντέλο ταξινόμησης που βασίζεται στην λογιστική συνάρτηση (logistic function). Ως πρόβλημα βελτιστοποίησης, η λογιστική παλινδρόμηση με ποινή ελαχιστοποιεί την ακόλουθη συνάρτηση κόστους (λογαριθμική πιθανοφάνεια):

$$-\log \mathcal{L}(\mathbf{b}, \mathbf{y}) = -\frac{1}{N} \left[\left(\sum_{i=0}^N y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right) - R(\mathbf{b}, \gamma) \right], \quad (4.2)$$

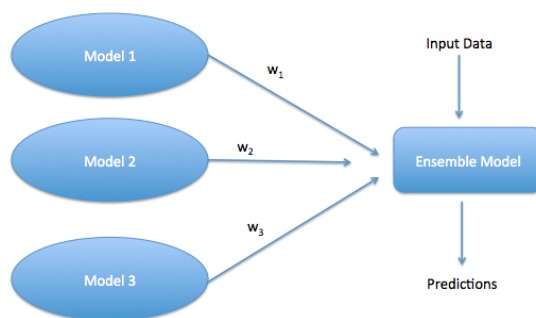
όπου

$$R(\mathbf{b}, \gamma) = \begin{cases} \frac{\gamma}{2} \|\mathbf{beta}\|_2^2 & : \text{'l2' ποινή} \\ \gamma \|\beta\|_1 & : \text{'l1' ποινή} \end{cases} \quad (4.3)$$

Για λόγους πληρότητας θα αναφερθούν κάποιες βασικές μεταβλητές αυτού του αλγορίθμου. Η παράμετρος C σχετίζεται με την αντίστροφη τιμή της ισχύος ομαλοποίησης των δεδομένων. Μικρότερες τιμές της οδηγούν σε ισχυρότερη ομαλοποίηση. Μέσω της μεταβλητής 'solver' ορίζεται ο αλγόριθμος που θα χρησιμοποιηθεί για την επίλυση του προβλήματος βελτιστοποίησης. Οι τιμές που μπορεί να έχει η μεταβλητή 'solver' είναι 'newton-cg', 'lbfgs', 'liblinear', 'sag' και 'saga'. Ο 'liblinear' είναι καλή επιλογή για μικρά σύνολα δεδομένων. Ο 'sag' και 'saga' είναι γρηγορότεροι για μεγάλα σύνολα δεδομένων. Οι 'newton-cg', 'lbfgs', 'sag' και 'saga' μπορούν να χειριστούν ποινή L_2 ή καθόλου ποινή (όπου η ποινή, όπως και προηγουμένως, αναφέρεται σε λανθασμένες ταξινομήσεις). Η ρύθμιση της παραμέτρου `multi_class` σε πολυωνυμική με αυτούς τους επιλυτές εκπαιδεύει ένα πραγματικό μοντέλο πολυωνυμικής λογιστικής παλινδρόμησης, το οποίο σημαίνει ότι οι εκτιμήσεις πιθανότητας θα βαθμονομηθούν καλύτερα από την προεπιλεγμένη ρύθμιση 'one-vs-rest' [45]. Ακόμα μια σημαντική μεταβλητή αποτελεί η 'max_iter', όπου μέσω της τιμής αυτής της παραμέτρου καθορίζεται ο μέγιστος αριθμός επαναλήψεων για σύγκλιση των επιλυτών (καθώς η επίλυση γίνεται με επαναληπτικούς αλγορίθμους αφού το πρόβλημα δεν επιλύεται αλγεβρικά).

4.2.2 Μοντέλα Συνόλων - Ensemble models

Στόχος των μοντέλων συνόλων είναι να συνδυάσουν τις προβλέψεις αρκετών βασικών εκτιμητών που έχουν κατασκευαστεί με έναν δεδομένο αλγόριθμο εκμάθησης προκειμένου



Σχήμα 4.1: Μοντέλα Συνόλων

να βελτιωθεί η γενικευσιμότητα / ευρωστία έναντι ενός μόνο εκτιμητή. Διακρίνονται συνήθως δύο οικογένειες μεθόδων συνόλου :

Στις μεθόδους μέσου όρου, η βασική αρχή είναι να δημιουργηθούν διάφοροι εκτιμητές ανεξάρτητα και στη συνέχεια να μετρηθούν οι προβλέψεις τους. Κατά μέσο όρο, ο συνδυασμένος εκτιμητής είναι συνήθως καλύτερος από οποιονδήποτε απλό εκτιμητή βάσης, επειδή η διακύμανση του είναι μειωμένη. Παραδείγματα συνιστούν οι μέθοδοι bagging, τα δάση τυχαίων δέντρων κ.α.

Στις μεθόδους ενίσχυσης, αντίθετα, οι βασικοί εκτιμητές κατασκευάζονται διαδοχικά με σκοπό κάποιος εξ αυτών να προσπαθεί να μειώσει την προκατάληψη του συνδυασμένου εκτιμητή. Το κίνητρο είναι ο συνδυασμός αρκετών αδύναμων μοντέλων για την δημιουργία ενός ισχυρού συνόλου. Παραδείγματα συνιστούν ο αλγόριθμος AdaBoost καθώς και ο αλγόριθμος Ενίσχυσης Σύστασης Δένδρων (Gradient Tree Boosting).

AdaBoost

Ο ταξινομητής AdaBoost βασίζεται, για την παραγωγή της τελικής πρόβλεψης, στην ιδέα της δημιουργίας ενός κανόνα υψηλής ακρίβειας πρόβλεψης συνδυάζοντας πολλούς σχετικά αδύναμους και ανακριβείς κανόνες, όπως αυτούς που δημιουργούνται από μικρά δέντρα αποφάσεων, που προέκυψαν από επανειλημμένα τροποποιημένες εκδόσεις των δεδομένων [46]. Πρακτικά γίνονται πολλαπλές εκτελέσεις ενός αλγορίθμου ταξινόμησης ρυθμίζοντας κάθε φορά τις παραμέτρους του. Για να παραχθεί η τελική πρόβλεψη, οι προβλέψεις από όλους τους συνδυασμούς ομαδοποιούνται με βάρη (με ειδικό βάρος για την κάθε επανάληψη) [47]. Για κάθε διαδοχική επανάληψη μέσα στο μοντέλο, τα βάρη του δείγματος επαναπροσδιορίζονται ανάλογα με την επιτυχία ή την αποτυχία της ταξινόμησης. Τα βάρη των εσφαλμένα ταξινομημένων περιπτώσεων ρυθμίζονται έτσι ώστε οι επόμενοι ταξινομητές να εστιάζουν περισσότερο σε δύσκολες περιπτώσεις.

Ο μέγιστος αριθμός εκτιμητών στους οποίους τερματίζεται η ενίσχυση μπορεί να προσδιοριστεί, αν είναι επιθυμητό. Σε περίπτωση τέλει τοποθέτησης, η διαδικασία της μάθησης διακόπτεται νωρίτερα. Ως «εκτιμητής βάσης» ορίζεται ο αλγόριθμος ταξινόμησης που χρησιμοποιείται κατά την εκτέλεση του ταξινομητή AdaBoost.

Bagging

Στους αλγορίθμους συνόλων, οι μέθοδοι bagging σχηματίζουν μια κατηγορία αλγορίθμων που δημιουργούν πολλές εμφανίσεις ενός εκτιμητή μαύρου κουτιού (black-box estimator) σε τυχαία υποσύνολα του αρχικού σετ εκπαίδευσης και στη συνέχεια συγκεντρώνουν τις μεμονωμένες προβλέψεις τους για να σχηματίσουν μια τελική πρόβλεψη. Αυτές οι μέθοδοι χρησιμοποιούνται ως τρόπος μείωσης της διακύμανσης ενός βασικού εκτιμητή (π.χ. ενός δέντρου αποφάσεων), εισάγοντας τυχαιοποίηση στη διαδικασία κατασκευής του και στη συνέχεια δημιουργώντας ένα σύνολο από αυτό. Σε πολλές περιπτώσεις, οι μέθοδοι συσκευασίας αποτελούν έναν πολύ απλό τρόπο βελτίωσης σε σχέση με ένα μόνο μοντέλο, χωρίς να απαιτείται η προσαρμογή του βασικού αλγορίθμου βάσης. Οι μέθοδοι bagging λειτουργούν καλύτερα με ισχυρά και πολύπλοκα μοντέλα (π.χ., πλήρως αναπτυγμένα δέντρα αποφάσεων) καθώς παρέχουν έναν τρόπο για τη μείωση της υπερβολικής προσαρμογής του μοντέλου, σε αντίθεση με τις μεθόδους ενίσχυσης που συνήθως λειτουργούν καλύτερα με αδύναμα μοντέλα (π.χ., ρηχά δέντρα αποφάσεων).

Υπάρχουν διάφορες εκδοχές μεθόδων bagging, οι οποίες διαφέρουν ως επί το πλείστον μεταξύ τους με τον τρόπο που σχεδιάζουν τυχαία υποσύνολα του σετ εκπαίδευσης. Για παράδειγμα:

- Όταν τυχαία υποσύνολα του συνόλου δεδομένων σχεδιάζονται ως τυχαία υποσύνολα των δειγμάτων, τότε αυτός ο αλγόριθμος είναι γνωστός ως (Pasting) [48].
- Όταν τα δείγματα λαμβάνονται με αντικατάσταση, τότε η μέθοδος είναι γνωστή ως Bagging [49].
- Όταν τυχαία υποσύνολα του συνόλου δεδομένων σχεδιάζονται ως τυχαία υποσύνολα των χαρακτηριστικών, τότε η μέθοδος είναι γνωστή ως Τυχαία Υποσύνολα (Random Subspaces) [50].
- Τέλος, όταν οι εκτιμητές βάσης βασίζονται σε υποσύνολα τόσο δειγμάτων όσο και χαρακτηριστικών, τότε η μέθοδος είναι γνωστή ως Τυχαία Τμήματα (Random Patches) [51].

Τυχαία Δάση (Random Forest)

Τα Τυχαία Δάση (Random Forest) είναι ένας συνδυασμός δέντρων πρόβλεψης (tree predictors), όπως είναι τα δέντρα απόφασης, όπου κάθε δέντρο εξαρτάται από τις τιμές ενός τυχαίου διανύσματος, ανεξάρτητου από τα υπόλοιπα διανύσματα αλλά με την ίδια κατανομή τιμών με αυτά [52]. Η τυχαιότητα στα δάση παράγει δέντρα απόφασης με ελαφρώς αποσυνδεδεμένα σφάλματα πρόβλεψης αντιμετωπίζοντας το πρόβλημα ότι τα δέντρα απόφασης εμφανίζουν συνήθως μεγάλη διακύμανση και τείνουν να έχουν προβλήματα υπερπροσαρμογής (over-fitting). Ένα τυχαίο δάσος ταιριάζει έναν αριθμό ταξινομητών δέντρων απόφασης στα δεδομένα και χρησιμοποιεί τον μέσο όρο για τη βελτίωση της ακρίβειας πρόβλεψης και να ελέγξει το over-fitting.

Ο αριθμός των δέντρων του δάσους και το μέγιστο βάθος κάθε δέντρου συνιστούν ρυθμιζόμενες παράμετροι από τον χρήστη. Το μέγιστο βάθος του δέντρου μπορεί να οριστεί

ως κανένα (‘one’) ή ως ένας ακέραιος αριθμός. Ειδικά για την πρώτη περίπτωση, οι κόμβοι επεκτείνονται έως ότου όλα τα φύλλα είναι «καθαρά» (δηλαδή όταν κάθε ένα αντιστοιχεί σε μία μόνο κλάση) ή έως ότου όλα τα φύλλα περιέχουν λιγότερα από τον ελάχιστο αριθμό δειγμάτων που απαιτείται για τον διαχωρισμό ενός εσωτερικού κόμβου, όπου τα δείγματα είναι οι παρατηρήσεις (άτομα) και κόμβοι είναι σημεία διαχωρισμού των παρατηρήσεων σε υποσύνολα βάσει της τιμής ενός χαρακτηριστικού.

4.2.3 Μηχανές Διανυσμάτων Υποστήριξης - (Support Vector Machines (SVM))

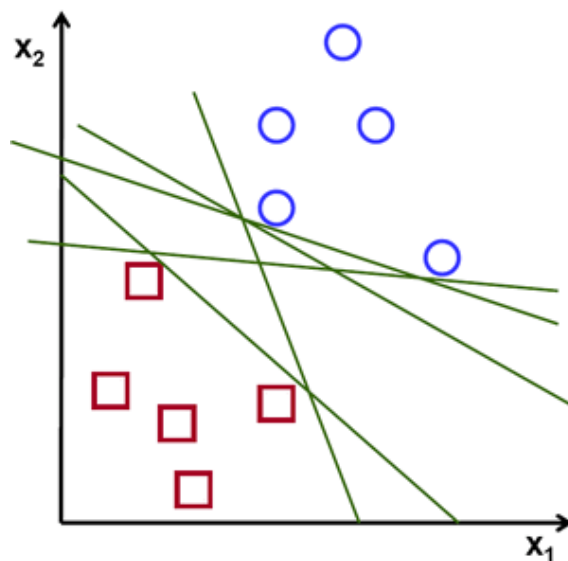
Οι Μηχανές Διανυσμάτων Υποστήριξης είναι μια χρήσιμη τεχνική για την ταξινόμηση δεδομένων. Παρά την φαινομενική ευκολία στην χρήση των Νευρωνικών Δικτύων, ωστόσο, μερικές φορές οδηγούμαστε σε μη ικανοποιητικά αποτελέσματα [53]. Μια εργασία ταξινόμησης συνήθως περιλαμβάνει την εκπαίδευση και τη δοκιμή δεδομένων που αποτελούνται από ορισμένες παρουσίες δεδομένων [54]. Κάθε παρουσία στο σύνολο εκπαίδευσης περιέχει μία τιμή στόχου και πολλά χαρακτηριστικά. Ο στόχος του SVM είναι να παράγει ένα μοντέλο που προβλέπει την τιμή-στόχο (target value) των παρουσιών δεδομένων στο σύνολο δοκιμών που δίνεται μόνο στα χαρακτηριστικά [55]. Αποτελούν υπολογιστικούς αλγορίθμους που κατασκευάζουν ένα υπερεπίπεδο σε άπειρο διαστατικό χώρο. Τα SVM μπορούν να χρησιμοποιηθούν για ταξινόμηση, παλινδρόμηση ή άλλες εργασίες. Διαισθητικά, ένας διαχωρισμός μεταξύ δύο γραμμικά διαχωρίσιμων τάξεων επιτυγχάνεται από οποιοδήποτε υπερεπίπεδο που δεν παρέχει κακή ταξινόμηση σε όλα τα σημεία δεδομένων οποιασδήποτε από τις θεωρούμενες κατηγορίες, δηλαδή, όλα τα σημεία που ανήκουν στην κατηγορία A επισημαίνονται ως +1, για παράδειγμα, και όλα τα σημεία που ανήκουν στην κατηγορία B φέρουν την ένδειξη -1 [56].

Υπάρχουν πολλά υπερεπίπεδα που μπορούν να ταξινομήσουν το ίδιο σύνολο δεδομένων όπως φαίνεται στο (Σχ.4.2) παρακάτω. Οι Μηχανές Διανυσμάτων Υποστήριξης είναι μια προσέγγιση όπου ο στόχος είναι να βρεθεί το καλύτερο υπερεπίπεδο διαχωρισμού, δηλαδή το υπερεπίπεδο που παρέχει την υψηλότερη απόσταση περιθωρίου μεταξύ των πλησιέστερων σημείων των δύο κατηγοριών (ονομάζεται λειτουργικό περιθώριο). Αυτή η προσέγγιση, γενικά, εγγυάται ότι όσο μεγαλύτερο είναι το περιθώριο τόσο χαμηλότερο είναι το σφάλμα γενίκευσης του ταξινομητή. Εάν υπάρχει τέτοιο υπερεπίπεδο, είναι σαφές ότι παρέχει το καλύτερο όριο διαχωρισμού μεταξύ των δύο κατηγοριών και είναι γνωστό ως υπερεπίπεδο μέγιστου περιθωρίου και ένας τέτοιος γραμμικός ταξινομητής είναι γνωστός ως ταξινομητής μέγιστου περιθωρίου.

Μαθηματική Διατύπωση

Διαθέτοντας l δείγματα εκπαίδευσης $\{x_i, y_i\}$, $i = 1, \dots, l$, όπου κάθε δείγμα έχει d εισόδους ($x_i \in R^d$), και μία κλάση(ετικέτα) που λαμβάνει μία ή δύο τιμές ($y_i \in \{-1, 1\}$). Τα υπερεπίπεδα στο R^d παραμετροποιούνται από ένα διάνυσμα (w), και μία σταθερά (b), όπως ορίζονται στην εξίσωση:

$$w \cdot x + b = 0 \quad (4.4)$$



Σχήμα 4.2: Υπερεπίπεδα διαχωρισμού. Πλήθος υπερεπιπέδων διαχωρισμού με διαφορετικό λειτουργικό περιθώριο το καθένα.

(Το ω είναι ένα ορθογώνιο ως προς το υπερεπίπεδο διάνυσμα.) Δοθέντος του υπερεπιπέδου (ω, b) που διαχωρίζει τα δεδομένα, προκύπτει η συνάρτηση:

$$f(\mathbf{x}) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b) \quad (4.5)$$

που διαχωρίζει -δυνητικά- σωστά τα δεδομένα. Ωστόσο το δοθέν υπερεπίπεδο που αναπαρίσταται από το (ω, b) εκφράζεται εξίσου από όλα τα ζεύγη $\{\beta\omega, \beta b\}$ για $\beta \in \mathbb{R}^+$. Ορίζουμε το *κανονικό υπερεπίπεδο* ως εκείνο που διαχωρίζει τα δεδομένα από το υπερεπίπεδο με απόσταση τουλάχιστον ίση με ακριβώς 1. Συνεπώς για ορισμένο υπερεπίπεδο, η κλιμάκωση (το β) είναι έμμεσα ορισμένη 1. Θεωρούμε οπότε τις ακόλουθες συνθήκες:

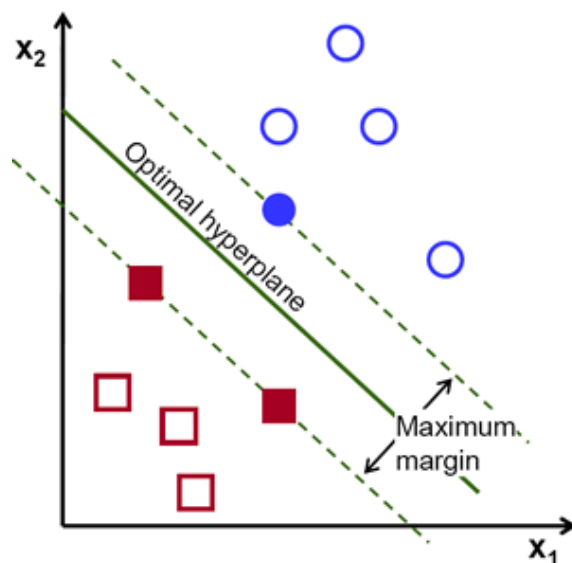
$$\mathbf{x}_i \cdot \mathbf{w} + b \geq +1 \quad \text{όταν } y_i = +1 \quad (4.6)$$

$$\mathbf{x}_i \cdot \mathbf{w} + b \leq -1 \quad \text{όταν } y_i = -1 \quad (4.7)$$

ή πιο συνοπτικά:

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 \quad \forall i \quad (4.8)$$

Τα υπερεπίπεδα αυτά έχουν μια «λειτουργική απόσταση» ≥ 1 (στην πράξη η τιμή της είναι ≥ 1), η οποία δεν θα πρέπει να συγχέεται με την «γεωμετρική απόσταση» ή την «Ευκλείδεια απόσταση». Για ορισμένο υπερεπίπεδο (ω, b) , όλα τα ζεύγη $\{\beta\omega, \beta b\}$ καθορίζουν το ακριβώς ίδιο υπερεπίπεδο, αλλά κάθε ζεύγος έχει την δική του «λειτουργική απόσταση» για δεδομένο σημείο. Για να αποκτήσουμε την γεωμετρική απόσταση από το υπερεπίπεδο προς το ορισμένο σημείο, πρέπει αρχικά να κανονικοποιήσουμε το πλάτος ω . Η απόσταση αυτή



Σχήμα 4.3: Υπερεπίπεδα διαχωρισμού. Αναπαράσταση μεγιστοποίησης της γεωμετρικής απόστασης μεταξύ εγγύτερων σημείων των κλάσεων και υπερεπιπέδου.

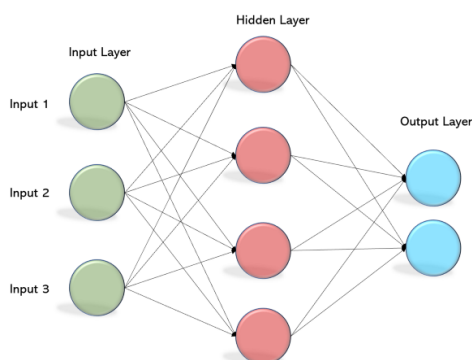
είναι:

$$d((\mathbf{w}, b), \mathbf{x}_i) = \frac{y_i(\mathbf{x}_i \cdot \mathbf{w} + b)}{\|\mathbf{w}\|} \geq \frac{1}{\|\mathbf{w}\|} \quad (4.9)$$

Διαισθητικά επιθυμούμε να μεγιστοποιηθεί η γεωμετρική απόσταση μεταξύ του υπερεπιπέδου με τα εγγύτερα σημεία του συνόλου δεδομένων.

4.2.4 Πολυεπίπεδα Νευρωνικά Δίκτυα (Multilayer Perceptrons (MLP))

Τα Πολυεπίπεδα Νευρωνικά Δίκτυα είναι ένας ταξινομητής πολλαπλών επιπέδων που βασίζεται στον αλγόριθμο Perceptron. Το νευρωνικό δίκτυο Perceptron είναι παρόμοιο με τα μοντέρνα νευρωνικά δίκτυα, με την διαφορά ότι έχει μόνο ένα κρυφό στρώμα με ρυθμιζόμενα βάρη και κατώφλια μεταξύ των στρωμάτων εισόδου και εξόδου όπως φαίνεται και στο (Σχ.4.4) [57]. Τα MLP ανήκουν στον κλάδο των εμπρόσθιας τροφοδότησης (feedforward) τεχνητών νευρωνικών δικτύων [47]. Η βιβλιοθήκη scikit-learn παρέχει τον ταξινομητή MLPClassifier για δίκτυα MLP. Ο MLPClassifier εκπαιδεύεται επαναληπτικά αφού σε κάθε βήμα του χρόνου υπολογίζονται οι μερικοί παράγωγοι της συνάρτησης απώλειας σε σχέση με τις παραμέτρους του μοντέλου για την ενημέρωση αυτών. Το μοντέλο αυτό βελτιστοποιεί την συνάρτηση της λογαριθμικής απώλειας (log-loss function) ή της στοχαστικής κλίσης καθόδου (stochastic gradient descent) ανάλογα με τον επιλυτή που έχει οριστεί από το χρήστη (αν είναι ο 'lbfgs' συμβαίνει το πρώτο ενδεχόμενο, ενώ αν είναι ο 'sgd' ή ο 'adam' το δεύτερο). Πολύ σημαντικό ρόλο στην λειτουργία και στην επίδοση του MLP έχουν οι παράμετροι που χρησιμοποιούνται για την εκπαίδευσή του.



Σχήμα 4.4: Δομή ενός ταξινομητή MLP.

Μαθηματική Διατύπωση

Δίνεται ένα σύνολο δειγμάτων εκπαίδευσης $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ που $x_i \in \mathbf{R}^n$ και $y_i \in \{0, 1\}$, είναι ένα κρυφό στρώμα ενός κρυφού νευρώνα MLP που «μαθαίνει» τη συνάρτηση $f(x) = W_2 g(W_1^T x + b_1) + b_2$ όπου $W_1 \in \mathbf{R}^m$ και $W_2, b_1, b_2 \in \mathbf{R}$ είναι παράμετροι του μοντέλου. Τα W_1, W_2 αντιπροσωπεύουν τα βάρη του στρώματος εισόδου και του κρυφού στρώματος, αντίστοιχα. Τα b_1, b_2 αντιπροσωπεύουν την πόλωση (bias) που προστίθεται στο κρυφό στρώμα και το επίπεδο εξόδου, αντίστοιχα. $g(\cdot) : \mathbf{R} \rightarrow \mathbf{R}$ είναι η συνάρτηση ενεργοποίησης, που ορίζεται συνήθως ως η υπερβολική εφαπτομένη. Δίνεται ως,

$$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (4.10)$$

Για δυαδική ταξινόμηση, η $f(x)$ περνά μέσα από τη λογιστική συνάρτηση $g(z) = 1/(1 + e^{-z})$ για τη λήψη τιμών εξόδου μεταξύ 0 και 1. Ως κατώτατο όριο, το 0,5, κατατάσσει δείγματα εξόδου μεγαλύτερα ή ίσα του 0,5 στην θετική κλάση και τα υπόλοιπα στην αρνητική κλάση. Εάν υπάρχουν περισσότερες από δύο κλάσεις, το ίδιο θα ήταν ένα διάνυσμα μεγέθους (n_{classes}).

Αντί της λογιστικής συνάρτησης γίνεται χρήση της συνάρτησης softmax, η οποία γράφεται ως,

$$\text{softmax}(z)_i = \frac{\exp(z_i)}{\sum_{i=1}^k \exp(z_i)} \quad (4.11)$$

όπου το z_i αντιπροσωπεύει το i -στο στοιχείο της εισόδου στην συνάρτηση softmax, το οποίο αντιστοιχεί στην κλάση i με K τον αριθμό των κλάσεων. Το αποτέλεσμα είναι ένα διάνυσμα που περιέχει τις πιθανότητες το x δείγμα να ανήκει σε κάθε κλάση. Η έξοδος είναι η κλάση με την υψηλότερη πιθανότητα. Στην παλινδρόμηση, η έξοδος παραμένει ως $f(x)$. Επομένως, η συνάρτηση ενεργοποίησης της εξόδου είναι μόνο η ταυτοτική συνάρτηση.

Το MLP χρησιμοποιεί διαφορετικές συναρτήσεις απώλειας ανάλογα με τον τύπο του προβλήματος. Η συνάρτηση απώλειας για το πρόβλημα της ταξινόμησης είναι η Διασταυρωμένη Εντροπία (Cross-Entropy), η οποία στη δυαδική περίπτωση είναι,

$$\text{Loss}(\hat{y}, y, W) = -y \ln \hat{y} - (1 - y) \ln (1 - \hat{y}) + a \|W\|_2^2 \quad (4.12)$$

όπου $a\|W\|_2^2$ είναι ένας όρος κανονικοποίησης L_2 (γνωστός και ως ποινή) που «τιμωρεί» πολύπλοκα μοντέλα. Το $a > 0$ είναι μία μη αρνητική υπερ-παράμετρος που ελέγχει το μέγεθος της ποινής.

Ο αλγόριθμος σταματά όταν φτάσει σε έναν προκαθορισμένο μέγιστο αριθμό επαναλήψεων ή όταν η βελτίωση της απώλειας είναι κάτω από έναν συγκεκριμένο, μικρό αριθμό.

4.2.5 Δένδρα Αποφάσεων (Decision Trees)

Τα Δένδρα Αποφάσεων αποτελούν μια πολύ γνωστή τεχνική ταξινόμησης σε διάφορα θέματα αναγνώρισης προτύπων [58]. Ένα δέντρο αποφάσεων αποτελείται από κόμβους για έλεγχο και απόφαση σε χαρακτηριστικά, άκρα για διακλάδωση βάσει τιμών του επιλεγμένου χαρακτηριστικού και «φύλλα» όπου επισυνάπτεται μια μοναδική κλάση ανά φύλλο [59]. Η λειτουργία ενός δέντρου αποφάσεων μπορεί να χωριστεί σε κάποια στάδια. Στο πρώτο στάδιο γίνεται η κατασκευή του και στη συνέχεια γίνεται η αποκοπή (αναρρόφηση) των λιγότερο χρήσιμων κλάδων (όπου κλάδοι είναι οι κανόνες για την κατάταξη των δεδομένων σε κλάσεις) του δέντρου για την καλύτερη εξισορρόπηση της ακρίβειας των δεδομένων εκπαίδευσης με την πολυπλοκότητα του μοντέλου [60]. Το τελικό δέντρο χωρίζει το χώρο των χαρακτηριστικών σε έναν αριθμό επισημασμένων περιοχών.

Το μέγιστο βάθος του δέντρου (`max_depth`) μπορεί να οριστεί ως κανένα (`None`) ή ως ένας ακέραιος αριθμός. Ειδικότερα, για την πρώτη περίπτωση, οι κόμβοι επεκτείνονται έως ότου όλα τα φύλλα είναι «καθαρά» (δηλαδή όταν κάθε ένα αντιστοιχεί σε μία μόνο κλάση) ή έως ότου όλα τα φύλλα περιέχουν λιγότερα από τον ελάχιστο αριθμό δειγμάτων που απαιτείται για τον διαχωρισμό ενός εσωτερικού κόμβου (όπου κόμβος νοείται ένα κανόνας που κατατάσσει βάσει χαρακτηριστικών, μια παρατήρηση σε κλάση).

Μαθηματική Διατύπωση

Με δεδομένα τα διανύσματα εκπαίδευσης $x_i \in R^n$, $i = 1, \dots$, και ένα διάνυσμα ετικέτας $y \in R^l$, ένα δέντρο αποφάσεων χωρίζει αναδρομικά το χώρο των χαρακτηριστικών έτσι ώστε τα δείγματα με τις ίδιες ετικέτες ή παρόμοιες τιμές-στόχους (target values) να ομαδοποιούνται.

Θεωρούμε ότι τα δεδομένα στον κόμβο m αντιπροσωπεύονται από τα σύνολα \mathcal{Q}_m με N_m δείγματα. Για κάθε υποψήφια διαίρεση $\partial = (j, t_m)$ που αποτελείται από ένα χαρακτηριστικό j και ένα κατώφλι t_m , διαχωρίζει τα δεδομένα σε $\mathcal{Q}_m^{left}(\partial)$ και $\mathcal{Q}_m^{right}(\partial)$ υποσύνολα, όπου,

$$\begin{aligned}\mathcal{Q}_m^{left}(\partial) &= \{(x, y) | x_j \leq t_m\} \\ \mathcal{Q}_m^{right}(\partial) &= \mathcal{Q}_m \setminus \mathcal{Q}_m^{left}(\partial)\end{aligned}\tag{4.13}$$

Στη συνέχεια υπολογίζεται η ποιότητα ενός υποψήφιου διαχωρισμού του κόμβου m χρησιμοποιώντας μια συνάρτηση απώλειας (loss function) $H()$, η επιλογή της οποίας εξαρτάται από την φύση του προβλήματος (ταξινόμηση ή παλινδρόμηση).

