



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Τομέας Συστημάτων Μετάδοσης Πληροφορίας Τεχνολογίας Υλικών
Εργαστήριο Βιοϊατρικών Προσομοιώσεων και Απεικονιστικής Τεχνολογίας (BIOSIM)

Ανάπτυξη ερμηνεύσιμων μοντέλων μηχανικής μάθησης με
σκοπό την εκτίμηση της επικινδυνότητας αθηρωματικών
πλακών σε ασθενείς με καρωτιδική νόσο

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Παναγιώτα Καρατζά

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

Αθήνα, Νοέμβριος 2021



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Τομέας Συστημάτων Μετάδοσης Πληροφορίας

Τεχνολογίας Υλικών

Εργαστήριο Βιοϊατρικών Προσομοιώσεων και Απεικονιστικής Τεχνολογίας (BIOSIM)

Ανάπτυξη ερμηνεύσιμων μοντέλων μηχανικής μάθησης με σκοπό την εκτίμηση της επικινδυνότητας αθηρωματικών πλακών σε ασθενείς με καρωτιδική νόσο

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Παναγιώτα Καρατζά

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

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 4^η Νοεμβρίου 2021.

.....
Κωνσταντίνα Σ. Νικήτα
Καθηγήτρια ΕΜΠ

.....
Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής ΕΜΠ

.....
Γεώργιος Στάμου
Καθηγητής ΕΜΠ

Αθήνα, Νοέμβριος 2021.

.....

Παναγιώτα Καρατζά

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

© Παναγιώτα Καρατζά, 2021

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Το εγκεφαλικό επεισόδιο συνεχίζει να αποτελεί μία σημαντική αιτία νοσηρότητας και θνησιμότητας σε παγκόσμιο επίπεδο. Περίπου το 85% των εγκεφαλικών επεισοδίων είναι ισχαιμικά από τα οποία τουλάχιστον το 20–30% μπορούν να αποδοθούν στην καρωτιδική νόσο. Καρωτιδική νόσος ονομάζεται η στένωση ή η απόφραξη των καρωτιδικών αρτηριών. Την αιτιολογία της καρωτιδικής νόσου κατά κύριο λόγο και σε ποσοστό 90% αποτελεί η αθηροσκλήρωση, δηλαδή η δημιουργία αθηρωματικής πλάκας στον έσω χιτώνα των αρτηριών. Η κλινική εκδήλωση της καρωτιδικής νόσου μπορεί να είναι είτε συμπτωματική είτε ασυμπτωματική. Πιο συγκεκριμένα στα αρχικά της στάδια η καρωτιδική νόσος είναι ασυμπτωματική, κατά την εξέλιξή της όμως το πρώτο σύμπτωμα μπορεί να είναι ένα εγκεφαλικό ή ένα παροδικό ισχαιμικό επεισόδιο. Προκειμένου να προβλεφθεί έγκαιρα και να αποτραπεί μία τέτοια κατάσταση, διερευνάται, στο πλαίσιο της παρούσας διπλωματικής εργασίας, η δημιουργία ερμηνεύσιμων μοντέλων βαθιάς μάθησης, ικανά να κατηγοριοποιήσουν τους ασθενείς ανάλογα με τον τρόπο θεραπείας που θα πρέπει να ακολουθήσουν. Η κατηγοριοποίηση των ασθενών γίνεται σε δύο κλάσεις: στην κλάση χαμηλού κινδύνου όπου οι ασθενείς καλούνται να ακολουθήσουν συντηρητική αγωγή και στην κλάση υψηλού κινδύνου όπου οι ασθενείς καλούνται να ακολουθήσουν επεμβατική θεραπεία. Για το σκοπό αυτό χρησιμοποιήθηκαν υπερηχογραφικές απεικονιστικές καταγραφές β-σάρωσης των καρωτιδικών αρτηριών σε διαμήκη τομή καθώς και ένα διάλυμα 13 διαφορετικών αιματολογικών και δημογραφικών χαρακτηριστικών από 51 ασθενείς με καρωτιδική νόσο. Για τις ανάγκες της εργασίας, αρχικά υλοποιήθηκαν δύο διαφορετικά μοντέλα : ένα απλό νευρωνικό δίκτυο και ένα συνελικτικό νευρωνικό δίκτυο. Το πρώτο μοντέλο δέχεται ως είσοδο τα 13 κλινικά χαρακτηριστικά και το δεύτερο τις απεικονιστικές υπερηχογραφικές καταγραφές. Για την αποτελεσματική διαχείριση της μη ισορροπημένης φύσης του συνόλου δεδομένων επιλέξαμε να χρησιμοποιήσουμε την τεχνική της συλλογικής μάθησης κατά την δημιουργία και των δύο μοντέλων. Στη συνέχεια τα μοντέλα εκπαιδεύτηκαν και αξιολογήθηκαν με τη μέθοδο της διασταυρωμένη επικύρωσης, επιτυγχάνοντας αξιόλογα αποτελέσματα. Κατόπιν, τρεις διαφορετικές μέθοδοι ερμηνευσιμότητας και πιο συγκεκριμένα οι Global Surrogate, Individual Conditional Expectation και Shapley Va-lues χρησιμοποιήθηκαν με σκοπό να ερμηνεύσουν τις προβλέψεις των μοντέλων. Πειραματικά οι μέθοδοι ερμηνευσιμότητας ανέδειξαν πως τα χαρακτηριστικά HDL και LDL χοληστερόλη, τριγλυκερίδια, γλυκόζη αίματος και ουρία επηρεάζουν σε σημαντικό βαθμό τις τελικές προβλέψεις του απλού συλλογικού μοντέλου. Αντίστοιχα, για το συνελικτικό μοντέλο προέκυψε ότι τα εικονοστοιχεία με μικρότερη ενδιάμεση τιμή φωτεινότητας ωθούν το μοντέλο να κατατάξει τον ασθενή με μεγαλύτερη πιθανότητα στην κατηγορία υψηλού κινδύνου.

Λέξεις Κλειδιά: καρωτιδική νόσος, αθηρωμάτωση, μηχανική μάθηση, νευρωνικά δίκτυα, συνελικτικά δίκτυα, συλλογική μάθηση, ανισόρροπα δεδομένα, ερμηνευσιμότητα, Shapley values, ICE plots, Global Surrogate model.

Abstract

Stroke continues to be a major cause of morbidity and mortality worldwide. About 85% of strokes are ischemic, and at least 20 – 30% of them can be attributed to carotid artery disease. Carotid artery disease occurs when the blood vessels that deliver blood to your brain and head, narrow or clog. The main cause of carotid disease, at a rate of 90%, is atherosclerosis, ie the buildup of fats, cholesterol, and other substances in and on artery walls. The clinical form of the carotid disease can be either symptomatic or asymptomatic. More specifically, in its early stages, the carotid disease is asymptomatic, but as it grows, the first symptom may be a stroke or a transient ischemic attack. In order to timely predict and prevent such a situation, the development of interpretable deep learning models, capable of categorizing patients according to the treatment they should follow, is being investigated in this thesis. Patients have been categorized into two classes: the low-risk class where patients are required to follow conservative treatment and the high-risk class where patients are required to follow invasive treatment. For this purpose, ultrasound imaging recordings of a longitudinal section of the carotid arteries were used, as well as a vector of 13 features (demographic, blood biomarkers) from 51 patients with carotid disease. Two different models have been implemented: a simple neural network and a convolutional neural network. The first model takes as input the 13 clinical features and the second the images of the ultrasound recordings. In order to handle the unbalanced nature of the used dataset, we chose to use an ensemble learning technique for the creation of both models. The models were then trained and evaluated by the cross-validation method, achieving good performance. Then, three different interpretability techniques, namely Global Surrogate, Individual Conditional Expectation, and Shapley Values, were used to interpret the model predictions. The interpretability results have shown that the characteristics of HDL and LDL cholesterol, triglycerides, blood glucose, and urea significantly affect the final predictions of the simple ensemble model. Respectively, for the convolutional ensemble model, it turned out that the pixels with a lower median intensity value drive the model to classify the patient with a higher probability in the high-risk class.

Keywords: carotid artery disease, atheromatosis, machine learning, neural networks, convolution neural networks, ensemble learning, imbalanced data, interpret ability, Shapley values, ICE plots, Global Surrogate model

Ευχαριστίες

Ολοκληρώνοντας τη διπλωματική μου εργασία θα ήθελα να ευχαριστήσω θερμά την καθηγήτρια κα Κωνσταντίνα Νικήτα τόσο για την εμπιστοσύνη που μου έδειξε όσο και για την καθοδήγηση που μου παρείχε στα πρώτα μου βήματα στον τομέα της έρευνας.

Επίσης θα ήθελα να ευχαριστήσω την ερευνήτρια μεταδιδάκτορα κα Καλλιόπη Δαλακλείδη για τον χρόνο που αφιέρωσε ώστε να με βοηθήσει καθώς και για όλες τις πολύτιμες συμβουλές της όλον αυτόν τον καιρό.

Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου Γιάννα και Νίκο για την απόλυτη στήριξη και εμπιστοσύνη που μου έδειξαν όλο αυτό το διάστημα καθώς και τον Πάνο με τον οποίο μοιράστηκα όλες τις όμορφες αλλά και δύσκολες στιγμές των φοιτητικών μου χρόνων και αποτέλεσε τον καλύτερο συνοδοιπόρο που θα μπορούσα να έχω.

Περιεχόμενα

1	Καρωτιδική Νόσος	19
1.1	Αγγειακά Εγκεφαλικά Επεισόδια	20
1.2	Αρτηρίες	21
1.3	Ανατομία Καρωτιδικών Αρτηριών	21
1.4	Αθηρωσκήλυνση - Αθηρωματική Πλάκα	22
1.5	Καρωτιδική Νόσος	24
1.6	Παράγοντες κινδύνου εμφάνισης Καρωτιδικής Νόσου	25
1.7	Διάγνωση Καρωτιδικής Νόσου	26
1.8	Τρόποι αντιμετώπισης ΚΝ	28
2	Μηχανική μάθηση	31
2.1	Τεχνητή Νοημοσύνη και Μηχανική Μάθηση	31
2.1.1	Τεχνητή Νοημοσύνη	31
2.1.2	Μηχανική Μάθηση	31
2.2	Νευρωνικά Δίκτυα	32
2.2.1	Δομή Νευρωνικών Δικτύων	33
2.2.2	Εκπαίδευση Νευρωνικών Δικτύων	34
2.3	Συνελικτικά Νευρωνικά Δίκτυα	35
2.4	Μετρικές αξιολόγησης των μοντέλων μηχανικής μάθησης	36
2.5	Διασταυρωμένη Επικύρωση	38
2.6	Διαχείριση μη ισορροπημένων δεδομένων	38
2.6.1	Συλλογική Μάθηση	38
2.6.2	Εκπαίδευση με ευαισθησία κόστους	39
2.6.3	Υπερδειγματοληψία	40
2.6.4	Υποδειγματοληψία	40
2.7	Εφαρμογές της μηχανικής μάθησης στην καρωτιδική νόσο	40
3	Ερμηνευσιμότητα μηχανικής μάθησης	43
3.1	Χρησιμότητα της Ερμηνευσιμότητας	43
3.2	Τεχνικές Ερμηνευσιμότητας	44
3.2.1	Εγγενώς ερμηνεύσιμα μοντέλα	45
3.2.2	Model agnostic μέθοδοι	45
3.3	Εφαρμογές της ερμηνευσιμότητας στον κλάδο της βιοϊατρικής επιστήμης	47
4	Υλικό και Μέθοδοι	49
4.1	Σύνολο Δεδομένων	49
4.1.1	Περιγραφή υπερηχογραφικών απεικονίσεων β-σάρωσης των καρωτιδικών αρτηριών	50

4.1.2	Περιγραφή αριθμητικών κλινικών δεδομένων	51
4.1.3	Ιδιαιτερότητες του συνόλου δεδομένων	54
4.2	Μεθοδολογία κατασκευής αλγορίθμων μηχανική μάθησης	55
4.2.1	Σχήμα Συλλογικής Μάθησης	55
4.2.2	Συλλογικό Συνελικτικό Νευρωνικό Δίκτυο	55
4.2.3	Συλλογικό Απλό Νευρωνικό Δίκτυο	57
4.3	Μεθοδολογία κατασκευής μεθόδων ερμηνευσιμότητας	59
4.3.1	Global Surrogate	59
4.3.2	Individual Conditional Expectation	60
4.3.3	Shapley Values	60
5	Ανάλυση Αποτελεσμάτων	61
5.1	Αποτελέσματα μοντέλων μηχανικής μάθησης	61
5.2	Αποτελέσματα μεθόδων ερμηνευσιμότητας	61
5.2.1	Ερμηνευσιμότητα Απλού Συλλογικού μοντέλου	62
5.2.2	Ερμηνευσιμότητα CNN μοντέλου	76
6	Συμπεράσματα και μελλοντικές προεκτάσεις	83
6.1	Συμπεράσματα	83
6.2	Μελλοντικές Προεκτάσεις	84

Κατάλογος Σχημάτων

1.1	Κατηγορίες Αγγειακών Εγκεφαλικών Επεισοδίων (AEE)	20
1.2	Φυσιολογική (Υγιής) καρωτίδα	22
1.3	Στάδια δημιουργίας της αθηρωματικής πλάκας	22
1.4	Εμβολισμός αθηρωματικής πλάκας	23
1.5	Αθηρωματική πλάκα καρωτίδας στον καρωτικό διχασμό της αρτηρίας	23
1.6	Καρωτιδική Νόσος που οδήγησε σε AEE	25
1.7	Duplex υπερηχογράφημα έγχρωμης ροής φυσιολογικής καρωτίδας	27
1.8	Απεικόνιση δεξιάς σοβαρής στένωσης της καρωτίδας μέσω CTA	27
1.9	Απεικόνιση δεξιάς σοβαρής στένωσης της καρωτίδας μέσω MRI	28
1.10	Ενδαρτηρεκτομή καρωτίδας	29
1.11	Αγγειοπλαστική καρωτίδας με τοποθέτηση stent	29
1.12	Αλγόριθμος εύρεσης κατάλληλης θεραπείας για την ΚΝ. BMT = Καλύτερη Φαρμακευτική Θεραπεία, CAS = Carotid Artery Stenting (Τοποθέτηση stent), CEA = Carotid Endarterectomy (Ενδαρτηρεκτομή)	30
2.1	Παράδειγμα CNN αποτελούμενο από: Επίπεδο εισόδου (Input), Επίπεδο δισδιάστατης συνέλιξης (Conv2D), Επίπεδο ενεργοποίησης (ReLU), Συγκεντρωτικό Επίπεδο (MaxPool) και Πλήρως συνδεδεμένο επίπεδο (FC)	36
2.2	Καμπύλη Roc	38
3.1	Εννοιολογικό σχήμα μη ερμηνεύσιμων μοντέλων και εγγενώς ερμηνεύσιμων μοντέλων	45
4.1	Κάθε ασθενής αντιπροσωπεύεται από μία ή περισσότερες υπερηχογραφικές απεικονιστικές καταγραφές και από ένα διάλυμα δεκατριών κλινικών χαρακτηριστικών	49
4.2	Υπέρηχογραφικές απεικονίσεις β-σάρωσης των αθηρωμάτων που υπάρχουν στις καρωτιδικές αρτηρίες οι οποίες ανήκουν σε: (α) Ασθενή υψηλού κινδύνου (b) Ασθενή χαμηλού κινδύνου	51
4.3	(a) Αρχικό ανισόρροπο σύνολο δεδομένων, όπου με Class 1 συμβολίζουμε την κλάση των ασθενών Υψηλού κινδύνου (High risk) και με Class 0 την κλάση των ασθενών Χαμηλού κινδύνου (Low risk) (b) Διαχωρισμός του αρχικού ανισόρροπου συνόλου δεδομένων σε 3 διαφορετικά ισορροπημένα υποσύνολα, χωρίζοντας την Class 1 σε τρία ανεξάρτητα υποσύνολα.	54
4.4	Αρχιτεκτονική Συλλογικού μοντέλου μηχανικής μάθησης. Κάθε ισορροπημένο υποσύνολο του αρχικού συνόλου δεδομένων αποτελεί είσοδο σε κάθε ένα από τα επιμέρους μοντέλα τα οποία στη συνέχεια εκπαιδεύονται ξεχωριστά. Η τελική απόφαση του Συλλογικού μοντέλου λαμβάνεται μέσω του συμφητισμού των προβλέψεων των τριών μοντέλων (επίπεδο voting).	55