Έτσι, επιλέγουμε τις παραμέτρους που ελαχιστοποιούν την συνάρτηση απώλειας,

$$\partial^* = \operatorname{argmin}_{\partial} G(\mathcal{Q}_m, \partial)\tag{4.14}$$

Τέλος επαναλαμβάνουμε για τα υποσύνολα $\mathcal{Q}_m^{left}(\partial^*)$ και $\mathcal{Q}_m^{right}(\partial^*)$ έως ότου επιτευχθεί το

μέγιστο επιτρεπόμενο βάθος, $N_m < \min_{samples}$ ή $N_m = 1$.

Κεφάλαιο 5

Πειραματική Διαδικασία και Μεθοδολογία Επεξεργασίας και Ανάλυσης Δεδομένων

Στο κεφάλαιο αυτό περιγράφεται η υλοποίηση του συστήματος, με βάση τη θεωρητική μελέτη που παρουσιάστηκε στο προηγούμενο κεφάλαιο. Αρχικά παρουσιάζεται το πρωτόκολλο, η πειραματική διαδικασία και τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν. Στη συνέχεια δίνονται οι λεπτομέρειες υλοποίησης για τους βασικούς αλγορίθμους του συστήματος καθώς και η δομή του κώδικα.

5.1 Σχεδίαση Κλινικού Πρωτοκόλλου

5.1.1 Κριτήρια Εισαγωγής και Εξαιρέσης Ασθενών

Κριτήρια Εισαγωγής Ασθενών

Στο εν λόγω έργο συμμετέχουν ασθενείς, ανεξαρτήτου ηλικίας με κλινικά (φυσική/γυναικολογική εξέταση), βιοχημικά (ειδικοί καρκινικοί δείκτες, πχ. Ca125) και απεικονιστικά (υπερηχοτομογραφικά, υπολογιστική τομογραφία) ευρήματα συμβατά με καρκίνο των ωοθηκών. Θα συμπεριλαμβάνονται και ασθενείς με μεταστατική νόσο κατά την αρχική διάγνωση. Καμία εκ των ασθενών της μελέτης δεν θα πρέπει να έχει λάβει θεραπεία για τη νόσο στο παρελθόν. Όλες οι ασθενείς θα έχουν εκτιμηθεί κλινικά και θα παραπέμπονται από το τμήμα Γυναικολογικής Ογκολογίας της Α' Μαιευτικής Γυναικολογικής Κλινικής ΕΚΠΑ, του Νοσοκομείου 'Αλεξάνδρα', το οποίο θα είναι και υπεύθυνο για την τήρηση και φύλαξη του φακέλου νοσηλείας τους. Η ένταξη τους στην μελέτη θα πραγματοποιείται μετά από κατάλληλη ενημέρωση από τον θεράποντα ιατρό, και μόνο μετά από γραπτή συγκατάθεση. Στην Εικόνα 1 παρουσιάζεται αναλυτικά το έντυπο συγκατάθεσης που θα κληθούν να συμπληρώσουν οι εξεταζόμενες και το οποίο εγκρίθηκε από την Επιτροπή Ηθικής και Δεοντολογίας του Αρεταιείου Νοσοκομείου [Αρ. Πρωτοκόλλου: 127/17-04-2019]. Όλες οι ασθενείς θα ενταχθούν σε κλινικό πρωτόκολλο παρακολούθησης και θα επανελέγχονται κλινικά, βιοχημικά και με Μαγνητική Τομογραφία, για την καταγραφή της πορείας της νόσου.

Κριτήρια Εξαιρέσης

Εξαιρέθηκαν εκείνες οι ασθενείς:

- στις οποίες αντενδείκνυται η διενέργεια ΜΤ λόγω παρουσίας στο σώμα συσκευών/προθεμάτων μη συμβατών με το μαγνητικό πεδίο (π.χ. βηματοδότες, καρδιακές βαλβίδες, εμφυτευμένοι απινιδωτές, κοχλιακά εμφυτεύματα κ.α.),
- στις οποίες αντενδείκνυται η χρήση ενδοφλεβίου παραμαγνητικού μέσου είτε λόγω γνωστής υπερευαισθησίας σε σκευάσματα γαδολίνιου είτε λόγω νεφρικής ανεπάρκειας,
- στις οποίες, η ιστολογική διάγνωση θα αναδείξει άλλο ιστολογικό τύπο εκτός του υψηλόβαθμου επιθηλιακού καρκίνου ωοθηκών (high grade epithelial ovarian cancer) και συγκεκριμένα χαμηλής κακοηθείας επιθηλιακό καρκίνο ή μη επιθηλιακό νεόπλασμα.

5.1.2 Περιγραφή Διαδικασίας Κλινικού Χειρισμού Ασθενών

Οι ασθενείς θα υποβάλλονται αρχικά σε ΜΤ άνω και κάτω κοιλίας με την εφαρμογή αυστηρά επιλεγμένων και ειδικά προσαρμοσμένων ακολουθιών παλμών για τη μελέτη των απεικονιστικών δεδομένων του όγκου, οι οποίες αναφέρονται αναλυτικά στον Πίνακα 5.1. Μετά το πέρας της εξέτασης, τα δεδομένα που θα συλλεχθούν, οδηγούνται σε κατάλληλη βάση δεδομένων διαθέσιμη στο Αρεταίειο Νοσοκομείο και στο Εθνικό Μετσόβιο Πολυτεχνείο η οποία θα αναπτυχθεί για το σκοπό αυτό. Στη συνέχεια, με ειδικό λογισμικό, τα δεδομένα θα επεξεργάζονται σε κατάλληλο σταθμό εργασίας με τελικό σκοπό την ανάδειξη περιοχών διαφορετικής επιθετικότητας εντός του όγκου, βάσει προκαθορισμένου συστήματος βαθμονόμησης, το οποίο θα προκύψει από τη συνεργασία των ιατρών ακτινολόγων, των ακτινοφυσικών και των εταιρειών που θα συμμετέχουν στην μελέτη. Έτσι θα προκύπτει ένας χρωματικός χάρτης του όγκου (και των μεταστάσεων εφόσον υπάρχουν), στον οποίο θα κωδικοποιούνται με διαφορετικό χρώμα οι περιοχές με επιθετικά απεικονιστικά χαρακτηριστικά από αυτές με λιγότερο επιθετικά απεικονιστικά χαρακτηριστικά.

Στη συνέχεια και κατόπιν συνεννόησης με τους χειρουργούς του τμήματος Γυναικολογικής Ογκολογίας του Νοσοκομείου «Αλεξάνδρα», οι ασθενείς θα υποβάλλονται σε χειρουργική σταδιοποίηση (διενεργούμενη πρώτη χειρουργική διερεύνηση). Τα βήματα που θα ακολουθούνται με ιδιαίτερη επιμέλεια, κατά σειρά, για τη σωστή χειρουργική σταδιοποίηση, θα περιλαμβάνουν: μέση κάθετη τομή για ικανοποιητική πρόσβαση στην άνω κοιλία, λήψη ελεύθερου ασκτικού υγρού ή εκπλύματος περιτοναϊκής κοιλότητας, προσεκτική επισκόπηση και ψηλάφηση όλης της περιτοναϊκής κοιλότητας, επιμελή έλεγχο του πρωτοπαθούς όγκου, εξαίρεση των ύποπτων περιοχών και έλεγχος για εξωπυελική νόσο. Η χειρουργική επέμβαση θα περιλαμβάνει ολική υστερεκτομή μετά των εξαρτημάτων και επιπλοεκτομή για την ακριβή σταδιοποίηση της νόσου. Εάν υπάρχει εμφανής μεταστατική νόσος αυτή θα αφαιρείται. Στη συνέχεια θα γίνεται δειγματοληπτική εξαίρεση πυελικών λεμφαδένων και επί θετικής για κακοήθεια ταχείας βιοψίας, συστηματική λεμφαδεκτομή στην πύελο και παραορτικά. Η χειρουργική διερεύνηση θα καθορίζει το στάδιο της νόσου σύμφωνα με τα κριτηρια της FIGO (Federation International of Gynecology and Obstetrics). Στα προχωρημένα στάδια, θα καταβάλλεται η μεγίστη προσπάθεια για χειρουργική εξαίρεση όσο το δυνατόν περισσότερο όγκου, με σκοπό τη μηδενική μακροσκοπικά υπολειπόμενη νόσο (complete cytoreduction, residual disease 0, R0). Στην προσπάθεια αυτή επιτρέπεται η εφαρμογή ιδιαίτερα επιθετικής χειρουργικής (εντερεκτομή, σπληνεκτομή, ηπατεκτομή, πολλαπλές περιτοναϊκές

εκτομές κ.λπ.)

Σε συνεννόηση με τους χειρουργούς και τον παθολογοανατόμο που θα λαμβάνει το χειρουργικό παρασκεύασμα, τα υποδεικνυόμενα από την απεικόνιση τμήματα του όγκου θα αντιστοιχίζονται στο παρασκεύασμα (βάσει εκτυπωμένου σε τρισδιάστατη μορφή χρωματικού χάρτη του όγκου), θα εξαιρούνται, θα μονιμοποιούνται και θα αποστέλλονται για εξέταση του γονιδιώματος στο μοριακό βιολόγο. Θα ακολουθείται παρόμοια διαδικασία και για τις επιλεγμένες μεταστάσεις. Λόγω του αυξημένου κόστους της γονιδιακής ανάλυσης, θα λαμβάνονται μόνο δύο δείγματα από τον πρωτοπαθή όγκο και ένα από μετάσταση, ενώ θα μελετηθεί και το γονιδίωμα από το περιφερικό αίμα της ασθενούς.

Οι ασθενείς θα λαμβάνουν στη συνέχεια χημειοθεραπευτική αγωγή και θα τίθενται σε παρακολούθηση που θα περιλαμβάνει κλινική εξέταση και έλεγχο καρκινικών δεικτών (CEA, Ca125, Ca19-9, Ca15-3) μετά το τέλος του τρίτου σχήματος της χημειοθεραπείας, στο τέλος αυτής και κάθε τρεις μήνες μετέπειτα, ως το τέλος της μελέτης. Στις ασθενείς που θα διαπιστωθεί υποτροπή, η νόσος θα εκτιμηθεί ξανά με Μαγνητική Τομογραφία (Πίν.5.2), και θα εξαιρεθεί χειρουργικά με σκοπό την εκ νέου ανάλυση του γονιδιώματος αυτών. Σκοπός της φάσης αυτής είναι η σύγκριση των απεικονιστικών χαρακτηριστικών και του γονιδιώματος της υποτροπής με αυτά του αρχικού όγκου για την ανάδειξη τυχόν ενδιάμεσων μεταβολών. Στο τέλος της μελέτης θα πραγματοποιηθεί στατιστική ανάλυση των δεδομένων.

Στόχος μας είναι να μελετήσουμε περίπου 75-80 ασθενείς, συμπεριλαμβανομένων και των ασθενών με υποτροπή.

5.1.3 Πρωτόκολλο Μαγνητικής Τομογραφίας (MT)

Οι εξετάσεις θα πραγματοποιηθούν σε υψηλού πεδίου Μαγνητικό Τομογράφο στη Μονάδα Μαγνητικού Τομογράφου του Α' Εργαστηρίου Ακτινολογίας του ΕΚΠΑ, στο Αρεταίειο Νοσοκομείο. Το πρωτόκολλο ακολουθιών παλμών αποτυπώνεται στους τρεις ακόλουθους πίνακες: Πίνακας.5.1, Πίνακας.5.2, Πίνακας.5.3.

5.1.4 Διαδικασία Επισήμανσης Δειγμάτων

Στη μελέτη θα ενταχθούν γυναίκες με καρκίνο των ωοθηκών, με μακροσκοπικά μετρήσιμη νόσο οι οποίες και θα υποβληθούν σε πρωτογενή κυτταρομειωτική επέμβαση. Η προεγχειρητική εκτίμηση και διεγχειρητική λήψη ιστού θα πραγματοποιείται με βάση το ως άνω περιγραφόμενο πρωτόκολλο. Από τις ασθενείς αυτές, εκείνες οι οποίες θα υποτροπιάζουν και θα αντιμετωπιστούν μετά τη χορήγηση χημειοθεραπείας χειρουργικά, θα υποβληθούν για δεύτερη φορά σε ανάλυση των ιδιαιτεροτήτων του όγκου και θα γίνει σύγκριση με την πρωτογενή νόσο. Η επισήμανση του όγκου (και των περιτοναϊκών μεταστάσεων, εφόσον υπάρχουν) θα πραγματοποιείται διεγχειρητικά με βάση τις ανατομικές πληροφορίες από τον προεγχειρητικό απεικονιστικό έλεγχο. Κατά την εξαίρεση του όγκου θα τοποθετούνται ράμματα τα οποία θα προσδιορίζουν την πρόσθια- οπίσθια επιφάνεια και τον άνω-κάτω πόλο του όγκου και ενδεχομένως περιοχές του όγκου πλησίον συγκεκριμένων σταθερών ανατομικών δομών, ώστε να διευκολύνεται η αναγνώριση της θέσης του όγκου στο χώρο. Ανάλογα θα επισημανθούν και οι επιλεγμένες μεταστάσεις.

5.1.5 Παθολογοανατομική Ανάλυση Δειγμάτων

Θα παραλαμβάνεται το αφαιρεθέν εξάρτημα και οι σεσημασμένες περιτοναϊκές εμφυτεύσεις από το χειρουργείο προκειμένου να γίνει ταχεία βιοψία, αν απαιτείται (για να αποκλεισθεί η πιθανότητα μεταστατικού στην ωοθήκη όγκου). Το παρασκεύασμα θα προσανατολίζεται βάσει των διεγχειρητικών πληροφοριών, των ανατομικών σχέσεων ωοθήκης - σάλπιγγας και με τη βοήθεια της τρισδιάστατης εικόνας που θα το συνοδεύει. Θα λαμβάνονται ιστικές τομές διαστάσεων 20x10x3 mm από τις περιοχές ενδιαφέροντος οι οποίες θα μονιμοποιούνται σε ουδέτερη φορμολή 10% και ακολούθως θα δημιουργούνται κύβοι παραφίνης. Θα ακολουθήσει εξέταση τομής αιματοξυλίνης-ηωσίνης για επιβεβαίωση της διάγνωσης και θα λαμβάνονται τομές από κάθε κύβο οι οποίες θα αποστέλλονται για μοριακό έλεγχο.

5.2 Πειραματική Διαδικασία

Σύμφωνα με το ιατρικό πρωτόκολλο που θα εκτελεστεί στα πλαίσια του έργου, θα συλλεχθεί πλήθος απεικονιστικών δεδομένων για κάθε ασθενή. Η εξαγωγή και σύνθεση των πληροφοριών από τους διάφορους τύπους δεδομένων ανά ασθενή θα προσδώσει τα επιθυμητά ραδιώματα (radiomics), βάσει των οποίων στη συνέχεια θα πραγματοποιηθεί η στατιστική ανάλυση και η ανάπτυξη του απαιτούμενου προγνωστικού μοντέλου.

5.2.1 Συλλεχθέντα Δεδομένα

Σύμφωνα με τις προδιαγραφές του έργου, για τη σύνθεση των επιθυμητών ραδιωμάτων, σε κάθε εξεταζόμενη περίπτωση, θα πρέπει να συλλέγονται οι παρακάτω τύποι απεικονιστικών δεδομένων:

- Υψηλής ανάλυσης τρισδιάστατες ακολουθίες T1 και T2 προσανατολισμού.
 - Έγχρωμους παραμετρικούς χάρτες T1, υψηλής χωρικής διακριτικής ικανότητας,
 - Έγχρωμους παραμετρικούς χάρτες T2, υψηλής χωρικής διακριτικής ικανότητας.
- Ακολουθίες μοριακής διάχυσης.
 - Έγχρωμους παραμετρικούς χάρτες ADC,
 - Έγχρωμους παραμετρικούς χάρτες μικρο-αιματικής διήθησης,
 - Έγχρωμους παραμετρικούς χάρτες καθαρής μοριακής διάχυσης.
- Δυναμική μελέτη σκιαγραφικής ενίσχυσης με έγχυση ενδοφλεβίως παραμαγνητικής ουσίας.
 - Εκτίμηση ημιποσοτικών δεικτών αιματικής διήθησης (Wash-in, Wash-out, Time to Peak, Mean Transit Time, Area under the curve),
 - Εκτίμηση ποσοτικών δεικτών αιματικής διήθησης (Ktrans, Kep, Ve).
- Παραμετρικούς χάρτες θερμοκρασίας.



Σχήμα 5.1: Σύστημα Απεικόνισης Μαγνητικού Συντονισμού το οποίο χρησιμοποιήθηκε για τη συλλογή των δεδομένων του έργου.

5.2.2 Περιγραφή Υλικού Συλλογής Δεδομένων

Για την συλλογή των παραπάνω απαιτούμενων δομών δεδομένων, στα πλαίσια του έργου θα χρησιμοποιηθούν οι υποδομές του Α' Εργαστηρίου Ακτινολογίας της Ιατρικής Σχολής του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών (ΕΚΠΑ), το οποίο στεγάζεται στο Αρεταίειο Νοσοκομείο. Συγκεκριμένα, θα χρησιμοποιηθεί το σύστημα απεικόνισης μαγνητικού συντονισμού (ΑΜΣ), το οποίο φαίνεται στην (Σχ.5.1). Το σύστημα εγκαταστάθηκε το 2017 και είναι κλειστού τύπου με στατικό μαγνητικό πεδίο H_0 έντασης 3 Tesla (T). Πρόκειται για μοντέλο της εταιρίας PHILIPS (INGENIA 3T, MultiTransmit/Dstream). Το σύστημα αυτό επικοινωνεί με ηλεκτρονικά συστήματα αρχειοθέτησης εικόνων (PACS) και με ρομποτικά εκτυπωτικά μέσα (CD/DVD) για την εκτύπωση των εικόνων των εξεταζομένων.

Το εν λόγω σύστημα λειτουργεί βάσει των παρακάτω τεχνικών προδιαγραφών, η ονομαστική τιμή του στατικού μαγνητικού πεδίου είναι 3T μετρημένο στο ισόκεντρο του συστήματος, και η ομοιογένειά του είναι < 1.1 ppm (V_{rms}) για σφαιρικό όγκο διαμέτρου 45cm. Οι μετρήσεις ομοιογένειας αναφέρονται στην μέθοδο μέτρησης με φασματοσκοπία NMR. Η θερμοκρασία του δωματίου του συστήματος διατηρείται σταθερή τεχνητά μέσω κλιματισμού στους $+20^\circ$. Για τις ανάγκες του έργου θα χρησιμοποιηθούν τα στοιχεία του συστήματος απεικόνισης μαγνητικού συντονισμού που περιγράφονται στις ακόλουθες υποενότητες.

Σύστημα βαθμιδωτών Πεδίων

Το σύστημα βαθμιδωτών πεδίων είναι αγωγίμου τύπου με ενεργή αυτοθωράκιση και διατίθεται για το συγκεκριμένο σύστημα στην βασική έκδοσή του με τα εξής χαρακτηριστικά :

- Omega,
- High Power Gradients,

- Gradient maximum Strength: 45 mT/m,
- Slew Rate: 200 mT/m/ms.

Οι παραπάνω τιμές αναφέρονται σε κάθε ορθογώνιο άξονα βαθμίδωσης με κύκλο λειτουργίας 100% ξεχωριστά για κάθε άξονα. Τα αγωγάματα πηνία οδηγούνται από τρία ξεχωριστά τροφοδοτικά και ψύχονται από κατάλληλο κλειστού κύκλου υδρόψυκτο σύστημα ψύξης.

Σύστημα Ραδιοσυχνοτήτων

Το βασικό σύστημα ραδιοσυχνοτήτων (πηνίο σώματος) αποτελείται από τέσσερα (4) κανάλια εκπομπής RF (Οκταπολική η διπλή κυκλική πόλωση εκπομπής) με δύο τελεστικούς ενισχυτές και μέγιστη ηλεκτρική ισχύ κορυφής 2x18 KW. Η τεχνολογία εκπομπής RF είναι ψηφιακού τύπου, με χρήση υδρόψυκτης ψηφιακής συσκευής εκπομπής. Η εκπομπή γίνεται από το πηνίο σώματος με χρήση της ειδικής τεχνολογίας πολλαπλής εκπομπής (MultiTransmit, διπλή κυκλική πόλωση εκπομπής, 2 ενισχυτές) για την βελτιστοποίηση της ομοιογένειας εκπομπής της RF. Με την χρήση της τεχνολογίας αυτής μειώνεται σημαντικά το τεχνικό σφάλμα της διηλεκτρικής σκίασης στις εικόνες σώματος στα συστήματα 3T. Πάντοτε στην εκπομπή δεσμεύονται τέσσερα κανάλια από το πηνίο εκπομπής. Η λήψη γίνεται με τα υπόλοιπα κανάλια του σχετικού πηνίου λήψης. Αν το πηνίο είναι μόνο πηνίο λήψης τότε χρησιμοποιούνται όλα τα κανάλια που συνδέονται με το συγκεκριμένο πηνίο λήψης. Για τη λήψη το σύστημα διαθέτει προενισχυτή και μετατροπέα αναλογικού σε ψηφιακό σήματος (ADC converter) ενσωματωμένους στα ίδια τα πηνία λήψης (τεχνολογία Dstream). Με τον τρόπο αυτόν επιτυγχάνεται σημαντική αύξηση του τελικού λόγου σήματος προς θόρυβο γιατί πρακτικά λόγω των μικρών αποστάσεων (όλες οι συσκευές βρίσκονται κοντά στα πηνία λήψης), μειώνονται σημαντικά οι θόρυβοι μεταφοράς των αναλογικών σημάτων από τα πηνία λήψης στους προενισχυτές και τους ADCs, που στους περισσότερους κατασκευαστές βρίσκονται στα μηχανοστάσια, έξω από τους ακτινολογικούς θαλάμους λειτουργίας των συστημάτων ΑΜΣ.

Στο ψηφιακό σύστημα Dstream είναι εφικτή η ευρυζωνική πολυπλεξία σημάτων από τα πηνιοστοιχεία που αποτελούν ένα κύριο πηνίο προς τους προενισχυτές και τους ADCs που βρίσκονται μέσα ή κοντά σε αυτό το πηνίο και η μετέπειτα οδήγησή τους μέσα από καλώδια οπτικών ινών (optical fibers) στους υπολογιστές ανακατασκευής των εικόνων. Στο σύστημα Dstream δεν υπάρχουν καλώδια μεταφοράς αναλογικών σημάτων από τα πηνιοστοιχεία/πηνία στους υπολογιστές ανακατασκευής των εικόνων.

Λογισμικό Συστήματος ΑΜΣ

Η πλατφόρμα του λογισμικού λειτουργίας του συστήματος απεικόνισης μαγνητικού σπιντονομορίου και μετα-επεξεργασίας των παραγόμενων εικόνων είναι τεχνολογίας PHILIPS (W2008). Το λογισμικό επιτρέπει την επικοινωνία πρωτοκόλλου DICOM 3 με όλες τις επιμέρους υπηρεσίες (Query, retrieve, store, print) για όλα τα περιφερικά συνδεδεμένα εκτυπωτικά μηχανήματα και εξωτερικά συστήματα αρχειοθέτησης εικόνων. Διατίθενται όλες σχεδόν οι ακολουθίες για τον συγκεκριμένο τύπο μηχανήματος όπως SE (SE, TSE, DRIVE, single

shot TSE) και GRE (FFE, turboFE, THRIVE, Balanced FFE, FLAIR, 3D TFE, EPI, PCA), οι περισσότερες με 2D και 3D δυνατότητες λήψης.

Πηνία Συστήματος ΑΜΣ

Στο συγκεκριμένο σύστημα απεικόνισης μαγνητικού συντονισμού που θα χρησιμοποιηθεί για τις ανάγκες του έργου, υπάρχει πάντα μόνιμα εγκατεστημένο το κεντρικό πηνίο εκπομπής/λήψης που βρίσκεται ενσωματωμένο μέσα στον κύλινδρο εισαγωγής εξεταζομένων του συστήματος. Για τις εξετάσεις καρκίνου ωοθηκών θα χρησιμοποιηθεί ως πηνίο λήψης το πηνίο της κοιλιάς (dS Anterior 3.0T). Στο τμήμα διατίθενται δύο τέτοια πηνία. Επίσης υπάρχει συνοδευτικός εξοπλισμός από ομοιώματα και μετρητικά όργανα για τους τακτικούς ποιοτικούς ελέγχους στο σύστημα.

5.2.3 Λειτουργικές Προδιαγραφές Λογισμικού για τα Δεδομένα Εισόδου

Το λογισμικό που θα αναπτυχθεί στα πλαίσια του έργου θα πρέπει να είναι σε θέση να επεξεργαστεί και να διαχειριστεί όλους τους τύπους δεδομένων που αναφέρθηκαν στο υποκεφάλαιο 5.1.4. Για να επιτευχθεί η μέγιστη δυνατή συμβατότητα με το υλικό συλλογής των δεδομένων (υποκεφάλαιο 5.1.4), θα υποστηρίζονται τρισδιάστατα ψηφιακά απεικονιστικά δεδομένα συμβατά με το πρότυπο ιατρικών δεδομένων DICOM.

Συγκεκριμένα, θα υποστηρίζονται τόσο σειρές δεδομένων DICOM, όπου τα δεδομένα αποδίδονται χρησιμοποιώντας μια εικόνα ανά τομή της απεικονιζόμενης ανατομικής περιοχής, όσο και τρισδιάστατες δομές DICOM, των οποίων η ανάγνωση επιτυγχάνεται χρησιμοποιώντας τις πληροφορίες που αποθηκεύονται σε περιγραφικά αρχεία κειμένου DICOMDIR. Όπως και στις περισσότερες περιπτώσεις δεδομένων DICOM, το λογισμικό θα υποστηρίζει εύρος χρώματος 16-bit τόνων του γκρι, το οποίο παρέχει τη δυνατότητα απόδοσης 65536 μοναδικών αποχρώσεων του γκρι για κάθε εικονοστοιχείο των απεικονιζόμενων ανατομικών περιοχών.

Τα δεδομένα θα πρέπει να κατηγοριοποιούνται αρχικά ανά εξεταζόμενη περίπτωση (π.χ. βάσει του ονόματος του ασθενούς και την ημερομηνία λήψης) και στη συνέχεια ανά μέθοδο απεικόνισης (T1, T2, κλπ). Ο διαχωρισμός αυτός θα βοηθήσει στην τήρηση αρχείου εξετάσεων, αλλά και στη βέλτιστη επεξεργασία των δεδομένων από τους υπό ανάπτυξη αλγόριθμους.

5.3 Προεπεξεργασία Δεδομένων Εισόδου

Ο καρκίνος των ωοθηκών είναι ο πέμπτος σε προκλήσεις θανάτων [4] και ευθύνεται για τους περισσότερους θανάτους συγκρινόμενος με τους λοιπούς καρκίνους του αναπαραγωγικού συστήματος της γυναίκας [5]. Παρ' όλα αυτά και παρά το γεγονός ότι με τη χρήση ηλεκτρονικών υπολογιστών το κόστος για την αναζήτηση αυτόματων τρόπων διάγνωσης στηριζόμενων σε επεξεργασία εικόνας είναι ιδιαίτερα μικρό, η βιβλιογραφική έρευνα οδηγεί στο συμπέρασμα ότι δεν υπάρχουν πολλά συστήματα επεξεργασίας ιατρικών εικόνων με εφαρμογή στον καρκίνο των ωοθηκών.

Σκοπός του έργου Destiny είναι με βάση την περιορισμένη βιβλιογραφία να αναπτύξει ένα τέτοιο σύστημα κάνοντας συνδυαστική χρήση εξελιγμένων τεχνικών και αλγορίθμων.

Ειδικότερα, για την υλοποίηση της προδιαγεγραμμένης λειτουργικότητας του προτεινόμενου συστήματος, απαιτείται η χρήση εξελιγμένων τεχνικών επεξεργασίας ψηφιακών, ιατρικών δεδομένων και πιο συγκεκριμένα :

- τεχνικών ευθυγράμμισης τρισδιάστατων ιατρικών δεδομένων,
- τεχνικών κατάτμησης τρισδιάστατων ιατρικών δεδομένων,
- τεχνικών ψηφιακής ανακατασκευής, σύντηξης και απεικόνισης τρισδιάστατων ιατρικών δεδομένων και
- συστημάτων υποστήριξης αποφάσεων και μηχανικής μάθησης.

5.3.1 Ευθυγράμμιση Τρισδιάστατων Ιατρικών Δεδομένων - Registration

Σημαντικό τμήμα της λειτουργικότητας του προτεινόμενου συστήματος, είναι η εφαρμογή κατάλληλου αλγορίθμου ευθυγράμμισης των δεδομένων που επεξεργάζεται το σύστημα. Ένας από τους πρωταρχικούς στόχους του υπό ανάπτυξη λογισμικού, είναι η σύντηξη πληροφοριών από διάφορα απεικονιστικά συστήματα και μεθόδους και η παρουσίασή τους στον τελικό χρήστη με τη μορφή ραδιωμάτων. Για να γίνει αυτό δυνατό, τα δεδομένα που θα προέρχονται από τα διάφορα απεικονιστικά συστήματα και τις διαφορετικές τεχνικές απεικόνισης, θα πρέπει να είναι απολύτως ευθυγραμμισμένα. Πιθανά σφάλματα, αστοχίες ή ανακρίβειες στην ευθυγράμμιση των υπό σύντηξη δεδομένων, είναι πολύ πιθανό να αλλοιώσουν την παρουσιαζόμενη πληροφορία σε τέτοιο βαθμό που να μην μπορεί πλέον να αποτιμηθεί κλινικά. Για το λόγο αυτό, απαιτείται η εφαρμογή αλγορίθμου ευθυγράμμισης των τρισδιάστατων δεδομένων που θα προκύψουν στο πλαίσιο του έργου, ο οποίος θα επιτυγχάνει τη μεγαλύτερη δυνατή ακρίβεια.