4.5	Αρχιτεκτονική CNN μοντέλων. Στα συνελικτικά επίπεδα (Conv) ο πρώτος αριθμός μέσα στην παρένθεση αντιπροσωπεύει τον αριθμό των εικόνων που λαμβάνονται ως είσοδος στο εκάστοτε επίπεδο, ο δεύτερος τον αριθμό των εικόνων που προκύπτουν ως έξοδος από το επίπεδο και ο τρίτος την διάσταση του συνελικτικού πυρήνα που χρησιμοποιείται.	56
4.6	Αρχιτεκτονική των επιμέρους SNN μοντέλων. Στα γραμμικά μέρη σύνδεσης (Linear) ο πρώτος αριθμός μέσα στην παρένθεση αναφέρεται στον αριθμό των νευρώνων της εισόδου του εκάστοτε layer ενώ ο δεύτερος στον αριθμό νευρώνων της εξόδου.	58
5.1	Ερμηνευσιμότητα Απλού Συλλογικού μοντέλου μέσω ενός Δέντρου Απόφασης (Decision Tree) που λειτουργεί ως υποκατάστατο μοντέλο. Στο διάγραμμα οι κόμβοι που είναι χρωματισμένοι με αποχρώσεις του πορτοκαλί δηλώνουν ότι τα περισσότερα δείγματα, πριν να πραγματοποιηθεί ο διαχωρισμός, ανήκουν στην κλάση των ασθενών Υψηλού κινδύνου (High risk) ενώ με αποχρώσεις του μπλε σημειώνονται οι κόμβοι που υποδηλώνουν το αντίθετο, δηλαδή την κλάση Χαμηλού κινδύνου (Low risk). Οι πιο σκούρες αποχρώσεις υποδηλώνουν πως ο κόμβος είναι τερματικός (φύλλο).	63
5.2	ICE plot 1ου επιμέρους μοντέλου για το χαρακτηριστικό της HDL χοληστερόλης. Το αριστερό διάγραμμα αποτελείται από καμπύλες που η κάθε μία αντιπροσωπεύει έναν ξεχωριστό ασθενή. Το δεξί διάγραμμα αποτελεί τον μέσο όρο αυτών των καμπυλών	64
5.3	ICE plot 2ου επιμέρους μοντέλου για το χαρακτηριστικό της HDL χοληστερόλης .	64
5.4	ICE plot 3ου επιμέρους μοντέλου για το χαρακτηριστικό της HDL χοληστερόλης .	65
5.5	ICE plot 1ου επιμέρους μοντέλου για το χαρακτηριστικό της LDL χοληστερόλης .	65
5.6	ICE plot 2ου επιμέρους μοντέλου για το χαρακτηριστικό LDL χοληστερόλη	66
5.7	ICE plot 3ου επιμέρους μοντέλου για το χαρακτηριστικό LDL χοληστερόλη	66
5.8	ICE plot 1ου επιμέρους μοντέλου για το χαρακτηριστικό της γλυκόζης αίματος, Blood Glucose	67
5.9	ICE plot 2ου επιμέρους μοντέλου για το χαρακτηριστικό της γλυκόζης αίματος, Blood Glucose	67
5.10	ICE plot 3ου επιμέρους μοντέλου για το χαρακτηριστικό της γλυκόζης αίματος, Blood Glucose	68
5.11	ICE plot 1ου επιμέρους μοντέλου για το χαρακτηριστικό των τριγλυκεριδίων, TG .	68
5.12	ICE plot 2ου επιμέρους μοντέλου για το χαρακτηριστικό των τριγλυκεριδίων, TG .	69
5.13	ICE plot 3ου επιμέρους μοντέλου για το χαρακτηριστικό των τριγλυκεριδίων, TG .	69
5.14	ICE plot απλού συλλογικού μοντέλου για το χαρακτηριστικό της HDL χοληστερόλης	70
5.15	ICE plot απλού συλλογικού μοντέλου για το χαρακτηριστικό της γλυκόζης αίματος, Blood Glucose	70
5.16	ICE plot απλού συλλογικού μοντέλου για το χαρακτηριστικό της LDL χοληστερόλης	71
5.17	ICE plot απλού συλλογικού μοντέλου για το χαρακτηριστικό των τριγλυκεριδίων, TG	71
5.18	Summary plot 1ου επιμέρους μοντέλου. Κάθε κουκίδα αντιπροσωπεύει την τιμή ενός χαρακτηριστικού ενός συγκεκριμένου δείγματος. Οι κόκκινες κουκκίδες αντιστοιχούν σε υψηλές τιμές του χαρακτηριστικού ενώ οι μπλε σε χαμηλές.	72
5.19	Summary plot 2ου επιμέρους μοντέλου	73
5.20	Summary plot 3ου επιμέρους μοντέλου	74

5.21	Summary plot απλού συλλογικού μοντέλου. Οι shapley values τιμές για το απλό συλλογικό μοντέλο προέκυψαν από το μέσο όρο των shapley value τιμών των τριών επιμέρους μοντέλων	75
5.22	Ερμηνευσιμότητα 1ου επιμέρους μοντέλου μέσω της μεθόδου Shapley Values, για δείγμα ασθενούς που ανήκει στην κατηγορία υψηλού κινδύνου. Τα εικονοστοιχεία που χρωματίζονται με κόκκινο χρώμα ωθούν το μοντέλο να λάβει την απόφαση ότι το δείγμα ανήκει στην κατηγορία του υψηλού κινδύνου ενώ τα μπλε το απωθούν. Το εικονιζόμενο δείγμα κατατάσσεται εσφαλμένα από το 1ο επιμέρους μοντέλο στην κατηγορία του χαμηλού κινδύνου καθώς η τελική απόφασή του επηρεάζεται σημαντικά από την έντονα λευκή περιοχή στο κέντρο του αθηρώματος	77
5.23	Ερμηνευσιμότητα 2ου επιμέρους μοντέλου με SV	78
5.24	Ερμηνευσιμότητα 3ου επιμέρους μοντέλου με SV	78
5.25	Ερμηνευσιμότητα 1ου επιμέρους μοντέλου με SV	79
5.26	Ερμηνευσιμότητα 2ου επιμέρους μοντέλου με SV	79
5.27	Ερμηνευσιμότητα 3ου επιμέρους μοντέλου με SV	80
5.28	Ερμηνευσιμότητα 1ου επιμέρους μοντέλου με SV	80
5.29	Ερμηνευσιμότητα 2ου επιμέρους μοντέλου με SV	81
5.30	Ερμηνευσιμότητα 3ου επιμέρους μοντέλου με SV	81

Κατάλογος Πινάκων

1.1	Ταξινόμηση και πορεία εξέλιξης των αθηρωματικών πλακών σύμφωνα με την Αμερικανική Ένωση Καρδιολογίας.	24
2.1	Συναρτήσεις ενεργοποίησης	33
4.1	Τρέχουσα κλινική πρακτική για την επιλογή της θεραπείας ενός ασθενή με καρωτιδική αθηρωμάτωση. Με Θ1 συμβολίζουμε την συντηρητική αγωγή, με Θ2 την επεμβατική μέθοδο θεραπείας και με ΒΣ τον βαθμό στένωσης της καρωτίδας. . . .	50
4.2	Ρυθμίσεις της διάταξης υπερήχων για την καταγραφή πραγματικών χρονικών ακολουθιών εικόνων υπερήχων β-σάρωσης της καρωτίδας σε διαμήκη τομή	51
4.3	Υπολογισμός του βαθμού στένωσης της καρωτίδας με βάση το πρότυπο που ακολουθείται στη μελέτη ACSRS. Όπου με PSV αναφερόμαστε στη Μέγιστη συστολική ταχύτητα σε cm/sec, με EDV στην Τελοδιαστολική ταχύτητα σε cm/sec, με ICA στη έσω κοινή αρτηρία και με CCA στην κοινή καρωτιδική αρτηρία	52
5.1	Αποτελέσματα συνελικτικού συλλογικού μοντέλου	61
5.2	Αποτελέσματα απλού συλλογικού μοντέλου	61
5.3	Αποτελέσματα της αξιολόγησης του Δέντρου απόφασης κατά τη λειτουργία του ως global surrogate μοντέλο	62

Κεφάλαιο 1

Καρωτιδική Νόσος

Η καρωτιδική νόσος αποτελεί μία από τις κύριες αιτίες για την πρόκληση ισχαιμικών εγκεφαλικών επεισοδίων. Πιο συγκεκριμένα το 87% όλων των εγκεφαλικών επεισοδίων είναι ισχαιμικά [1] και από αυτά αν και αρκετά χαρακτηρίζονται κρυπτογενή, τουλάχιστον το 20%–30% των ισχαιμικών επεισοδίων μπορούν να αποδοθούν στην νόσο των καρωτιδών [2]. Ο ασθενής που πάσχει από την νόσο των καρωτιδών τις περισσότερες φορές δεν εκδηλώνει συμπτώματα και δυστυχώς συχνά η πρώτη εκδήλωση είναι ένα αγγειακό εγκεφαλικό επεισόδιο [3]. Παρόλα αυτά η πρόωρη και κατάλληλη αντιμετώπιση μπορούν να μειώσουν αυτόν τον κίνδυνο. Σε αυτό μπορεί να συμβάλει σημαντικά η δημιουργία υπολογιστικών μοντέλων, ικανών να κατηγοριοποιήσουν την κατάσταση του ασθενούς προτείνοντάς του την κατάλληλη θεραπεία. Πιο συγκεκριμένα η ανάπτυξη των αλγορίθμων μηχανικής μάθησης έχει αποτελέσει πολύ σημαντικό σταθμό στη βελτίωση πολλών ιατρικών διαγνώσεων. Στην παρούσα διπλωματική εργασία θα χρησιμοποιήσουμε ένα ανισόρροπο (unbalanced) σύνολο δεδομένων το οποίο αποτελείται από εικόνες υπερήχων β-σάρωσης της καρωτίδας αλλά και από διανύσματα αιματολογικών χαρακτηριστικών για κάθε ασθενή.

Μέχρι και σήμερα αρκετές μελέτες έχουν ασχοληθεί με την δημιουργία μοντέλων για την κατάταξη των ασθενών με καρωτιδική νόσο. Ειδικότερα στις μελέτες [4], [5], [6], [7], [8], [9], [10] χρησιμοποιήθηκαν υπερηχογραφικές απεικονίσεις β-σάρωσης, όπως και στην παρούσα εργασία, με σκοπό την κατάταξη της αθηρωματικής πλάκας ανάλογα με την επικινδυνότητάς της μέσω μοντέλων μηχανικής μάθησης και ειδικότερα συνελικτικών νευρωνικών δικτύων.

Όλες οι προαναφερθείσες εργασίες κατάφεραν να επιτύχουν πολύ υψηλά επίπεδα διαγνωστικής ακρίβειας. Παρόλα αυτά, λόγω της εσωτερικής δομής τους δουλεύουν ως μαύρα κουτιά (black boxes), δηλαδή οι λόγοι που οδήγησαν τα υπολογιστικά μοντέλα τους να λάβουν μία συγκεκριμένη απόφαση δεν είναι γνωστοί. Η συμπεριφορά αυτή αποτελεί εμπόδιο για την ευρεία υιοθέτησή τους στην κλινική πρακτική, όπου οι αποφάσεις αφορούν ανθρώπινες ζωές. Σε αυτό το πρόβλημα έρχονται να δώσουν λύση οι μέθοδοι ερμηνευσιμότητας, με τις οποίες και θα ασχοληθούμε στην παρούσα εργασία. Ο όρος ερμηνευσιμότητα αναφέρεται στο βαθμό που μπορούν να γίνουν κατανοητές από τον άνθρωπο, οι προβλέψεις ενός μοντέλου [11]. Μέσω των μεθόδων ερμηνευσιμότητας καταφέρνουμε να ερμηνεύσουμε σε μεγάλο βαθμό τις προβλέψεις των μοντέλων δίνοντας μια ευρύτερη εικόνα στον θεράποντα ιατρό αλλά και τον μηχανικό για την συμπεριφορά του μοντέλου. Οι μέθοδοι ερμηνευσιμότητας μπορούν να χωριστούν σε δύο κατηγορίες: στα εγγενώς ερμηνεύσιμα μοντέλα και στις τεχνικές ερμηνευσιμότητας ανεξάρτητες του μοντέλου (model agnostic methods) [12]. Στην παρούσα εργασία θα ασχοληθούμε με τις model agnostic μεθόδους.

Άλλες εργασίες που χρησιμοποίησαν παρόμοιες μεθόδους για την ερμηνευσιμότητα των μοντέλων που αφορούν κλινικές αποφάσεις είναι η εργασία [13] κατά την οποία δημιουργήθηκαν

εγγενώς ερμηνεύσιμα μοντέλα, όπως τα δέντρα απόφασης, με σκοπό τη διάγνωση του καρκίνου του μαστού. Ακόμα στην εργασία [14] δημιουργήθηκαν τρία μοντέλα μηχανικής μάθησης τα οποία ερμηνεύτηκαν με χρήση τριών μεθόδων ερμηνευσιμότητας οι οποίες είναι ανεξάρτητες του μοντέλου (model agnostic methods).

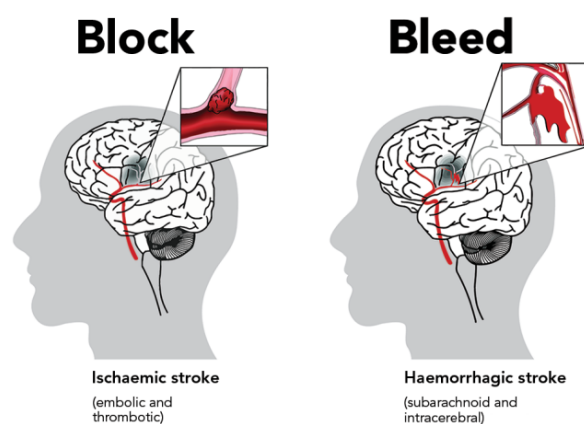
Εν κατακλείδι η παρούσα εργασία έχει ως σκοπό τη διαχείριση ενός ανισόρροπου συνόλου δεδομένων και τη δημιουργία ερμηνεύσιμων μοντέλων βαθιάς μάθησης, με απώτερο σκοπό τη σωστή κατηγοριοποίηση των ασθενών με καρωτιδική νόσο ανάλογα με την θεραπεία που καλούνται να ακολουθήσουν.

1.1 Αγγειακά Εγκεφαλικά Επεισόδια

Τα Αγγειακά Εγκεφαλικά Επεισόδια (ΑΕΕ) καταγράφονται ως η τρίτη μεγαλύτερη απειλή για την υγεία μας, μετά τις καρδιακές παθήσεις και τον καρκίνο. Ταυτόχρονα αποτελούν τη δεύτερη αιτία καρδιαγγειακού θανάτου, ενώ αποτελούν την πιο συχνή αιτία θανάτου ως αποτέλεσμα νευρολογικών διαταραχών [15]. Κάθε χρόνο, η πιθανότητα που αναλογεί για ένα νέο ΑΕΕ είναι περίπου 160 ανά 100000 ανθρώπους.

Τα ΑΕΕ προκαλούνται από εστιακές βλάβες του εγκεφάλου οποιασδήποτε αγγειακής αιτιολογίας και διάρκειας μεγαλύτερης των 24 ωρών. Διακρίνονται σε δύο κατηγορίες, τα ισχαιμικά και τα αιμορραγικά (Σχήμα. 1.1).

Τα ισχαιμικά επεισόδια, τα οποία αποτελούν το 87% των ΑΕΕ, προκαλούνται όταν ένα αγγείο του εγκεφάλου, που τον τροφοδοτεί με οξυγόνο και θρεπτικά συστατικά, αποφράσσεται λόγω θρομβώσεων ή εμβολών. Και αυτά με τη σειρά τους διαχωρίζονται σε τρεις υποκατηγορίες, τα εμβολικά ΑΕΕ, τα θρομβωτικά ΑΕΕ και τα κενοδοπιώδη ΑΕΕ. Τα εμβολικά αποτελούν το 20% των ισχαιμικών ΑΕΕ και είναι καρδιακής ή αρτηριακής προέλευσης [1]. Τα θρομβωτικά αφορούν κυρίως μεγάλα αγγεία και οφείλονται σε αθηρωμάτωση των καρωτιδικών και ενδοκράνιων αρτηριών. Τέλος, τα κενοδοπιώδη προκαλούνται λόγω του αποκλεισμού διαττηραίνουσών αρτηριών της βασικής και σπονδυλικής αρτηρίας και κλάδων της μέσης εγκεφαλικής αρτηρίας.



Σχήμα 1.1: Κατηγορίες Αγγειακών Εγκεφαλικών Επεισοδίων (ΑΕΕ)

Τα αιμορραγικά ΑΕΕ προκύπτουν όταν κάποιο από τα αγγεία του εγκεφάλου υφίσταται διαρροή ή ρήξη, προκαλώντας αιμορραγία μέσα ή γύρω από τον εγκέφαλο. Οφείλονται κυρίως σε υπερτασική αγγειακή νόσο, σε εγκεφαλικά ανευρίσματα και σε αρτηριοφλεβώδεις δυσπλασίες.

1.2 Αρτηρίες

Οι αρτηρίες αποτελούν ελαστικούς και συσταλτικούς σωλήνες που μεταφέρουν οξυγονωμένο αίμα από την καρδιά προς τα υπόλοιπα όργανα. Ανάλογα με την διάμετρό τους χωρίζονται σε μεγάλες, μεσαίες, μικρές και αρτηρίδια, ενώ ανάλογα με την κατασκευή του τοιχώματός τους, διακρίνονται σε δύο τύπους, τον ελαστικό και τον μυϊκό, σύμφωνα με την κατάταξη Ranvier. Το φυσιολογικό τοίχωμα των αρτηριών αποτελείται από τρεις σαφώς διακρινόμενες και ομόκεντρες στοιβάδες: έσω, μέσο και έξω χιτώνα .

Ο **έσω χιτώνας** αποτελείται από μία στοιβάδα ενδοθηλιακών κυττάρων που επενδύουν την εσωτερική επιφάνεια των αγγείων και από ποικίλη ποσότητα ενδοθηλιακού συνδετικού ιστού. Είναι αυτός που περιβάλλει τον αυλό του αγγείου, την περιοχή δηλαδή από την οποία μεταφέρεται το αίμα και ταυτόχρονα αποτελεί το μέρος του αγγειακού τοιχώματος όπου εμφανίζεται η αθηροσκλήρωση [16].

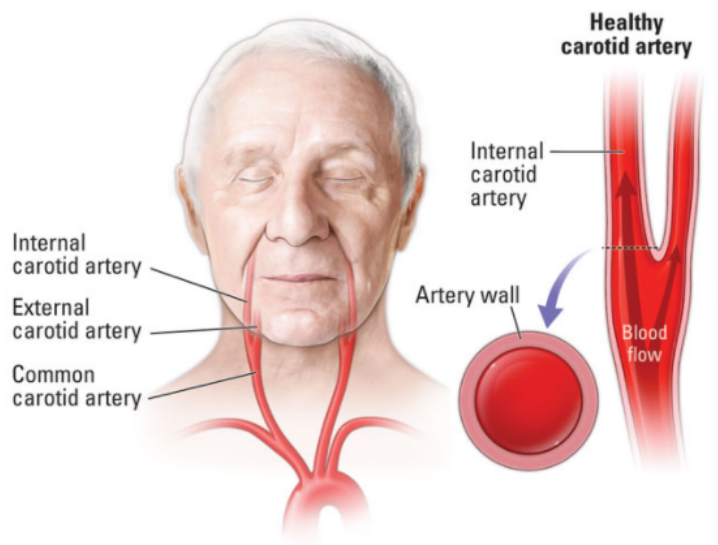
Ο **μέσος χιτώνας** είναι συνήθως ο παχύτερος από τους τρεις χιτώνες, αποτελείται από λεία μυϊκά κύτταρα και ινοελαστικό συνδετικό ιστό.

Τέλος ο **έξω χιτώνας** αποτελείται από ινοελαστικό συνδετικό ιστό που είναι πλούσιος σε κολλαγόνο και αποτελεί σημαντικό υποστηρικτικό στοιχείο του αγγειακού τοιχώματος.

1.3 Ανατομία Καρωτιδικών Αρτηριών

Στην παρούσα διπλωματική εργασία θα ασχοληθούμε με τις καρωτιδικές αρτηρίες οι οποίες είναι ελαστικού τύπου και μεγάλου μεγέθους αρτηρίες (Σχήμα. 1.2).

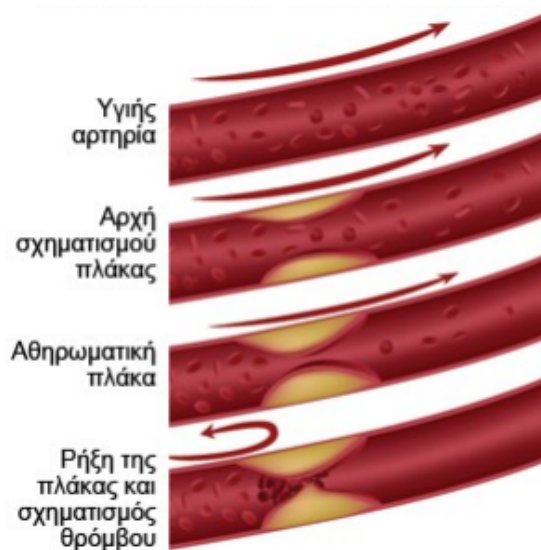
Οι καρωτιδικές αρτηρίες ξεκινούν από την αορτή, και εισέρχονται στο κρανίο διατρέχοντας όλο το μήκος της πλάγιας τραχηλικής χώρας. Αποτελούν τις κύριες αρτηρίες που παρέχουν οξυγονωμένο αίμα στον εγκέφαλο (μία για κάθε εγκεφαλικό ημισφαίριο). Πιο συγκεκριμένα οι κοινές καρωτιδικές αρτηρίες υπάρχουν στην αριστερή και την δεξιά πλευρά του τραχήλου. Οι δύο αρτηρίες ξεκινούν από διαφορετικά σημεία της αορτής και πιο συγκεκριμένα του αορτικού τόξου, αλλά ακολουθούν συμμετρικές πορείες προς τον εγκέφαλο. Στο άνω άκρο του θυρωειδή χόνδρου οι κοινές αρτηρίες χωρίζονται σε έσω και έξω καρωτιδικές αρτηρίες αντίστοιχα. Η έσω καρωτίδα παρέχει αίμα στον εγκέφαλο ενώ η έξω καρωτίδα μετά την έκφυσή της από την κοινή καρωτίδα δίνει 8 κλάδους και παρέχει αίμα στο πρόσωπο και στο τριχωτό της κεφαλής [17]. Όπως βλέπουμε και στο Σχήμα. 1.2 υπό κανονικές συνθήκες το εσωτερικό τοίχωμα των αρτηριών (έσω χιτώνας) είναι ομαλό, με την πάροδο του χρόνου όμως μπορεί να δημιουργηθεί μία ανώμαλη προβολή η οποία ονομάζεται αθηρωματική πλάκα.



Σχήμα 1.2: Φυσιολογική (Υγιής) καρωτίδα

1.4 Αθηρωσκλήρυνση - Αθηρωματική Πλάκα

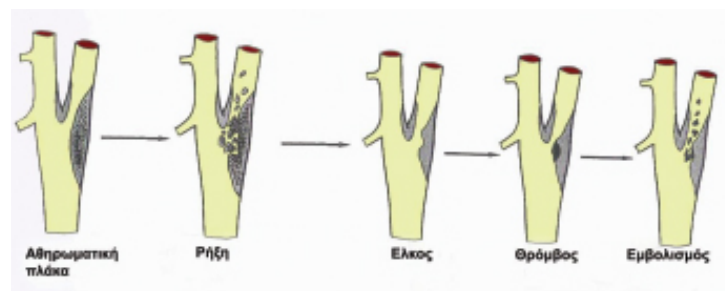
Η αθηροσκλήρωση είναι μία νόσος ή οποία εξελίσσεται από την παιδική ηλικία όμως οι κλινικές τις εκδηλώσεις εμφανίζονται κυρίως κατά τη μέση ηλικία ή και αργότερα. Η βασική μορφολογική αλλοίωση θεωρείται η αθηρωματική πλάκα ή το αθήρωμα, η οποία σχηματίζεται στον έσω χιτώνα των αρτηριών. Αναπτύσσεται αργά και χωρίς κλινικά συμπτώματα για πολλά χρόνια (Σχήμα 1.3). Ειδικότερα, η αθηρωματική πλάκα δημιουργείται από την εναπόθεση λιπιδίων, κυρίως χοληστερόλης χαμηλής πυκνότητας (LDL), στον έσω χιτώνα της αρτηρίας και πρόκληση φλεγμονώδους αντίδρασης που συνοδεύεται από ανάπτυξη ινοβλαστών [18].



Σχήμα 1.3: Στάδια δημιουργίας της αθηρωματικής πλάκας

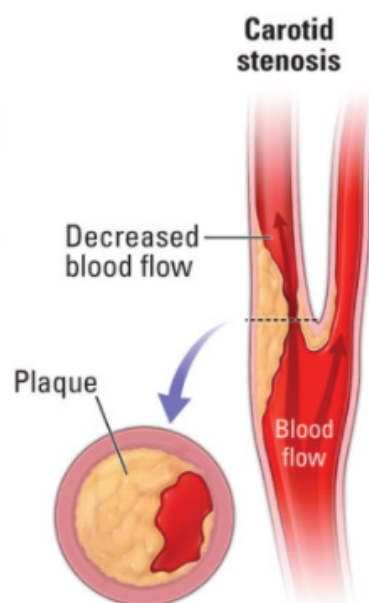
Η εξέλιξη του αθηρώματος μπορεί να γίνει με διαφορετικούς τρόπους. Ορισμένες φορές οι

πλάκες μεγαλώνουν σταδιακά εμποδίζοντας όλο και περισσότερο την προσαγωγή του αίματος στον εγκέφαλο. Ενώ άλλες φορές οι πλάκες αιμορραγούν στο εσωτερικό είτε αναπτύσσεται θρόμβος στην επιφάνεια τους. Η αύξηση του μεγέθους της βλάβης μπορεί να είναι είτε συνεχής είτε να μεταβληθεί από πρόκληση ενδοπλαστικής αιμορραγίας ή οποία μπορεί να οδηγήσει σε απότομη αύξηση του μεγέθους της βλάβης, μέχρι και την απόφραξη του αγγείου. Αν σε οποιαδήποτε χρονική στιγμή προκληθεί αποκόλληση του θρόμβου (αθηρωματικού υλικού) από το ενδοθήλιο του έσω χιτώνα της καρωτιδικής αρτηρίας, μία ανοιχτή κοιλότητα δημιουργείται στο κεντρικό τμήμα της αθηρωματικής πλάκας και ονομάζεται έλκος. Το έλκος μπορεί να προκαλέσει την συσσώρευση και ενεργοποίηση των αιμοπεταλίων με αποτέλεσμα να εμφανιστεί θρόμβος είτε η δημιουργία δευτερογενούς αθηρωματικής πλάκας. Ο θρόμβος ή τμήματα αυτού, μπορεί να αποσπαστεί και να απελευθερωθεί στην κυκλοφορία, προκαλώντας αρτηριακή εμβολή (Σχήμα. 1.4).



Σχήμα 1.4: Εμβολισμός αθηρωματικής πλάκας

Η εμφάνιση του αθηρώματος είναι πιο συχνή σε σημεία εκφύσεις κλάδων ή διχασμού. Πιο συγκεκριμένα τα πιο συχνά σημεία δημιουργίας αθηρωματικής πλάκας είναι τα εξής: καρωτικός διχασμός 38% (Σχήμα. 1.5), καρωτιδικό σιφώνιο, αρχή μέσης η πρόσθιας εγκεφαλικής 33%, ostium σπονδυλικών 20% και προσβολή κλάδων αορτικού τόξου 9% [18].



Σχήμα 1.5: Αθηρωματική πλάκα καρωτίδας στον καρωτικό διχασμό της αρτηρίας

1.5 Καρωτιδική Νόσος

Καρωτιδική Νόσος (ΚΝ) ονομάζεται η στένωση ή η απόφραξη των καρωτιδικών αρτηριών. Την αιτιολογία της καρωτιδικής νόσου κατά κύριο λόγο και σε ποσοστό περίπου 90% αποτελεί η αθηροσκλήρωση - αθηρωματική πλάκα [19]. Η κλινική εκδήλωση της καρωτιδικής νόσου μπορεί να είναι συμπτωματική και ασυμπτωματική. Πιο συγκεκριμένα στα αρχικά της στάδια η ΚΝ είναι ασυμπτωματική, κατά την εξέλιξή της όμως το πρώτο σύμπτωμα μπορεί να είναι ένα ΑΕΕ. Μάλιστα, το 80% όλων των εγκεφαλικών επεισοδίων οφείλονται σε βλάβες των καρωτίδων και μόνο το 20% σε αλλοιώσεις των σπονδυλικών αρτηριών. Παρόλα αυτά ο ασθενής ενδέχεται να βιώσει ορισμένα προειδοποιητικά συμπτώματα ενός ΑΕΕ. Η κατάσταση αυτή ονομάζεται Παροδικό Εγκεφαλικό Επεισόδιο (ΠΕΕ) και τα συμπτώματα της διαρκούν από λίγα λεπτά έως και μία ώρα και περιλαμβάνουν:

- Αίσθημα αδυναμίας, αιμωδίας ή καύσους στη μία πλευρά του σώματος
- Αδυναμία ελέγχου των κινήσεων των άνω ή κάτω άκρων
- Απώλεια όρασης από το ένα μάτι
- Αδυναμία στην ομιλία

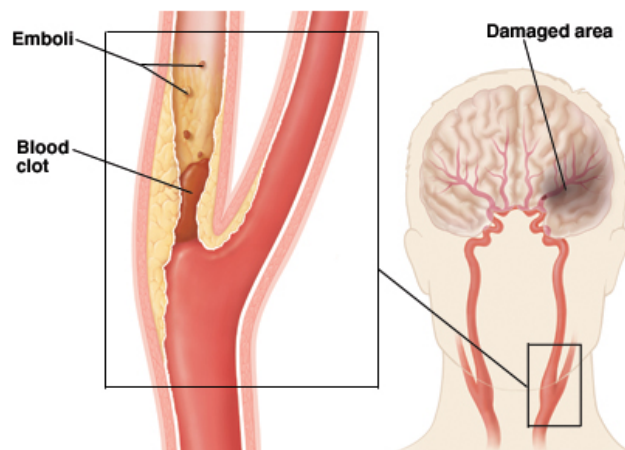
Μία καρωτιδική στένωση χαρακτηρίζεται συμπτωματική όταν σχετίζεται με ένα τουλάχιστον περιστατικό ΑΕΕ ή ΠΕΕ κατά τη διάρκεια των τελευταίων έξι μηνών. Σε κάθε άλλη περίπτωση η ΚΝ θεωρείται ασυμπτωματική και της περισσότερες φορές ανακαλύπτεται τυχαία. Όπως συμπεραίνεται και από τα παραπάνω η ΚΝ εκδηλώνεται όχι μόνο στο χώρο αλλά και στο χρόνο. Με βάση την εξέλιξη της επικινδυνότητάς-σοβαρότητάς τους η Αμερικανική Ένωση Καρδιολογίας (American Heart Association) διέκρινε έξι τύπους αθηρωματικών πλακών οι οποίες παρουσιάζονται αναλυτικά στον Πίνακα. 1.1 [20].

Πορεία Εξέλιξης	Τύπος αθηρωματικής πλάκας	Χαρακτηρισμός
I ↓	Τύπος I: Αρχική βλάβη • Μεμονωμένα λιπώδη μακροφάγα (αφρώδη κύτταρα)	Πρώιμες αλλοιώσεις
II ↓	Τύπος II: Λιπώδης Γράμμωση • Συσσώρευση αφρωδών κυττάρων	
III ↓	Τύπος III: Προ-αθήρωμα • Αλλοιώσεις τύπου II και εξωκυττάρια συσσωρεύσεις λιπιδίων μεταξύ κυττάρων λείου μυός	Ενδιάμεσες αλλοιώσεις
IV ↓	Τύπος IV: Αθήρωμα • Σαφώς καθορισμένος λιπιδικός πυρήνας	Προχωρημένες αλλοιώσεις
V ↓	Τύπος V: Ινωδοαθήρωμα • Αθήρωμα με ινώδη κάψα (Va) • Επεκτεταμένη αποτιάνωση (Vb) • Επεκτεταμένη ίνωση (Vc)	
VI	Τύπος VI: Περίπλοκη αθηρωματική βλάβη • Ραγείσα αλλοίωση τύπου IV ή V (VIa) • Ενδοτοιχωματική αιμορραγία (VIb) • Υπερκείμενη θρόμβωση (VIc)	

Πίνακας 1.1: Ταξινόμηση και πορεία εξέλιξης των αθηρωματικών πλακών σύμφωνα με την Αμερικανική Ένωση Καρδιολογίας.

Στις περιπτώσεις λοιπόν που η κατάσταση είναι αρκετά σοβαρή η ΚΝ έχει ως αποτέλεσμα τη μείωση της παροχής αίματος στον εγκέφαλο και συνεπώς τον δυνητικό κίνδυνο εμφάνισης ΑΕΕ (Σχήμα. 1.6). Υπάρχουν τρεις βασικοί τρόποι με τους οποίους η ΚΝ μπορεί να οδηγήσει σε ΑΕΕ [18].

1. Οι εναποθέσεις αθηρωματικής πλάκας στα τοιχώματα της αρτηρίας μπορούν να προκαλέσουν σημαντική μείωση της ροής του αίματος στον εγκέφαλο μέχρι και την πλήρη διακοπή παροχής αίματος, εάν η πλάκα φράξει εντελώς την αρτηρία.
2. Πολλές φορές η πλάκα είναι αρκετά μαλακή με αποτέλεσμα να σπάει. Αυτό ο οργανισμός το αντιλαμβάνεται ως εσωτερικό τραύμα και προσπαθεί να το επουλώσει με τη δημιουργία θρόμβου μέσω των αιμοπεταλίων. Αν ο θρόμβος αυτός μεγαλώσει αρκετά μπορεί να οδηγήσει σε απόφραξη της αρτηρίας.
3. Πολύ συχνά ένα κομμάτι της πλάκας ή του θρόμβου μπορεί να αποσπαστεί και μέσω της ροής του αίματος να οδηγηθεί σε άλλες μικρότερες αρτηρίες του εγκεφάλου και να τις φράξει, διακόπτοντας τη ροή του αίματος στον εγκέφαλο και προκαλώντας ΕΕ.