Γενικά, ως ευθυγράμμιση ορίζεται η διαδικασία εύρεσης κατάλληλου γεωμετρικού μετασχηματισμού, μέσω του οποίου εξομαλύνονται οι γεωμετρικές αποκλίσεις ανάμεσα σε διαφορετικές απεικονίσεις του ίδιου αντικειμένου ή ανάμεσα σε απεικονίσεις διαφορετικών αντικειμένων. Στην περίπτωση του έργου, οι απεικονίσεις αυτές είναι η λήψεις απεικονιστικών δεδομένων από την περιοχή ενδιαφέροντος (ανατομική περιοχή της μήτρας).

Πρακτικά, η ευθυγράμμιση είναι μια επαναληπτική διαδικασία δοκιμών γεωμετρικών μετασχηματισμών, μέχρι να επιτευχθεί αποδεκτή χωρική σύμπτωση των δεδομένων αναφοράς με τα υπό ευθυγράμμιση δεδομένα. Οι περισσότεροι αλγόριθμοι ευθυγράμμισης απεικονιστικών δεδομένων απαρτίζονται από τα ακόλουθα τρία στοιχεία [61, 62]: (α) μια τεχνική βελτιστοποίησης, η οποία σε κάθε επανάληψη της διαδικασίας υπολογίζει νέες παραμέτρους για τον προτιμώμενο γεωμετρικό μετασχηματισμό, (β) έναν γεωμετρικό μετασχηματισμό, ο οποίος μετασχηματίζει τις συντεταγμένες των σημείων των υπό ευθυγράμμιση δεδομένων σε κάθε επανάληψη και, (γ) μια συνάρτηση σφάλματος/ομοιότητας, η οποία σε κάθε επανάληψη της διαδικασίας υπολογίζει και ποσοτικοποιεί την ακρίβεια του γεωμετρικού μετασχηματισμού.

Στη βιβλιογραφία υπάρχει πλήθος αλγορίθμων βελτιστοποίησης, οι οποίοι μπορούν να ταξινομηθούν σε τρεις γενικές κατηγορίες, ανάλογα με την υπερκείμενη λογική του αλγορίθμου [63, 64]: (α) Μαθηματικά μοντέλα, (β) επαναληπτικοί αλγόριθμοι και, (γ) ευρετικοί

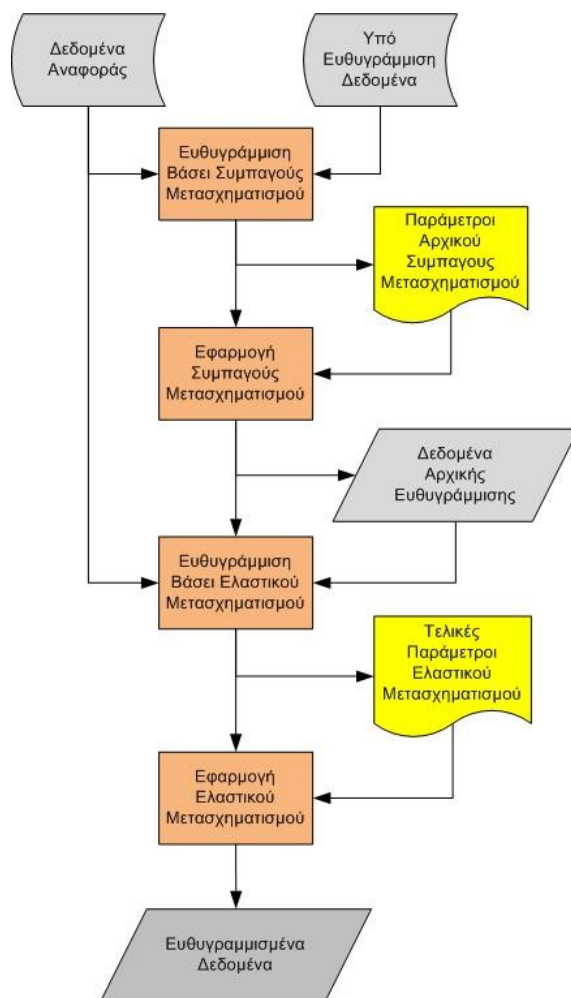
αλγόριθμοι. Στην πρώτη κατηγορία εμπίπτουν οι πλέον διαδεδομένες τεχνικές, όπως η μέθοδος του Βρεντ, η τεχνική εύρεσης χρυσής τομής (Golden Section Search), η μέθοδος μονοδιάστατης εύρεσης πρώτων παραγώγων (One-Dimensional Search with First Derivatives), η μέθοδος Downhill Simplex και η μέθοδος Powell. Στην κατηγορία των επαναληπτικών αλγορίθμων περιλαμβάνονται τεχνικές όπως η μέθοδος του Newton (και οι διάφορες παραλλαγές της), η μέθοδος συζευγμένης κλίσης (Conjugate Gradient), η μέθοδος καθόδου κλίσης (Gradient Descent), τεχνικές εσωτερικών σημείων (Interior Point Methods), τεχνικές υπό-κλίσης (Subgradient Methods), η ελλειψοειδής μέθοδος, διάφορες τεχνικές παρεμβολών (Interpolation Methods) και διάφορες μέθοδοι εύρεσης προτύπων (Pattern Search Methods). Τέλος, σπανιότερα χρησιμοποιούνται πιο πολύπλοκοι ευρετικοί αλγόριθμοι, όπως, ο μιμητικός αλγόριθμος (Memetic algorithm), η τεχνική διαφορικής εξέλιξης (Differential Evolution), η μέθοδος δυναμικής χαλάρωσης (Dynamic Relaxation), γενετικοί αλγόριθμοι (Genetic Algorithms), ο αλγόριθμος Nelder-Mead, ο αλγόριθμος σμήνους σωματιδίων (Particle Swarm Optimization) και ο αλγόριθμος προσομοιωμένης ανόπτωσης (Simulated Annealing).

Στην περίπτωση του προτεινόμενου συστήματος, όπου λόγω της φυσιολογίας των απεικονιζόμενων περιοχών παρατηρούνται ελαστικές παραμορφώσεις ανάμεσα στις συγκρινόμενες δομές δεδομένων (μαλακοί ιστοί), απαιτήθηκε η χρήση ενός ελαστικού γεωμετρικού μετασχηματισμού, προκειμένου να επιτευχθεί η μέγιστη δυνατή ακρίβεια στην ευθυγράμμιση των δεδομένων. Κλασικά παραδείγματα τέτοιων μετασχηματισμών περιλαμβάνουν τους μετασχηματισμούς Thin Plate Splines, B-Splines Demons [65, 66]. Συνήθως, η εφαρμογή του ελαστικού μετασχηματισμού έπεται μιας αρχικής διαδικασίας ταχείας ευθυγράμμισης περιορισμένης ακρίβειας η οποία βασίζεται σε συμπαγή γεωμετρικό μετασχηματισμό, όπως για παράδειγμα τον ευρέως διαδεδομένο αφινικό (affine) μετασχηματισμό [67].

Η προτιμώμενη συνάρτηση σφάλματος/ομοιότητας εξαρτάται από την εφαρμογή της διαδικασίας ευθυγράμμισης και πρωτίτως από το είδος των υπό επεξεργασία δεδομένων. Για το λόγο αυτό έχουν προταθεί πλήθος τέτοιων συναρτήσεων, οι περισσότερες εκ των οποίων προέρχονται από στατιστικές εφαρμογές. Τα σημαντικότερα κριτήρια ομοιότητας που χρησιμοποιούνται ευρέως στην ευθυγράμμιση ψηφιακών εικόνων περιλαμβάνουν: (α) τη μέση τιμή των απόλυτων διαφορών της φωτεινότητας των αντίστοιχων σημείων των υπό ευθυγράμμιση δεδομένων, (β) τον συντελεστή συσχέτισης (Correlation Coefficient) και, (γ) την ομοιομορφία των εικόνων αναλογίας και τον συντελεστή κοινής πληροφορίας (Mutual Information) [68].

Οι παραπάνω συναρτήσεις σχετίζονται άμεσα με τη φωτεινότητα των σημείων των υπό ευθυγράμμιση δεδομένων. Είναι επίσης δυνατή η χρήση κριτηρίων τα οποία δεν εξαρτώνται από τις χρωματικές ιδιότητες των δεδομένων, αλλά κυρίως από τον υπολογισμό αποστάσεων ανάμεσα στις ακμές των υπό ευθυγράμμιση δεδομένων [69]. Η γνωστότερη ίσως μέθοδος που εμπίπτει σε αυτή την κατηγορία είναι η μέθοδος των Χαρτών Απόστασης (Distance Maps) [70]. Οι τεχνικές αυτές, λόγω της αυξημένης ταχύτητας επεξεργασίας, είναι ιδανικές για χρήση σε δεδομένα μεγάλου όγκου, όπως για παράδειγμα στα τρισδιάστατα ιατρικά δεδομένα.

Η όλη διαδικασία της ευθυγράμμισης τρισδιάστατων απεικονιστικών δεδομένων συνοψίζεται στο διάγραμμα της (Σχ.5.2).



Σχήμα 5.2: Διαδικασία κλιμακωτής ευθυγράμμισης απεικονιστικών δεδομένων βάσει αρχικού συμπαγούς μετασχηματισμού και δευτερεύοντος ελαστικού μετασχηματισμού.

5.3.2 Κατάτμηση Τρισδιάστατων Ιατρικών Δεδομένων - Segmentation

Μια πολύ σημαντική προϋπόθεση της αποτίμησης των καρκινικών όγκων είναι η απομόνωση (εξαγωγή) του ίδιου του όγκου, έτσι ώστε η επεξεργασία να εστιαστεί σε πολύ συγκεκριμένη και περιορισμένη περιοχή ενδιαφέροντος. Πρακτικά, από το προτεινόμενο σύστημα είναι αδύνατο να προκύψει εξαγωγή της περιοχής του καρκίνου χωρίς προηγουμένως να έχει εφαρμοστεί κατάτμηση των ιατρικών εικόνων που ελήφθησαν από τον μαγνητικό τομογράφο, είτε απευθείας είτε κατόπιν σύνθεσης πληροφοριών από πλήθος εικόνων.

Τμηματοποίηση ή κατάτμηση εικόνας είναι η διαδικασία ομαδοποίησης των εικονοστοιχείων μιας ψηφιακής εικόνας σε περιοχές που διακρίνονται από αντίστοιχα επίπεδα φωτεινότητας ή φυσικά όρια. Το ζητούμενο από την εφαρμογή της τμηματοποίησης είναι ο εντοπισμός περιοχών ενδιαφέροντος και η διάκριση των ορίων των περιοχών αυτών ώστε να αναλυθούν ευκολότερα. Με άλλα λόγια, τμηματοποίηση ή κατάτμηση είναι η κατηγοριοποίηση των εικονοστοιχείων, ώστε εικονοστοιχεία της ίδιας κατηγορίας να έχουν αντίστοιχα χαρακτηριστικά.

Οι αλγόριθμοι κατάτμησης μπορούν να διακριθούν στις εξής τρεις γενικές κατηγορίες:

- ημιαυτόματοι (semi-automatic),
- αυτόματοι (automatic),
- χειροκίνητοι (manual).

Οι ημιαυτόματοι αλγόριθμοι απαιτούν επίβλεψη και χρειάζονται δεδομένα στα οποία έχει αποδοθεί ετικέτα με βάση την οποία κατηγοριοποιούνται οι περιοχές μιας εικόνας. Οι ετικέτες αποδίδονται στα εικονοστοιχεία των αρχικών εικόνων μέσω μιας διαδικασίας προεπεξεργασίας η οποία συνήθως βασίζεται στη χρήση κάποιου γρήγορου αλγορίθμου κατάτμησης, ή απ' ευθείας από τον ίδιο το χρήστη. Αντίθετα, οι αυτόματοι αλγόριθμοι δεν προϋποθέτουν πληροφορίες για τις εικόνες που πρόκειται να υποβληθούν σε κατάτμηση και παράγουν αποτέλεσμα δίχως να απαιτείται κάποια παρέμβαση. Η συγκεκριμένη κατηγορία αντιμετωπίζει δυσκολία στην εφαρμογή της και στο παραγόμενο αποτέλεσμα, όταν απαιτείται η επεξεργασία πολύπλοκων περιοχών ενδιαφέροντος, όπου δεν είναι διακριτά τα όρια ανάμεσα σε γειτονικές ανατομικές δομές. Οι χειροκίνητοι ή διαδραστικοί αλγόριθμοι είναι αυτοί που επικρατούν σε εφαρμογές που απαιτούν να εκτιμηθεί ένας όγκος ή να εκτιμηθούν οι διαστάσεις του και χρησιμοποιούνται συνήθως όταν είναι αδύνατη η χρήση αυτομάτων ή ημιαυτομάτων τεχνικών.

Γενικά, η κατάτμηση των ιατρικών δεδομένων έχει εξετασθεί ως θέμα εκτενώς τα τελευταία χρόνια, με αποτέλεσμα να έχουν προταθεί μεγάλο πλήθος σχετικών αλγορίθμων στη βιβλιογραφία. Μερικές μεθοδολογίες από τις πλέον διαδεδομένες περιλαμβάνουν τις τεχνικές Connected Threshold, Otsu, Neighborhood Connected, Confidence Connected και Isolated Connected. Τα παραπάνω παραδείγματα εμπίπτουν στην κατηγορία των αλγορίθμων ανάπτυξης περιοχής. Οι αλγόριθμοι ανάπτυξης περιοχής (region growing) ξεκινούν από ένα σημείο που βρίσκεται εντός της περιοχής ενδιαφέροντος και εκτελούν μια σύγκριση με τα γειτονικά εικονοστοιχεία ώστε να εκτιμηθεί αν αυτά θα πρέπει να ανήκουν στην περιοχή ενδιαφέροντος.

Μια άλλη κατηγορία αλγορίθμων που θα μπορούσαν να έχουν εφαρμογή στο προτεινόμενο σύστημα είναι αυτή των ενεργών περιγραμμάτων (active contours). Σε αυτή την κατηγορία αλγορίθμων ένα σύνολο σημείων δημιουργούν ένα αρχικό περίγραμμα (περιοχή ενδιαφέροντος) το οποίο μεταβάλλεται σε συνάρτηση κάποιων παραμέτρων-περιορισμών. Στο μοντέλο των snakes που θα μπορούσε να έχει εφαρμογή στην κατάτμηση προκειμένου να εξαχθεί ο καρκίνος των ωοθηκών, λαμβάνεται υπόψη η εσωτερική ενέργεια μιας περιοχής που την κάνει να κινείται σαν μεμβράνη κατά την εφαρμογή του αλγορίθμου, η εξωτερική ενέργεια που καθορίζει την κύρτωση της περιοχής ενδιαφέροντος και ένας τρίτος περιοριστικός παράγοντας (image term) που καθιστά το μοντέλο ανεπηρέαστο από το θόρυβο κατά την εφαρμογή σε διαδοχικές εικόνες [71]. Είναι αντιληπτό από την περιγραφή του εν λόγω αλγορίθμου ότι είναι ιδιαίτερα σημαντικό η αρχικοποίηση από τον χρήστη, ή μέσω κάποιας αυτοματοποιημένης διαδικασίας, να γίνει κοντά στην περιοχή ενδιαφέροντος.

Μία άλλη οικογένεια ευρέως χρησιμοποιούμενων αλγορίθμων είναι οι αλγόριθμοι watershed. Στο μοντέλο των watershed η γκρι κλίμακας εικόνα αντιμετωπίζεται ως μια τοπογραφική επιφάνεια στην οποία ξεκινώντας από ένα σημείο καλύπτεται όλη η επιφάνεια μέχρι ένα τοπικό ελάχιστο φωτεινότητας. Με αυτό τον τρόπο δημιουργούνται περιοχές-κοιλάδες της εικόνας με αντίστοιχα επίπεδα φωτεινότητας. Ένα πιθανό πρόβλημα που προκύπτει από την εφαρμογή αλγορίθμου watershed είναι η υπερ-κατάτμηση της εικόνας, πρόβλημα όμως το οποίο αντιμετωπίζεται με προεπεξεργασία της εικόνας ή αντίστοιχα εκ των υστέρων επεξεργασία με την εφαρμογή κάποιας πιο εξελιγμένης και ακριβούς τεχνικής κατάτμησης.

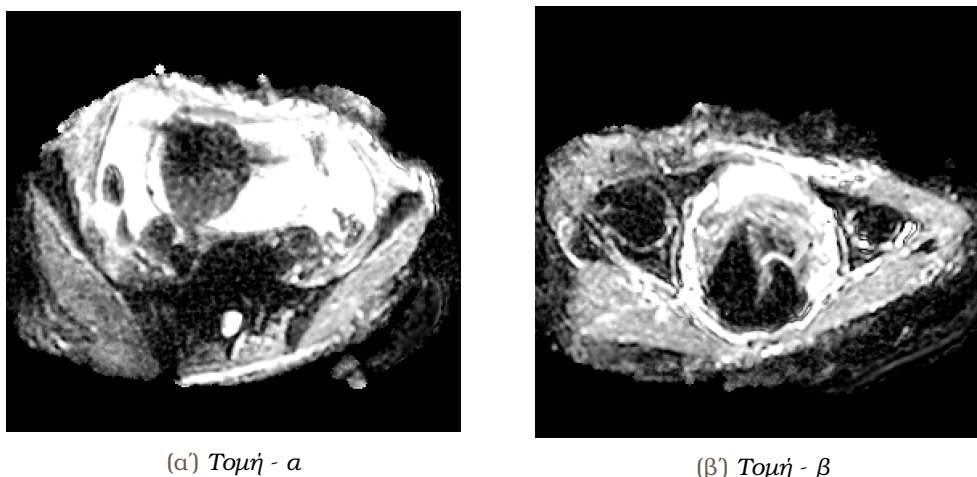
5.3.3 Ανακατασκευή, Σύντηξη και Απεικόνιση Τρισδιάστατων Ιατρικών Δεδομένων - Reconstruction, Subtraction, Visualization

Μετά από την επιτυχή ευθυγράμμιση των δεδομένων που προέκυψαν από τις διάφορες τεχνικές λήψης, είναι δυνατή η σύντηξη και παρουσίαση των πληροφοριών σε κοινό σύστημα συντεταγμένων. Η σύντηξη των πληροφοριών μπορεί να επιτευχθεί με ποικίλους τρόπους (από απλά υψιπερατά φίλτρα ως πολύπλοκους μετασχηματισμούς κυματίων) [72].

Η ανακατασκευή των ευθυγραμμισμένων δεδομένων και η παρουσίαση των παραπάνω πληροφοριών μπορεί να επιτευχθεί εφαρμόζοντας τεχνικές απόδοσης όγκου (volume rendering) [73]. Μερικές από τις πλέον διαδεδομένες τεχνικές απόδοσης όγκου περιλαμβάνουν: την τεχνική Volume Ray Casting, την τεχνική Splatting, τη μέθοδο Shear Warp, καθώς επίσης και διάφορες τεχνικές που βασίζονται στην υφή των απεικονιζόμενων αντικειμένων [74]. Στα Σχήματα (Σχ.5.3α') και (Σχ.5.3β') παρουσιάζονται παραδείγματα απεικόνισης τρισδιάστατων δεδομένων μαγνητικής τομογραφίας, μέσω εξελιγμένων τεχνικών απόδοσης όγκου.

5.4 Εξαγωγή Χαρακτηριστικών - Feature Extraction

Στον τομέα της ιατρικής, η ραδιωμική (radiomics) είναι μια μέθοδος που εξάγει μεγάλο αριθμό χαρακτηριστικών από ακτινογραφικές ιατρικές εικόνες χρησιμοποιώντας αλγόριθμους χαρακτηρισμού δεδομένων [31, 29]. Αυτά τα χαρακτηριστικά, που ονομάζονται ραδιολογικά χαρακτηριστικά, έχουν τη δυνατότητα να φέρουν στο φως χαρακτηριστικά ασθενειών τα οποία πιθανώς δεν είναι ανιχνεύσιμα με γυμνό μάτι [75]. Η υπόθεση της ραδιωμική ε-



Σχήμα 5.3: Τρισδιάστατες απεικονίσεις από ανακατασκευη δεδομένων μαγνητικής τομογραφίας, μέσω τεχνικών απόδοσης όγκου.

ίναι ότι ευδιάκριτα χαρακτηριστικά απεικόνισης μεταξύ μορφών ασθένειας μπορεί να είναι χρήσιμα τόσο για την πρόγνωση όσο και για την θεραπευτική απόκριση διαφόρων κλινικών καταστάσεων, παρέχοντας έτσι πολύτιμες πληροφορίες για εξατομικευμένη θεραπεία [76]. Η ραδιωμική προέκυψε από τον τομέα της ογκολογίας [77] και αποτελεί μία από τις πιο προηγμένες μεθόδους στον τομέα.

5.4.1 Κλάσεις Χαρακτηριστικών

Για τους σκοπούς του έργου ανατρέξαμε σε πλέον σύγχρονες υπολογιστικά εφαρμογές της ραδιωμικής, με σκοπό την εξαγωγή πληθώρας χαρακτηριστικών ικανών να διαχωρίσουν τα κλινικά δείγματα σε τρεις περιοχές καρκινικής επικινδυνότητας. Συγκεκριμένα, έγινε χρήση ενός εργαλείου ελεύθερου λογισμικού (open - source software), του pyradiomics [78] για την εξαγωγή, σε πρώτο βαθμό, των χαρακτηριστικών:

- Χαρακτηριστικά σχήματος (Shape features)
- Χαρακτηριστικά βάσει ιστογράμματος ή πρώτης τάξης χαρακτηριστικά (Histogram-based features or first-order statistics)
- Χαρακτηριστικά υφής ή στατιστικά δεύτερης τάξης (Textural features or second-order statistics)

Η διαδικασία έγινε αυτοματοποιημένα και περιλαμβάνει τα παρακάτω στάδια.

Αρχικά, στο **Βήμα 1** λαμβάνουμε ως είσοδο την ακολουθία των εικόνων της ασθενούς μαζί με ένα **.txt** αρχείο, το οποίο περιλαμβάνει τις συντεταγμένες των εικονοστοιχείων ανά περιοχή καρκινικής επικινδυνότητας (0, 1, 2).

Ακολούθως, στο **Βήμα 2** δημιουργούμε για κάθε περιοχή καρκινικής επικινδυνότητας μία λίστα εικόνων.

Παράλληλα, στο **Βήμα 3** κατασκευάζουμε για κάθε εικόνα την αντίστοιχη μάσκα βάσει των συντεταγμένων που διαθέτουμε. Οι διαστάσεις της μάσκας είναι ίδιες με εκείνες της αντίστοιχης εικόνας εισόδου και η κάθε τιμή της καθορίζεται από τον εξής κανόνα:

$$\text{μάσκα} = \begin{cases} +1 : \text{εικονοστοιχείο που περιλαμβάνεται στο .txt αρχείο} \\ 0 : \text{αλλού} \end{cases} \quad (5.1)$$

Στο **Βήμα 4** με τον συνδυασμό της εικόνας εισόδου και της μάσκας εξάγουμε τα προαναφερθέντα χαρακτηριστικά (Υποκεφάλαιο.3.1), τα οποία είναι στο σύνολο 104.

Από τα συνολικά χαρακτηριστικά στο **Βήμα 5** κρατάμε μόνο τα μη μηδενικά ή εκείνα των οποίων οι τιμές διαφέρουν από εικόνα (δείγμα) σε εικόνα ακριβώς για να υπάρχει στατιστική αξία όλων των χαρακτηριστικών.

Τα χαρακτηριστικά αυτά έπειτα στο **Βήμα 6** τα αποθηκεύουμε σε **.csv** αρχεία για την περαιτέρω στατιστική ανάλυση και αξιολόγηση τους.

Τα συνολικά χαρακτηριστικά, τα οποία προκύπτουν από τα **Βήματα 1 - 6** παρατίθενται στους πίνακες: Πίνακας(5.4) και Πίνακας(5.5).

5.4.2 Κλάσεις Φίλτρων

Ωστόσο, παρά την εξαντλητική εξαγωγή χαρακτηριστικών όλων των κατηγοριών όπως περιγράφηκαν στο Υποκεφάλαιο(3.1) κρίθηκε πως στο σύνολο τους τα χαρακτηριστικά αυτά πιθανώς να μην αρκούν για την επαρκή και ικανή διάκριση των καρκινικών περιοχών. Μάλιστα σε οποιοδήποτε πρόβλημα εξόρυξης δεδομένων, επιλογής χαρακτηριστικών και ταξινόμησης είναι θεμιτή αν όχι απαραίτητη η ύπαρξη όσων των δυνατών περισσότερων χαρακτηριστικών. Για τον σκοπό αυτό το ίδιο εργαλείο μας επιτρέπει να εφαρμόσουμε ποικίλα φίλτρα σε κάθε αρχική μη επεξεργασμένη εικόνα του συνόλου δεδομένων μας με αποτέλεσμα την λήψη νέας εικόνας, μετασχηματισμένης ως προς την αρχική. Έτσι πλέον χρησιμοποιούμε τις μετασχηματισμένες εικόνες ως νέα δεδομένα εισόδου λαμβάνοντας εκ νέου για κάθε μία εξ αυτών 104 ραδιολογικά χαρακτηριστικά. Μεταξύ άλλων χρησιμοποιήσαμε πυρήνες $\nabla^2 G(x, y, z)$ Laplacian of Gaussian (LoG) (που στην ουσία είναι η συνέλιξη του γκαουσιανού φίλτρου και του λαπλασιανού τελεστή για την δεύτερη παράγωγο) και μετασχηματισμούς κυματίου (wavelet decompositions) ως φίλτρα στις αρχικές εικόνες.

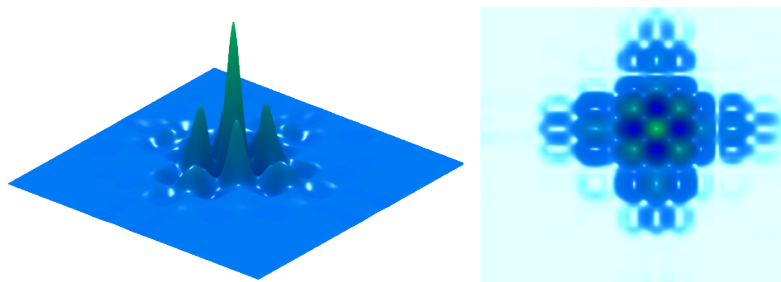
$$G(x, y, z, \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^3} e^{-\frac{x^2+y^2+z^2}{2\sigma^2}} \quad (5.2)$$

Χρησιμοποιήσαμε σειρά τιμών του σ για να επικεντρωθούμε τόσο σε πιο λεπτά όσο και πιο χονδροειδή χαρακτηριστικά υφής, που κυμαίνονται από 2 (λεπτά) έως 5 mm (χοντροειδή) προσαυξημένα κατά 1 mm.

Ο μετασχηματισμός κυματίου (Σχ.5.4) είναι μια πολύ γνωστή τεχνική επεξεργασίας εικόνας. Στην περίπτωση μας χρησιμοποιήθηκαν όλοι οι δυνατοί συνδυασμοί υπερπερατών και βαθυπερατών φίλτρων στα επίπεδα x, y .

Συνολικά έγινε εφαρμογή των κάτωθι φίλτρων:

- **Κυματίου (Wavelet)**, αποδίδει 4 αποσυνθέσεις ανά επίπεδο (όλοι οι πιθανοί συνδυασμοί εφαρμογής ενός υπερπερατού και βαθυπερατού φίλτρου στις δύο διαστάσεις).
- **Laplacian of Gaussian (LoG)**, φίλτρο ενίσχυσης άκρων. Δίνει έμφαση στις περιοχές αλλαγής του επιπέδου του γκρι, όπου το σ καθορίζει πόσο χονδροειδής θα πρέπει να είναι η αναφερόμενη υφή. Για χαμηλές τιμές του σ δίδεται έμφαση στις λεπτές



Σχήμα 5.4: 2-D Ορθογώνια Κυμάτια.

υφές (αλλαγή σε μικρή απόσταση), ενώ για υψηλές τιμές του σ δίδεται έμφαση στις χονδροειδέις υφές (αλλαγή επιπέδου του γκρι σε μεγάλη απόσταση).

- **Τετραγωνικό (Square)**, γίνεται γραμμική κλιμακοποίηση των τιμών από το τετράγωνο των εντάσεων της εικόνας πίσω στο αρχικό εύρος. Οι αρνητικές τιμές στην αρχική εικόνα θα γίνουν ξανά αρνητικές μετά την εφαρμογή του φίλτρου.
- **Τετραγωνικής Ρίζας (SquareRoot)**, παίρνει την τετραγωνική ρίζα των απόλυτων εντάσεων της εικόνας και τις επαναφέρει στην αρχική περιοχή. Οι αρνητικές τιμές στην αρχική εικόνα θα γίνουν ξανά αρνητικές μετά την εφαρμογή του φίλτρου.
- **Λογαριθμικό (Logarithm)**, λαμβάνει τον λογάριθμο της απόλυτης έντασης +1. Γίνεται κλιμακοποίηση των τιμών στο αρχικό εύρος ενώ οι αρνητικές αρχικές τιμές γίνονται ξανά αρνητικές μετά την εφαρμογή του φίλτρου.
- **Εκθετικό (Exponential)**, η φιλτραρισμένη ένταση προκύπτει από το εκθετικό της αρχικής απόλυτης έντασης της εικόνας. Γίνεται κλιμακοποίηση των τιμών στο αρχικό εύρος ενώ οι αρνητικές αρχικές τιμές γίνονται ξανά αρνητικές μετά την εφαρμογή του φίλτρου.
- **Κλίσης (Gradient)**, επιστρέφει το μέγεθος της κλίσης.

Έτσι καταλήγουμε σε ένα σύνολο ≈ 1240 χαρακτηριστικών για κάθε εικόνα εισόδου από τα οποία εν συνεχεία θα επιλεγούν εκείνα που συμβάλλουν στην βέλτιστη διάκριση των καρκινικών περιοχών.

5.4.3 Επιλογή Χαρακτηριστικών και Ταξινόμηση

Χρήση Εργαλείου `scikit - learn`

Μεταξύ των αλγόριθμων επιλογής χαρακτηριστικών οι οποίοι χρησιμοποιήθηκαν στα πλαίσια της παρούσας εργασίας, είναι αλγόριθμοι της βιβλιοθήκης `scikit - learn`. Θα γίνει αναφορά για κάθε αλγόριθμο σε μερικές από τις παραμέτρους του οι οποίες επηρεάζουν την λειτουργία του καθενός και αφορούν την παρούσα εργασία [79].