Σχήμα 1.6: Καρωτιδική Νόσος που οδήγησε σε ΑΕΕ

Συμπερασματικά λοιπόν η ΚΝ είναι μία νόσος που εξελίσσεται της περισσότερες φορές αθόρυβα οδηγώντας τον ασθενή απευθείας σε πολύ σοβαρά συμπτώματα (ΠΕΕ ή ΑΕΕ) χωρίς καμία προειδοποίηση. Πιο συγκεκριμένα ενδεικτικό της σοβαρότητας της νόσου αποτελεί το γεγονός ότι τα ΑΕΕ τα οποία οφείλονται στην ΚΝ, καταγράφονται ως η τρίτη μεγαλύτερη απειλή για την υγεία μας, μετά τις καρδιακές παθήσεις και τον καρκίνο. Η έγκαιρη διάγνωση της λοιπόν αποτελεί καίριο παράγοντα για την αντιμετώπισή της ακόμα και όταν η κατάσταση του ασθενούς είναι ασυμπτωματική και στα πρώιμα στάδια της ανάπτυξής της.

1.6 Παράγοντες κινδύνου εμφάνισης Καρωτιδικής Νόσου

Η ΚΝ λόγω αθηροσκλήρωσης, άσχετα εάν προκαλεί συμπτώματα ή όχι, συνήθως προσβάλλει άτομα ηλικίας άνω των 60 ετών και είναι συχνότερη στους άνδρες παρά στις γυναίκες. Μόνο το 1% των ασθενών ηλικίας 50-59 ετών εμφανίζει σοβαρού βαθμού στένωση των καρωτίδων,

ενώ το ποσοστό αυτό είναι 10% για τις ηλικίες 80-89 ετών. Μεταξύ όσων πάσχουν από στεφανιαία νόσο ή από περιφερική αποφρακτική αρτηριοπάθεια, το ποσοστό αυτών που έχουν και καρωτιδική νόσο ανέρχεται στο 20 – 30%.

Όπως αναφέραμε και προηγουμένως οι αθηροσκληρωτικές πλάκες προκαλούνται από την φλεγμονή που προκαλεί η είσοδος LDL χοληστερίνης αμέσως κάτω από το ενδοθήλιο. Βασικό ρόλο λοιπόν στην εμφάνισή τους έχει η αυξημένη ποσότητα LDL χοληστερίνης στο αίμα του ασθενή. Αυτό μπορεί να προκύψει λόγω κακής διατροφής (φαγητά με υψηλή χοληστερίνη, κεκορεσμένα και τρανς λιπαρά) αλλά και λόγω αυξημένης παραγωγής από το ήπαρ, που οφείλεται σε κληρονομικούς παράγοντες. Παρόλα αυτά η είσοδος της LDL χοληστερίνης στο εσωτερικό της αρτηρίας μπορεί να προκληθεί και εξαιτίας κάποιας βλάβης του ενδοθηλίου. Επομένως παράγοντες που μπορούν να προκαλέσουν τον τραυματισμό του ενδοθηλίου αποτελούν και παράγοντες κινδύνου για την εμφάνιση της ΚΝ. Αυτοί οι παράγοντες μπορεί να είναι: η ήδη αυξημένη LDL χοληστερίνη, το κάπνισμα, η υψηλή αρτηριακή πίεση (υπέρταση) και ο σακχαρώδης διαβήτης.

Τέλος άλλες σπάνιες αιτίες εμφάνισης ΚΝ είναι τα ανευρύσματα καρωτίδας και η ινομυϊκή δυσπλασία.

1.7 Διάγνωση Καρωτιδικής Νόσου

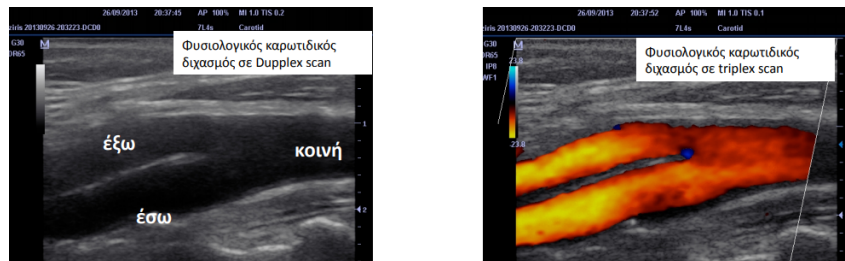
Η διάγνωση της ΚΝ μπορεί να γίνει με ποικίλους τρόπους. Εκτός από το ιστορικό που θα λάβει ο γιατρός από τον ασθενή, τον πιο σημαντικό ρόλο στην διάγνωση της ΚΝ έχουν η φυσική εξέταση και οι απεικονιστικές τεχνικές που επιτρέπουν την απεικόνιση των δομικών και μορφολογικών χαρακτηριστικών της αθηρωματικής πλάκας και μπορούν να είναι ιδιαίτερα χρήσιμες για τον προσδιορισμό των ευάλωτων πλακών και εν συνεχεία την πρόβλεψη της τελικής κλινικής έκβασης. Κατά την φυσική εξέταση ο γιατρός μπορεί να προβεί στις ακόλουθες ενέργειες:

- Γενική εξέταση αίματος και έλεγχος πήκτικότητας
- Ψηλάφηση σφίξεων
- Αχρόαση τραχήλου (φυσήματα)
- Οφθαλμολογική εκτίμηση
- Νευρολογική εκτίμηση
- Καρδιολογική εκτίμηση

Αντίστοιχα, οι πιο διαδεδομένες απεικονιστικές εξετάσεις [21] που μπορεί να χρησιμοποιήσει για την διάγνωση της ΚΝ είναι οι εξής:

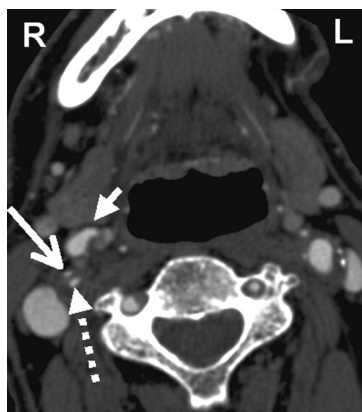
- **Duplex υπερηχογράφημα έγχρωμης ροής**, συνδυάζει τις πληροφορίες της συμβατικής υπερηχογραφικής απεικόνισης ενός αγγείου και τις πληροφορίες εξέτασης της ροής στο αγγείο με Doppler. Επίσης, με την εξέταση αυτή ο ιατρός βλέπει πώς ρέει το αίμα στο αγγείο και μετρά την ταχύτητα ροής [22]. Έτσι, μπορεί να ελέγξει την πορεία του αγγείου, να υπολογίσει το μέγεθός του και το βαθμό στένωσής του, αν υπάρχει. Το συμβατικό υπερηχογράφημα χρησιμοποιεί υπερήχους (κύματα ήχου τα οποία δεν μπορεί να ακούσει το ανθρώπινο αυτί). Ένας υπολογιστής μετατρέπει τα κύματα σε δισδιάστατες ασπρόμαυρες εικόνες, που λέγονται απεικόνιση B-mode. Το

υπερηχογράφημα με Doppler μετράει το πώς τα κύματα ήχου ανακλώνται σε κινούμενα αντικείμενα, εν προκειμένω στα ερυθρά αιμοσφαίρια του αίματος, και στέλνει τις πληροφορίες στον υπολογιστή, ο οποίος παράγει έγχρωμες εικόνες που απεικονίζουν τη ροή και ορισμένες αλλοιώσεις των αγγείων.



Σχήμα 1.7: Duplex υπερηχογράφημα έγχρωμης ροής φυσιολογικής καρωτίδας

- **Υπολογιστική Αξονική Αγγειογραφία (CTA - Computed Tomography Angiography)**, είναι μία μη επεμβατική σάρωση ακτίνων X η οποία παρέχει λεπτομερείς εικόνες ανατομικών δομών του εγκεφάλου μετά από ενδοφλέβια χορήγηση σκιαγραφικού υλικού, σε χρόνο ο οποίος εξαρτάται από την αρτηρία ή την φλέβα που επιθυμούμε να απεικονιστεί. Η συγκεκριμένη απεικονιστική μέθοδος παρέχει τις καλύτερα ποιοτικά απεικονίσεις τόσο των αιμοφόρων αγγείων, μέσω της αγγειογραφίας, όσο και των μαλακών ιστών μέσω της αξονικής τομογραφίας (CT) και δίνει στον γιατρό τη δυνατότητα να προσδιορίσει με ακρίβεια τη στένωση που έχει προκληθεί στην εκάστοτε αρτηρία.

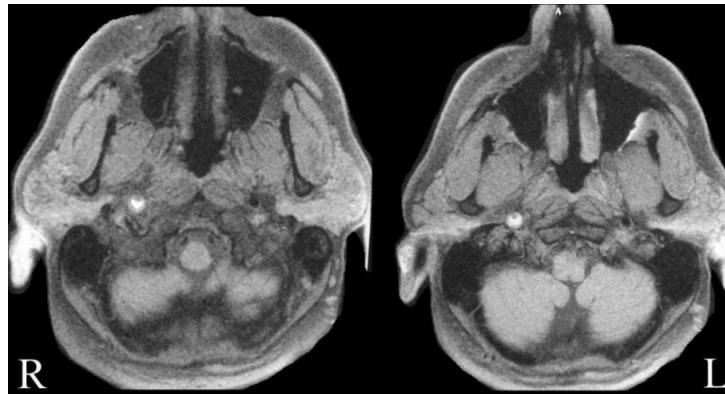


Σχήμα 1.8: Απεικόνιση δεξιάς σοβαρής στένωσης της καρωτίδας μέσω CTA

Στα αρνητικά της συγκεκριμένης μεθόδου συγκαταλέγονται η χρήση υψηλών δόσεων ακτινοβολίας μέσω των ακτίνων X, καθώς και η νεφροτοξικότητα των σκιαγραφικών υλικών που χορηγούνται στον ασθενή.

- **Μαγνητική Αγγειογραφία (MRA - Magnetic Resonance Angiography)**, πρόκειται για ακόμα μία μη επεμβατική απεικονιστική μέθοδο η οποία χρησιμοποιεί ένα ισχυρό μαγνητικό πεδίο, ραδιοκύματα και έναν ηλεκτρονικό υπολογιστή για να παράγει εικόνες του εγκεφάλου και των αρτηριών. Οι εικόνες που προκύπτουν από την MRA είναι πολύ λεπτομερείς και μπορούν να ανιχνεύσουν ακόμα και μικρά ΕΕ. Σε αντίθεση με την CTA η απεικονιστική μέθοδος MRA δεν επιβαρύνει τον ασθενή με

ακτινοβολία και το σκιαγραφικό υλικό που χρησιμοποιείται είναι λιγότερο επικίνδυνο στο να προκαλέσει κάποια αλλεργική αντίδραση.



Σχήμα 1.9: Απεικόνιση δεξιάς σοβαρής στένωσης της καρωτίδας μέσω MRI

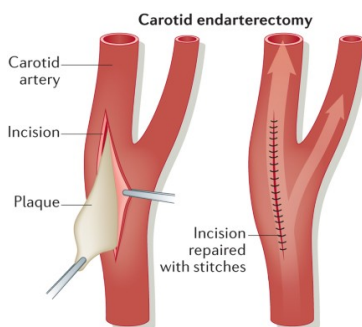
Παράλληλα υπάρχουν και άλλες όχι τόσο διαδεδομένες απεικονιστικές τεχνικές για την ανίχνευση της ευάλωτης πλάκας όπως η εγκεφαλική αγγειογραφία, το ευδαγγειακό υπερηχογράφημα (IVUS), η οπτική τομογραφία συνοχής (OCT) και η μοριακή απεικόνιση.

1.8 Τρόποι αντιμετώπισης ΚΝ

Με βάση τα παραπάνω στοιχεία συμπεραίνουμε πως η έγκυρη αξιολόγηση της επικινδυνότητας της ΚΝ για ΑΕΕ, είναι αναγκαία, έτσι ώστε να επιλέγεται για κάθε ασθενή η καλύτερη θεραπεία αντιμετώπισης της νόσου και να αποτρέπονται έγκαιρα ισχαιμικές επιπλοκές αυτής. Ανάλογα με την κατάσταση του κάθε ασθενή οι τρόποι της ΚΝ ποικίλουν. Ο θεράπων γιατρός καλείται να εκτιμήσει την εξέλιξη της νόσου αλλά και την φυσιολογική κατάσταση του ασθενούς και να επιλέξει ανάμεσα σε δύο θεραπευτικές μεθόδους. Αυτές είναι: η συντηρητική αγωγή με φάρμακα και βελτίωση του καθημερινού τρόπου ζωής (πιο υγιεινές συνήθειες) είτε η επεμβατική αποκατάσταση της βλάβης, η οποία μπορεί να επιτευχθεί μέσω ενδαρτηρεκτομής ή μέσω καρωτιδικής αγγειοπλαστικής με τοποθέτησης ενδονάρθηκα (stent). Οι τρεις μέθοδοι αντιμετώπισης περιγράφονται αναλυτικότερα στη συνέχεια.

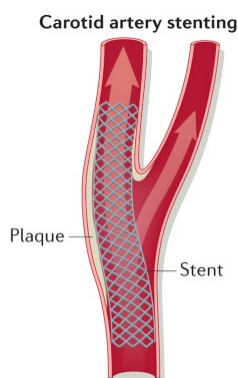
- Κατά την συντηρητική αγωγή ο ασθενής καλείται να λαμβάνει φάρμακα για την αντιμετώπιση της νόσου σε συνδυασμό με τη βελτίωση της καθημερινότητάς του. Τα φάρμακα περιλαμβάνουν τις στατίνες, τα αντιαιμοπεταλιακά και τα αντιυπερτασικά. Οι στατίνες σε συνδυασμό με μία υγιεινή διατροφή μπορούν να αντιμετωπίσουν την LDL χοληστερόλη κατά 25 – 30%. Αυτό θα έχει ως αποτέλεσμα τη μείωση της αθηρωματικής πλάκας καθώς όπως αναφέραμε και προηγουμένως το βασικό συστατικό για την δημιουργία της είναι η LDL χοληστερόλη. Τα αντιαιμοπεταλιακά φάρμακα (π.χ. ασπιρίνη) ενισχύουν την αραιώση του αίματος και αποτρέπουν την δημιουργία θρόμβων βοηθώντας έτσι στην φυσιολογική ροή του αίματος μέσα στις στενωμένες αρτηρίες. Τέλος τα αντιυπερτασικά φάρμακα βοηθούν στον έλεγχο της αρτηριακής πίεσης καθώς όπως αναφέραμε και προηγουμένως η υψηλή αρτηριακή πίεση είναι ένας σημαντικός παράγοντας για την εμφάνιση ΑΕΕ.
- Στην καρωτιδική ενδαρτηρεκτομή ο γιατρός διαχειρίζεται επεμβατικά την ΚΝ. Η επέμβαση γίνεται με γενική ή τοπική αναισθησία. Ο αγγειοχειρουργός μέσα από μία

τομή στον τράχηλο ανοίγει και καθαρίζει την καρωτίδα από την αθηρωματική πλάκα αναδημιουργώντας έναν λείο και βατό αυλό. Στη συνέχεια μέσω ενός εμφυτεύματος (patch) γίνεται σύγκλιση της αρτηριοτομής για την πρόληψη πιθανής μετεγχειρητικής στένωσης του αγγείου.



Σχήμα 1.10: Ενδαρτηρεκτομή καρωτίδας

- Η αγγειοπλαστική καρωτίδας με τοποθέτηση stent είναι μία σχετικά νέα και ελάχιστα επεμβατική τεχνική. Ο αγγειοχειρουργός με τοπική αναισθησία παρακεντεί τη μηριαία αρτηρία και με τη βοήθεια ειδικών συρμάτων και καθετήρων φτάνει στην περιοχή βλάβης της αρτηρίας. Εκεί με ειδικό μπαλόνι προκαλεί διαστολή της πλάκας και στη συνέχεια τοποθετεί έναν ενδονάρθηκα (stent) για να διατηρηθεί ανοικτή η καρωτιδική αρτηρία αποτρέποντα την φραγή του αγγείου λόγω της πλάκας.



Σχήμα 1.11: Αγγειοπλαστική καρωτίδας με τοποθέτηση stent

Σε σχέση με τις παραπάνω μεθόδους αντιμετώπισης, σύγχρονες μελέτες έδειξαν ότι οι δύο επεμβατικές τεχνικές μπορούν να οδηγήσουν σε ανεπιθύμητες επιπλοκές (ΑΕΕ, αιματώματα, καρδιακή προσβολή, κ.τ.λ.) τόσο κατά τη διάρκειά τους όσο και σε μετεγχειρητική φάση, έως και 30 μέρες μετά από την επέμβαση. Επομένως είναι πολύ σημαντικό ο εκάστοτε γιατρός να επιλέξει την σωστή θεραπεία στηριζόμενος σε όλες τις παραμέτρους που διαθέτει σχετικά με τον ασθενή του (εξέλιξη της νόσου, ηλικία, φυσικά κατάσταση κ.τ.λ.). Πιο συγκεκριμένα στις τελευταίες οδηγίες της Ευρωπαϊκής Καρδιολογικής Εταιρίας [23], το 2017 αναφέρονται τα εξής (Σχήμα. 1.12):

Όσον αφορά την αγγειοπλαστική καρωτίδας με τοποθέτηση stent:

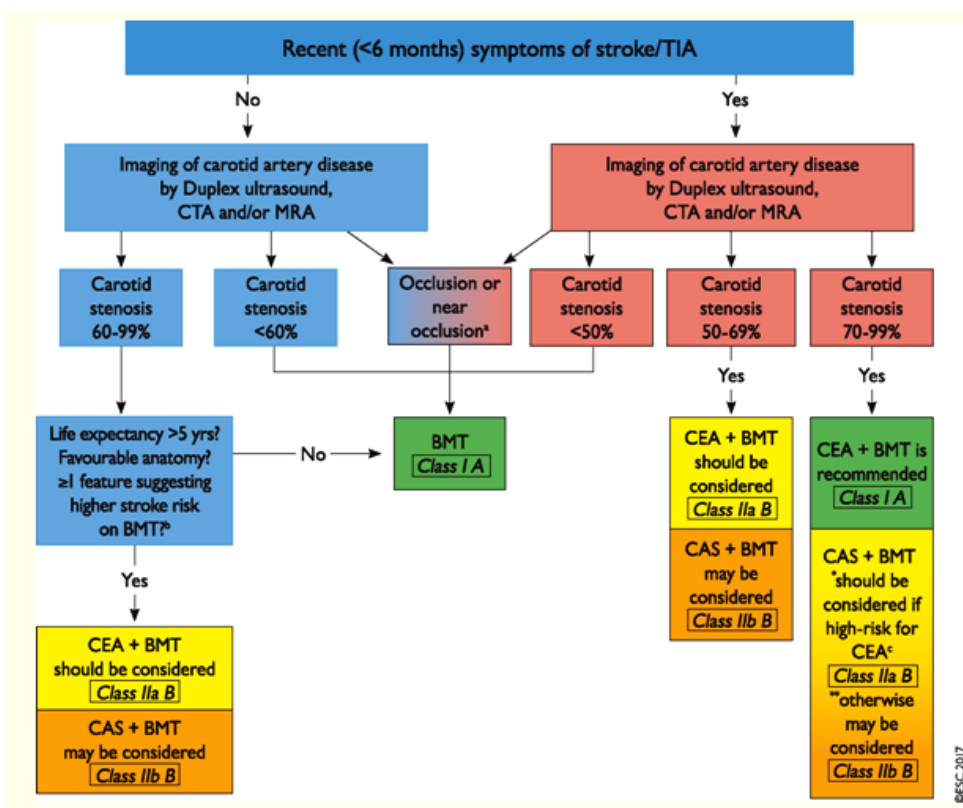
- Για τους ασυμπτωματικούς ασθενείς το όφελος είναι ελάχιστα μεγαλύτερο από τη ζημία. Επομένως ο ασθενής ακολουθεί φαρμακευτική αγωγή και η τοποθέτηση stent κρίνεται ανάλογα με τη φυσική κατάσταση του.

- Για τους συμπτωματικούς με στένωση της τάξεως 50% – 69% ισχύει ότι και για τους ασυμπτωματικούς ασθενείς.
- Για τους συμπτωματικούς με στένωση της τάξεως 70% – 99%, η τοποθέτηση stent μπορεί να διενεργηθεί σε ασθενείς που η ηλικία τους ξεπερνάει τα 80 έτη, που πάσχουν από σημαντική καρδιοπάθεια ή βαριά πνευμονοπάθεια, σε ύπαρξη προηγούμενης μεγάλης εγχείρησης ή ακτινοβολίας στο λαιμό αλλά και σε πλήρη απόφραξη της άλλης έσω καρωτίδας. Βέβαια όπως αναφέραμε και προηγουμένως υπάρχει ο κίνδυνος εμφράγματος και εγκεφαλικού μέσα στις επόμενες 30 ημέρες.

Όσον αφορά την ενδαρτηρεκτομή:

- Για τους ασυμπτωματικούς ασθενείς με στένωση της τάξεως 60% – 99%, μπορεί να διενεργηθεί αν το προσδόκιμο ζωής είναι μεγαλύτερο από 5 χρόνια, η ανατομία είναι κατάλληλη και υπάρχει μεγάλος κίνδυνος για ΑΕΕ αν ο ασθενής ακολουθήσει αποκλειστικά φαρμακευτική αγωγή.
- Για τους συμπτωματικούς με στένωση της τάξεως 70% – 99% η εν λόγω επεμβατική θεραπεία συνίσταται.
- Για τους συμπτωματικούς με στένωση της τάξεως 50% – 69%, θα μπορούσε να εφαρμοστεί αφού ο ασθενής ακολουθήσει πρώτα την καλύτερη φαρμακευτική αγωγή.

Σε κάθε άλλη περίπτωση η αντιμετώπιση της νόσου γίνεται αποκλειστικά με φαρμακευτική αγωγή.



Σχήμα 1.12: Αλγόριθμος εύρεσης κατάλληλης θεραπείας για την ΚΝ. BMT = Καλύτερη Φαρμακευτική Θεραπεία, CAS = Carotid Artery Stenting (Τοποθέτηση stent), CEA = Carotid Endarterectomy (Ενδαρτηρεκτομή)

Κεφάλαιο 2

Μηχανική μάθηση

2.1 Τεχνητή Νοημοσύνη και Μηχανική Μάθηση

2.1.1 Τεχνητή Νοημοσύνη

Ενας αλγόριθμος χαρακτηρίζεται ως αλγόριθμος Τεχνητής Νοημοσύνης (Artificial Intelligence - AI) όταν έχει την ικανότητα να αναπαράγει τις γνωστικές λειτουργίες ενός ανθρώπου. Δηλαδή να μπορεί να μαθαίνει, να σχεδιάζει, να αποφασίζει και να δημιουργεί. Μέσω της Τεχνητής Νοημοσύνης οι αλγόριθμοι "κατανοούν" το περιβάλλον τους και είναι σε θέση να επιλύσουν προβλήματα τα οποία αποσκοπούν στην επίτευξη ενός συγκεκριμένου στόχου. Πιο συγκεκριμένα, ο υπολογιστής λαμβάνει δεδομένα, τα επεξεργάζεται στηριζόμενος σε κάποιον αλγόριθμο και ανταποκρίνεται βάση αυτών με αυτονομία.

2.1.2 Μηχανική Μάθηση

Μηχανική μάθηση (Machine Learning - ML), είναι η ικανότητα ενός υπολογιστικού συστήματος να δημιουργεί μοντέλα ή πρότυπα από ένα σύνολο δεδομένων. Η μηχανική μάθηση αποτελεί κλάδο της Τεχνητής Νοημοσύνης και ασχολείται με τη μελέτη αλγορίθμων που βελτιώνουν τη συμπεριφορά τους σε κάποια εργασία που τους έχει ανατεθεί χρησιμοποιώντας την εμπειρία τους σε ένα συγκεκριμένο σύνολο δεδομένων και κάνοντας προβλέψεις σχετικά με αυτό. Οι αλγόριθμοι της μηχανικής μάθησης, στους οποίους στη συνέχεια θα αναφερόμαστε με τον όρο: μοντέλα, εκπαιδεύονται με τρόπο τέτοιο ώστε να είναι ικανοί να αναγνωρίζουν μοτίβα στα δεδομένα προκειμένου να λάβουν αποφάσεις, βασιζόμενες στην στατιστική, τη θεωρία πιθανοτήτων και τη βελτιστοποίηση. Ο τομέας της μηχανικής μάθησης αναπτύσσει τρεις μεθόδους μάθησης, οι οποίες προκύπτουν με βάση τους τρόπους με τους οποίους μαθαίνει ο ανθρώπινος εγκέφαλος: επιβλεπόμενη μάθηση, μη επιβλεπόμενη μάθηση και ενισχυτική μάθηση. Αναλυτικότερα:

- **Επιβλεπόμενη Μάθηση:** Σε αυτόν τον τρόπο μάθησης το μοντέλο για κάθε πιθανή είσοδο γνωρίζει την επιθυμητή έξοδο. Έτσι κατασκευάζει μία συνάρτηση η οποία απεικονίζει δεδομένες εισόδους σε γνωστές εξόδους με απώτερο σκοπό τη γενίκευση-βελτιστοποίηση της και σε εισόδους με άγνωστη έξοδο.
- **Μη Επιβλεπόμενη Μάθηση:** Σε αυτόν το τρόπο μάθησης το μοντέλο κατασκευάζει μία συνάρτηση για ένα σύνολο παρατηρήσεων χωρίς οι επιθυμητές έξοδοι να είναι γνωστές.

- **Ενισχυτική Μάθηση:** Σε αυτόν τον τρόπο μάθησης το μοντέλο μαθαίνει μία στρατηγική ενεργειών μέσα από άμεση αλληλεπίδραση με το περιβάλλον.

Στην παρούσα εργασία θα ασχοληθούμε με την πρώτη μέθοδο, την Επιβλεπόμενη Μάθηση.

Τέλος, αν και απέχουμε πάρα πολύ από τη δημιουργία αλγορίθμων που μαθαίνουν και αποφασίζουν τόσο καλά όσο ο άνθρωπος, τα μοντέλα που έχουν δημιουργηθεί μέχρι σήμερα έχουν επιτρέψει τη δημιουργία εμπορικών εφαρμογών, σε πολλαπλούς τομείς της καθημερινότητας, με σημαντική επιτυχία [24]. Πιο συγκεκριμένα, μεταξύ των κλάδων που επωφελούνται από την ανάπτυξη της Μηχανικής Μάθησης συγκαταλέγεται και η Ιατρική και πιο συγκεκριμένα επιστήμες όπως οι: Ακτινολογία, Ογκολογία, Νευρολογία, Καρδιολογία [25], [26] και πολλές ακόμα ειδικότητες [27].

2.2 Νευρωνικά Δίκτυα

Νευρωνικό Δίκτυο ονομάζεται ένα δίκτυο διασυνδεδεμένων νευρώνων. Στην περίπτωση των βιολογικών νευρώνων πρόκειται για ένα τμήμα νευρικού ιστού. Στην περίπτωση των τεχνητών νευρώνων πρόκειται για μία δομή η οποία αποτελείται από πολλούς νευρώνες οι οποίοι χωρίζονται σε επίπεδα και συνδέονται μεταξύ τους. Στην παρούσα εργασία θα ασχοληθούμε με τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks).

Κάθε ΤΝΔ είναι εμπνευσμένο από το Κεντρικό Νευρικό Σύστημα του ανθρώπινου οργανισμού και προσπαθεί να το προσομοιώσει. Τα βασικά δομικά στοιχεία ενός ΤΝΔ είναι οι νευρώνες, οι συνάψεις και τα βάρη τους, και οι συναρτήσεις ενεργοποίησης. Πιο συγκεκριμένα:

- Οι **Νευρώνες** αποτελούν το πιο σημαντικό δομικό στοιχείο ενός ΤΝΔ. Κάθε νευρώνας δέχεται ένα σύνολο αριθμητικών εισόδων, είτε από άλλους νευρώνες είτε από το περιβάλλον, επιτελεί έναν υπολογισμό με βάση αυτές και παράγει μία έξοδο η οποία με τη σειρά της είτε τροφοδοτείται σε έναν άλλο νευρώνα ως είσοδος, είτε αποτελεί την έξοδο του μοντέλου. Τα είδη των νευρώνων σε ένα δίκτυο-μοντέλο είναι τρία:
 - Οι *νευρώνες εισόδου*, οι οποίοι δεν επιτελούν κανέναν υπολογισμό και απλά μεσολαβούν ανάμεσα στα δεδομένα εισόδου και τους υπολογιστικούς νευρώνες.
 - Οι *υπολογιστικοί ή κρυμμένοι νευρώνες*, στους οποίους γίνεται η επεξεργασία των αριθμητικών δεδομένων.
 - Οι *νευρώνες εξόδου*, οι οποίοι διοχετεύουν στο περιβάλλον τις τελικές αριθμητικές εξόδους του δικτύου.

Σε κάθε μοντέλο, ένα σύνολο από νευρώνες με κοινά χαρακτηριστικά, ονομάζεται επίπεδο νευρώνων. Μεταξύ της εισόδου και της εξόδου του δικτύου μπορεί να έχουμε περισσότερα από ένα επίπεδα, τα λεγόμενα "κρυφά επίπεδα" (hidden layers) τα οποία αποτελούνται από τους υπολογιστικούς ή κρυμμένους νευρώνες.

- Οι **συνάψεις** και τα **βάρη** τους, μέσω των οποίων γίνεται η σύνδεση των νευρώνων μεταξύ τους. Με τον όρο σύναψη εννοούμε την σύνδεση ενός νευρώνα με έναν άλλο. Σε κάθε σύναψη αντιστοιχεί και ένα βάρος, το οποίο μεταβάλλεται με απώτερο σκοπό να οδηγήσει το μοντέλο στα βέλτιστα αποτελέσματα.
- Οι **συναρτήσεις ενεργοποίησης**, είναι οι λειτουργίες που εμπεριέχει και καλείται να εκτελέσει κάθε νευρώνας. Αναφέρονται επίσης και ως συναρτήσεις περιορισμού,

Γραμμική συνάρτηση (Linear)	$\phi(x) = x$
Βηματική συνάρτηση (Step)	$\phi(x) = \begin{cases} 1, & x > 0 \\ 0.5, & x = 0 \\ 0, & x < 0 \end{cases}$
Συνάρτηση "ράμπας" (ReLU)	$\phi(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$
Σιγμοειδής συνάρτηση (Sigmoid)	$\phi(x) = \frac{1}{1+e^{-x}}$
Συνάρτηση Υπερβολικής Εφαπτομένης (Hyperbolic Tangent)	$\phi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

Πίνακας 2.1: Συναρτήσεις ενεργοποίησης

καθώς περιορίζουν το επιτρεπτό εύρος πλάτους του σήματος εξόδου, του νευρώνα, σε κάποια πεπερασμένη τιμή. Η επιλογή της όποιας συνάρτησης ενεργοποίησης εξαρτάται από την εφαρμογή στην οποία αποσκοπεί η χρήση ενός ΤΝΔ καθώς επίσης και στην αρχιτεκτονική του συγκεκριμένου μοντέλου. Η συνάρτηση ενεργοποίησης μπορεί να είναι βηματική (step transfer function), γραμμική (linear transfer function), μη γραμμική (non-linear transfer function) και στοχαστική (stochastic transfer function). Στον Πίνακα. 2.1 φαίνονται ορισμένες από τις πιο γνωστές συναρτήσεις ενεργοποίησης.

Συμπερασματικά λοιπόν, το κύριο χαρακτηριστικό των νευρωνικών δικτύων που τα διαφοροποιεί από κάθε άλλο μηχανικό αλγόριθμο, είναι η ικανότητά τους να μαθαίνουν και να εξελίσσονται. Η διαδικασία της μάθησης επιτυγχάνεται μέσω της εκπαίδευσης (training), μια επαναληπτικής διαδικασίας που έχει ως σκοπό την σταδιακή προσαρμογή των παραμέτρων του δικτύου σε κατάλληλες τιμές ώστε να επιλύεται με επαρκή επιτυχία το εκάστοτε πρόβλημα. Αφού η εκπαίδευσή του ολοκληρωθεί οι τελικές τιμές των παραμέτρων διατηρούνται σταθερές και πλέον θεωρείται λειτουργικό. Δηλαδή, είναι σε θέση να δώσει ορθές εξόδους για εισόδους διαφορετικές από αυτές με τις οποίες εκπαιδεύτηκε.

2.2.1 Δομή Νευρωνικών Δικτύων

Όπως αναφέρθηκε και προηγουμένως ένα σύνολο από νευρώνες με κοινά χαρακτηριστικά, ονομάζεται επίπεδο νευρώνων. Κάθε μοντέλο ΤΝΔ αποτελείται από μια αλληλουχία επιπέδων παράλληλης διάταξης τα οποία δεν είναι ίδια μεταξύ τους και διαφοροποιούνται ως εξής:

- **Επίπεδο Εισόδου (input layer)**, αποτελεί το 1ο επίπεδο του μοντέλου και σε αυτό εμφανίζονται τα δεδομένα εισόδου. Ο αριθμός των νευρώνων του είναι ίσος με τις μεταβλητές εισόδου του μοντέλου.
- **Κρυφό Επίπεδο (hidden layer)**, ο αριθμός των κρυφών επιπέδων καθώς και ο αριθμός των νευρώνων κάθε κρυφού επιπέδου ορίζεται από τον κατασκευαστή του μοντέλου. Κάθε κρυφό επίπεδο συνδέει το προηγούμενο και το επόμενο κατά σειρά επίπεδο νευρώνων. Το εκάστοτε μοντέλο θεωρείται όλο και πιο πολύπλοκο ανάλογα με τον αριθμό των επιπέδων του. Όταν ο αριθμός των κρυφών επιπέδων είναι δύο

ή και παραπάνω αναφερόμαστε πλέον σε Βαθιά Νευρωνικά Δίκτυα (Deep Learning Models).