Αναδρομική Εξάλειψη Χαρακτηριστικών Διασταυρούμενης Επαλήθευσης - Recursive Feature Elimination Cross Validation (RFECV) Ο επιλογέας χαρακτηριστικών Recursive Feature Elimination Cross Validation (RFECV) έχει ως στόχο την εξάλειψη επαναλαμβανόμενων χαρακτηριστικών χρησιμοποιώντας ως κριτήριο κατάταξης μία βαθμολογία (βάρος) που έχει αποδοθεί σε κάθε χαρακτηριστικό από έναν εξωτερικό εκτιμητή με επιτήρηση (π.χ. SVM) [80] με βάση την ορισμένη στρατηγική διασταυρούμενης επαλήθευσης 'cv' διαχωρισμού των συνόλων σε σετ εκπαίδευσης και σε σετ ελέγχου. Για τον παραπάνω σκοπό πραγματοποιείται μία επαναλαμβανόμενη διαδικασία κατά την οποία ο αριθμός του συνόλου των χαρακτηριστικών μειώνεται καθώς αφαιρούνται τα λιγότερο σημαντικά χαρακτηριστικά σε κάθε επανάληψη μέχρι να καταλήξει στον επιθυμητό αριθμό χαρακτηριστικών. Στον RFECV που παρέχεται από τη βιβλιοθήκη scikit-learn ο αριθμός των επιθυμητών χαρακτηριστικών ορίζεται από την μεταβλητή 'n_features_to_select'. Ο αριθμός των χαρακτηριστικών που θα αφαιρεθούν σε κάθε επανάληψη ορίζεται από την μεταβλητή 'step'. Αν η τιμή αυτής είναι μεταξύ 0 και 1 αντιστοιχεί στο ποσοστό των χαρακτηριστικών (στρογγυλοποίηση προς τα κάτω) που θα αφαιρεθούν σε κάθε επανάληψη, αλλιώς αντιστοιχεί στο ακριβές νούμερο. Τέλος, ο εκτιμητής ορίζεται από την μεταβλητή 'estimator'.

Επιλογή βάσει Μοντέλου - Select From Model Ο επιλογέας χαρακτηριστικών Select From Model χρησιμοποιεί ως κριτήριο κατάταξης το βάρος που έχει αποδοθεί σε κάθε χαρακτηριστικό από έναν εκτιμητή. Ο εκτιμητής αυτός είναι απαραίτητο να περιέχει τη μεταβλητή της σημαντικότητας των χαρακτηριστικών (feature importances) ή συντελεστών (coefficients) γιατί έχουν καθοριστικό ρόλο στην επιλογή των σημαντικών χαρακτηριστικών. Εάν κάποια από τις δύο παραπάνω μεταβλητές ενός χαρακτηριστικού έχει τιμή μικρότερη από το προβλεπόμενο όριο τότε το χαρακτηριστικό αυτό δεν θεωρείται σημαντικό και αφαιρείται [81]. Σημαντικές παράμετροι αυτού του επιλογέα είναι, το όριο επιλογής χαρακτηριστικών ('threshold'), καθώς και η παράμετρος 'max_features' η οποία ορίζει τον μέγιστο αριθμό χαρακτηριστικών πάνω από το κατώφλι τα οποία επιτρέπεται να επιλεγθούν.

Επιλογή Εκατοστιαίας Τιμής - Select Percentile Ο επιλογέας χαρακτηριστικών Select Percentile επιλέγει τα χαρακτηριστικά με κριτήριο το ποσοστό των δεδομένων που έχουν την υψηλότερη βαθμολογία. Την τιμή του ποσοστού αυτού την ορίζει ο χρήστης μέσω της μεταβλητής percentile. Γίνεται χρήση συναρτήσεων για τον υπολογισμό των χαρακτηριστικών που έχουν την υψηλότερη βαθμολογία. Μερικά παραδείγματα συναρτήσεων που μπορούν να χρησιμοποιηθούν για αυτόν το σκοπό αποτελούν η συνάρτηση f_classif και η συνάρτηση mutual_info_classif [47]. Η συνάρτηση mutual_info_classif βασίζεται στις αμοιβαίες πληροφορίες (Mutual information - MI) μεταξύ δύο τυχαίων μεταβλητών, δηλαδή μελετά την εξάρτηση αυτών χρησιμοποιώντας την εκτίμηση της εντροπίας η οποία έχει προκύψει από τις k-πλησιέστερες αποστάσεις των γειτόνων [82].

Χρήση Εργαλείου scikit - feature

Αντίστοιχα χρησιμοποιήθηκαν αλγόριθμοι επιλογής χαρακτηριστικών, οι οποίοι ανήκουν στην βιβλιοθήκη scikit - feature. Θα γίνει αναφορά για την λειτουργία κάθε αλγόριθμου.

Αλγόριθμοι βασισμένοι στην Θεωρία της Πληροφορίας - Information Theoretical Based

Conditional Informative Feature Extraction - CIFE Ο επιλογέας χαρακτηριστικών Conditional Informative Feature Extraction (CIFE) είναι ένας αλγόριθμος με επιτήρηση που έχει ως βασική ιδέα την επιλογή χαρακτηριστικών τέτοιων ώστε, δοσμένων των πιθανών κατηγοριών, να ελαχιστοποιηθεί η αλληλεξάρτηση χαρακτηριστικών για το τελικό αποτέλεσμα ταξινόμησης δείγματος σε μια κλάση [83].

Conditional Mutual Information Maximisation - CMIM Ο επιλογέας χαρακτηριστικών Conditional Mutual Information Maximisation (CMIM) είναι ένας αλγόριθμος με επιτήρηση που βασίζεται στην επιλογή χαρακτηριστικών τα οποία ενώ είναι ανεξάρτητα μεταξύ τους βελτιώνουν την ταξινόμηση, αλλά και εξαρτώνται ελαφρά σε ζευγάρια το ένα από το άλλο. Το CMIM είναι μια προς τα εμπρός επιλογή των χαρακτηριστικών βασισμένη στο κριτήριο των Markov blankets των Koller και Sahami [84]. Εδώ, δημιουργούνται ομάδες χαρακτηριστικών βασισμένες σε ένα ήδη επιλεγμένο χαρακτηριστικό. Στη συνέχεια, γίνεται απόρριψη χαρακτηριστικών μέσω μιας διαδικασίας κατά την οποία κάθε χαρακτηριστικό X μπορεί να απορριφθεί εφόσον υπάρχει ήδη επιλεγμένο άλλο χαρακτηριστικό τέτοιο ώστε τα χαρακτηριστικά της ομάδας X, Y να είναι ανεξάρτητα μεταξύ τους.

Double Input Symmetrical Relevance - DISR Ο επιλογέας χαρακτηριστικών Double Input Symmetrical Relevance - DISR είναι ένας αλγόριθμος με επιτήρηση που βασίζεται σε δύο αρχές. Η πρώτη του αρχή αφορά στην έννοια της μεταβλητής συμπληρωματικότητας, δηλαδή βασίζεται στο γεγονός ότι η αντιμετώπιση των χαρακτηριστικών μεμονωμένα δεν αποφέρει πάντα τα αποδοτικότερα αποτελέσματα καθώς ο συνδυασμός μεταβλητών ίσως να αποφέρει καλύτερη επιλογή χαρακτηριστικών. Η δεύτερη αρχή στην οποία βασίζεται είναι ότι ο μέσος όρος των πληροφοριών όλων των υποσυνόλων προσδιορίζει το κατώτερο όριο στις πληροφορίες ενός συνόλου μεταβλητών [85].

Interactive Cluster Analysis Procedure - ICAP Ο επιλογέας χαρακτηριστικών Interactive Cluster Analysis Procedure (ICAP) είναι ένας αλγόριθμος με επιτήρηση που βασίζεται στη μελέτη της αλληλεξάρτησης χαρακτηριστικών και κυρίως στα χαρακτηριστικά που επηρεάζουν όμοια το τελικό αποτέλεσμα ταξινόμησης [86]. Στο αλγόριθμο επιλογής χαρακτηριστικών (ICAP) τα χαρακτηριστικά ταξινομούνται χρησιμοποιώντας την αλληλεπίδραση της ετικέτας της κλάσης με διαφορετικά σύνολα χαρακτηριστικών.

Αλγόριθμοι Ομοιότητας - Similarity based

Fisher Score Ο επιλογέας χαρακτηριστικών Fisher Score είναι ένας αλγόριθμος με επίβλεψη που χωρίζει τα χαρακτηριστικά σε κατηγορίες με στόχο οι τιμές χαρακτηριστικών των δειγμάτων μιας κατηγορίας να είναι παρόμοιες μεταξύ τους και να μην παρουσιάζουν ομοιότητα με αυτές από τις άλλες κλάσεις.

Laplacian score Ο επιλογέας χαρακτηριστικών Laplacian score είναι ένας αλγόριθμος χωρίς επιτήρηση που έχει ως αρχή ότι η ύπαρξη δυο σημείων που βρίσκονται κοντά μεταξύ τους στο χώρο των παρατηρήσεων, σημαίνει ότι αφορούν στην ίδια κατηγορία. Σημαντική βοήθεια προσφέρουν εδώ τα γραφήματα κοντινότερων γειτόνων, τα οποία διευκολύνουν την κατηγοριοποίηση σε ομάδες των παρατηρήσεων [87].

Spectral - Spec Ο επιλογέας χαρακτηριστικών Spectral είναι ένας αλγόριθμος χωρίς επιτήρηση ο οποίος έχει παρόμοια βασική ιδέα με το Laplacian Score [88].

Trace Ratio Ο επιλογέας χαρακτηριστικών Trace Ratio είναι ένας αλγόριθμος με επιτήρηση στον οποίο η επιλογή ενός υποσυνόλου χαρακτηριστικών στηρίζεται στη βαθμολογία του συνόλου των χαρακτηριστικών. Ο τρόπος λειτουργίας του περιλαμβάνει την δημιουργία δυο πινάκων συγγένειας για την περιγραφή των ομοιοτήτων των χαρακτηριστικών μεταξύ των κλάσεων και εντός μιας κλάσης [89].

Relief Ο επιλογέας χαρακτηριστικών Relief είναι ένας επιλογέας χαρακτηριστικών πολλαπλών κλάσεων με επιτήρηση που έχει ως βασική ιδέα τους πλησιέστερους γείτονες. Αναλυτικότερα, επιλέγει μια τυχαία περίπτωση και εντοπίζει τους k πλησιέστερους γείτονές του που ανήκουν στην ίδια κλάση με αυτή και τους k - πλησιέστερους γείτονες οι οποίοι δεν ανήκουν στην ίδια κλάση με αυτή, βάσει τυχαίας (αρχικά) βαθμολογίας του καθενός. Στη συνέχεια επαναυπολογίζει την βαθμολογία (βάρος) ώστε να αυξηθούν οι γείτονες που ανήκουν στην ίδια κατηγορία [90].

Αλγόριθμοι Αραιής Μάθησης - Sparse learning based

Discriminative Feature Selection - UDFS Ο επιλογέας χαρακτηριστικών Discriminative Feature Selection (UDFS) είναι ένας αλγόριθμος χωρίς επιτήρηση ο οποίος εστιάζει στην επιλογή των χαρακτηριστικών με βάση τα διακριτά χαρακτηριστικά [91]. Αναλυτικότερα επιδιώκει την μεγαλύτερη δυνατή απόσταση μεταξύ των χαρακτηριστικών δυο διαφορετικών κλάσεων και την ελάχιστη απόσταση μεταξύ των χαρακτηριστικών που ανήκουν στην ίδια κλάση.

Multi-Cluster Feature Selection - MCFS Ο επιλογέας χαρακτηριστικών Multi-Cluster Feature Selection (MCFS) είναι ένας αλγόριθμος χωρίς επιτήρηση. Ο MCFS έχει την δυνατότητα να παρουσιάζει καλό χειρισμό πολλαπλών ομαδοποιημένων δομών. Αυτό οφείλεται στο γεγονός ότι ο MCFS βασίζεται στη μέτρηση των συσχετίσεων μεταξύ διαφορετικών χαρακτηριστικών χωρίς πληροφορίες ετικέτας, δηλαδή χωρίς να λαμβάνει υπόψιν του την κλάση στην οποία ανήκει κάθε δείγμα [92].

Nonnegative Discriminative Feature Selection - NDFS Ο επιλογέας χαρακτηριστικών Nonnegative Discriminative Feature Selection (NDFS) είναι ένας αλγόριθμος χωρίς επιτήρηση. Η επιλογή ενός υποσυνόλου διακριτικών χαρακτηριστικών στηρίζεται στην από κοινού φασματική ομαδοποίηση και την επιλογή χαρακτηριστικών σε ένα κοινό πλαίσιο.

Δημιουργεί σε πρώτη φάση μια σειρά ψευδο-κατηγοριών μέσω της φασματικής ανάλυσης και επιβάλλει περιορισμούς σχετικά με την μη αρνητικότητα και ορθογωνιότητα κατά τον υπολογισμό της φάσης για τη φασματική ομαδοποίηση. Η φάση της επιλογής χαρακτηριστικών καθοδηγείται από τις μη αρνητικές ετικέτες ψευδο-κατηγορίας οι οποίες λειτουργούν ως περιορισμοί παλινδρόμησης. Σύμφωνα με τους δημιουργούς του, οι περιορισμοί αυτοί οδηγούν τις αρχικά δημιουργημένες ψευδείς κατηγορίες να προσεγγίσουν τις πραγματικές [93].

Στατιστικοί Αλγόριθμοι - Statistical based

Correlation based Feature selection - CFS Ο επιλογέας χαρακτηριστικών Correlation-based Feature Selection (CFS) είναι ένας αλγόριθμος με επίβλεψη. Ο CFS αξιολογεί και υπολογίζει τη αξία ενός χαρακτηριστικού εξετάζοντας τον βαθμό συσχέτισης και την δυνατότητα πρόβλεψης που εισάγει στο επιλεγμένο σύνολο χαρακτηριστικών. Οι συσχετίσεις υπολογίζονται με την χρήση της συμμετρικής αβεβαιότητας και στην συνέχεια επιλέγεται ένα υποσύνολο από τα χαρακτηριστικά, με κριτήριο διακοπής την ύπαρξη πέντε διαδοχικών πλήρως αναπτυγμένων υποσυνόλων στα οποία δεν παρουσιάζεται βελτίωση, χρησιμοποιώντας την Best First Search μεθοδολογία. Τα πλεονεκτήματα του CFS είναι ότι για την πραγματοποίηση της αξιολόγησής του δεν είναι απαραίτητο να κρατήσει κανένα τμήμα των δεδομένων εκπαίδευσης [89].

T-score Ο επιλογέας χαρακτηριστικών T-score είναι ένας αλγόριθμος με επίβλεψη ο οποίος βρίσκει εφαρμογή στην επίλυση δυαδικών προβλημάτων. Στον υπολογισμό του T-score στην περίπτωση που χρησιμοποιούνται άνισα μεγέθη δειγμάτων και η διακύμανση είναι άνιση. Εδώ ως δείγματα εννοούνται οι παρατηρήσεις (άτομα) για κάθε ομάδα στην οποία μπορούν να ταξινομηθούν, ενώ άνισα μεγέθη σημαίνει ότι ο πληθυσμός των ομάδων είναι άνισος, δηλαδή το πλήθος της μίας ομάδας είναι τουλάχιστον διπλάσιο από το άλλο. Το T-score βασίζεται στον λόγο μεταξύ της μέσης διαφοράς και της διακύμανσης δύο τάξεων, ώστε να εκτιμήσει αν ένα χαρακτηριστικό μεταξύ δυο κλάσεων έχει την δυνατότητα να παρουσιάσει στατιστική διαφορά στην μέση τιμή τους.

F-score Ο επιλογέας χαρακτηριστικών F-score είναι ένας αλγόριθμος με επίβλεψη. Ο αλγόριθμος αυτός κρίνει την δυνατότητα ενός χαρακτηριστικού να διαχωρίσει αποδοτικά δείγματα που ανήκουν σε διαφορετικές κατηγορίες ανεξάρτητα από τα υπόλοιπα χαρακτηριστικά. Δίνει δηλαδή, τη δυνατότητα αξιολόγησης και σύγκρισης της διακριτικής ικανότητας κάθε χαρακτηριστικού. Η τιμή F-score ενός χαρακτηριστικού είναι ανάλογη της διακριτικής ικανότητας προκειμένου να αποφανθεί ο αλγόριθμος ταξινόμησης ως προς την ομάδα στην οποία θα τοποθετήσει το δείγμα [94].

Gini-index Ο επιλογέας χαρακτηριστικών Gini βασίζεται στην στατιστική και είναι ένας αλγόριθμος με επίβλεψη. Ο Gini βασίζεται στην επιλογή χαρακτηριστικών με κριτήριο την ικανότητα ενός χαρακτηριστικού να ξεχωρίζει μεταξύ δυο κλάσεων. Η τιμή του δείκτη Gini κάθε χαρακτηριστικού είναι ανεξάρτητη από τα άλλα χαρακτηριστικά και υποδηλώνει την

σημαντικότητα ενός χαρακτηριστικού. Η τιμή του είναι αντιστρόφως ανάλογη της αξίας του γι' αυτό και πιο σημαντικά είναι τα χαρακτηριστικά με το μικρότερο δείκτη Gini.

Wrapper methods Ο επιλογέας χαρακτηριστικών Wrapper methods, που ονομάζονται επίσης πολυμεταβλητή μέθοδος, λαμβάνει υπόψη συσχετίσεις και αλληλεπιδράσεις μεταξύ των χαρακτηριστικών. Ενώ μονομεταβλητές μέθοδοι ερευνούν μόνο τη στατιστική σχέση μεταξύ των χαρακτηριστικών, οι Wrapper methods δημιουργούν ένα υποσύνολο χαρακτηριστικών, εφαρμόζουν αυτό το υποσύνολο σε ένα μοντέλο πρόβλεψης και αξιολογούν την ποιότητα της απόδοσής του. Στη συνέχεια, δοκιμάζεται ένα νέο υποσύνολο χαρακτηριστικών και το υποσύνολο των χαρακτηριστικών με την καλύτερη απόδοση αντιπροσωπεύει το τελικό σύνολο επιλεγμένων χαρακτηριστικών. Εξαιτίας της επαναληπτικής φύσης των Wrapper methods, αυτές οι μέθοδοι είναι υπολογιστικά δαπανηρές. Οι μέθοδοι Wrapper περιλαμβάνουν μεταξύ άλλων αμφίδρομη αναζήτηση, εξαντλητική επιλογή χαρακτηριστικών, εμπρόσθια επιλογή χαρακτηριστικών και οπίσθια κατάργηση χαρακτηριστικών [95].

Χρήση Εργαλείου `colin - mico`

Αντίστοιχα χρησιμοποιήθηκαν αλγόριθμοι επιλογής χαρακτηριστικών, οι οποίοι ανήκουν στην βιβλιοθήκη `colin - mico` που είναι βασισμένη στην βιβλιοθήκη `scikit - learn`. Θα γίνει αναφορά για την λειτουργία κάθε αλγόριθμου [96].

Αλγόριθμοι βασισμένοι στην Θεωρία της Πληροφορίας - **Information Theoretical Based**

Joint Mutual Information - JMI Ο επιλογέας χαρακτηριστικών Joint Mutual Information (JMI) είναι ένας αλγόριθμος με επιτήρηση που βασίζεται για την επιλογή των χαρακτηριστικών στην αμοιβαία πληροφόρηση μαζί με την εντροπία μεταξύ οποιωνδήποτε τυχαίων μεταβλητών [89].

Mutual Information Feature Selection - MIFS Ο επιλογέας χαρακτηριστικών Mutual Information Feature Selection (MIFS) είναι ένας αλγόριθμος με επιτήρηση που λαμβάνει υπόψη για την επιλογή των χαρακτηριστικών τη σχετικότητα των χαρακτηριστικών και τον πλεονασμό των χαρακτηριστικών [89]. Ο MIFS εστιάζει αποκλειστικά στις αμοιβαίες πληροφορίες που υπολογίζονται για κάθε στοιχείο - παρατήρηση του αρχικού συνόλου δεδομένων.

Mutual information aximisation - MIM Ο επιλογέας χαρακτηριστικών Mutual information Maximisation - MIM) είναι ένας αλγόριθμος με επιτήρηση που λαμβάνει υπόψη αποκλειστικά την συσχέτιση του χαρακτηριστικού με της ταμπέλες της κλάσης (αν δηλαδή, στην προκειμένη περίπτωση, κάποιος είναι υγιής ή ασθενής). Για αυτό τον επιλογέα η αξία του χαρακτηριστικού είναι ανάλογη με την τιμή της συσχέτισης του [89].

Minimal Redundancy Maximal Relevance Criterion - mRMR Ο επιλογέας χαρακτηριστικών Minimal Redundancy Maximal Relevance Criterion - mRMR είναι ένας αλγόριθμος με επιτήρηση. Ο τρόπος επιλογής των χαρακτηριστικών βασίστηκε σε μια σειρά διαισθητικών μέτρων πλεονασμού και σχετικότητας με στόχο την μείωση του πλεονασμού δηλαδή αφαίρεση χαρακτηριστικών των οποίων η επίδραση στο αποτέλεσμα ταξινόμησης υπερκαλύπτεται από άλλα χαρακτηριστικά [97].

Όνομα ακολουθίας	Επίπεδο	Χρήση	Πρωτόκολλο
T2W-TSE	Οβελιαίο	Βασική διαγνωστική ακολουθία με τη μεγαλύτερη αντίθεση (contrast resolution) μεταξύ όγκου και παρακείμενων φυσιολογικών ιστών	Συμβατικό πρωτόκολλο.
2W-DIXON	Εγκάρσιο	Βασική διαγνωστική ακολουθία με τη μεγαλύτερη αντίθεση (contrast resolution) μεταξύ	Συμβατικό πρωτόκολλο.
T1W-DIXON	Εγκάρσιο	Ανατομική πληροφορία, ανίχνευση αιμορραγίας λίπους	Συμβατικό πρωτόκολλο.
2W-3D(VISTA)	Εγκάρσιο	Υψηλής ανάλυσης ακολουθία με λεπτές τομές για μεγαλύτερη λεπτομέρεια και δυνατότητα ανασύνθεσης θα χρησιμοποιηθεί για την κατασκευή του τρισδιάστατου μοντέλου του όγκου (3D printing) βάση του οποίου θα γίνει η λήψη του ιστού ο οποίος θα σταλεί για γενετική ανάλυση από τους χειρουργούς/παθολογοανατόμους	Αντιστοίχιση με 3D printing.
DWI STIR (b=0,1200)	Εγκάρσιο	Χαρτογράφηση όγκου λεμφαδένων	Ακολουθίες για την εξαγωγή χαρακτηριστικών υφής(texture) του όγκου.
m-FFE-T2_R2_star(T2SC)	Εγκάρσιο	Για relaxometry T2* και ποσοτικοποίηση λιπους δημιουργία παραμετρικού χάρτη	Ακολουθίες για την εξαγωγή χαρακτηριστικών υφής(texture) του όγκου

Πίνακας 5.1: Πρωτόκολλο ακολουθιών παλμών Μαγνητικής Τομογραφίας για προεγχειρητική εκτίμηση ασθενών με Καρκίνο Ωοθηκών (DESTINY)

Όνομα ακολουθίας	Επίπεδο	Χρήση	Πρωτόκολλο
s2Ca1_TSE(T2C)	Εγκάρσιο	Για 2 δημιουργία relaxometry παραμετρικού χάρτη	Ακολουθίες για την εξαγωγή χαρακτηριστικών υφής(texture) του όγκου.
T1_native(T1C)	Εγκάρσιο	Για 1 δημιουργία relaxometry παραμετρικού χάρτη	Ακολουθίες για την εξαγωγή χαρακτηριστικών υφής(τεξτυρε) του όγκου.
T1_smDixon_FFE_Dynamic(T1DCE)	Στεφανιαίο	Ποιοτική, ημιποσοτική (καμπύλες αιμάτωσης) και ποσοτική ανάλυση αιμοδυναμικών χαρακτηριστικών του όγκου	Ακολουθίες για την εξαγωγή χαρακτηριστικών υφής(texture) του όγκου.
DWI_STIR_b0, 1200	Εγκάρσιο	Με πολλαπλά (8-10) b values - ανάλυση μικρομοριακής αιμάτωσης του όγκου (intravoxel incoherent motion)	Ακολουθίες για την εξαγωγή χαρακτηριστικών υφής(texture) του όγκου.
Ακολουθία μέτρησης της θερμοκρασίας του όγκου (RTEMP)	Εγκάρσιο	Πρωτότυπη ακολουθία παλμών για εκτίμηση της θερμοκρασίας του όγκου, συγκριτικά με τη θερμοκρασία του σώματος (υπό επεξεργασία)	Ακολουθίες για την εξαγωγή χαρακτηριστικών υφής(texture) του όγκου.
1W_mDixon_ax+GD	Εγκάρσιο	Συμβατική ακολουθία για συνολική εκτίμηση των υπολοίπων δομών της πυέλου	Συμβατικό πρωτόκολλο

Πίνακας 5.2: Πρωτόκολλο ακολουθιών παλμών Μαγνητικής Τομογραφίας μεταθεραπευτικής παρακολούθησης ασθενών με Καρκίνο Ωοθηκών (DESTINY)

Όνομα ακολουθίας	Επίπεδο	Χρήση	Πρωτόκολλο
T2_AX_MVXD_HR-RT	Εγκάρσιο	Συμβατική ακολουθία εκτίμησης των οργάνων της άνω κοιλίας.	Συμβατικό πρωτόκολλο προεγχειρητικής εκτίμησης των οργάνων της άνω κοιλίας.
DWI_3b_RT	Εγκάρσιο	Εντόπιση μεταστατικών εστιών - ποιοτική και ποσοτική ανάλυση των οργάνων δεδομένων χάρτη φαινομενικού συντελεστή διάχυσης (ADC).	Συμβατικό πρωτόκολλο προεγχειρητικής εκτίμησης των οργάνων της άνω κοιλίας.
1 DIXON+ GD	Εγκάρσιο	Συνολική εκτίμηση οργάνων άνω κοιλίας	Συμβατικό πρωτόκολλο προ-εγχειρητικής εκτίμησης των οργάνων της άνω κοιλίας

Πίνακας 5.3: Πρωτόκολλο ακολουθιών παλμών Μαγνητικής Τομογραφίας εκτίμησης άνω κοιλίας για τον εντοπισμό μεταστάσεων εκτός πυέλου, ασθενών με Καρκίνο Ωοθηκών (DESTINY)

Πρώτης Τάξης	Σχήματος	GLCM	GLSZM
Entropy	Mesh Surface	Autocorrelation	Small Area Emphasis (SAE)
Interquartile Rnage	Pixel Surface	Joint Average	Large Area Emphasis (LAE)
Total Energy	Perimeter	Cluster Prominence	Gray Level Non-Uniformity (GLN)
Maximum	Perimeter to Surface ratio	Cluster Shade	Gray Level Non-Uniformity Normalized (GLNN)
Energy	Sphericity	Cluster Tendency	Size-Zone Non-Uniformity (SZN)
Minimum	Spherical Disproportion	Contrast	Size-Zone Non-Uniformity Normalized (SZNN)
10th percentile	Maximum 2D diameter	Correlation	Zone Percentage (ZP)
90th percentile	Major Axis Length	Difference Average	Gray Level Variance (GLV)
Mean	Minor Axis Length	Difference Entropy	Zone Variance (ZV)
Median	Elongation	Difference Variance	Zone Entropy (ZE)
Range		Joint Energy	Low Gray Level Zone Emphasis (LGLZE)
Mean Absolute Deviation (MAD)		Joint Entropy	High Gray Level Zone Emphasis (HGLZE)
Robust Mean Absolute Deviation (rMAD)		Informational Measure of Correlation (IMC) 1	High Gray Level Run Emphasis (HGLRE)
Root Mean Squared (RMS)		Informational Measure of Correlation (IMC) 2	Small Area Low Gray Level Emphasis (SALGLE)
Standard Deviation		Inverse Difference Moment (IDM)	Small Area High Gray Level Emphasis (SAHGLE)
Skewness		Maximal Correlation Coefficient (MCC)	Large Area Low Gray Level Emphasis (LALGLE)
Kurtosis		Inverse Difference Moment Normalized (IDMN)	Large Area High Gray Level Emphasis (LAHGLE)
Variance		Inverse Difference (ID)	
Uniformity		Inverse Difference Normalized (IDN)	
		Inverse Variance	
		Maximum Probability	
		Sum Average	
		Sum Entropy	
		Sum of Squares	

Πίνακας 5.4: Χαρακτηριστικά: Πρώτης τάξης, Σχήματος, GLCM, GLSZM.

GLRLM	NGTDM	GLDM
Short Run Emphasis (SRE)	Coarseness	Large Dependence High Gray Level Emphasis (LDHGLE)
Long Run High Gray Level Emphasis (LRHGLE)	Contrast	Small Dependence Emphasis (SDE)
Long Run Emphasis (LRE)	Busyness	Large Dependence Emphasis (LDE)
Gray Level Non-Uniformity (GLN)	Complexity	Gray Level Non-Uniformity (GLN)
Gray Level Non-Uniformity Normalized (GLNN)	Strength	Dependence Non-Uniformity (DN)
Run Length Non-Uniformity (RLN)		Dependence Non-Uniformity Normalized (DNN)
Run Length Non-Uniformity Normalized (RLNN)		Gray Level Variance (GLV)
Run Percentage (RP)		Dependence Variance (DV)
Zone Variance (ZV)		Dependence Entropy (DE)
Run Variance (RV)		Low Gray Level Emphasis (LGLE)
Run Entropy (RE)		High Gray Level Emphasis (HGLE)
Low Gray Level Run Emphasis (LGLRE)		Small Dependence Low Gray Level Emphasis (SDLGLE)
Gray Level Variance (GLV)		Small Dependence High Gray Level Emphasis (SDHGLE)
Short Run Low Gray Level Emphasis (SRLGLE)		Large Dependence Low Gray Level Emphasis (LDLGLE)
Short Run High Gray Level Emphasis (SRHGLE)		
Long Run Low Gray Level Emphasis (LRLGLE)		

Πίνακας 5.5: Χαρακτηριστικά: GLRLM, NGTDM, GLDM.