- **Επίπεδο Εξόδου (output layer)**, αποτελεί το τελευταίο επίπεδο του μοντέλου και σε αυτό εμφανίζεται η τελική έξοδος. Ο αριθμός των νευρώνων αυτού του επιπέδου είναι ίσος με τις πιθανές μεταβλητές εξόδου.

Ανάλογα με την αρχιτεκτονική τους και τον τρόπο με τον οποίο συνδέονται οι νευρώνες τα ΤΝΔ κατηγοριοποιούνται σε: Νευρωνικά Δίκτυα με Πρόσθια Τροφοδότηση και Νευρωνικά Δίκτυα με Ανατροφοδότηση. Αναλυτικότερα:

- **Νευρωνικά Δίκτυα με Πρόσθια Τροφοδότηση (feedforward)**, σε αυτά τα μοντέλα το σήμα μεταφέρεται από την είσοδο έως την έξοδο μέσω των νευρώνων, χωρίς να μπορεί να συμβεί το αντίστροφο.
- **Νευρωνικά Δίκτυα με Ανατροφοδότηση (feedback)**, όπου το ΤΝΔ περιέχει τουλάχιστον έναν βρόγχο ανατροφοδότησης. Δηλαδή, τουλάχιστον ένα επίπεδο νευρώνων έχει το ρόλο να ανατροφοδοτεί το σήμα εξόδου πίσω στις εισόδους όλων των άλλων νευρώνων.

Στην παρούσα εργασία θα ασχοληθούμε αποκλειστικά με μοντέλα πρόσθιας τροφοδότησης.

2.2.2 Εκπαίδευση Νευρωνικών Δικτύων

Προηγουμένως εξετάσαμε τη δομή των νευρωνικών δικτύων και τον τρόπο επικοινωνίας μεταξύ των νευρώνων. Στη συνέχεια είναι σημαντικό να επεξηγηθεί και ο τρόπος εκπαίδευσής τους. Καθ' όλη τη διαδικασία εκπαίδευσης ενός ΤΝΔ, χρησιμοποιείται ένας αλγόριθμος διόρθωσης του σφάλματος που είναι γνωστός ως Αλγόριθμος Σύγκλισης ή Αλγόριθμος Εκπαίδευσης. Ο αλγόριθμος αυτός καθορίζει τον τρόπο με τον οποίο μεταβάλλονται τα συναπτικά βάρη του δικτύου σε κάθε επανάληψη, καθώς και τον τρόπο με τον οποίο βελτιστοποιείται το παραγόμενο αποτέλεσμα. Ο συνηθέστερος αλγόριθμος εκπαίδευσης είναι ο αλγόριθμος Οπισθοδιάδοσης του Σφάλματος (Backpropagation Error Algorithm), ο οποίος χρησιμοποιείται κυρίως στις εφαρμογές επιβλεπόμενης μάθησης και ειδικά στα ΤΝΔ πρόσθιας τροφοδότησης πολλών επιπέδων. Για να χρησιμοποιηθεί ο αλγόριθμος της οπισθοδιάδοσης του σφάλματος πρέπει το ΤΝΔ να αποτελείται από τουλάχιστον ένα κρυφό επίπεδο και η συνάρτηση ενεργοποίησης που χρησιμοποιεί να είναι μη γραμμική (π.χ. η σιγμοειδής συνάρτηση ενεργοποίησης). Τα βασικά δομικά στοιχεία του αλγόριθμου οπισθοδιάδοσης είναι η συνάρτηση κόστους (loss function), ο βελτιστοποιητής (optimizer) που έχει ως παράμετρο τον ρυθμό εκπαίδευσης (learning rate) και ο αριθμός των εποχών εκπαίδευσης (epochs). Τα βήματα που ακολουθεί ο αλγόριθμος περιγράφονται στη συνέχεια:

- Αρχικά τα συναπτικά βάρη του δικτύου λαμβάνουν τυχαίες τιμές.
- Στη συνέχεια, σε κάθε εποχή του αλγορίθμου η διαδικασία που ακολουθείται είναι η εξής:
 - Υπολογίζεται το σφάλμα που προκύπτει μέσω της συνάρτησης κόστους.
 - Έπειτα τα βάρη του δικτύου ανανεώνονται με τρόπο που υποδεικνύεται από τον βελτιστοποιητή που έχει επιλέξει ο κατασκευαστής του δικτύου. Η ανανέωση των βαρών εξαρτάται από την τιμή του σφάλματος και του βήματος εκπαίδευσης και

έχει ως στόχο την ελαχιστοποίηση του σφάλματος του ανανεωμένου, πλέον δικτύου. Για παράδειγμα, στην περίπτωση του Gradient Descent optimizer τα βάρη μεταβάλλονται κατά την αρνητική κατεύθυνση της παραγώγου του σφάλματος πολλαπλασιασμένη με το βήμα εκπαίδευσης.

- Τα βήματα αυτά επαναλαμβάνονται ανάλογα με τον αριθμό των εποχών που έχουμε επιλέξει. Ο αριθμός των εποχών επιλέγεται έτσι ώστε η παραπάνω διαδικασία να συγκλίνει σε σχεδόν μηδενικό σφάλμα είτε σε σφάλμα που μπορεί να θεωρηθεί ανεκτό.

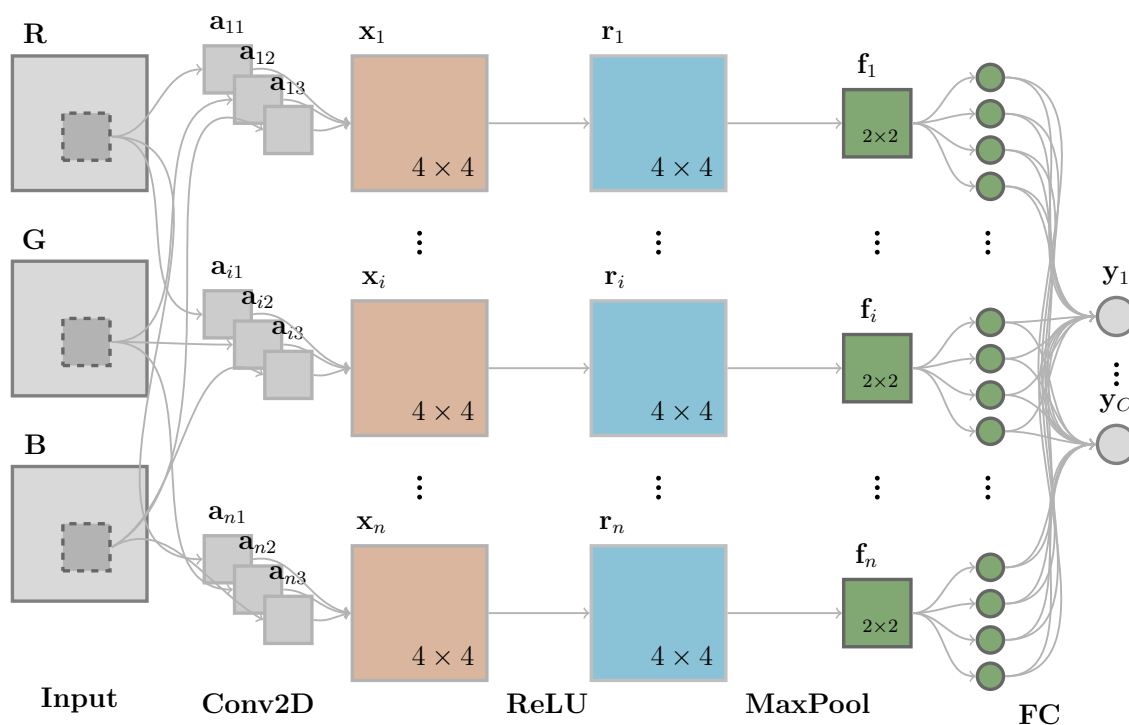
2.3 Συνελικτικά Νευρωνικά Δίκτυα

Τα τελευταία χρόνια έχει επιτευχθεί μεγάλη πρόοδος στην αναγνώριση προτύπων και την όραση υπολογιστών μέσω των αρχιτεκτονικών δικτύων βαθιάς μάθησης. Η έρευνα έχει στραφεί σε δίκτυα που μπορούν να διαχειριστούν αποτελεσματικά δεδομένα εικόνων. Τέτοια δίκτυα είναι τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN). Ένα CNN ορίζεται ένα δίκτυο ΤΝΔ το οποίο περιέχει τουλάχιστον ένα συνελικτικό επίπεδο. Ο όρος "συνελικτικό" προέρχεται από την γραμμική μαθηματική πράξη της συνέλιξης. Μάλιστα, σε ένα συνελικτικό επίπεδο πραγματοποιείται διδιάστατη συνέλιξη μεταξύ των εικόνων εισόδου και των διδιάστων συνελικτικών φίλτρων του επιπέδου. Τα φίλτρα αυτά ονομάζονται αλλιώς και πυρήνες (kernels). Η έξοδος του επιπέδου εξαρτάται από τις παραμέτρους που αφορούν το μέγεθος των συνελικτικών φίλτρων, το βήμα της συνέλιξης, το padding των εικόνων, κλπ. Τα CNN αποτελούν επέκταση των βαθιών νευρωνικών δικτύων αφού στο τελικό τους τμήμα περιέχουν πλήρως συνδεδεμένα επίπεδα. Στο αρχικό τους τμήμα τα συνελικτικά επίπεδα αποτελούν ένα πολύ αποτελεσματικό τρόπο εξαγωγής χαρακτηριστικών των εικόνων εισόδου.

Για την κατασκευή ενός CNN, χρησιμοποιούνται τρεις τύποι επιπέδων: το Συνελικτικό Επίπεδο (Convolutional Layer), το επίπεδο Ενεργοποίησης (Activation Layer), το Συγκεντρωτικό Επίπεδο (Pooling Layer) και Πλήρως Συνδεδεμένο Επίπεδο (Fully-Connected Layer). Στο παρακάτω σχήμα φαίνεται ένα παράδειγμα ενός απλοποιημένου δικτύου CNN για συγκεκριμένα αριθμητικά δεδομένα το οποίο δέχεται ως είσοδο μία εικόνα με διαστάσεις 4×4 αποτελούμενη από τρία κανάλια χρώματος R,G,B και δίνει ως έξοδο ένα διάνυσμα $\mathbf{y}_1, \dots, \mathbf{y}_C$.

Αναλυτικά, τα επίπεδα του δικτύου του σχήματος 2.1 περιγράφονται ως εξής:

- Το επίπεδο εισόδου **Input**, περιλαμβάνει την εικόνα εισόδου που αποτελείται από 3 κανάλια (R, G, B).
- Στη συνέχεια στο συνελικτικό επίπεδο **Conv2D** πραγματοποιείται συνέλιξη των φίλτρων \mathbf{a}_{i1} , \mathbf{a}_{i2} , \mathbf{a}_{i3} με τα τρία κανάλια εισόδου για κάθε $i = 1, \dots, n$. Τελικά προκύπτουν n καινούριες εικόνες ίδιας διάστασης.
- Στη συνέχεια ακολουθεί το επίπεδο που εφαρμόζεται η συνάρτηση ενεργοποίησης, που εν προκειμένω έχει επιλεγεί η **ReLU** ενεργοποίηση. Κάθε pixel των εικόνων εξόδου δίνεται από την σχέση $\mathbf{r}_i = \max\{\mathbf{x}_i, 0\}$. Το επίπεδο αυτό διατηρεί τις διαστάσεις αναλλοίωτες.
- Το συγκεντρωτικό επίπεδο, που στην προκειμένη περίπτωση είναι ένα **MaxPooling** επίπεδο, πραγματοποιεί υποδειγματοληψία στις χωρικές διαστάσεις της εικόνας ανάλογα με το "βήμα" που θα ορίσουμε. Στην περίπτωσή μας θεωρούμε ότι το "βήμα"



Σχήμα 2.1: Παράδειγμα CNN αποτελούμενο από: Επίπεδο εισόδου (Input), Επίπεδο διαστάσης συνέλιξης (Conv2D), Επίπεδο ενεργοποίησης (ReLU), Συγκεντρωτικό Επίπεδο (MaxPool) και Πλήρως συνδεδεμένο επίπεδο (FC)

είναι 2, επομένως οι τελικές διαστάσεις της εξόδου θα είναι (2×2) . Τα κανάλια παραμένουν σε πλήθος n .

- Τέλος, οι εικόνες αφού πρώτα ξεδιπλωθούν (flatten) σε ένα ενιαίο διάνυσμα, δίνουν την έξοδο y μέσω ενός πλήρους συνδεδεμένου επιπέδου **FC**. Τελικά παίρνουμε τις εξόδους y_1, \dots, y_C ως την τελική πρόβλεψη του μοντέλου, όπου C είναι το πλήθος των κλάσεων του προβλήματος που καλούμαστε να επιλύσουμε.

2.4 Μετρικές αξιολόγησης των μοντέλων μηχανικής μάθησης

Για την αξιολόγηση του κάθε μοντέλου επιλέξαμε να χρησιμοποιήσουμε τις μετρικές Accuracy, Sensitivity, Specificity, F1-score και RoC/AUC.

Η μετρική Accuracy αντιπροσωπεύει το ποσοστό των δειγμάτων τα οποία ταξινομήθηκαν σωστά. Δηλαδή, προσδιορίζει τον λόγο που προκύπτει από τη διαίρεση των συνολικών σωστών προβλέψεων του μοντέλου προς τον αριθμό των συνολικών προβλέψεων που έγιναν. Επομένως δίνεται από τον τύπο:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

όπου :

$$TP = TruePositive$$

$$TN = TrueNegative$$

$$FP = FalsePositive$$

$$FN = FalseNegative$$

Η μετρική Sensitivity ή αλλιώς Recall προσδιορίζει το ποσοστό των αληθώς θετικών δειγμάτων προς όλα τα θετικά δείγματα και υπολογίζεται με βάση τον παρακάτω τύπο.

$$Sensitivity = \frac{TP}{TP + FN}$$

Εδώ ως θετικά δείγματα αναφέρουμε αυτά στα οποία ο ασθενής ανήκει στην κατηγορία του υψηλού κινδύνου. Πιο απλά θα μπορούσαμε να πούμε πως η μετρική Sensitivity καθορίζει πόσο καλά έχει προβλέψει το εκάστοτε μοντέλο τις περιπτώσεις όπου ο ασθενής θεωρείται υψηλού κινδύνου και καλείται να ακολουθήσει επεμβατική θεραπεία.

Η μετρική Specificity προσδιορίζει το ποσοστό των αληθώς αρνητικών δειγμάτων προς όλα τα αρνητικά δείγματα και υπολογίζεται με βάση τον παρακάτω τύπο.

$$Specificity = \frac{TN}{TN + FP}$$

Στο πρόβλημά μας ως αρνητικά δείγματα θεωρούμε τους ασθενείς που ανήκουν στην κατηγορία του χαμηλού κινδύνου και καλούνται να ακολουθήσουν συντηρητική αγωγή.

Ακόμα μία μετρική που θα χρησιμοποιήσουμε είναι η F1-score η οποία προκύπτει από τον παρακάτω μαθηματικό τύπο:

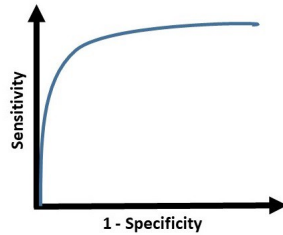
$$F1score = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

Ορισμός της μετρικής recall έγινε παραπάνω, αντίστοιχα η μετρική precision προσδιορίζει το ποσοστό των αληθώς θετικών δειγμάτων προς όλα τα δείγματα που προβλέφθηκαν θετικά και υπολογίζεται με βάση τον παρακάτω τύπο.

$$Recall = \frac{TP}{TP + FN}$$

Η μετρική F1-score αποτελεί τον αρμονικό μέσο των μετρικών recall και precision και χρησιμοποιείται όταν θέλουμε να ελέγξουμε αν υπάρχει "ισορροπία" ανάμεσα σε αυτές τις δύο μετρικές. Επίσης η μετρική F1-score είναι ευρέως χρησιμοποιούμενη και δίνει αντιπροσωπευτικά αποτελέσματα σε προβλήματα που διαχειρίζονται σύνολα δεδομένων με μεγάλες διαφορές στην κατανομή των κλάσεων, δηλαδή imbalanced data όπως και αυτό που θα χρησιμοποιήσουμε στην παρούσα εργασία.

Τέλος θα χρησιμοποιήσουμε τη μετρική RoC/AUC. Η καμπύλη RoC προσδιορίζει την ικανότητα του μοντέλου να διαχωρίζει τις κλάσεις στα διάφορα κατώφλια (threshold). Η AUC προσδιορίζει το εμβαδόν της περιοχής που σχηματίζεται κάτω από τη καμπύλη RoC. Πιο απλά θα μπορούσαμε να πούμε πως η μετρική RoC/AUC μας φανερώνει την ικανότητα του μοντέλου να διαχωρίζει τις δύο κλάσεις στις οποίες χωρίζονται τα δεδομένα μας. Όσο μεγαλύτερη είναι η τιμή της AUC τόσο καλύτερα γίνεται ο διαχωρισμός ανάμεσα στους ασθενείς που ανήκουν στην κλάση χαμηλού κινδύνου από αυτούς που ανήκουν στην κλάση του υψηλού κινδύνου. Η καμπύλη RoC/AUC προκύπτει από τη γραφική παράσταση των δύο παραπάνω μετρικών όπως φαίνεται στο Figure 5.30.



Σχήμα 2.2: Καμπύλη Roc

2.5 Διασταυρωμένη Επικύρωση

Η διασταυρωμένη επικύρωση ή αλλιώς cross validation (CV) είναι μία τεχνική που χρησιμοποιείται για την αξιολόγηση των μοντέλων μηχανικής μάθησης [28]. Το αρχικό σύνολο δεδομένων χωρίζεται σε k ίσα μέρη και κάθε μέρος χρησιμοποιείται διαδοχικά για την αξιολόγηση του μοντέλου (test) ενώ τα υπόλοιπα μέρη για την εκπαίδευση και την επικύρωσή του (train, validation). Η τελική αξιολόγηση του μοντέλου προκύπτει από το μέσο όρο των k εκτιμήσεων που προέκυψαν οδηγώντας μας σε μία αξιόπιστη εκτίμηση για το τελικό σφάλμα του εκάστοτε μοντέλου μας.

Στην παρούσα εργασία έχουμε 51 διαφορετικά δείγματα ασθενών και στο σύνολο των υπερήχων β-σάρωσης και στα κλινικά δεδομένα. Με βάση το μέγεθος του συνόλου μας επιλέξαμε να εφαρμόσουμε διασταυρωμένη επικύρωση 5 μερών (5 cross validation - 5CV). Για να διατηρήσουμε τις αναλογίες των δύο κλάσεων σταθερές σε κάθε ένα από τα σύνολα εκπαίδευσης, επικύρωσης και αξιολόγησης ακολουθήσαμε τον διαχωρισμό που θα περιγράψουμε στη συνέχεια, ξεχωριστά σε κάθε μία από αυτές.

Όσον αφορά το σύνολο των υπερήχων β-σάρωσης, όπως αναφέραμε και προηγουμένως, διαθέτουμε δεδομένα από 51 διαφορετικούς ασθενείς, όμως οι απεικονίσεις που έχουμε είναι περισσότερες καθώς σε ορισμένους ασθενείς αντιστοιχούν παραπάνω από μία εικόνες. Στην παρούσα εργασία επιλέξαμε να εκπαιδύσουμε και να αξιολογήσουμε τα μοντέλα μας με όλες τις διαθέσιμες εικόνες. Κατά τον διαχωρισμό των δεδομένων οι εικόνες που ανήκουν στον ίδιο ασθενή θα κατανεμηθούν είτε στο σύνολο εκπαίδευσης (training) είτε στο σύνολο αξιολόγησης (test) είτε στο σύνολο επικύρωσης (validation).

2.6 Διαχείριση μη ισορροπημένων δεδομένων

Τα μη ισορροπημένα δεδομένα είναι ένα πρόβλημα στην μηχανική μάθηση. Με τον όρο μη ισορροπημένα δεδομένα εννοούμε ότι τα στιγμιότυπα ενός εκ των δύο τάξεων είναι περισσότερα σε αριθμό, δηλαδή το πλήθος των παρατηρήσεων δεν είναι ίδιο για όλες τις τάξεις σε ένα σύνολο δεδομένων. Το πρόβλημα αυτό μπορεί να εμφανιστεί και σε προβλήματα δύο κλάσεων (binary class) και σε προβλήματα με μεγαλύτερο αριθμό τάξεων (multi class). Στη συνέχεια θα αναλύσουμε τις πιο γνωστές τεχνικές για την αντιμετώπιση των μη ισορροπημένων δεδομένων.

2.6.1 Συλλογική Μάθηση

Η βασική ιδέα της Συλλογικής Μάθησης (Ensemble Learning) είναι να δημιουργήσει έναν αλγόριθμο μάθησης (Ensemble Neural Networks - ENN) συνδυάζοντας τις προβλέψεις τουλάχισ-

τον δύο μεμονωμένων εκπαιδευμένων μοντέλων [29]. Πιο συγκεκριμένα κάθε μεμονωμένο μοντέλο εκπαιδεύεται αυτόνομα στο σύνολο εκπαίδευσης και στη συνέχεια συνεργάζονται για την τελική απόφαση-έξοδο του ENN. Η μέθοδος της συλλογικής μάθησης συμβάλλει στο να μειωθεί η διακύμανση της πρόβλεψης και να μην υπάρχει μεγάλη ευαισθησία σε μικρές αλλαγές της εισόδου. Η λογική πίσω από τη μέθοδο της Συλλογικής Μάθησης στηρίζεται πως το από κοινού αποτέλεσμα του ENN θα έχει καλύτερη ακρίβεια κατά μέσο όρο, από τα αποτελέσματα των επιμέρους μοντέλων που το απαρτίζουν. Η παραπάνω πρόταση επιβεβαιώνεται από πολλαπλές θεωρητικές και πειραματικές έρευνες.

Παρόλο που υπάρχει πληθώρα μεθοδολογιών μέσω των οποίων μπορεί να δημιουργηθεί ένα ENN, δύο είναι αυτές οι οποίες χρησιμοποιούνται συνηθέστερα. Τις μεθόδους αυτές θα τις αναλύσουμε στη συνέχεια.

- **Bagging:** Η τεχνική Bagging αναπτύχθηκε επίσημα από τον Leo Breiman το 1996. Στην τεχνική αυτή κάθε μεμονωμένο μοντέλο του ENN εκπαιδεύεται αυτόνομα με ένα υποσύνολο των δεδομένων εκπαίδευσης και η τελική πρόβλεψη του ENN προκύπτει από μία απλή ψηφοφορία των μοντέλων που το απαρτίζουν. Πιο συγκεκριμένα αν το αρχικό σύνολο εκπαίδευσης αποτελείται από n δείγματα και το ENN από 3 αυτόνομα μοντέλα, τότε κάθε ένα από αυτά τα μοντέλα θα εκπαιδευτεί με $\frac{n}{3}$ στο πλήθος δείγματα τα οποία θα αποτελούν υποσύνολο των n αρχικών δειγμάτων εκπαίδευσης. Με αυτόν τον τρόπο επιτυγχάνεται η εκπαίδευση ενός συνόλου ανεξάρτητων ταξινομητών, οι οποίοι έχουν κατασκευαστεί με βάση διαφορετικά σύνολα εκπαίδευσης.
- **Boosting:** Στην τεχνική Boosting όπως και στην τεχνική Bagging η βασική ιδέα είναι ο συνδυασμός μεμονωμένων μοντέλων για την βελτιστοποίηση της ακρίβειας του τελικού ENN. Η κεντρική ιδέα της τεχνικής Boosting είναι η εφαρμογή διαδοχικών ομοιογενών μοντέλων μηχανικής μάθησης, όπου το καθένα από αυτά προσπαθεί να βελτιώσει την ακρίβεια του τελικού μοντέλου εστιάζοντας στα σφάλματα του προηγούμενου μοντέλου. Αυτό έχει ως αποτέλεσμα κάθε νέο μοντέλο να δίνει μεγαλύτερη βαρύτητα στις λανθασμένες, από τα προηγούμενα μοντέλα, παρατηρήσεις με απώτερο σκοπό να μειώσει το σφάλμα όσο περισσότερο γίνεται. Πιο συγκεκριμένα, ο αλγόριθμος boosting εκπαιδεύει τον αρχικό μοντέλο με ίσα βάρη για όλα τα δεδομένα εκπαίδευσης και στην συνέχεια εκπαιδεύει τα υπόλοιπα μοντέλα ανακατανέμοντας τα βάρη στο σύνολο εκπαίδευσης με τέτοιο τρόπο ώστε σε κάθε επόμενο γύρο τα βάρη των λάθους ταξινομημένων δειγμάτων να αυξάνονται. Η τελική πρόβλεψη προκύπτει από ψηφοφορία μεταξύ των μοντέλων. Σε αυτήν την κατηγορία μοντέλων ανήκουν αλγόριθμοι όπως ο XGBoost και ο AdaBoost.
- **Stacking:** Η κεντρική ιδέα πίσω από την τεχνική Stacking είναι ο συνδυασμός διαφορετικών μοντέλων μηχανικής μάθησης και ενός επιπλέον μοντέλου το οποίο εκπαιδεύεται κατάλληλα ώστε να συνδυάσει τις προβλέψεις των μεμονωμένων μοντέλων και να προκύψει η τελική έξοδος. Τα μοντέλα που απαρτίζουν το ENN μοντέλο ονομάζονται Μοντέλα Επιπέδου-0 (level-0 models) και το μοντέλο που χρησιμοποιείται για τον συνδυασμό των προβλέψεων αναφέρεται ως Μοντέλο Επιπέδου-1 (level-1 model).

2.6.2 Εκπαίδευση με ευαισθησία κόστους

Τα περισσότερα μοντέλα μηχανικής μάθησης υποθέτουν πως κάθε λανθασμένη πρόβλεψη αντιστοιχεί στην ίδια "ποινή"-κόστος. Η υπόθεση αυτή δεν ισχύει στις συνθήκες του πραγματικού κόσμου. Για παράδειγμα σε ένα πρόβλημα ταξινόμησης ασθενών με καρκίνο, μία

λάθος αρνητική (False Negative - FN) διάγνωση είναι πολύ πιο επικίνδυνη για τη ζωή του ασθενή, άρα θα πρέπει να επιδέχεται και μεγαλύτερη ποινή, από μία λάθος θετική πρόβλεψη (False Positive - FP). Κατά την εκπαίδευση με ευαισθησία κόστους (Cost-Sensitive Learning - CSL), η "ποινή" που επιβάλλεται στο μοντέλο σε περίπτωση λανθασμένης ταξινόμησης ενός δείγματος που ανήκει στην κλάση της μειοψηφίας είναι σημαντικά μεγαλύτερη από την αντίστοιχη ποινή που αποδίδεται για ένα λάθος ταξινομημένο δείγμα που ανήκει στην τάξη της πλειοψηφίας. Με τον τρόπο αυτό αντισταθμίζεται το μικρό πλήθος των παραδειγμάτων εκπαίδευσης, με την προσπάθεια αποφυγής λάθους που θα επιφέρει μεγάλη ποινή.

2.6.3 Υπερδειγματοληψία

Όπως και οι προηγούμενες τεχνικές που αναφέραμε έτσι και η τεχνική της υπερδειγματοληψίας χρησιμοποιείται για την εξισορρόπηση των άνισων κατηγοριών δεδομένων. Όταν το πλήθος των δεδομένων μίας κλάσης είναι ανεπαρκές, τότε η μέθοδος της υπερδειγματοληψίας προσπαθεί να τα εξισορροπήσει αυξάνοντας το πλήθος των "σπάνιων" δειγμάτων. Μία βασική τεχνική της υπερδειγματοληψίας είναι η τεχνική SMOTE (Synthetic Minority Over-sampling Technique). Στην τεχνική SMOTE η κλάση της μειοψηφίας εμπλουτίζεται με συνθετικά δεδομένα τα οποία δημιουργούνται εκτελώντας συγκεκριμένες διεργασίες πάνω στα δεδομένα της κλάσης μειοψηφίας.

2.6.4 Υποδειγματοληψία

Σε αντίθεση με την μέθοδο της υπερδειγματοληψίας αυτή η τεχνική προσπαθεί να εξισορροπήσει το σύνολο των δεδομένων μειώνοντας τον αριθμό των δειγμάτων που ανήκουν στην τάξη της πλειοψηφίας, αφαιρώντας τυχαία σύνολα δειγμάτων από αυτό. Η μέθοδος Cluster Centroid και η μέθοδος Tomek Link, είναι δύο από τις μεθόδους που μπορούν να χρησιμοποιηθούν για την εφαρμογή της υποδειγματοληψίας σε ένα σύνολο δεδομένων.

2.7 Εφαρμογές της μηχανικής μάθησης στην καρωτιδική νόσο

Όπως αναφέραμε και στο 1ο Κεφάλαιο, η καρωτιδική νόσος αποτελεί την βασικότερη αιτία των ΑΕΕ σε παγκόσμιο επίπεδο. Επομένως η έγκαιρη και σωστή διάγνωση της αθηροσκλήρωσης των καρωτίδων είναι πολύ σημαντική για την αξιολόγηση της σοβαρότητας της αγγειακής νόσου καθώς και για την εύρεση της κατάλληλης θεραπευτικής λύσης και κατ' επέκταση για την πρόληψη ενός ΑΕΕ. Είτε με συντηρητική αγωγή είτε επεμβατικά μέσω ενδαρτηρεκτομής ή αγγειοπλαστικής με τοποθέτηση stent. Η αποτελεσματική εφαρμογή μηχανικών μοντέλων πρόβλεψης είναι απαραίτητη για την έγκαιρη και εύστοχη αντιμετώπιση της [30], [31], [32]. Η ραγδαία ανάπτυξη της τεχνητής νοημοσύνης και της μηχανικής μάθησης αποτέλεσαν σημαντικούς παράγοντες για τη δημιουργία μοντέλων που εκπληρώνουν αυτό τον σκοπό [33]. Στη συνέχεια γίνεται αναφορά σε κάποιες από τις μελέτες που έγιναν με σκοπό την αντιμετώπιση και τη διαχείριση της καρωτιδικής νόσου.

Ο Lekadir και άλλοι [4] πρότειναν ένα συνελικτικό νευρωνικό δίκτυο, CNN, με σκοπό την κατάταξη των μερών της αθηρωματικής πλάκας σε λιπιδικό πυρήνα, ινώδες κάλυμμα και ασβεστοποιημένο ιστό. Αναλυτικότερα, οι συγγραφείς χρησιμοποίησαν 90.000 patches (μέρη της εικόνας) από 50 υπερηχογραφικές απεικονίσεις καρωτίδας για την εκπαίδευση και την αξιολόγηση του μοντέλου τους και η τελική συσχέτιση ανάμεσα στα αποτελέσματα που

προέκυψαν από το CNN και από την κλινική εκτίμηση των περιοχών του αθηρώματος, είναι της τάξεως του 90%.

Στην μελέτη [6] ο Skandha και άλλοι χρησιμοποίησαν τεχνικές βαθιάς μάθησης με σκοπό την κατηγοριοποίηση 346 υπερηχογραφικών απεικονίσεων των καρωτιδικών αρτηριών σε αθηρώματα ασυμπτωματικών και συμπτωματικών ασθενών. Τα τελικά αποτελέσματα ήταν αρκετά ικανοποιητικά και πιο συγκεκριμένα η μετρική Accuracy ήταν ίση με 95.66% και η RoC/AUC ίση με 0.956.

Μια ακόμα εργασία που ασχολήθηκε με την καρωτιδική νόσο ήταν η [5]. Στην εργασία αυτή οι συγγραφείς δημιούργησαν ένα νέο σύστημα διάγνωσης με τη βοήθεια υπολογιστή για την διάγνωση της καρωτιδικής νόσου. Ειδικότερα, χρησιμοποίησαν εικόνες υπερήχων της καρωτίδας με στόχο τον εντοπισμό και την κατάτμηση της αθηρωματικής πλάκας σε εικόνες υπερήχων β-σάρωσης αλλά και την κατηγοριοποίηση αυτής σε φυσιολογική ή μη φυσιολογική, με βάση το πάχος του έσω-μέσου χιτώνα της κοινής καρωτιδικής αρτηρίας (IMT), χρησιμοποιώντας τη μεταφερόμενη μάθηση (transfer learning) και πιο συγκεκριμένα το προεκπαιδευμένο συνελικτικό δίκτυο Alex Net.