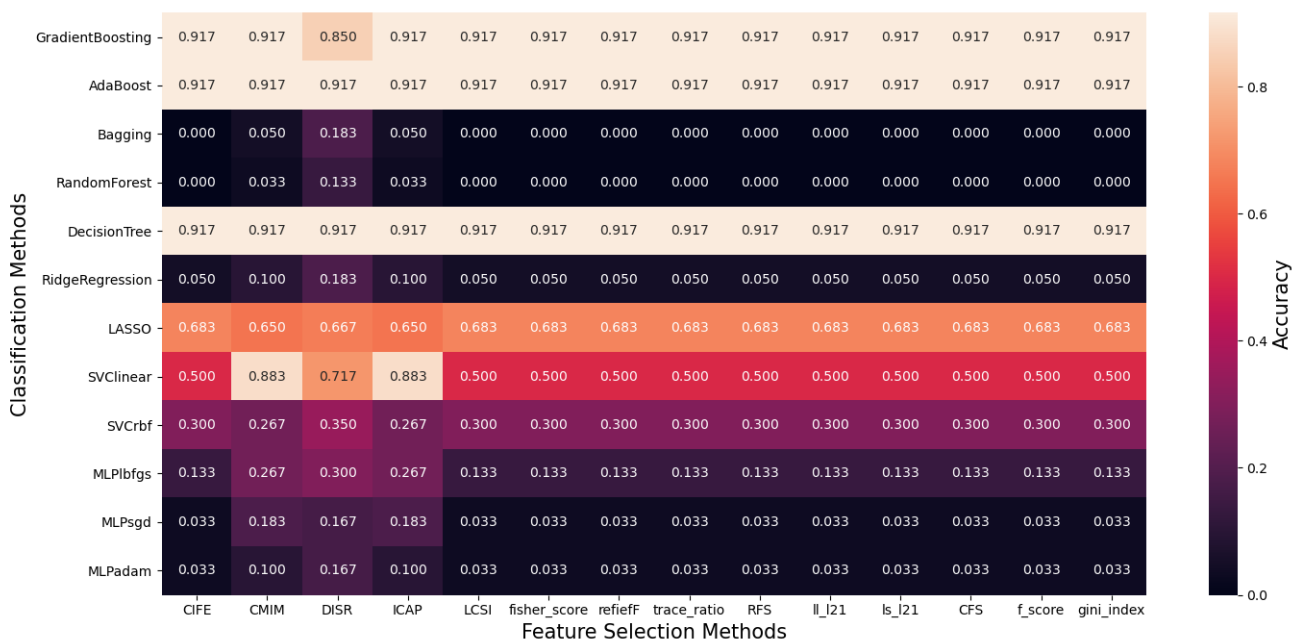
Κεφάλαιο 6

Αποτελέσματα

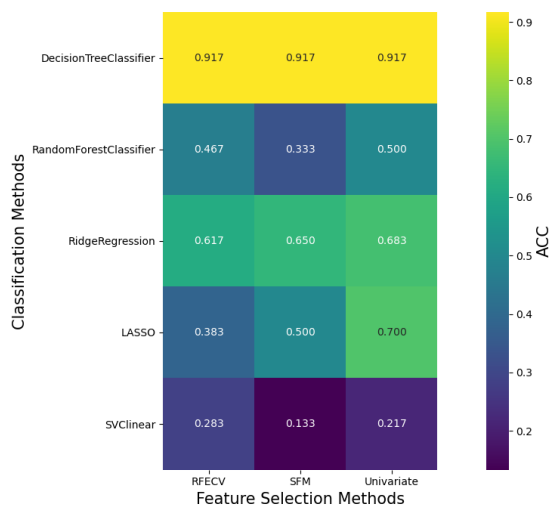
Στο κεφάλαιο αυτό παρατίθενται τα αποτελέσματα των προσομοιώσεων του συστήματος για τις περιπτώσεις της: (α) Επιβλεπόμενης Μάθησης, (β) Μη Επιβλεπόμενης Μάθησης για προκαθορισμένο αριθμό επιλεγέντων χαρακτηριστικών (100, 200). Στην περίπτωση της βιβλιοθήκης `colin-mico` ο αριθμός των επιλεγέντων χαρακτηριστικών επιλέγεται αυτόματα κατά την εκτέλεση του προγράμματος.

6.1 Απόδοση Ταξινομητών - Ασθενής 6

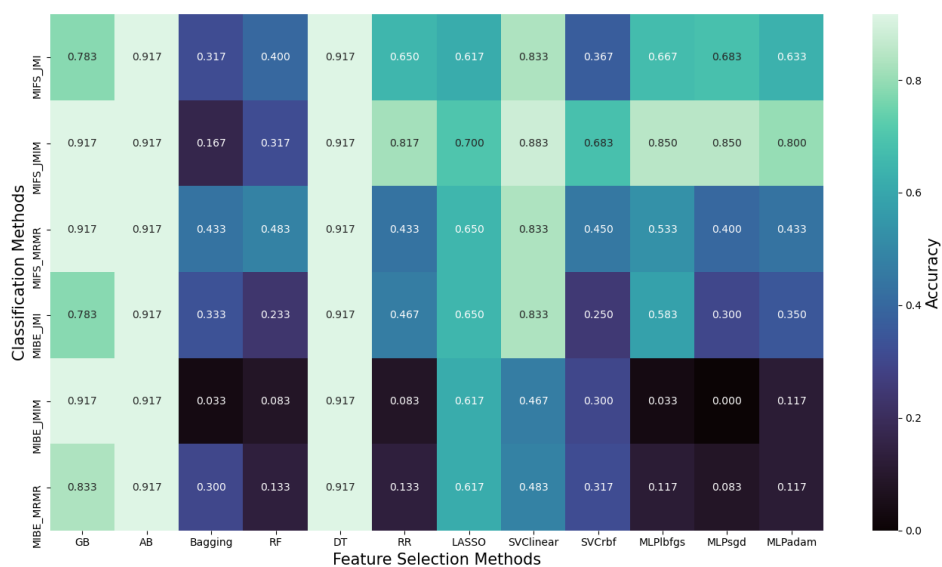
Επιβλεπόμενη Μάθηση



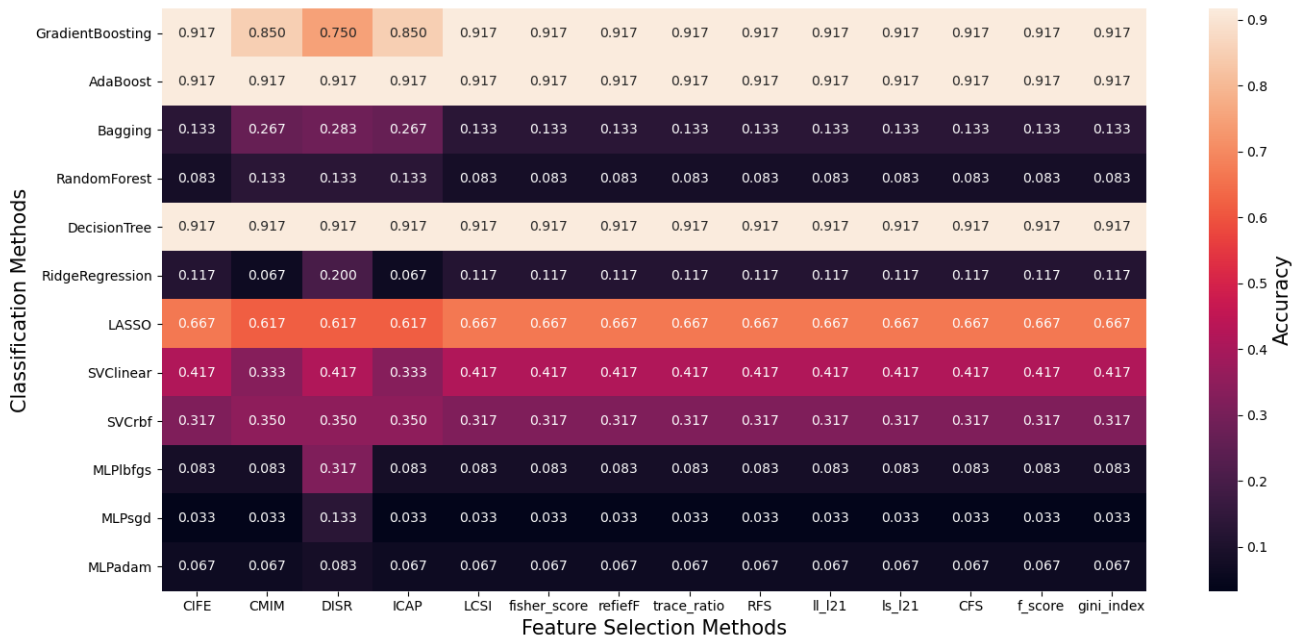
Εικόνα 6.1: **scikit-feature** - 100 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.2: **scikit-learn**. Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).

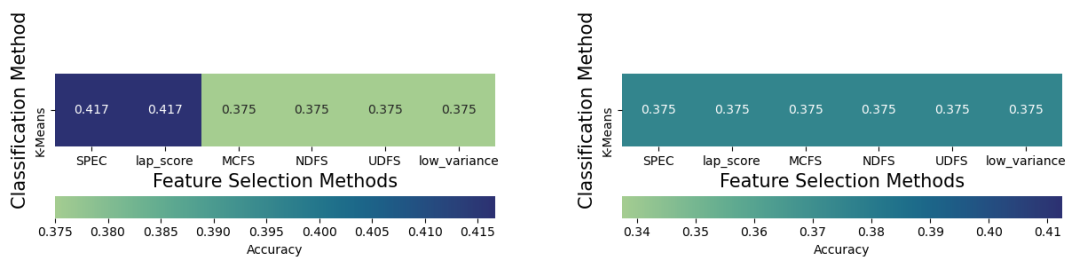


Εικόνα 6.3: **colin-mico**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.4: **scikit-feature** - 200 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

Μη Επιβλεπόμενη Μάθηση



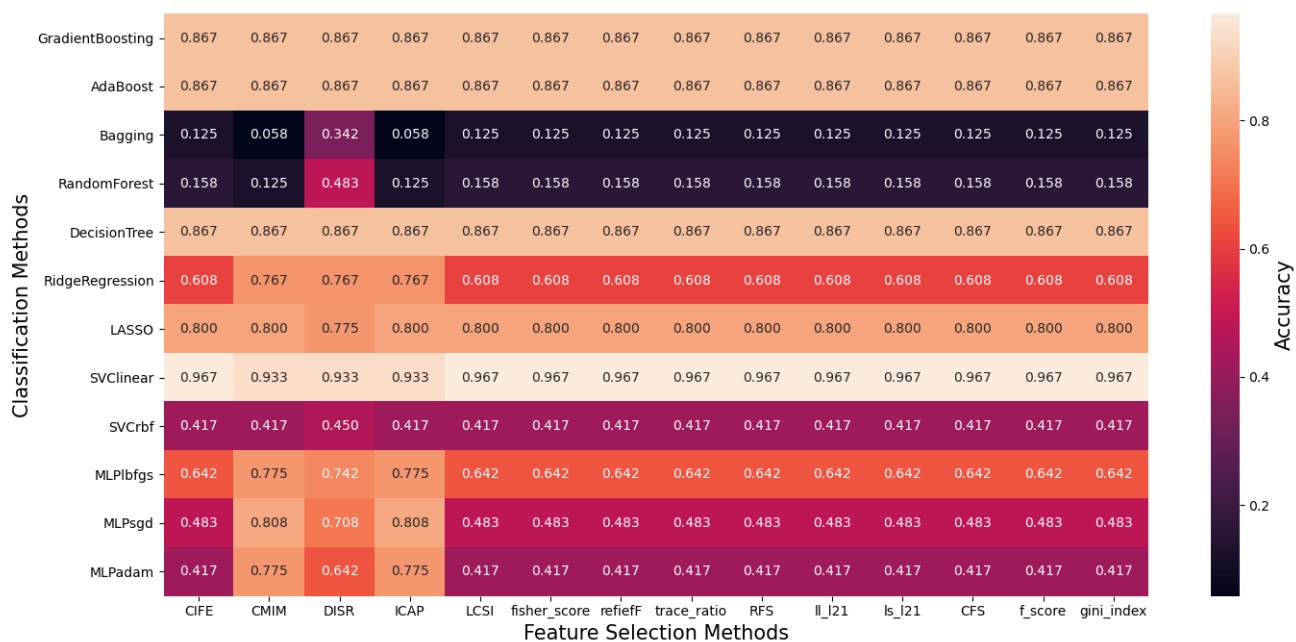
(α) 100 επιλεγμένα χαρακτηριστικά

(β) 200 επιλεγμένα χαρακτηριστικά

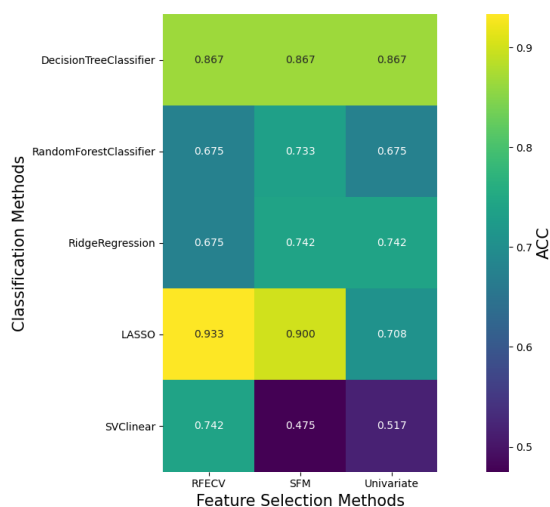
Σχήμα 6.1: **scikit-feature**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

6.2 Απόδοση Ταξινομητών - Ασθενής 12

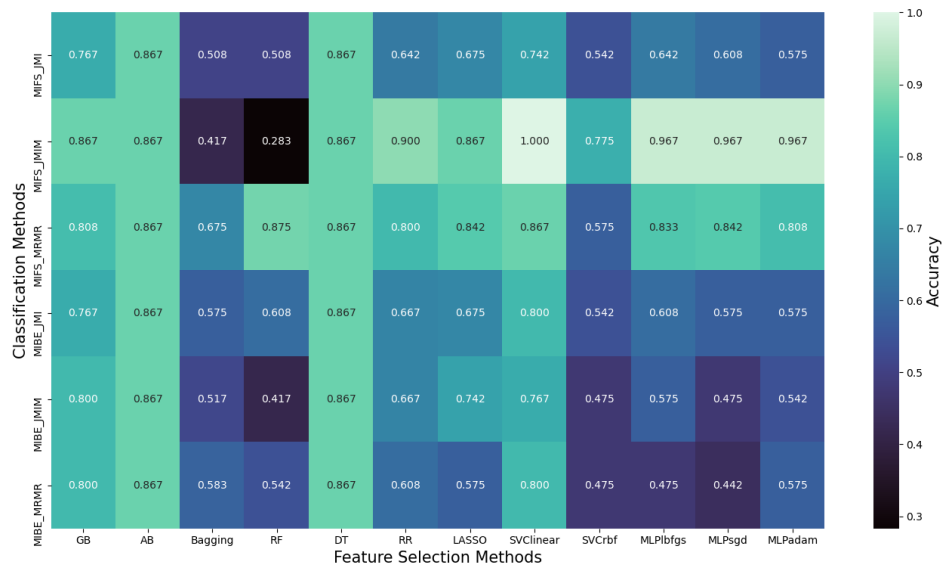
Επιβλεπόμενη Μάθηση



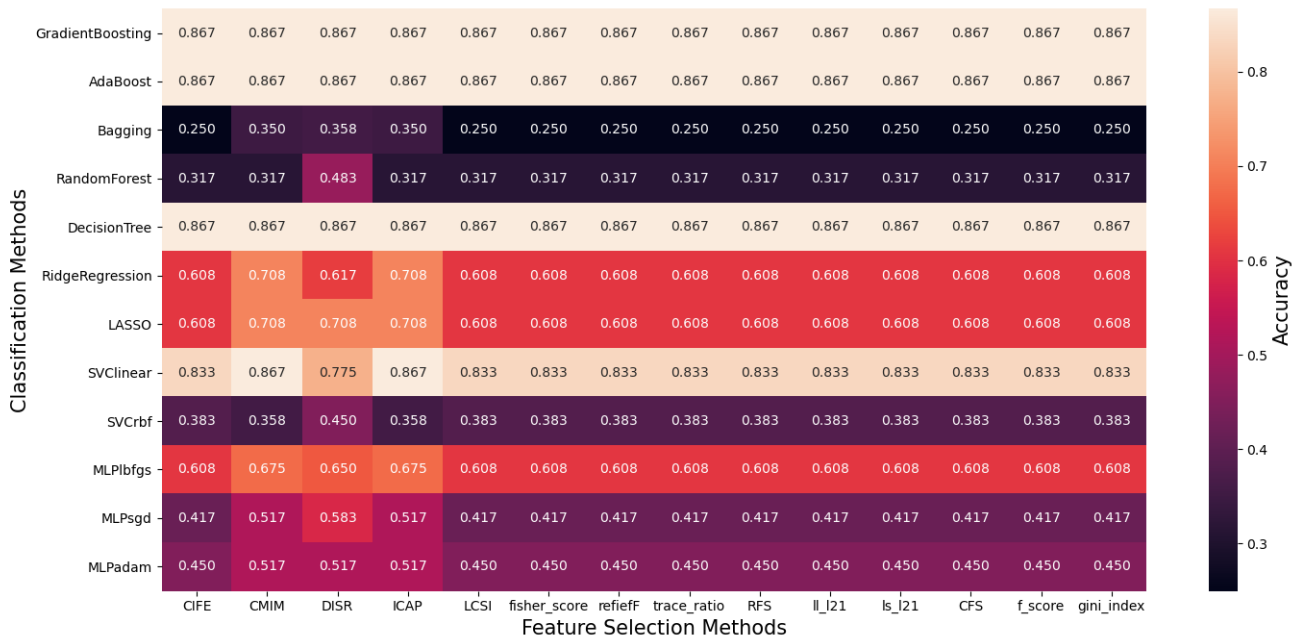
Εικόνα 6.5: **scikit-feature** - 100 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.6: **scikit-learn**. Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).

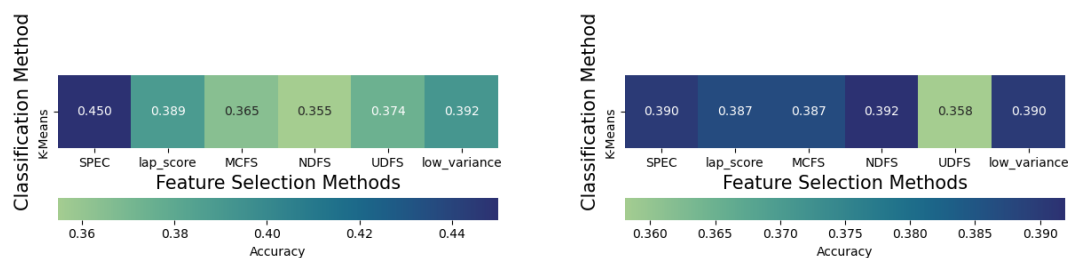


Εικόνα 6.7: **colin-mico**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.8: **scikit-feature - 200 επιλεγμένα χαρακτηριστικά**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

Μη Επιβλεπόμενη Μάθηση



(α) 100 επιλεγμένα χαρακτηριστικά

(β) 200 επιλεγμένα χαρακτηριστικά

Σχήμα 6.2: **scikit-feature**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

6.3 Απόδοση Ταξινομητών - Ασθενής 16

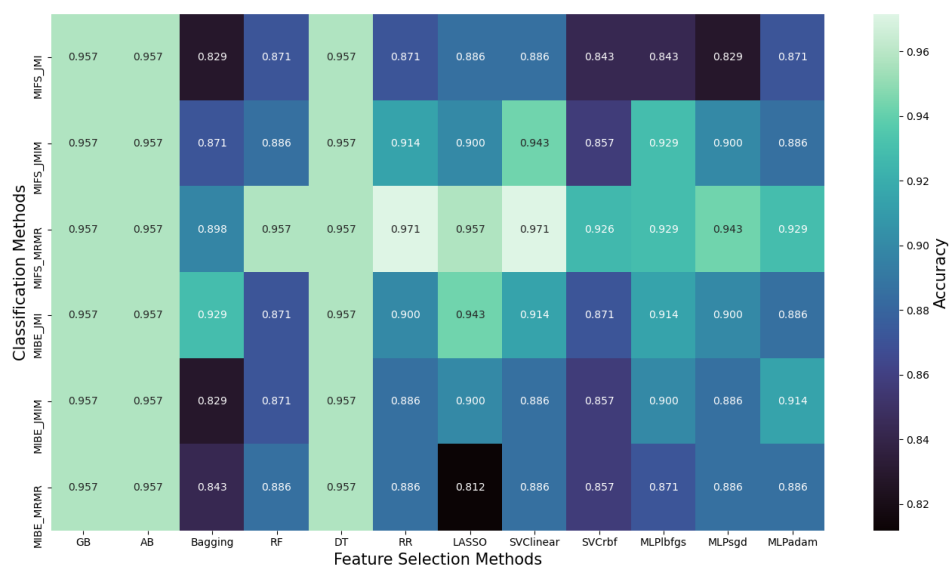
Επιβλεπόμενη Μάθηση



Εικόνα 6.9: **scikit-feature** - 100 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.10: **scikit-learn**. Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).

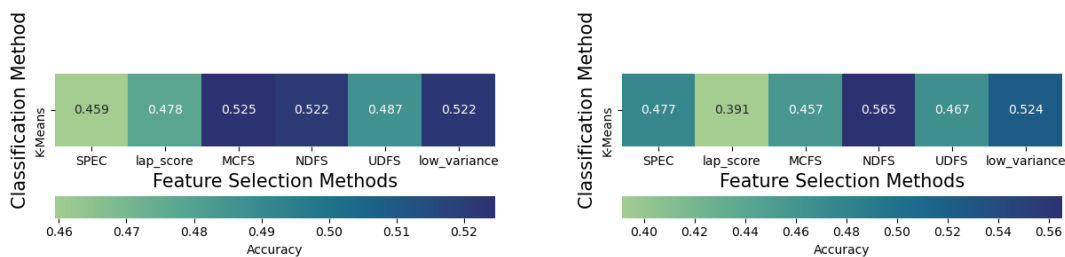


Εικόνα 6.11: **colin-mico**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.12: **scikit-feature** - 200 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

Μη Επιβλεπόμενη Μάθηση



(α) 100 επιλεγμένα χαρακτηριστικά

(β) 200 επιλεγμένα χαρακτηριστικά

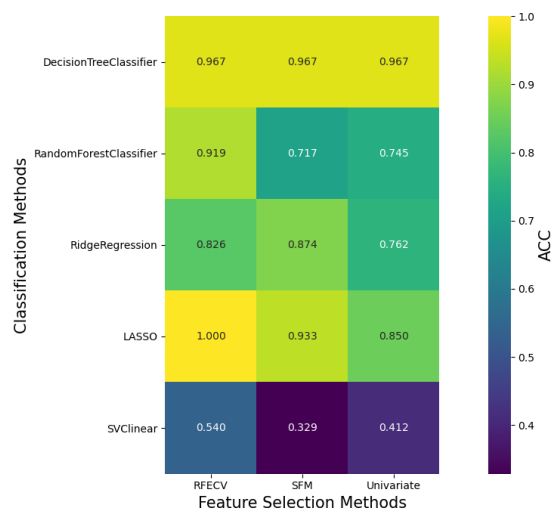
Σχήμα 6.3: **scikit-feature**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

6.4 Απόδοση Ταξινομητών - Ασθενής 18

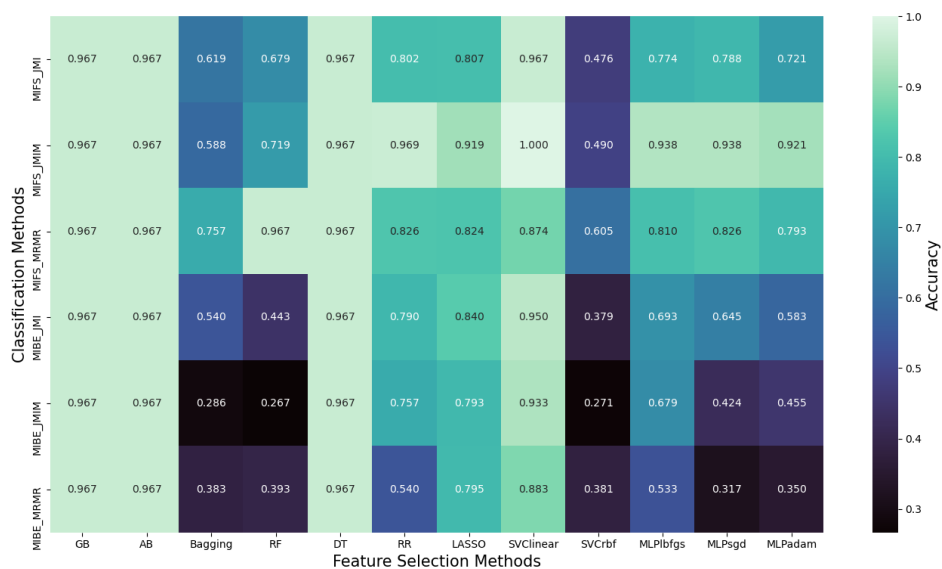
Επιβλεπόμενη Μάθηση



Εικόνα 6.13: **scikit-feature** - 100 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.14: **scikit-learn**. Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).

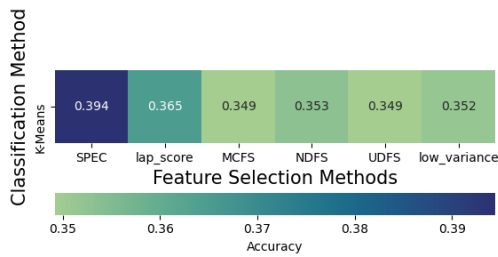


Εικόνα 6.15: **colin-mico**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

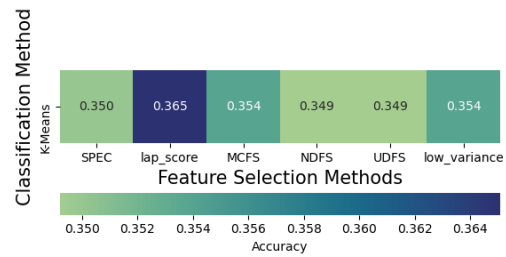


Εικόνα 6.16: **scikit-feature - 200 επιλεγμένα χαρακτηριστικά**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

Μη Επιβλεπόμενη Μάθηση



(α') 100 επιλεγμένα χαρακτηριστικά

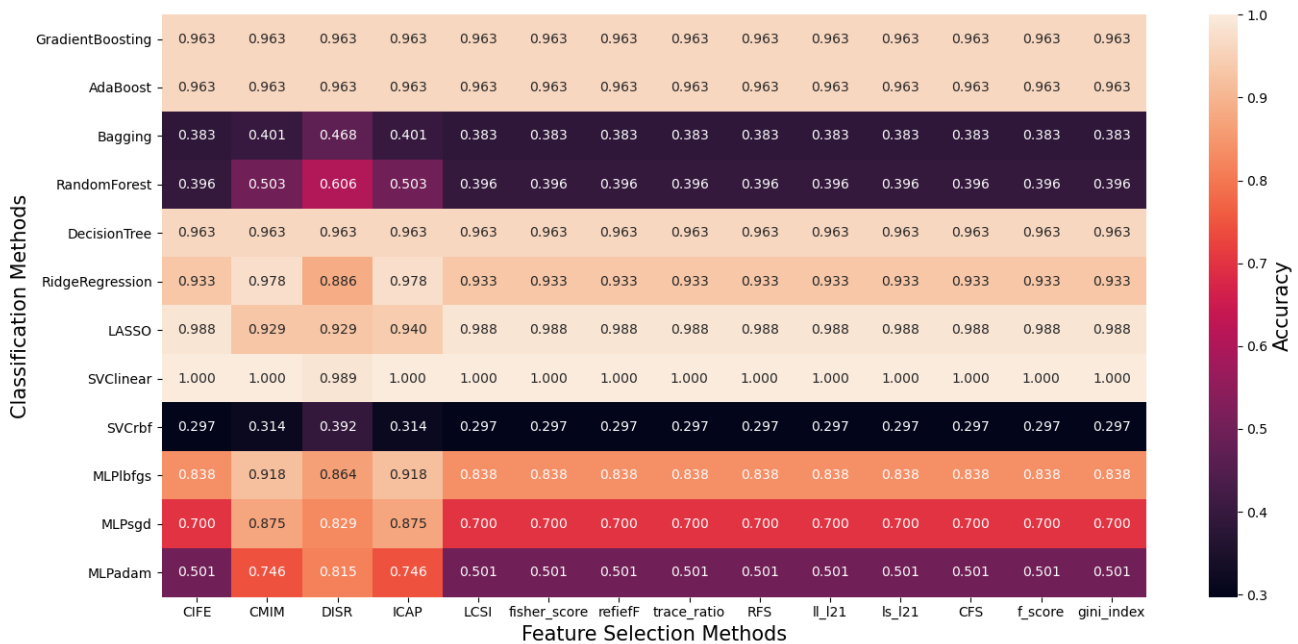


(β') 200 επιλεγμένα χαρακτηριστικά

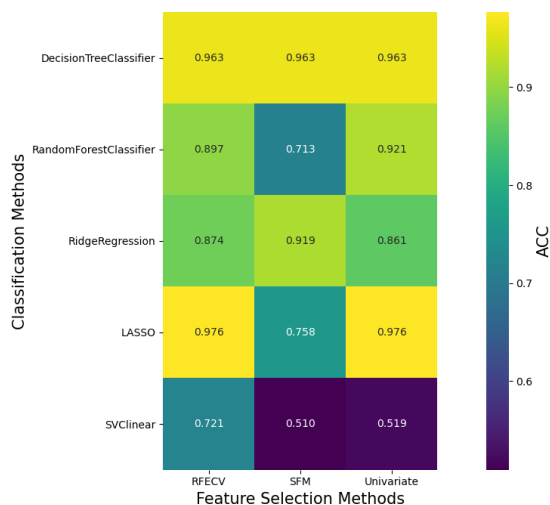
Σχήμα 6.4: **scikit-feature**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

6.5 Απόδοση Ταξινομητών - Ασθενής 22

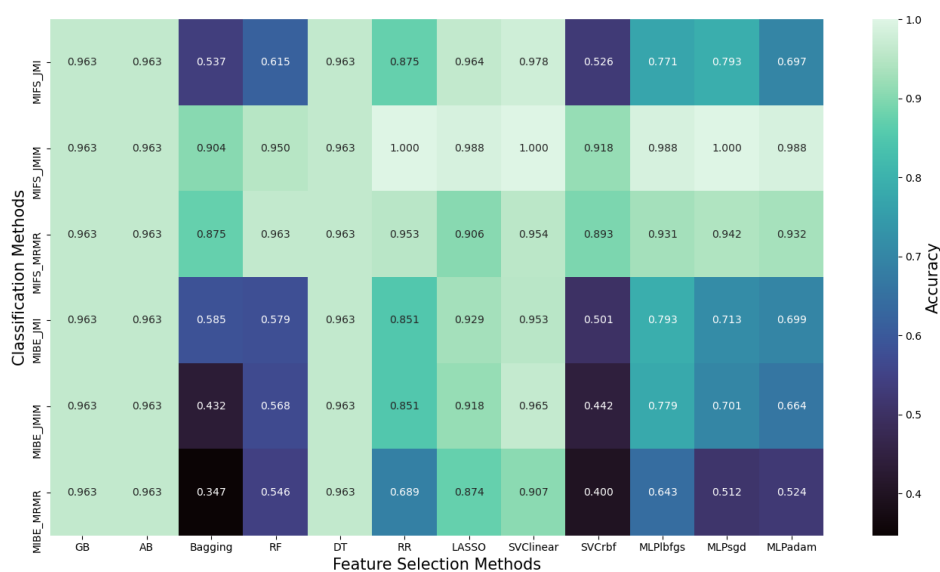
Επιβλεπόμενη Μάθηση



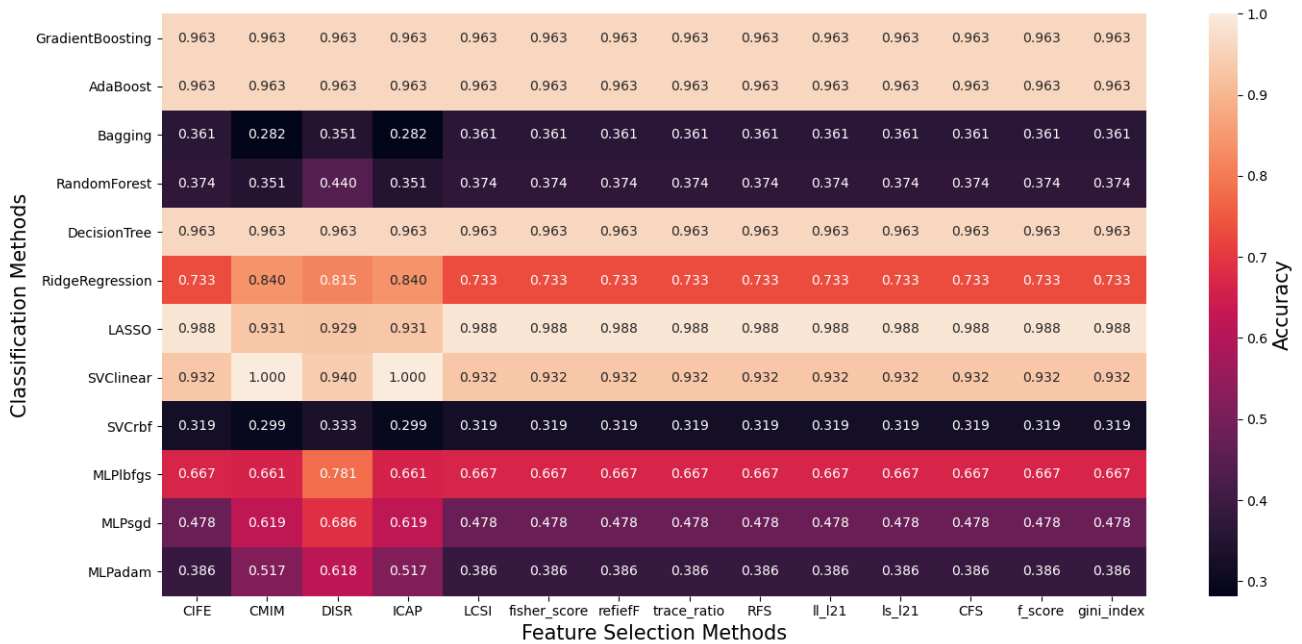
Εικόνα 6.17: **scikit-feature** - 100 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.18: **scikit-learn**. Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).

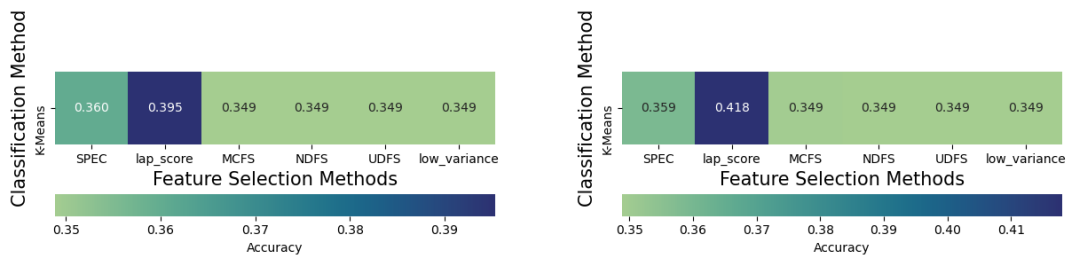


Εικόνα 6.19: **colin-mico**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.20: **scikit-feature** - 200 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

Μη Επιβλεπόμενη Μάθηση



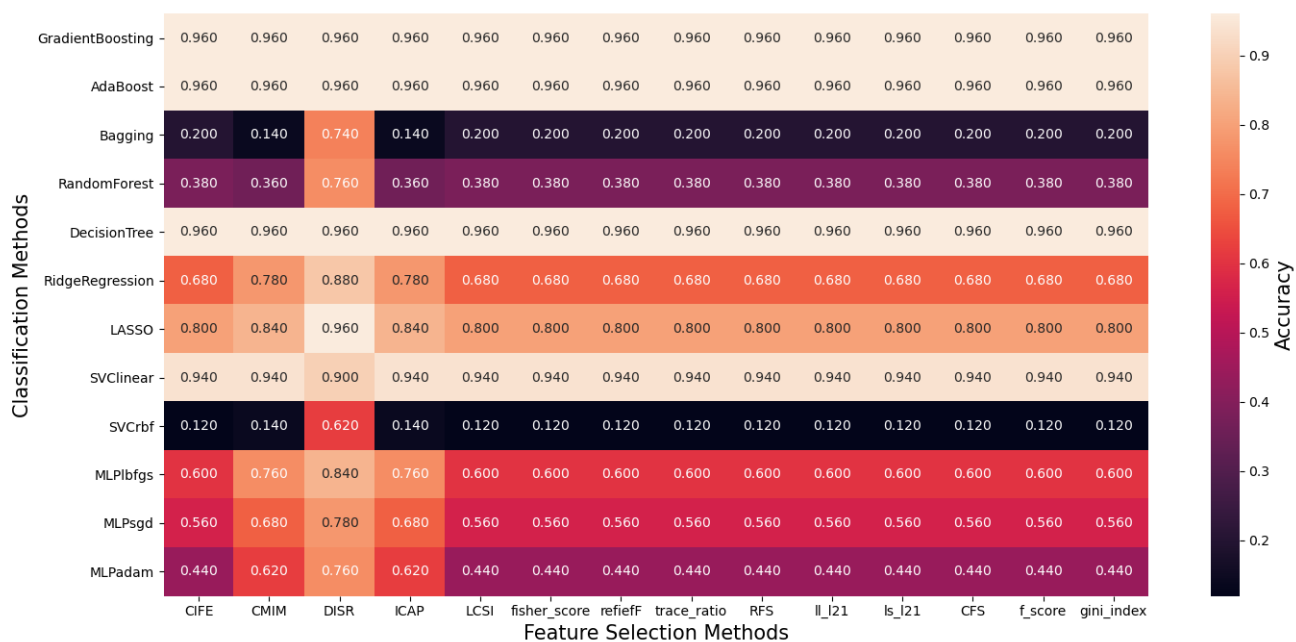
(α) 100 επιλεγμένα χαρακτηριστικά

(β) 200 επιλεγμένα χαρακτηριστικά

Σχήμα 6.5: **scikit-feature**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

6.6 Απόδοση Ταξινομητών - Ασθενής 23

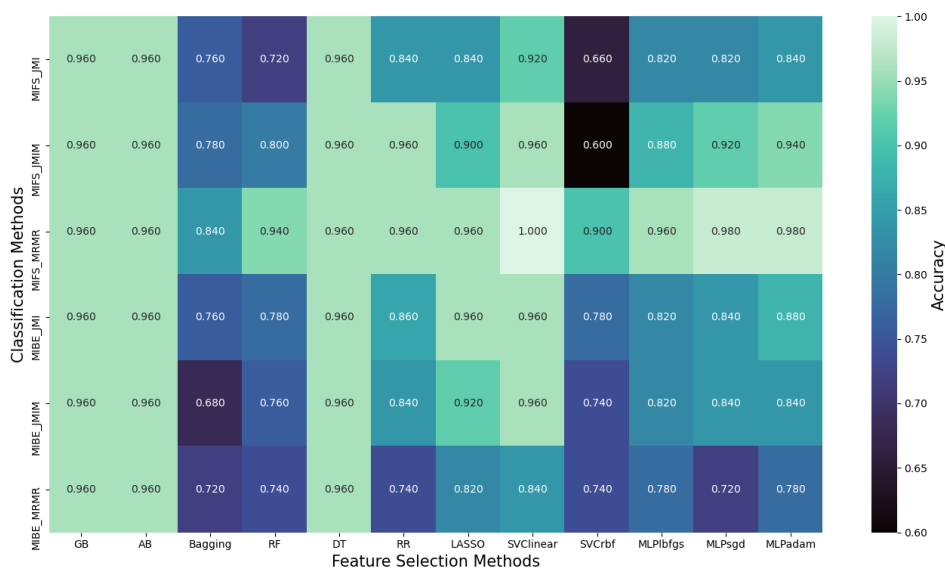
Επιβλεπόμενη Μάθηση



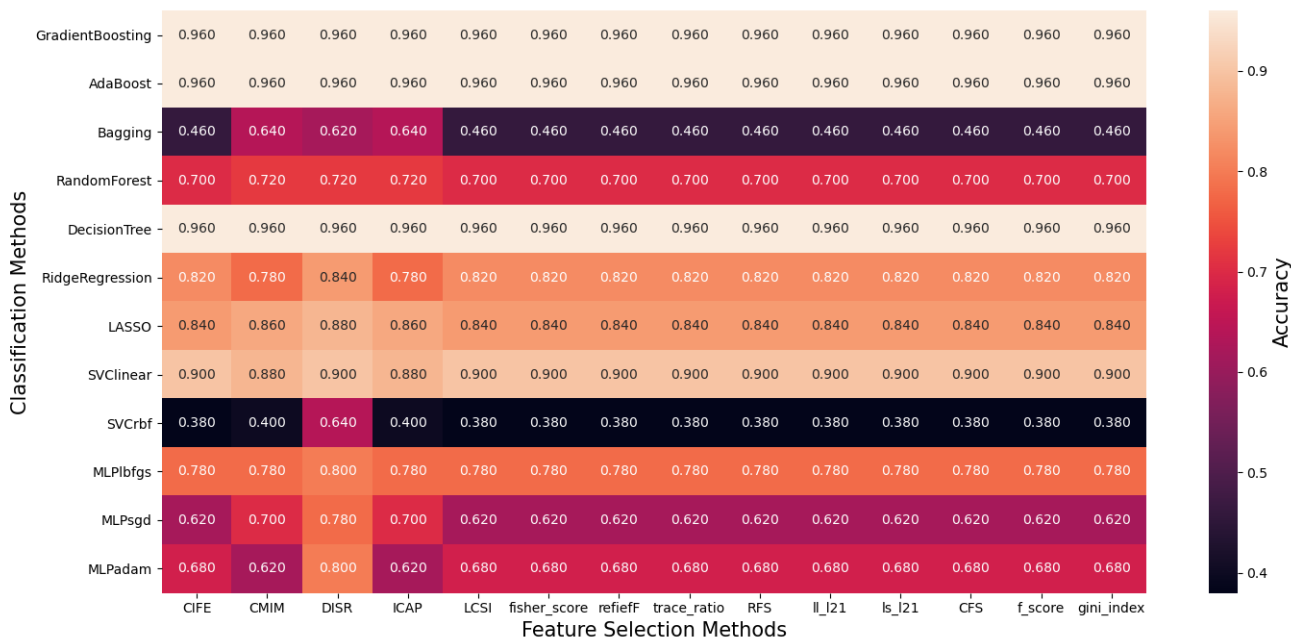
Εικόνα 6.21: **scikit-feature** - 100 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.22: **scikit-learn**. Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).

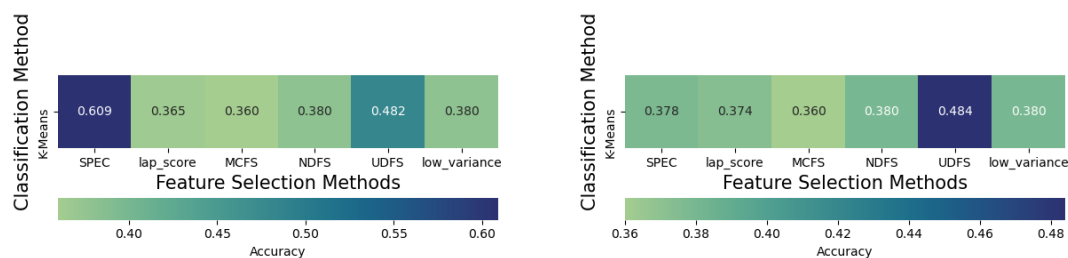


Εικόνα 6.23: **colin-mico**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.24: **scikit-feature - 200 επιλεγμένα χαρακτηριστικά**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

Μη Επιβλεπόμενη Μάθηση



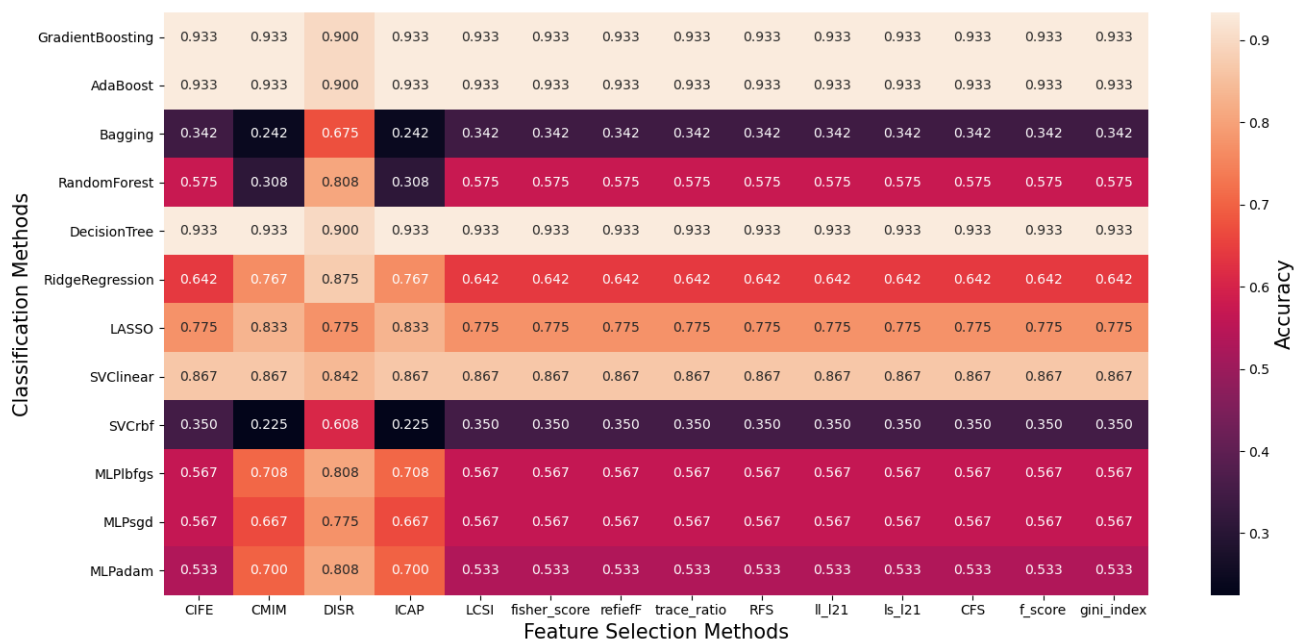
(α) 100 επιλεγμένα χαρακτηριστικά

(β) 200 επιλεγμένα χαρακτηριστικά

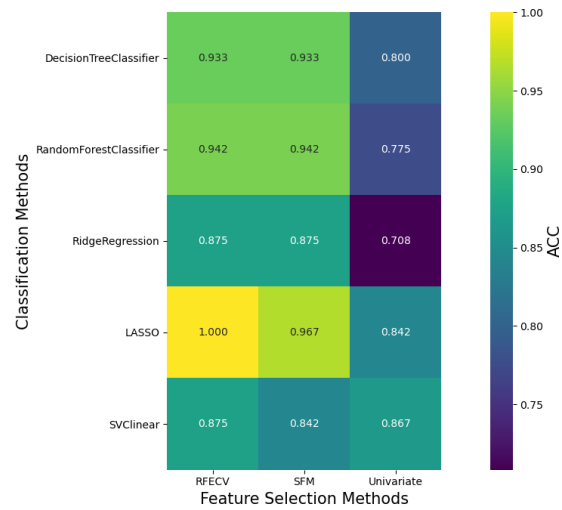
Σχήμα 6.6: **scikit-feature**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

6.7 Απόδοση Ταξινομητών - Ασθενής 27

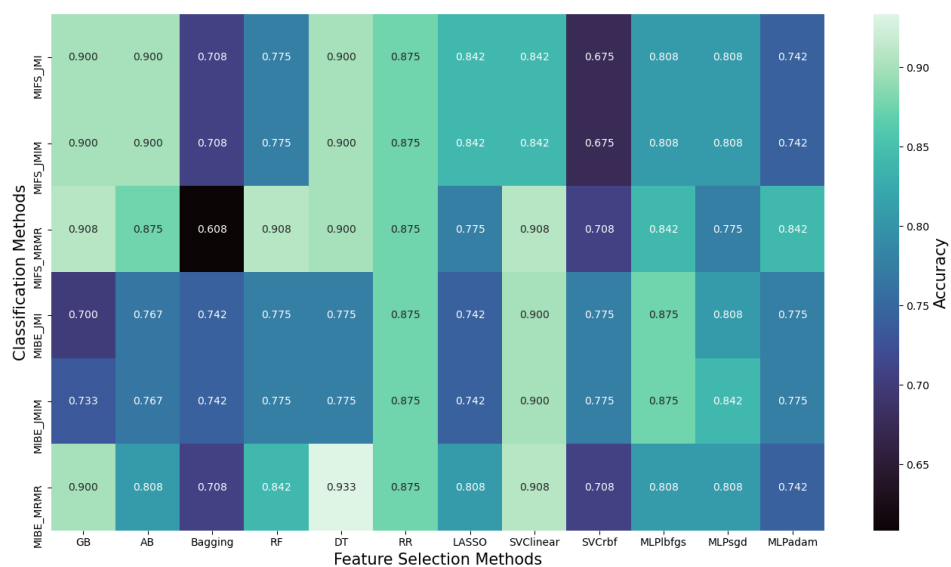
Επιβλεπόμενη Μάθηση



Εικόνα 6.25: **scikit-feature** - 100 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.26: **scikit-learn**. Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).

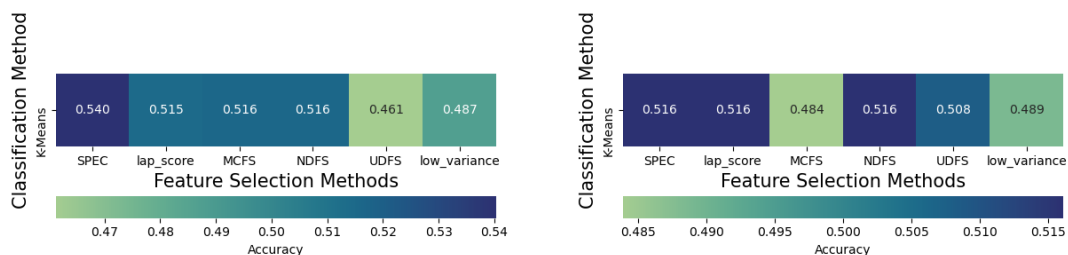


Εικόνα 6.27: **colin-mico**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.28: **scikit-feature** - 200 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

Μη Επιβλεπόμενη Μάθηση



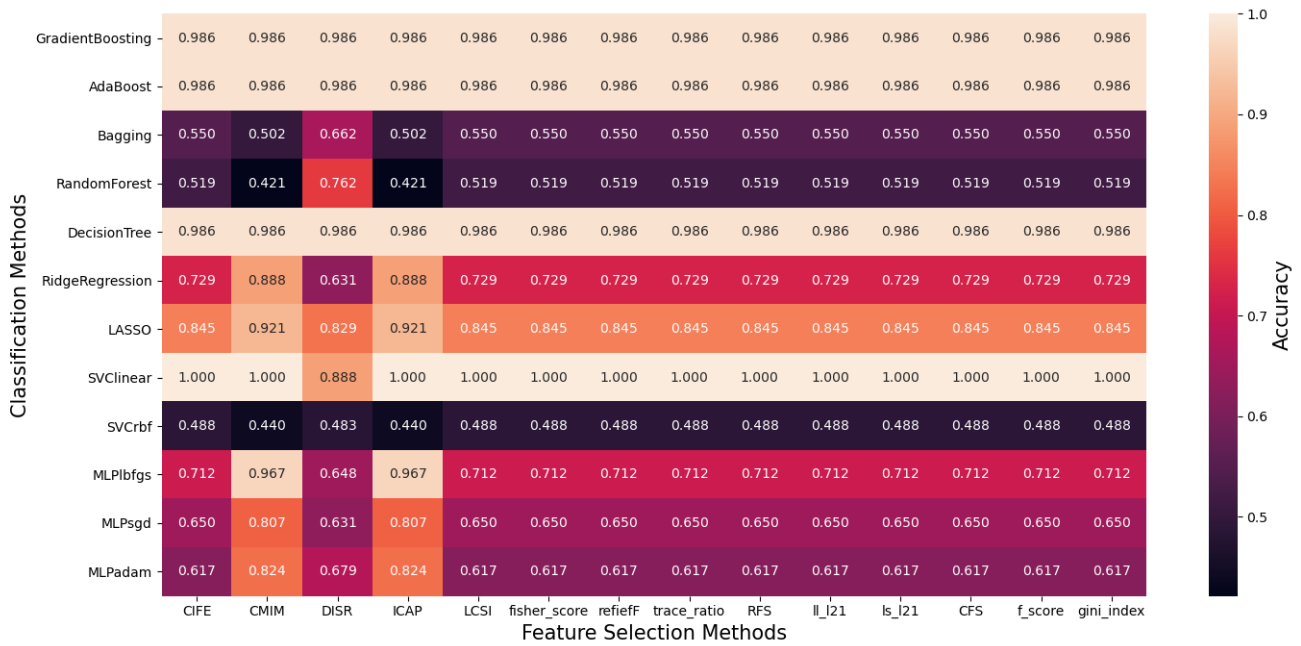
(α) 100 επιλεγμένα χαρακτηριστικά

(β) 200 επιλεγμένα χαρακτηριστικά

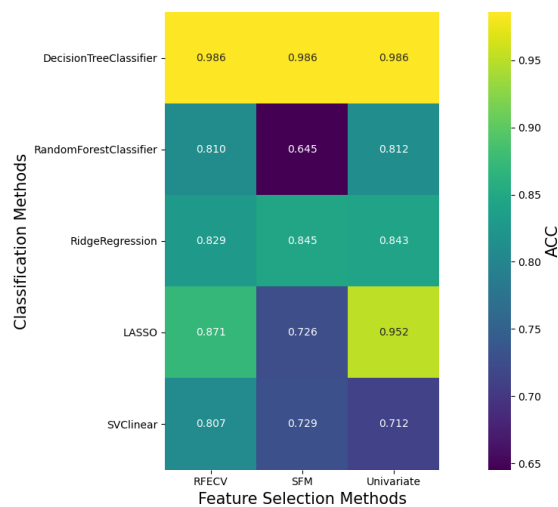
Σχήμα 6.7: **scikit-feature**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

6.8 Απόδοση Ταξινομητών - Ασθενής 28

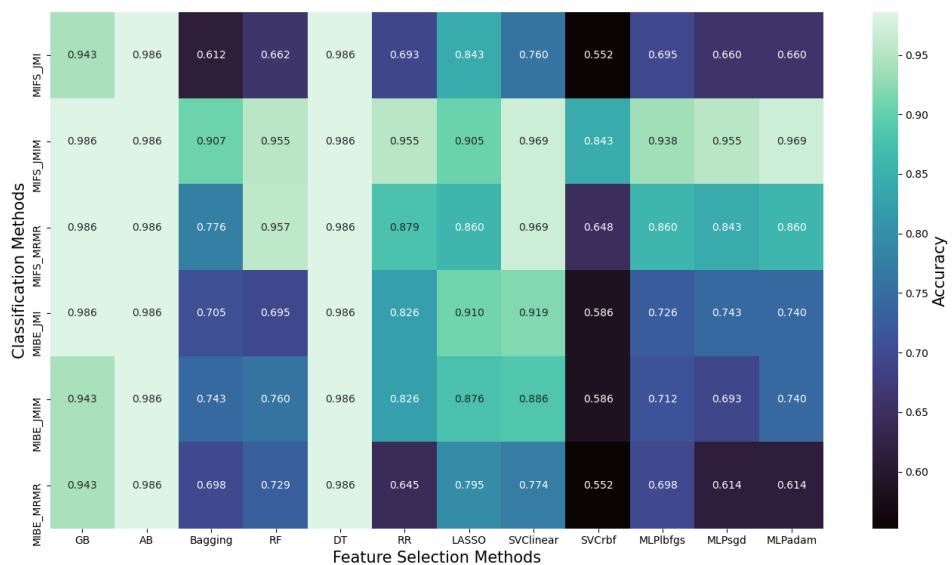
Επιβλεπόμενη Μάθηση



Εικόνα 6.29: **scikit-feature** - 100 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.30: **scikit-learn**. ΠΠρογνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).

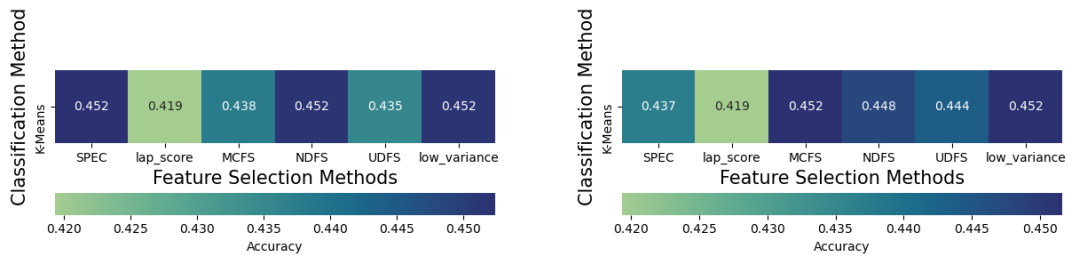


Εικόνα 6.31: **colin-mico**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.32: **scikit-feature - 200 επιλεγμένα χαρακτηριστικά**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

Μη Επιβλεπόμενη Μάθηση



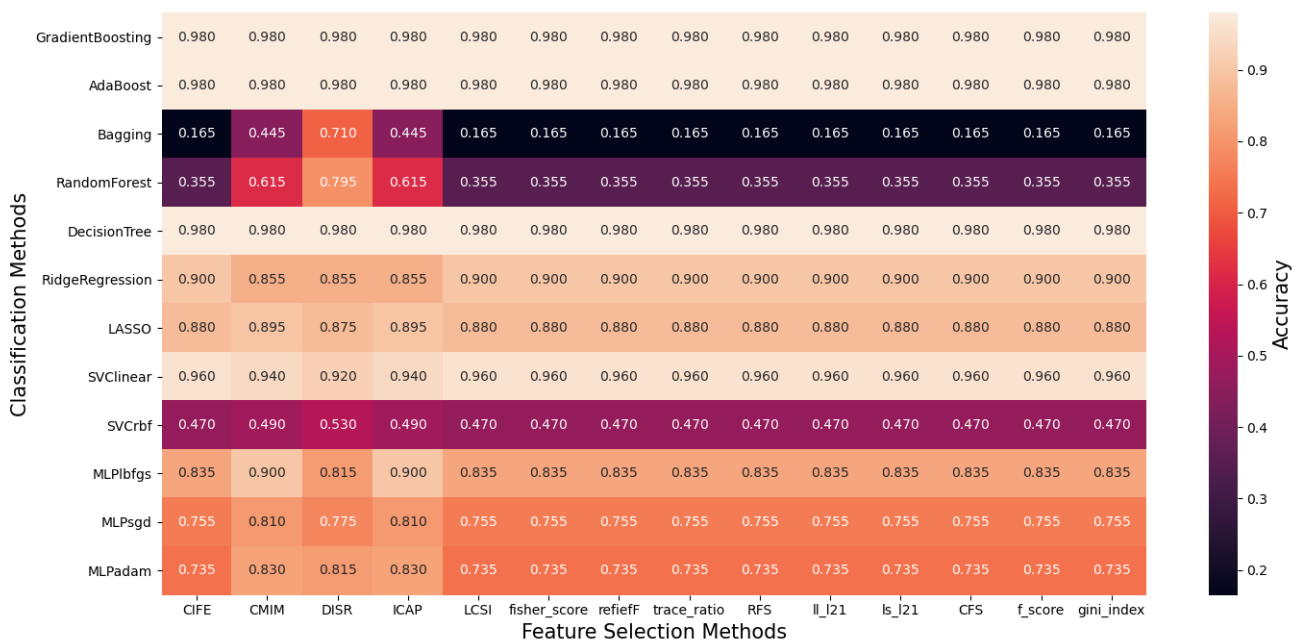
(α) 100 επιλεγμένα χαρακτηριστικά

(β) 200 επιλεγμένα χαρακτηριστικά

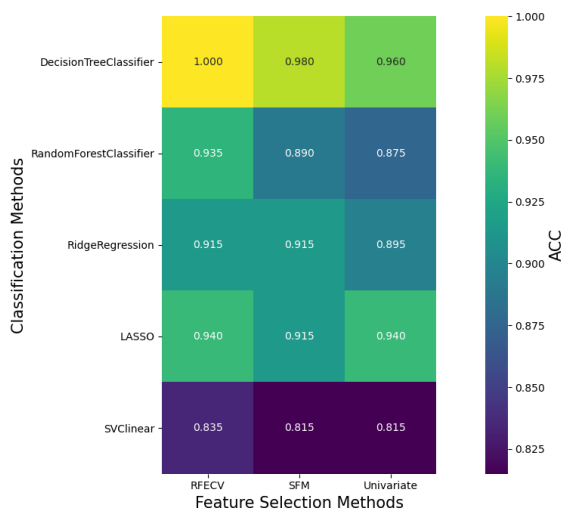
Σχήμα 6.8: **scikit-feature**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

6.9 Απόδοση Ταξινομητών - Ασθενής 29

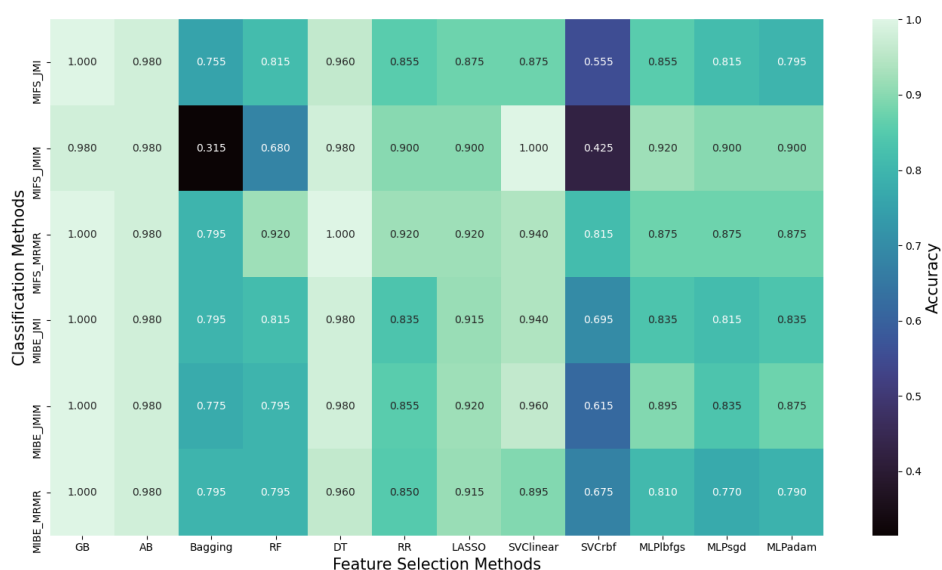
Επιβλεπόμενη Μάθηση



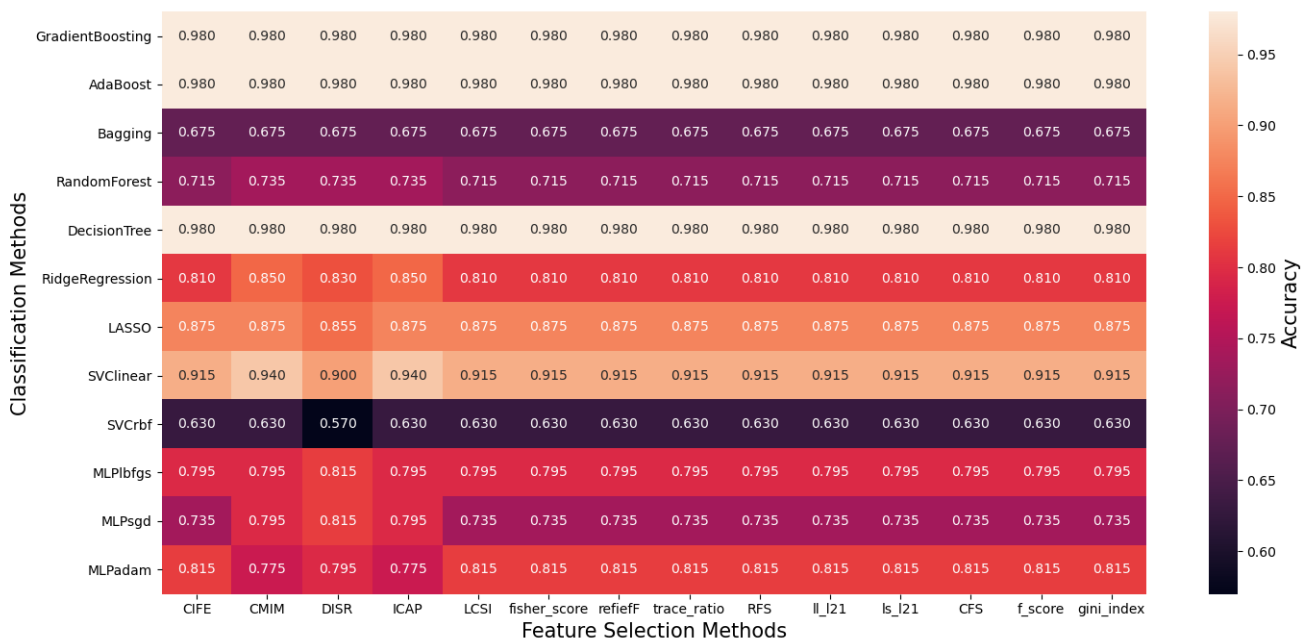
Εικόνα 6.33: **scikit-feature** - 100 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.34: **scikit-learn**. Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).

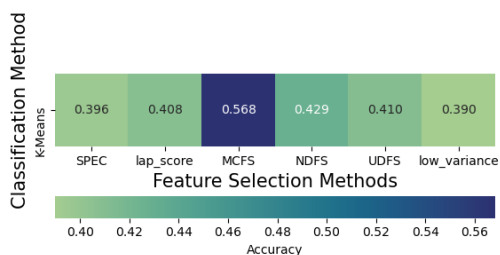


Εικόνα 6.35: **colin-mico**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

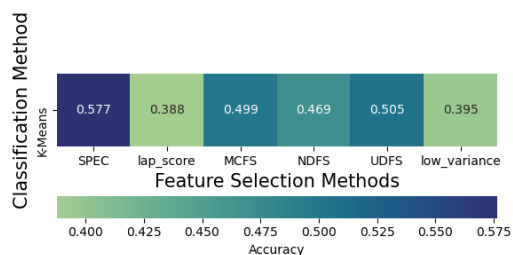


Εικόνα 6.36: **scikit-feature** - 200 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

Μη Επιβλεπόμενη Μάθηση



(α) 100 επιλεγμένα χαρακτηριστικά

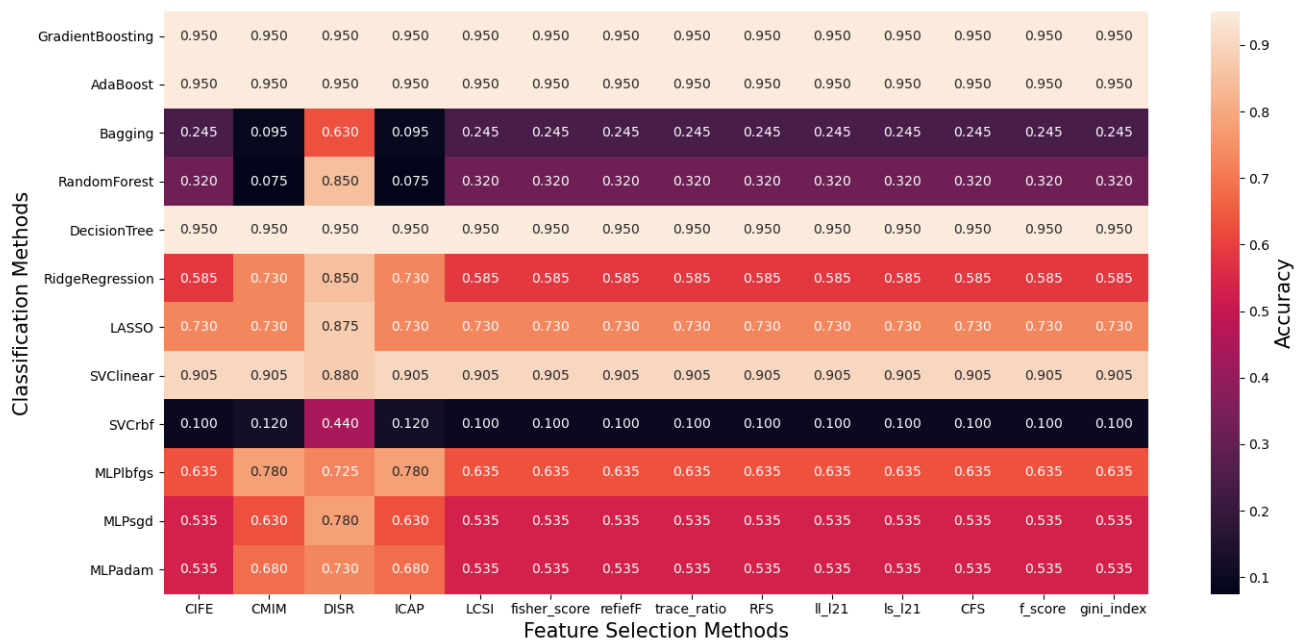


(β) 200 επιλεγμένα χαρακτηριστικά

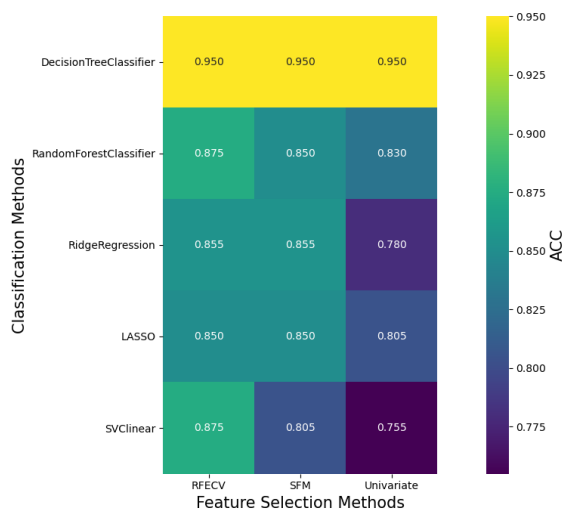
Σχήμα 6.9: **scikit-feature**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

6.10 Απόδοση Ταξινομητών - Ασθενής 30

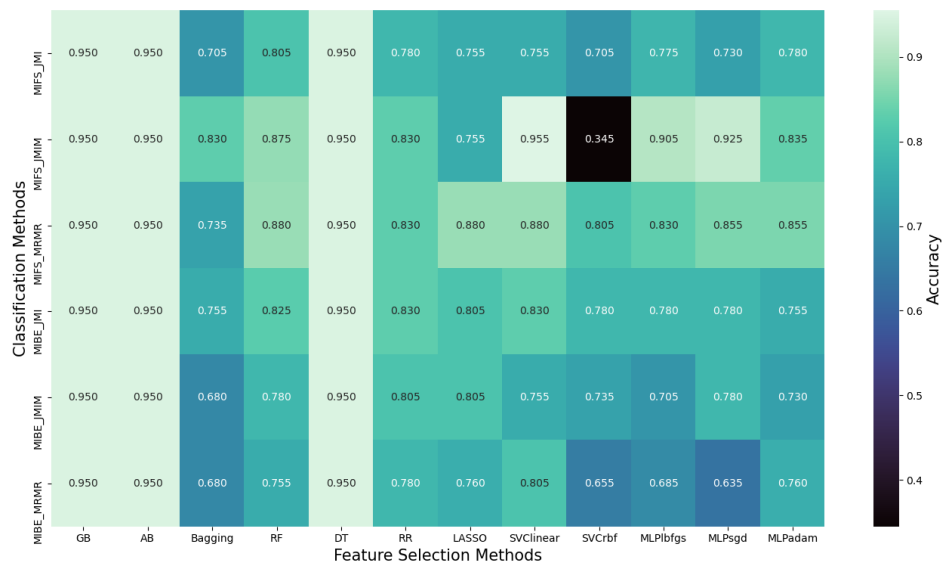
Επιβλεπόμενη Μάθηση



Εικόνα 6.37: **scikit-feature** - 100 επιλεγμένα χαρακτηριστικά. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.38: **scikit-learn**. Προγνωστική ακρίβεια (ACC) της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) και της μεθόδου ταξινόμησης (σε σειρές).

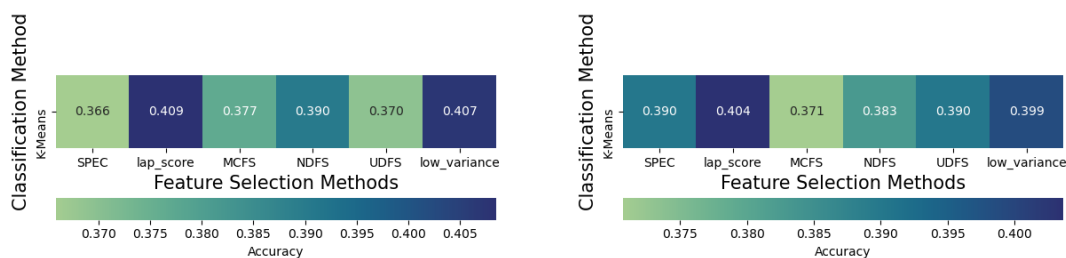


Εικόνα 6.39: **colin-mico**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).



Εικόνα 6.40: **scikit-feature - 200 επιλεγμένα χαρακτηριστικά**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

Μη Επιβλεπόμενη Μάθηση



(α) 100 επιλεγμένα χαρακτηριστικά

(β) 200 επιλεγμένα χαρακτηριστικά

Σχήμα 6.10: **scikit-feature**. Προγνωστική ακρίβεια (Accuracy) της μεθόδου επιλογής χαρακτηριστικών (σε σειρές) και της μεθόδου ταξινόμησης (σε στήλες).

6.11 Στατιστική Αξιολόγηση - Τυπική Απόκλιση

Ασθενής	GB	AD	BG	RF	DT	RR	LASSO	SVC1	SVCr	MLP1	MLPs	MLPa
A06_100	0.18	0.17	0.04	0.03	0.17	0.16	0.34	0.33	0.30	0.22	0.13	0.12
A06_200	0.18	0.17	0.04	0.03	0.17	0.16	0.34	0.33	0.30	0.22	0.13	0.12
A12_100	0.22	0.22	0.16	0.17	0.22	0.36	0.30	0.10	0.20	0.36	0.30	0.29
A12_200	0.22	0.22	0.23	0.28	0.22	0.33	0.38	0.22	0.30	0.28	0.30	0.27
A16_100	0.07	0.07	0.14	0.16	0.07	0.09	0.14	0.07	0.23	0.13	0.09	0.14
A16_200	0.07	0.07	0.22	0.17	0.07	0.09	0.18	0.08	0.18	0.11	0.12	0.12
A18_100	0.07	0.07	0.16	0.19	0.07	0.18	0.15	0.10	0.22	0.14	0.20	0.25
A18_200	0.07	0.07	0.16	0.19	0.07	0.18	0.15	0.10	0.22	0.14	0.20	0.25
A22_100	0.08	0.08	0.19	0.14	0.08	0.11	0.05	0.002	0.18	0.10	0.12	0.12
A22_200	0.08	0.08	0.17	0.16	0.08	0.16	0.05	0.07	0.15	0.13	0.15	0.14
A23_100	0.08	0.08	0.14	0.16	0.08	0.15	0.16	0.13	0.17	0.20	0.16	0.17
A23_200	0.08	0.08	0.21	0.19	0.08	0.15	0.17	0.14	0.22	0.14	0.13	0.14
A27_100	0.13	0.13	0.23	0.17	0.13	0.19	0.16	0.16	0.29	0.21	0.28	0.30
A27_200	0.15	0.15	0.23	0.14	0.15	0.15	0.20	0.16	0.28	0.06	0.18	0.18
A28_100	0.04	0.04	0.20	0.18	0.04	0.17	0.13	0.007	0.14	0.16	0.20	0.17
A28_200	0.04	0.04	0.20	0.17	0.04	0.23	0.16	0.18	0.14	0.20	0.18	0.18
A29_100	0.06	0.06	0.18	0.19	0.06	0.10	0.14	0.09	0.23	0.09	0.17	0.15
A29_200	0.06	0.06	0.21	0.20	0.06	0.21	0.14	0.10	0.18	0.15	0.15	0.16
A30_100	0.10	0.10	0.21	0.21	0.10	0.27	0.25	0.13	0.12	0.20	0.27	0.27
A30_200	0.10	0.10	0.22	0.25	0.10	0.27	0.24	0.20	0.16	0.26	0.27	0.28

Πίνακας 6.1: **scikit-feature** - **Επιβλεπόμενη Μάθηση**. Τυπική Απόκλιση (Standard Deviation) προγνωστικής ακρίβειας της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) της κάθε ασθενούς (σε σειρές).

Ασθενής	GB	AD	BG	RF	DT	RR	LASSO	SVC1	SVCr	MLPI	MLPs	MLPa
A06	0.23	0.17	0.24	0.23	0.17	0.25	0.33	0.22	0.30	0.22	0.22	0.25
A12	0.28	0.22	0.28	0.27	0.22	0.28	0.31	0.21	0.30	0.25	0.26	0.25
A16	0.07	0.07	0.14	0.12	0.07	0.10	0.10	0.09	0.14	0.10	0.12	0.11
A18	0.07	0.07	0.20	0.21	0.07	0.16	0.17	0.08	0.18	0.16	0.18	0.19
A22	0.08	0.08	0.15	0.14	0.08	0.11	0.08	0.05	0.16	0.12	0.12	0.13
A23	0.08	0.08	0.20	0.21	0.08	0.13	0.11	0.09	0.19	0.13	0.13	0.11
A27	0.17	0.15	0.19	0.17	0.15	0.17	0.20	0.15	0.21	0.21	0.21	0.21
A28	0.09	0.04	0.17	0.16	0.04	0.17	0.17	0.12	0.21	0.15	0.16	0.16
A29	0.01	0.06	0.15	0.14	0.06	0.13	0.15	0.08	0.16	0.13	0.12	0.13
A30	0.10	0.10	0.20	0.21	0.10	0.17	0.18	0.16	0.20	0.21	0.19	0.20

Πίνακας 6.2: **colin-mico - Επιβλεπόμενη Μάθηση**. Τυπική Απόκλιση (Standard Deviation) προγνωστικής ακρίβειας της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) της κάθε ασθενούς (σε σειρές).

Ασθενής	DT	RF	RR	LASSO	SVC1
A06	0.17	0.32	0.36	0.24	0.22
A12	0.22	0.27	0.32	0.23	0.27
A16	0.07	0.06	0.10	0.06	0.10
A18	0.07	0.13	0.16	0.10	0.21
A22	0.08	0.12	0.12	0.09	0.15
A23	0.08	0.14	0.12	0.12	0.23
A27	0.16	0.15	0.18	0.09	0.16
A28	0.04	0.22	0.19	0.12	0.18
A29	0.05	0.13	0.13	0.10	0.10
A30	0.10	0.16	0.17	0.16	0.22

Πίνακας 6.3: **scikit-learn - Επιβλεπόμενη Μάθηση**. Τυπική Απόκλιση (Standard Deviation) προγνωστικής ακρίβειας της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) της κάθε ασθενούς (σε σειρές).

Ασθενής	K-Means_100	K-Means_200
A06	0.001	0.001
A12	0.01	0.01
A16	0.01	0.01
A18	0.003	0.003
A22	0.001	0.001
A23	0.01	0.01
A27	0.01	0.02
A28	0.002	0.005
A29	0.03	0.01
A30	0.008	0.008

Πίνακας 6.4: **scikit-feature - Μη Επιβλεπόμενη Μάθηση**. Τυπική Απόκλιση (Standard Deviation) προγνωστικής ακρίβειας της μεθόδου επιλογής χαρακτηριστικών (σε στήλες) της κάθε ασθενούς (σε σειρές).

Κεφάλαιο 7

Συζήτηση

Στόχος της διπλωματικής εργασίας ήταν η εξαγωγή χαρακτηριστικών από ιατρικές εικόνες μαγνητικού τομογράφου για τον σχεδιασμό ενός συστήματος υποστήριξης κλινικών αποφάσεων προκειμένου να συγκριθεί η απόδοση διαφορετικών μοντέλων επιλογής χαρακτηριστικών και κατηγοριοποίησης της καρκινικής επικινδυνότητας σε ασθενείς με καρκίνων ωοθηκών. Κατά τη διάρκεια των προσομοιώσεων, αναλύσαμε μια σειρά ερωτημάτων σχετικά με τα βήματα προ-επεξεργασίας, τη χρήση διαφορετικών συνόλων δεδομένων και ταξινομητών και την απόδοση των μοντέλων. Ακολουθεί ο σχολιασμός των αποτελεσμάτων.

7.1 Συμπεράσματα

7.1.1 Προεπεξεργασία Δεδομένων

Η ύπαρξη όσο των δυνατών περισσότερων χαρακτηριστικών από τα χρησιμοποιούμενα μοντέλα βελτιώνει την ευρωστία και αποδοτικότητα του συστήματος καθότι η επιλογή και η κατηγοριοποίηση γίνεται από ένα διευρυμένο σύνολο δεδομένων. Με βάση την γενική αυτή αρχή, προσπαθήσαμε να αυξήσουμε τεχνητά τον αρχικό αριθμό των εξαγόμενων χαρακτηριστικών (100 στην περίπτωση μας) της βιβλιοθήκης *pyradiomics*, προκειμένου να αυξήσουμε την απόδοση διαφόρων ταξινομητών. Για να το επιτύχουμε αυτό (επαύξηση των δεδομένων), εφαρμόσαμε διαφορετικά φίλτρα στις εικόνες εισόδου μετασχηματίζοντας την αρχική πληροφορία σε νέα στατιστικά αξιοποιήσιμη πληροφορία για τα μοντέλα του συστήματος. Με τον τρόπο αυτό δεκαπλασιάσαμε τα εξαγόμενα χαρακτηριστικά, φτάνοντας σε σύνολο τα ≈ 1200 . Στην συνέχεια φροντίσαμε να ελέγξουμε την πληρότητα των χαρακτηριστικών αφαιρώντας τα χαρακτηριστικά εκείνα που δεν διαφοροποιούνταν από εικόνα σε εικόνα διατηρώντας σταθερή τιμή καθώς και εκείνα τα χαρακτηριστικά τα οποία δεν εμφανίζονταν σε όλες τις εικόνες μιας ακολουθίας για να εξασφαλίσουμε πως το σύνολο δεδομένων περιελάμβανε μόνο τα πλήρη εξαγόμενα χαρακτηριστικά κάθε εικόνας. Καθότι τα δεδομένα μεταξύ περιοχών ήταν ισορροπημένα δεν χρειάστηκε να σταθμίσουμε επιπρόσθετα το σύνολο των δεδομένων μας. Η επιλογή των φίλτρων τα οποία αναφέρονται στο Υποκεφάλαιο 5.4.2 έγινε με βάση την σχετική βιβλιογραφία και εξαντλήθηκαν όλοι οι συνδυασμοί που παρέχονται από το *pyradiomics*. Οι προσομοιώσεις έδειξαν ότι η πλειονότητα των επιλεγέντων χαρακτηριστικών ανήκει σε μετασχηματισμένες εικόνες υπό την εφαρμογή φίλτρου.

7.1.2 Αλγόριθμοι Επιλογής Χαρακτηριστικών

Παρατηρώντας αρχικά τα γραφήματα (heatmaps) που αφορούν την βιβλιοθήκη scikit feature, συνολικά για όλες τις περιπτώσεις ασθενών με βάση την προγνωστική ακρίβεια βλέπουμε την υπεροχή των μοντέλων επιβλεπόμενης μάθησης των οποίων η υψηλή απόδοση διατηρείται αναλογικά σε όλο το φάσμα των ταξινομητών. Παράλληλα βλέπουμε ότι οι αλγόριθμοι βασισμένοι στην θεωρία της πληροφορίας (CIFE, CMIM, DISR, ICAP, LCS1) δίνουν σταθερά υψηλές τιμές που κυμαίνονται στο 80-90% κατά περίπτωση ταξινομητή. Ακολουθούν οι αλγόριθμοι ομοιότητας, αραιής μάθησης καθώς και οι στατιστικοί αλγόριθμοι με μέση απόδοση της τάξης του 70-80%. Αντίστοιχα τα μοντέλα μη επιβλεπόμενης μάθησης παρουσιάζουν χαμηλότερη απόδοση ανεξαρτήτως ταξινομητή με την πλειονότητα αυτών να δίδουν ποσοστά της τάξης του 40-50%. Αξίζει να αναφέρουμε πως στα πλαίσια του έργου διαθέτουμε εξ αρχής τις ακριβείς συντεταγμένες όλων των καρκινικών περιοχών με βάση τα προτεινόμενα κλινικά όρια και σκοπός ήταν ο έλεγχος επαρκούς στατιστικής διακριτικής ικανότητας.

Τα παραγόμενα γραφήματα από τα μοντέλα της βιβλιοθήκης scikit learn εμφανίζουν σταθερή απόδοση ως προς την χρήση ταξινομητή με κυρίαρχη την μονομεταβλητή (univariate) μέθοδο που χρησιμοποιεί στην αρχή ένα στατιστικό μοντέλο (π.χ Anova) πριν την χρήση του ταξινομητή - επιλογή χαρακτηριστικών. Η μέθοδος αυτή βιβλιογραφικά αποδεικνύεται πως αυξάνει την απόδοση του μοντέλου γεγονός που επαληθεύτηκε και πειραματικά. Ακολουθούν οι αλγόριθμοι αναδρομικής επιλογής χαρακτηριστικών καθώς και ταξινομητές - επιλογείς χαρακτηριστικών των οποίων η απόδοση κυμαίνεται γύρω στο 70-80% ανάλογα με την υπό εξέταση ασθενή ενώ όλα τα μοντέλα δίδουν σε ορισμένες περιπτώσεις αποδόσεις άνω του 90%.

Στην περίπτωση του εργαλείου colin - mico εφαρμόζονται αλγόριθμοι βασισμένοι στην θεωρία της πληροφορίας. Παρατηρούμε πως τα αποτελέσματα που λαμβάνουμε είναι σταθερά υψηλά ειδικά στις περιπτώσεις εμπρόσθιας επιλογής χαρακτηριστικών με ποσοστά άνω του 70-80% και για ορισμένες ασθενείς άνω του 90%. Ακολουθούν οι υλοποιήσεις αναδρομικής οπίσθιας επιλογής χαρακτηριστικών με ανάλογα υψηλά ποσοστά. Η μέθοδος Minimum Redundancy Maximum Relevance (mRMR) στα πλαίσια της εμπρόσθιας αναζήτησης εμφανίζει υψηλή απόδοση η οποία κυμαίνεται άνω του 90% σε όλες τις περιπτώσεις ασθενών (επαναληψιμότητα). Η παρατήρηση αυτή ταυτίζεται με πληθώρα πειραματικών αποτελεσμάτων από βιβλιογραφικές πηγές και συνδέεται με την τάση του αλγόριθμου να επιλέγει ένα υποσύνολο χαρακτηριστικών που έχουν τη μεγαλύτερη συσχέτιση με μια κατηγορία (συνάφεια) και τη μικρότερη συσχέτιση μεταξύ τους (πλεονασμός) [97]. Αντίστοιχα πολύ υψηλή είναι και η απόδοση της μεθόδου Joint Mutual Information Maximisation.

7.1.3 Ταξινομητές

Στην περίπτωση των ταξινομητών παρατηρούμε πως συνολικά για όλες τις κατηγορίες αλγορίθμων επιλογής χαρακτηριστικών οι ταξινομητές AdaBoost και GradientBoosting εμφανίζουν την υψηλότερη απόδοση με σταθερά υψηλά ποσοστά άνω του 80-90%. Οι αλγόριθμοι αυτοί ανήκουν στην κατηγορία των μοντέλων συνόλων και δη στις μεθόδους ενίσχυσης. Ο συνδυασμός αρκετών αδύναμων μοντέλων για την δημιουργία ενός ισχυρού συνόλου δείχνει

να ταξινομεί με μεγάλη συνέπεια και αποδοτικότητα τις επιμέρους καρκινικές επικινδυνότητες [47]. Ανάλογα υψηλά ποσοστά δίδουν και τα δένδρα αποφάσεων στην πλειονότητα των περιπτώσεων. Αντίθετα παρατηρούμε μειωμένη απόδοση και μεταβλητή συμπεριφορά στην λειτουργία των πολυεπίπεδων νευρωνικών δικτύων καθώς και των αλγορίθμων Bagging και RandomForest, των οποίων η τυχαία επιλογή υποσυνόλων εντός του συνόλου δεδομένων οδηγεί σε λανθασμένη κατηγοριοποίηση των δειγμάτων. Η λογιστική παλινδρόμηση με ποινή είτε l_2 (RidgeRegression) είτε l_1 (Lasso) φαίνεται να λειτουργούν αποτελεσματικά στα μικρά σύνολα δεδομένων που διαχειριζόμαστε, αποβάλλοντας χαρακτηριστικά που δεν διαφοροποιούνται από περιοχή σε περιοχή. Τέλος οι μηχανές διανυσμάτων υποστήριξης και ιδίως ο γραμμικός πυρήνας παρουσιάζει αποδόσεις της τάξης του 80-90% γεγονός που πιθανώς να συνηγορεί πως τα δεδομένα είναι γραμμικά διαχωρίσιμα.

7.2 Περιορισμοί

Κατά την εκπόνηση αυτής της εργασίας, ήρθαμε αντιμέτωποι με αρκετούς περιορισμούς, που αφορούν τόσο συνολικά το έργο όσο και ειδικά την προτεινόμενη υλοποίηση.

7.2.1 Φύση των δεδομένων

Στα πλαίσια του έργου, συλλέχθηκε πλήθος απεικονιστικών δεδομένων για κάθε ασθενή. Τα προκύπτοντα ραδιώματα είναι αποτέλεσμα της εξαγωγής και σύνθεσης των πληροφοριών από τους διάφορους τύπους δεδομένων ανά ασθενή, βάσει των οποίων πραγματοποιήθηκε η στατιστική ανάλυση και η ανάπτυξη του απαιτούμενου προγνωστικού μοντέλου. Το πρωτόκολλο που εφαρμόστηκε στους ασθενείς είναι πειραματικό και η φύση του είναι διερευνητική σκοπεύοντας στην εύρεση, κατά το δυνατόν, βέλτιστων ορίων των περιοχών καρκινικής επικινδυνότητας εντός των βλαβών. Συγκεκριμένα ο ορισμός των κατωφλίων για τα όρια των περιοχών στις περιπτώσεις των πρώτων ασθενών (Ασθενής 6, Ασθενής 12) δεν ήταν σαφής αλλά δυναμικός και σταδιακά από την ασθενή 16 και έκτοτε ξεκίνησε η συστηματοποίηση του. Σκοπός της παρούσας εργασίας είναι να αξιολογήσει υπολογιστικά τα προτεινόμενα κλινικά όρια για το αν διακρίνουν ικανοποιητικά τις περιοχές ενδιαφέροντος και σε περίπτωση που απαιτηθεί να γίνει η αναπροσαρμογή των ορίων ή ακόμα και η θέσπιση δύο αντί τριών περιοχών καρκινικής επικινδυνότητας. Σε κάθε περίπτωση η εργασία αυτή μπορεί να χρησιμοποιηθεί μελλοντικά για τον έλεγχο νέων προτεινόμενων κλινικών ορίων.

7.2.2 Υπολογιστική διαδικασία

Το πρόβλημα της ταξινόμησης ιατρικών δεδομένων, βασίζεται σε τεχνικές μηχανικής μάθησης η λειτουργία των οποίων εμφανίζει δομικούς περιορισμούς. Ειδικότερα, για την ευρωστία και επαναληψιμότητα του συστήματος είναι θεμιτή αν όχι απαραίτητη η χρήση μεγάλων συνόλων δεδομένων διαφορετικών ασθενών για να γίνει αποδοτικότερα η εκπαίδευση των μοντέλων. Για να αντιμετωπίσουμε το ζήτημα αυτό οδηγηθήκαμε σε επαύξηση των δεδομένων με την χρήση στατιστικών φίλτρων και μετασχηματίζοντας τις αρχικές εικόνες. Ωστόσο καθότι το έργο βρίσκεται ακόμα σε εξέλιξη νέα δεδομένα προκύπτουν τα οποία θα βελτιώσουν δυνητικά την γνώση μας τόσο για την λειτουργία του συστήματος υποστήριξης όσο και

την ακρίβεια των κλινικών ορίων. Παράλληλα ένα συχνό φαινόμενο συστημάτων που βασίζονται σε τεχνικές μηχανικής μάθησης είναι η υπερπροσαρμογή (over-fitting). Συγκεκριμένα, λέγοντας υπερπροσαρμογή εννοούμε το φαινόμενο κατά το οποίο η τεχνική κατηγοριοποίηση ταιριάζει ακριβώς στα δεδομένα εκπαίδευσης και ίσως να μη μπορεί να εφαρμοστεί σε πιο ευρύ πληθυσμό δεδομένων. Για παράδειγμα, ας υποθέσουμε ότι τα δεδομένα εκπαίδευσης περιέχουν λανθασμένα δεδομένα ή δεδομένα με θόρυβο ή μη ισορροπημένα, ως προς τις τρεις κατηγορίες, δεδομένα. Σε αυτή την περίπτωση, το ακριβές ταίριασμα των δεδομένων δεν είναι επιθυμητό. Για να αποφύγουμε την υπερπροσαρμογή [98] :

- Κανονικοποιήσαμε εξ αρχής τα δεδομένα σε νέα εύρη τιμών κάνοντας χρήση της συνάρτησης StandardScaler αφαιρώντας από κάθε χαρακτηριστικό την μέση τιμή και διαιρώντας με την τυπική απόκλιση.
- Αφαιρέσαμε χαρακτηριστικά με επαναλαμβανόμενες σταθερές τιμές ή χαρακτηριστικά με ελλειπείς τιμές.
- Επαυξήσαμε το σύνολο δεδομένων κατά δέκα φορές με τον μετασχηματισμό των αρχικών εικόνων.
- Η προγνωστική ακρίβεια όλων των μοντέλων αποτυπώθηκε μετά από διαδοχικές συνεχείς εκτελέσεις ενώ σε όλες τις περιπτώσεις διενεργήθηκε διασταυρούμενη επαλήθευση της ακρίβειας δέκα πτυχών (10-fold, cross validation).
- Χρησιμοποιήσαμε μεταξύ άλλων μοντέλα συνόλων ως ταξινομητές τα οποία μειώνουν την υπερπροσαρμογή και αυξάνουν την προγνωστική ευελιξία των μοντέλων.

Παρόλα αυτά, η υπερπροσαρμογή είναι ένας δομικός περιορισμός και πρέπει να ληφθεί υπόψιν στην αξιολόγηση του συστήματος.

7.3 Μελλοντικές Προεκτάσεις

Το σύστημα που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής εργασίας θα μπορούσε να βελτιωθεί και να επεκταθεί περαιτέρω, τουλάχιστον ως προς τρεις κατευθύνσεις. Συγκεκριμένα, αναφέρονται τα ακόλουθα :

- Ενσωμάτωση περισσότερων κλινικών δειγμάτων από διαφορετικές ασθενείς συμβάλλοντας στην ευρωστία και επαναληψιμότητα του συστήματος.
- Ορισμός ακριβέστερων κλινικών ορίων των περιοχών επικινδυνότητας.
- Χρήση μοντέλων βαθιάς μάθησης στην εξαγωγή χαρακτηριστικών για την περαιτέρω επαύξηση του συνόλου δεδομένων (data augmentation). Η χρησιμότητα των μεθόδων βαθιάς μάθησης για την αυτοματοποιημένη αναγνώριση και τμηματοποίηση καρκινικών αλλοιώσεων στην θεραπευτική ακτινολογία αποτελεί ερευνητικό αντικείμενο με μεγάλες προοπτικές διερεύνησης [99].

- Χρήση μοντέλων επιλογής χαρακτηριστικών για την συστηματοποίηση του βέλτιστου αριθμού επιλεγμένων χαρακτηριστικών, με στόχο αφενός την μείωση διαστατικότητας του συστήματος και αφετέρου την εγκυρότερη προγνωστική ακρίβεια.

Βιβλιογραφία

- [1] George Kolostoumpis και Makrygiannaki Kleanthi. *Clinical Decision Support Systems: A Useful Tool in Clinical Practice*. *Health Science Journal*, 4:9-12, 2012.
- [2] Reed Sutton, David Pincock, Daniel Baumgart, Daniel Sadowski, Richard Fedorak και Karen Kroeker. *An overview of clinical decision support systems: benefits, risks, and strategies for success*. 3, 2020.
- [3] Stephen A. Cannistra. *Cancer of the Ovary*. *New England Journal of Medicine*, 351(24):2519-2529, 2004. PMID: 15590954.
- [4] Permuth Jennifer B Sellers Thomas A Reid, Brett M. *Epidemiology of ovarian cancer: a review*. 4(1):9-32, 2017.
- [5] JA Ledermann, FA Raja, C Fotopoulou, A Gonzalez-Martin, N Colombo, C Sessa και ESMO Guidelines Working Group. *Newly diagnosed and relapsed epithelial ovarian carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. *Annals of oncology : official journal of the European Society for Medical Oncology*, 24 Suppl 6:vi24-32, 2013.
- [6] Claudia Allemani, Hannah K Weir, Helena Carreira, Rhea Harewood, Devon Spika, Xiao Si Wang, Finian Bannon, Jane V Ahn, Christopher J Johnson, Audrey Bonaventure, Rafael Marcos-Gragera, Charles Stiller, GulnarAzevedo e Silva, Wan Qing Chen, Olufemi J Ogunbiyi, Bernard Rachet, Matthew J Soeberg, Hui You, Tomohiro Matsuda, Magdalena Bielska-Lasota, Hans Storm, Thomas C Tucker και Michel P Coleman. *Global surveillance of cancer survival 1995-2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2)*. *The Lancet*, 385(9972):977-1010, 2015.
- [7] Martin Pölcher, Steffen Hauptmann, Christina Fotopoulou, Barbara Schmalfeldt, Ivo Meinhold-Heerlein, Alexander Mustea, Ingo Runnebaum και Jalid Sehoul. *Opportunistic salpingectomies for the prevention of a high-grade serous carcinoma: a statement by the Kommission Ovar of the AGO*. *Archives of Gynecology and Obstetrics*, 292(1):231-234, 2015.
- [8] R Tyler Hillman, Gary B Chisholm, Karen H Lu και P Andrew Futreal. *Genomic Rearrangement Signatures and Clinical Outcomes in High-Grade Serous Ovarian Cancer*. *JNCI: Journal of the National Cancer Institute*, 110(3):265-272, 2018.
- [9] The Cancer Genome Atlas Research Network. *Integrated genomic analyses of ovarian carcinoma*. *Nature*, 474(7353):609-615, 2011.

- [10] Barbara M. Norquist, Maria I. Harrell, Mark F. Brady, Tom Walsh, Ming K. Lee, Suleyman Gulsuner, Sarah S. Bernards, Silvia Casadei, Qian Yi, Robert A. Burger, John K. Chan, Susan A. Davidson, Robert S. Mannel, Paul A. DiSilvestro, Heather A. Lankes, Nilsa C. Ramirez, Mary Claire King, Elizabeth M. Swisher και Michael J. Birrer. *Inherited Mutations in Women With Ovarian Carcinoma*. *JAMA Oncology*, 2(4):482-490, 2016.
- [11] Kathryn P. Pennington, Tom Walsh, Maria I. Harrell, Ming K. Lee, Christopher C. Pennil, Mara H. Rendi, Anne Thornton, Barbara M. Norquist, Silvia Casadei, Alexander S. Nord, Kathy J. Agnew, Colin C. Pritchard, Sheena Scroggins, Rochelle L. Garcia, Mary Claire King και Elizabeth M. Swisher. *Germline and Somatic Mutations in Homologous Recombination Genes Predict Platinum Response and Survival in Ovarian, Fallopian Tube, and Peritoneal Carcinomas*. *Clinical Cancer Research*, 20(3):764-775, 2014.
- [12] Jonathan S. Berek, Sean T. Kehoe, Lalit Kumar και Michael Friedlander. *Cancer of the ovary, fallopian tube, and peritoneum*. *International Journal of Gynecology & Obstetrics*, 143(Σ2):59-78, 2018.
- [13] Penelope Moyle, Helen Addley και Evis Sala. *Radiological Staging of Ovarian Carcinoma*. *Seminars in ultrasound, CT, and MR*, 31:388-98, 2010.
- [14] Rosemarie Forstner. *Radiological staging of ovarian cancer: Imaging findings and contribution of CT and MRI*. *European radiology*, 17:3223-35, 2008.
- [15] SA Sohaib, T Mills, Anju Sahdev, Judith Webb, P Vantrappen, Ian Jacobs και R Reznick. *The role of magnetic resonance imaging and ultrasound in women with adnexal masses*. *Clinical radiology*, 60:340-8, 2005.
- [16] S. Nougaret, M. Tardieu, H.A. Vargas, Caroline Md, Saskia Vande Perre, N. Bonanno, E. Sala και I. Thomassin-Naggara. *Ovarian cancer: An update on imaging in the era of radiomics*. *Diagnostic and Interventional Imaging*, 100, 2018.
- [17] Evis Sala, Masako Kataoka, Andrew Priest, Andrew Gill, Mary Mclean, Ilse Joubert, Martin Graves, Robin Crawford, Mercedes Jimenez-Linan, Helena Earl, Charlotte Hodgkin, John Griffiths, David Lomas και James Brenton. *Advanced Ovarian Cancer: Multiparametric MR Imaging Demonstrates Response- and Metastasis-specific Effects*. *Radiology*, 263:149-59, 2012.
- [18] Hebert Vargas, Maura Miccò, Seong Hong, Debra Goldman, Fanny Dao, Britta Weigelt, Robert Soslow, Hedvig Hricak, Douglas Levine και Evis Sala. *Association between Morphologic CT Imaging Traits and Prognostically Relevant Gene Signatures in Women with High-Grade Serous Ovarian Cancer: A Hypothesis-generating Study*. *Radiology*, 274:141477, 2014.
- [19] Stephanie Nougaret, Yulia Lakhman, Mithat Gönen, Debra Goldman, Maura Miccò, Melvin D'Anastasi, Sarah Johnson, Krishna Juluru, Angela Arnold, Ramon Sosa,

- Robert Soslow, Hebert Vargas, Hedvig Hricak, Noah Kauff και Evis Sala. *High-Grade Serous Ovarian Cancer: Associations between BRCA Mutation Status, CT Imaging Phenotypes, and Clinical Outcomes*. *Radiology*, 285:161697, 2017.
- [20] Hebert Vargas, Harini Veeraraghavan, Maura Miccò, Stephanie Nougaret, Yulia Lakhman, Andreas Meier, Ramon Sosa, Robert Soslow, Douglas Levine, Britta Weigelt, Carol Aghajanian, Hedvig Hricak, Joseph Deasy, Alexandra Snyder και Evis Sala. *A novel representation of inter-site tumour heterogeneity from pre-treatment computed tomography textures classifies ovarian cancers by clinical outcome*. *European Radiology*, 27:1–11, 2017.
- [21] Stefania Rizzo, Francesca Botta, Sara Raimondi, Daniela Origgi, Valentina Buscarino, Anna Colarieti, Federica Tomao, Giovanni Aletti, Vanna Zanagnolo, Maria Grande, Nicoletta Colombo και Massimo Bellomi. *Radiomics of high-grade serous ovarian cancer: association between quantitative CT features, residual tumour and disease progression within 12 months*. *European Radiology*, 28, 2018.
- [22] Miriam Mvan Heeswijk, Doenja M J Lambregts, Joost J Mvan Griethuysen, Stanley Oei, Sheng Xiang Rao, Carla A Mde Graaff, Roy F A Vliegen, Geerard L Beets, Nikos Papanikolaou και Regina G H Beets-Tan. *Automated and Semiautomated Segmentation of Rectal Tumor Volumes on Diffusion-Weighted MRI: Can It Replace Manual Volumetry?* *International journal of radiation oncology, biology, physics*, 94(4):824–831, 2016.
- [23] Philipp Lohmann, Khaled Bousabarah, Mauritius Hoevens και Harald Treuer. *Radiomics in radiation oncology—basics, methods, and limitations*. *Strahlentherapie und Onkologie*, 196, 2020.
- [24] Philipp Lohmann, Martin Kocher, G Ceccon, EK Bauer, G Stoffels, S Viswanathan, Maximilian Ruge, Bernd Neumaier, N. Shah, GR Fink, KJ Langen και Norbert Gall-diks. *Combined FET PET/MRI radiomics differentiates radiation injury from recurrent brain metastasis*. τόμος 58, 2019.
- [25] Mohammadreza Soltaninejad, Guang Yang, Tryphon Lambrou, Nigel Allinson, Timothy Jones, Thomas Barrick, Franklyn Howe και Xujiong Ye. *Supervised Learning based Multimodal MRI Brain Tumour Segmentation using Texture Features from Supervoxels*. *Computer Methods and Programs in Biomedicine*, 157, 2018.
- [26] Alex Zwanenburg, Martin Vallières, Mahmoud Abdalah, Hugo Aerts, Vincent Andrearczyk, Aditya Apte, Saeed Ashrafinia, Spyridon Bakas, Roelof Beukinga, Ronald Boellaard, Marta Bogowicz, Luca Boldrini, Irene Buvat, Gary Cook, Christos Davatzikos, Adrien Depeursinge, Marie Charlotte Desseroit, Nicola Dinapoli, Cuong Dinh και Steffen Löck. *The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping*. *Radiology*, 295:191145, 2020.

- [27] Russell Shinohara, Elizabeth Sweeney, Jeff Goldsmith, Navid Shiee, Farrah Mateen, Peter Calabresi, Samson Jarso, Dzung Pham, Daniel Reich και Ciprian Crainiceanu. *Statistical normalization techniques for magnetic resonance imaging*. *NeuroImage. Clinical*, 6:9–19, 2014.
- [28] Benjamin M. Ellingson, Taryar Zaw, Timothy F. Cloughesy, Kouros M. Naeini, Shadi Lalezari, Sandy Mong, Albert Lai, Phioanh L. Nghiemphu και Whitney B. Pope. *Comparison between intensity normalization techniques for dynamic susceptibility contrast (DSC)-MRI estimates of cerebral blood volume (CBV) in human gliomas*. *Journal of Magnetic Resonance Imaging*, 35(6):1472–1477, 2012.
- [29] Virendra Kumar, Yuhua Gu, Satrajit Basu, Anders Berglund, Steven Eschrich, Matthew Schabath, Kenneth Forster, Hugo Aerts, André Dekker, David Fenstermacher, Dmitry Goldgof, Lawrence Hall, Philippe Lambin, Yoganand Balagurunathan, Robert Gatenby και Robert Gillies. *Radiomics: The process and the challenges*. *Magnetic resonance imaging*, 30:1234–48, 2012.
- [30] Donghui Xu, Arati S. Kurani, Jacob D. Furst και Daniela S. Raicu. *Run-length Encoding for Volumetric Texture*. *The 4th IASTED International Conference on Visualization, Imaging, and Image Processing*, 2004.
- [31] Philippe Lambin, Ralph Leijenaar, Timo Deist, Jurgen Peerlings, Evelynde Jong, Janita Van Timmeren, Sebastian Sanduleanu, Ruben Larue, Aniek Even, Arthur Jochems, Yvonka Wijk, Henry Woodruff, Johan Soest, Tim Lustberg, Erik Roelofs, Wouter Elmpt, André Dekker, Felix Mottaghy, Joachim Wildberger και Seán Walsh. *Radiomics: The bridge between medical imaging and personalized medicine*. *Nature Reviews Clinical Oncology*, 14, 2017.
- [32] M. Amadasun και R. King. *Textural features corresponding to textural properties*. *Trans. Syst. Man. Cybern.*, 19:1264–1274, 1998.
- [33] Robert Haralick, K. Shanmugam και Itshak Dinstein. *Haralick RM, Shanmuga K, Dinstein I* *Textural features for image classification*. *IEEE Trans Syst Man Cybern 3: 610-621*. *Systems, Man and Cybernetics, IEEE Transactions on*, ΣΜ³:610 – 621, 1973.
- [34] Guillaume Thibault, Bernard FERTIL, Claire Navarro, Sandrine Pereira, Nicolas Lévy, Jean Sequeira και Jean Luc MARI. *Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification*. 2009.
- [35] Chengjun Sun και William Wee. *Neighboring Gray Level Dependence Matrix for Texture Classification*. *Computer Vision, Graphics, and Image Processing*, 23:341–352, 1983.
- [36] M. Zhou, J. Scott, Baishali Chaudhury, L. Hall, Dmitry Goldgof, K.W. Yeom, Michael Iv, Yangming Ou, Jayashree Kalpathy-Cramer, S. Napel, Robert Gillies, Olivier

- Gevaert και Robert Gatenby. *Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches*. *American Journal of Neuroradiology*, 39, 2017.
- [37] Alex Krizhevsky, I Sutskever και G Hinton. *Imagenet classification with deep convolutional neural networks*. σελίδες 1097–1105, 2012.
- [38] A Selvapandian και K Manivannan. *Fusion Based Glioma Brain Tumor Detection and Segmentation using ANFIS Classification*. *Computer Methods and Programs in Biomedicine*, 166, 2018.
- [39] Farzad Khalvati, Junjie Zhang, Audrey Chung, Mohammad Javad Shafiee, Alexander Wong και Masoom Haider. *MPCaD: A multi-scale radiomics-driven framework for automated prostate cancer localization and detection*. *BMC Medical Imaging*, 18, 2018.
- [40] Yoonkyung Cha, WON JANG, MI SOOK KIM, HYUNG YOO, Eun Kyung Paik, HEE JEONG και SANG MIN YOUN. *Prediction of Response to Stereotactic Radiosurgery for Brain Metastases Using Convolutional Neural Networks*. *Anticancer Research*, 38:5437–5445, 2018.
- [41] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang και Chunfang Liu. *A Survey on Deep Transfer Learning: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III*, σελίδες 270–279. 2018.
- [42] Tom M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997.
- [43] Ian Goodfellow, Yoshua Bengio και Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [44] Julianna Czum. *Dive Into Deep Learning*. *Journal of the American College of Radiology*, 17, 2020.
- [45] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1η έκδοση, 2007.
- [46] Bernhard Schölkopf, Zhiyuan Luo και Vladimir Vovk. *Empirical inference. Festschrift in honor of Vladimir N. Vapnik*. 2013.
- [47] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Edouard Duchesnay και Gilles Louppe. *Scikit-learn: Machine Learning in Python*. *Journal of Machine Learning Research*, 12, 2012.
- [48] Leo Breiman. *Pasting Small Votes for Classification in Large Databases and On-Line*. *Machine Learning*, 36:85–103, 1999.
- [49] L. Breiman. *Bagging predictors" Machine Learning*. *Machine learning*, 24, 1996.

- [50] Tin Ho. *The Random Subspace Method for Constructing Decision Forests*. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20:832 – 844, 1998.
- [51] Toon Calders, Floriana Esposito, Eyke Hüllermeier και Rosa Meo. *Machine Learning and Knowledge Discovery in Databases*. 2014.
- [52] L. Breiman. *Random forests, machine learning 45*. *Journal of Clinical Microbiology*, 2:199-228, 2001.
- [53] Vikramaditya Jakkula. *Tutorial on Support Vector Machine (SVM)*. 2021.
- [54] Yu Shiang Chien. *Pattern classification and scene analysis*. *Automatic Control, IEEE Transactions on*, 19:462- 463, 1974.
- [55] Nello Christianini και John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel Based Learning Methods*. 2000.
- [56] L. Ladický και Philip Torr. *Locally linear support vector machines*. *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, σελίδες 985-992, 2011.
- [57] Quanshi Zhang, Yu Yang, Yuchen Liu, Yingnian Wu και Song Zhu. *Unsupervised Learning of Neural Networks to Explain Neural Networks*. 2018.
- [58] Konstantinos Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson και Dionysis Bochtis. *Machine Learning in Agriculture: A Review*. *Sensors*, 18:2674, 2018.
- [59] N. B. Amor, S. Benferhat και Zied Elouedi. *Qualitative classification and evaluation in possibilistic decision trees*. *2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542)*, 2:653-657 ολ.2, 2004.
- [60]
- [61] Daeseok Yoon. *A new method for the automated alignment of dental radiographs for digital subtraction radiography*. *Dento maxillo facial radiology*, 29:11-9, 2000.
- [62] Avril Behan. *2-D and 3-D Image Registration For Medical, Remote Sensing, and Industrial Applications*. *The Photogrammetric Record*, 21, 2006.
- [63] J.B.Antoine Maintz και Max Viergever. *A Survey of Medical Image Registration*. *Medical image analysis*, 2:1-36, 1998.
- [64] Barbara Zitová και Jan Flusser. *Image Registration Methods: A Survey*. *Image and Vision Computing*, 21:977-1000, 2003.
- [65] Carles Matabosch, Joaquim Salvi, Xavier Pinsach και Jordi Es. *A comparative survey on free-form surface registration*. 2004.
- [66] Mark Holden. *A Review of Geometric Transformations for Nonrigid Body Registration*. *IEEE transactions on medical imaging*, 27:111-28, 2008.

- [67] Rafael Gonzalez και Richard Woods. *Digital Image Processing (2nd Edition)*. 2002.
- [68] Roger Woods, Simon Cherry και John Mazziotta. *Rapid Automated Algorithm for Aligning and Reslicing PET Images*. *Journal of computer assisted tomography*, 16:620–33, 1992.
- [69] France Laliberté, Langis Gagnon και Yunlong Sheng. *Registration and fusion of retinal images - An evaluation study*. *IEEE transactions on medical imaging*, 22:661–73, 2003.
- [70] PeiFeng Zeng και Tomio Hirata. *Distance Map Based Enhancement for Interpolated Images*, σελίδες 429–449. 2003.
- [71] Michael Kass, Andrew Witkin και Demetri Terzopoulos. *Active contour models*. *Int J Comput Vis. International Journal of Computer Vision*, 1:321–331, 1988.
- [72] Alex James και Belur V Dasarathy. *Medical Image Fusion: A survey of the state of the art*. *Information Fusion*, 2013.
- [73] Tomas Akenine-Möller και Eric Haines. *Real-time rendering (2. ed.)*. 2002.
- [74] Elliot K. Fishman, Derek R. Ney, David G. Heath, Frank M. Corl, Karen M. Horton και Pamela T. Johnson. *Volume Rendering versus Maximum Intensity Projection in CT Angiography: What Works Best, When, and Why*. *RadioGraphics*, 26(3):905–922, 2006. ΠΜΙΔ: 16702462.
- [75] Stephen Yip, Ying Liu, Chintan Parmar, Qian Li, Shichang Liu, Fangyuan Qu, Zhao Xiang Ye, Robert Gillies και Hugo Aerts. *Associations between radiologist-defined semantic and automatically computed radiomic features in non-small cell lung cancer*. *Scientific Reports*, 7, 2017.
- [76] Philippe Lambin, Emmanuel Rios Velazquez, Ralph Leijenaar, Sara Carvalho, Ruud Stiphout, Patrick Granton, Karen Zegers, Robert Gillies, Ronald Boellaard, André Dekker και Hugo Aerts. *Radiomics: Extracting more information from medical images using advanced feature analysis*. *European journal of cancer (Oxford, England : 1990)*, 48:441–6, 2012.
- [77] Robert Gillies, Paul Kinahan και Hedvig Hricak. *Radiomics: Images Are More than Pictures, They Are Data*. *Radiology*, 278:151169, 2015.
- [78] Joost J.M.van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina G.H. Beets-Tan, Jean Christophe Fillion-Robin, Steve Pieper και Hugo J.W.L. Aerts. *Computational Radiomics System to Decode the Radiographic Phenotype*. *Cancer Research*, 77(21):ε104–ε107, 2017.
- [79] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake Vanderplas, Arnaud Joly, Brian Holt και Gael Varoquaux. *API*

- design for machine learning software: Experiences from the scikit-learn project. API Design for Machine Learning Software: Experiences from the Scikit-learn Project*, 2013.
- [80] Isabelle Guyon, Jason Weston, Stephen Barnhill και Vladimir Vapnik. 2002.
- [81] Annapurna Samantary και Satya Dash. *Feature Selection Techniques to Predict the Religion of a Country from Its Flag*, σελίδες 191–201. 2020.
- [82] Alexander Kraskov, Harald Stögbauer και Peter Grassberger. *Erratum: Estimating mutual information [Phys. Rev. E 69, 066138 (2004)]*. *Physical Review E*, 83, 2011.
- [83] Dahua Lin και Xiaoou Tang. *Conditional Infomax Learning: An Integrated Framework for Feature Extraction and Fusion*. σελίδες 68–82, 2006.
- [84] Daphne Koller και Mehran Sahami. *Toward Optimal Feature Selection. Proceedings of the 13th International Conference on Machine Learning (ICML-1996)*, 96, 2000.
- [85] Patrick Meyer, Colas Schretter και Gianluca Bontempi. *Information-Theoretic Feature Selection in Microarray Data Using Variable Complementarity. Selected Topics in Signal Processing, IEEE Journal of*, 2:261 – 274, 2008.
- [86] Adjimi Ahlem, Abdenour Hacine-Gharbi, Philippe Ravier και Messaoud Mostefai. *Extraction and selection of binarised statistical image features for fingerprint recognition. International Journal of Biometrics*, 9:67, 2017.
- [87] Krzysztof Cios και Mark Shields. *Advances in neural information processing systems 7: By G. Tesauro, D. S. Touretzky and T. K. Leen (Eds.)*, MIT Press, Cambridge, MA, 1995, ISBN 0-262-20104-6, 1143 pp. *Neurocomputing*, 16:263, 1997.
- [88] Zheng Zhao και Huan Liu. *Spectral feature selection for supervised and unsupervised learning*. τόμος 227, σελίδες 1151–1157, 2007.
- [89] Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P. Trevino, Jiliang Tang και Huan Liu. *Feature Selection: A Data Perspective*. 50(6), 2017.
- [90] Marko Robnik-Šikonja και Igor Kononenko. *[No title found]. Machine Learning*, 53(1/2):23–69, 2003.
- [91] Yi Yang, Heng Shen, Zhigang Ma, Zi Huang και Xiaofang Zhou. *L21-norm regularized discriminative feature selection for unsupervised learning*. σελίδες 1589–1594, 2011.
- [92] Balaji Krishnapuram, Andrew Tomkins και Qiang Yang. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010*. 2010.
- [93] Z. Li, Y. Yang, Jiuru Liu, X. Zhou και H. Lu. *Unsupervised feature selection using nonnegative spectral analysis. Proceedings of the National Conference on Artificial Intelligence*, 2:1026–1032, 2012.

- [94] Pengwei Yang, Ryan Abo, Chang Liu, Zehua Chen, Haiguo Wu, Jike Cui, Chandri Yandava, Shannon Bailey, Curt Balch, Jeffery Gulcher και Thomas Chittenden. *Abstract 4539: Novel feature selection strategies for enhanced predictive modeling and deep learning in the biosciences*. *Cancer Research*, 77:4539–4539, 2017.
- [95] Kyle Noll, David Sabsevitz, Sujit Prabhu και Jeffrey Wefel. *Neuropsychology in the Neurosurgical Management of Primary Brain Tumors*. *Neurosurgical Neuropsychology*, σελίδες 157–183. Elsevier, 2019.
- [96] *MICO*. <http://jupiters1117.github.io/mico/>.
- [97] J. Berrendero, Antonio Cuevas και José Torrecilla. *The mRMR variable selection method: a comparative study for functional data*. *Journal of Statistical Computation and Simulation*, 86:1–17, 2015.
- [98] Xue Ying. *An Overview of Overfitting and its Solutions*. *Journal of Physics: Conference Series*, 1168:022022, 2019.
- [99] Vishwa Parekh και Michael Jacobs. *Deep learning and radiomics in precision medicine*. *Expert Review of Precision Medicine and Drug Development*, 4:1–14, 2019.