Ακόμα μία ενδιαφέρουσα μελέτη σχετικά με την κατηγοριοποίηση της αθηρωματικής πλάκας ήταν του Wei Ma και άλλων [7]. Στην εργασία αυτή ο στόχος ήταν η κατάταξη του αθηρώματος μέσω συνελικτικών δικτύων με βάση την ηχογένειά του. Πιο συγκεκριμένα, συλλέχθηκαν 1463 απεικονίσεις αθηρωματικής πλάκας της καρωτίδας από τις οποίες οι 335 πλάκες χαρακτηρίστηκαν ως πλούσιες σε ηχογένεια, οι 405 ενδιάμεσης ηχογένειας ενώ οι 723 ήταν ηχολιτικές (echolucent) πλάκες. Οι καταγραφές αυτές χρησιμοποιήθηκαν για την εκπαίδευση και την αξιολόγηση ενός CNN μοντέλου βασισμένου σε multilevel strip pooling (MSP) του οποίου η τιμή της μετρικής Accuracy είναι ίση με 92.1%.

Τέλος, στις μελέτες [34], [35] αναπτύχθηκαν υπολογιστικά συστήματα ικανά να διαχωρίζουν με επιτυχία τις συμπτωματικές από τις ασυμπτωματικές καρωτιδικές αθηρωματικές πλάκες χρησιμοποιώντας υπερηχογραφικές απεικονιστικές καταγραφές.

Κεφάλαιο 3

Ερμηνευσιμότητα μηχανικής μάθησης

Τα τελευταία χρόνια η μηχανική μάθηση και η τεχνητή νοημοσύνη έχουν χρησιμοποιηθεί σημαντικά στον τομέα της βιοϊατρικής και ειδικότερα στην ανάπτυξη καινοτόμων μεθόδων με σκοπό την αντιμετώπιση κλινικών ζητημάτων, όπως η κατηγοριοποίηση ασθενών ανάλογα με τον βαθμό τους προβλήματός τους είτε η εύρεση της κατάλληλης θεραπείας ανάλογα με τα κλινικά χαρακτηριστικά του ασθενή. Παρ' όλα αυτά οι περισσότεροι αλγόριθμοι τεχνητής νοημοσύνης λειτουργούν σαν μαύρα κουτιά (black boxes). Λόγω του σχεδιασμού τους, αρκούνται στο να δώσουν μία απλή απάντηση χωρίς να επεξηγήσουν τις αιτίες που τους οδήγησαν σε αυτή, με αποτέλεσμα να χάνουν την αξιοπιστία τους σε εφαρμογές που χειρίζονται κρίσιμες αποφάσεις. Πιο συγκεκριμένα, σε ευαίσθητα ζητήματα όπως αυτά που σχετίζονται με την υγεία, στα οποία επικεντρώνεται και η παρούσα εργασία, δεν μας αρκεί μια απλή πρόβλεψη, ακόμα και στην περίπτωση που οι μετρικές απόδοσης του μοντέλου μας είναι ενθαρρυντικές. Συνεπώς, δεν μπορούμε να εμπιστευτούμε το μοντέλο "τυφλά" χωρίς να γνωρίζουμε τους βασικούς λόγους που το οδήγησαν στην τελική του απόφαση, καθώς ένα λάθος μπορεί να αποβεί μοιραίο για τον ασθενή. Λύση για την επίλυση αυτού του προβλήματος αποτελεί η ερμηνευσιμότητα, δηλαδή η διαδικασία ανάδειξης της αιτίας μίας απόφασης ενός μοντέλου μηχανικής μάθησης, ούτως ώστε να γίνει κατανοητή από τον άνθρωπο.

3.1 Χρησιμότητα της Ερμηνευσιμότητας

Πολλές κρίσιμες αποφάσεις λαμβάνονται με τη βοήθεια συστημάτων μηχανικής μάθησης. Συνεπώς είναι κομβικής σημασίας να γνωρίζουμε τους λόγους που οδήγησαν το σύστημα να λάβει μια συγκεκριμένη απόφαση. Είναι γεγονός ότι όσο περισσότερο η απόφαση ενός συστήματος επηρεάζει την ανθρώπινη ζωή, τόσο σημαντικότερο είναι το σύστημα να μας εξηγεί τον λόγο για τον οποίο κατέληξε σε μια συγκεκριμένη απόφαση. Έτσι μπορούμε να ελέγξουμε την απόφαση του αλγορίθμου και με βάση τη δική μας λογική καταλήγοντας σε ένα ακόμα πιο αξιόπιστο αποτέλεσμα.

Μέσω της ερμηνευσιμότητας των αλγορίθμων μηχανικής μάθησης ο γιατρός έχει τη δυνατότητα να ελέγξει με βάση τις ιατρικές του γνώσεις και την κλινική του εμπειρία, τις τελικές αποφάσεις που έλαβε ο αλγόριθμος καταλήγοντας σε ένα ακόμα πιο σίγουρο αποτέλεσμα. Με αυτόν τον τρόπο, ο συνδυασμός των αποφάσεων του μοντέλου μηχανικής μάθησης αλλά και της ανάλυσης του γιατρού, δίνουν μία πιο έμπιστη και με μικρότερη πιθανότητα λάθους διάγνωση.

Εκτός από την ιατρική γνωμάτευση, η ερμηνευσιμότητα, μπορεί να συμβάλει και στη

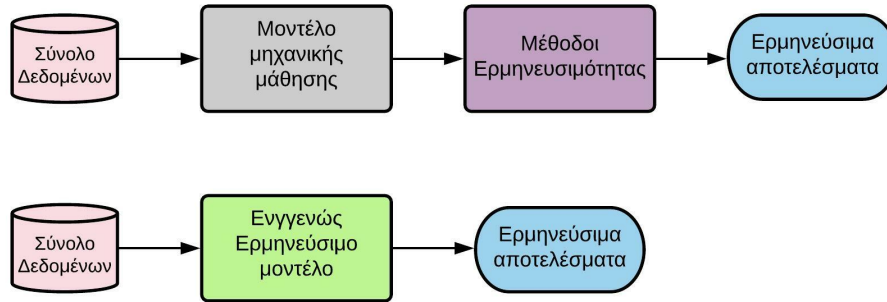
βελτίωση της δουλειάς του μηχανικού ο οποίος οφείλει να εμβαθύνει στα προβλήματα που χειρίζεται και να μην αρκείται στην απλή υλοποίηση των αλγορίθμων στηριζόμενος στην απόδοση των μετρικών τους. Έτσι μέσω της ερμηνευσιμότητας των αλγορίθμων μηχανικής μάθησης μπορεί να διακρίνει την εγκυρότητα των αποτελεσμάτων του μοντέλου και να κάνει τις κατάλληλες μετατροπές με σκοπό να το βελτιώσει.

Τέλος η χρήση της βελτιώνει σημαντικά και τις συνθήκες που ισχύουν για τον ασθενή. Έμμεσα η θετική επίδραση της για τον γιατρό και τον μηχανικό που αναλύσαμε παραπάνω επηρεάζει και τον ασθενή. Καθώς ο τελευταίος δέχεται μία διπλά επικυρωμένη διάγνωση η οποία έχει προκύψει μέσω της συνεργασίας των δύο επιστημόνων περιορίζοντας την πιθανότητα λάθους. Άμεσα, μπορεί να επηρεάσει τον ασθενή στην καλύτερη κατανόηση της ασθένειάς του καθώς και στους παράγοντες που συμβάλουν στην εμφάνισή της.

3.2 Τεχνικές Ερμηνευσιμότητας

Οι μέθοδοι ερμηνευσιμότητας που έχουμε μέχρι και σήμερα στη διάθεσή μας ποικίλουν ανάλογα με το μοντέλο μηχανικής μάθησης που χρησιμοποιούμε για την επίλυση του προβλήματός μας, με το στόχο που θέλουμε να πετύχουμε αλλά και με τον τρόπο που επιθυμούμε να ερμηνεύσουμε το μοντέλο. Οι δύο βασικές κατηγορίες της ερμηνευσιμότητας είναι η εγγενής και η εκ των υστέρων ερμηνευσιμότητα.

- Η **εγγενής ερμηνευσιμότητα** αφορά μοντέλα μηχανικής μάθησης που είναι ερμηνεύσιμα λόγω της αρχιτεκτονικής τους (Σχήμα 3.1), δηλαδή έχουν εγγενή ερμηνεία και μπορούν να ερμηνευτούν από μόνα τους χωρίς την συμβολή κάποιας άλλης μεθόδου. Συνήθως τα μοντέλα αυτά υστερούν στην πολυπλοκότητά τους και δεν μπορούν να διαχειριστούν αποδοτικά πιο σύνθετα προβλήματα.
- Η δεύτερη κατηγορία, αφορά τα μοντέλα που δεν είναι από την φύση τους ερμηνεύσιμα και για αυτό το λόγο χρειάζονται την συμβολή κάποιας άλλης μεθόδου για να ερμηνευτούν οι αποφάσεις τους. Η **εκ των υστέρων ή αγνωστική (model agnostic) ερμηνευσιμότητα**, αποτελείται από μεθόδους που εφαρμόζονται ξεχωριστά από τον αλγόριθμο μηχανικής μάθησης που χρησιμοποιείται για την επίλυση του εκάστοτε προβλήματος (Σχήμα. 3.1). Αποτελούν ξεχωριστό αλγόριθμο ο οποίος μπορεί να εφαρμοστεί σε οποιοδήποτε μη ερμηνεύσιμο μοντέλο που θα δημιουργήσουμε με σκοπό να το ερμηνεύσει όσο το δυνατόν καλύτερα. Συνήθως τα μη ερμηνεύσιμα μοντέλα που λειτουργούν ως "μαύρα κουτιά", είναι νευρωνικά δίκτυα πολλών επιπέδων (δίκτυα βαθιάς μάθησης) και διαχειρίζονται πιο πολύπλοκα προβλήματα και σύνολα δεδομένων. Συμπερασματικά, η ευελιξία της model-agnostic ερμηνευσιμότητας να εφαρμόζεται σε οποιοδήποτε μοντέλο, αποτελεί ένα από τα μεγαλύτερα πλεονεκτήματα της καθώς δίνει την ελευθερία στον μηχανικό να χρησιμοποιήσει όποιο μοντέλο θεωρεί καταλληλότερο για κάθε σύνολο δεδομένων χωρίς να περιορίζεται αποκλειστικά στα εγγενώς ερμηνεύσιμα μοντέλα. Στη συνέχεια αξίζει να αναφερθούμε σε κάποια βασικά κριτήρια που θα ήταν ιδανικό να πληρούν οι Model Agnostic μέθοδοι [11].
 - **Model flexibility**, δηλαδή η μέθοδος ερμηνευσιμότητας που θα δημιουργήσουμε να μπορεί να εφαρμοστεί σε οποιοδήποτε μοντέλο μηχανικής μάθησης χωρίς περιορισμούς.
 - **Explanation flexibility**, δηλαδή η δυνατότητα να λαμβάνουμε την ερμηνευσιμότητα των μοντέλων με διαφορετικές μορφές ανάλογα με τις ανάγκες μας.



Σχήμα 3.1: Εννοιολογικό σχήμα μη ερμηνεύσιμων μοντέλων και εγγενώς ερμηνεύσιμων μοντέλων

Στη συνέχεια επιλέξαμε να περιγράψουμε κάποιες από τις μεθόδους ερμηνευσιμότητας και των δύο κατηγοριών που περιγράψαμε προηγουμένως.

3.2.1 Εγγενώς ερμηνεύσιμα μοντέλα

- **Δέντρα αποφάσεων - Decision trees.** Τα οποία ταξινομούν τα δείγματα σε κλάσεις βάση ενός συγκεκριμένου κριτηρίου αξιολόγησης. Κάθε δέντρο απόφασης αποτελείται από τρία βασικά δομικά στοιχεία, τα οποία είναι:
 - **Κόμβοι**, στους οποίους εφαρμόζεται το κριτήριο που έχουμε επιλέξει και γίνεται η σύγκριση των δειγμάτων με βάση ένα συγκεκριμένο χαρακτηριστικό (feature) των δεδομένων.
 - **Ακμές-Κλαδιά**, οδηγούν τα δείγματα στον επόμενο κόμβο με βάση την απόφαση του κόμβου που βρίσκονται.
 - **Φύλλα**, είναι οι τερματικοί κόμβοι στους οποίους εμφανίζεται η τελική πρόβλεψη σχετικά με την κλάση στην οποία ανήκει το κάθε δείγμα.

Για να ερμηνεύσουμε αν τα τελικά αποτελέσματα του δέντρου απόφασης στηρίζονται σε θεμελιώδεις αποφάσεις ή πιο αποδοτική μέθοδος ερμηνευσιμότητας που μπορούμε να εφαρμόσουμε είναι η απεικόνιση του δέντρου.

- **Ταξινομητής k-κοντινότερων γειτόνων (k-Nearest neighbors).** Ταξινομεί κάθε δείγμα του συνόλου δεδομένων στηριζόμενος στα k κοντινότερα γειτονικά δείγματα. Πιο συγκεκριμένα αφού εντοπίσει τα k κοντινότερα δείγματα σε αυτό που εξετάζει το κατατάσσει στην κατηγορία όπου ανήκει η πλειοψηφία των γειτόνων του. Επομένως, ο τρόπος με τον οποίο καταλήγει στις τελικές του προβλέψεις είναι γνωστός και μπορεί να ελεγχθεί ανα πάσα στιγμή.

3.2.2 Model agnostic μέθοδοι

Οι παρακάτω model-agnostic μέθοδοι ερμηνευσιμότητας είναι αυτές με τις οποίες εργαστήκαμε στην παρούσα διπλωματική εργασία.

Global Surrogate

Η πρώτη μέθοδος ερμηνευσιμότητας με την οποία επιλέξαμε να ασχοληθούμε είναι η Global Surrogate. Με τη μέθοδο Global Surrogate δημιουργούμε ένα ερμηνεύσιμο μοντέλο (π.χ

k-nn ή decision tree) το οποίο έχει εκπαιδευτεί έτσι ώστε να προσεγγίζει τις προβλέψεις ενός νευρωνικού δικτύου (black box). Με αυτόν τον τρόπο πετυχαίνουμε την εξαγωγή συμπερασμάτων σχετικά με την ερμηνευσιμότητα του black box, μέσω του εγγενώς ερμηνεύσιμου μοντέλου της μεθόδου Global Surrogate [11]. Με άλλα λόγια, θα μπορούσαμε να πούμε πως χρησιμοποιούμε επιπλέον μηχανική μάθηση για να ερμηνεύσουμε μοντέλα μηχανικής μάθησης.

Το μεγαλύτερο πλεονέκτημα της Global Surrogate μεθόδου είναι η μεγάλη ευελιξία που παρουσιάζει. Συγκεκριμένα, έχει την ικανότητα να εφαρμόζεται σε κάθε είδους αλγόριθμο μηχανικής μάθησης χρησιμοποιώντας το εγγενώς ερμηνεύσιμο μοντέλο που ταιριάζει καλύτερα στην κάθε περίπτωση. Παρ' όλα αυτά δεν είναι πάντα σίγουρο ότι το εγγενώς ερμηνεύσιμο μοντέλο της Global Surrogate μεθόδου θα είναι "άξιο" υποκατάστατο του αρχικού μοντέλου. Για να υπολογίσουμε, λοιπόν, πόσο καλά υποκαθιστά το black box model το εγγενώς ερμηνεύσιμο μοντέλο χρησιμοποιούμε τη μετρική R-square που προκύπτει από τον παρακάτω μαθηματικό τύπο:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^n (\hat{y}_*^{(i)} - \hat{y}^{(i)})^2}{\sum_{i=1}^n (\hat{y}^{(i)} - \bar{\hat{y}})^2}$$

όπου $\hat{y}_*^{(i)}$ είναι η πρόβλεψη του Global Surrogate μοντέλου για το i-οστο δείγμα, $\hat{y}^{(i)}$ η πρόβλεψη του black box μοντέλου για το i-οστο δείγμα και $\bar{\hat{y}}$ ο μέσος όρος των προβλέψεων του black box μοντέλου.

Individual Conditional Expectation

Η επόμενη μέθοδος ερμηνευσιμότητας που επιλέξαμε να υλοποιήσουμε είναι η Individual Conditional Expectation (ICE) [11]. Η μέθοδος αυτή εκφράζει την ερμηνευσιμότητα των μοντέλων μέσω διαγραμμμάτων. Πιο συγκεκριμένα, για κάθε δείγμα του συνόλου δεδομένων δημιουργείται μια ξεχωριστή καμπύλη η οποία περιγράφει τον τρόπο με τον οποίο αλλάζει η πρόβλεψη του μοντέλου ανάλογα με τις μεταβολές της τιμής ενός επιλεγμένου χαρακτηριστικού (feature). Οι παραπάνω καμπύλες για ένα συγκεκριμένο χαρακτηριστικό συγκεντρώνονται σε ένα διάγραμμα το οποίο ονομάζεται ICE plot. Οι μεταβολές της πρόβλεψης, που αποτελούν το βασικό συστατικό δημιουργίας των καμπυλών ICE, μπορούν να υπολογιστούν μεταβάλλοντας την τιμή ενός συγκεκριμένου χαρακτηριστικού και διατηρώντας τις τιμές των υπολοίπων σταθερές. Με αυτόν τον τρόπο ερμηνεύουμε την συμπεριφορά του μοντέλου σε σχέση με τις διάφορες τιμές που μπορεί να λάβει ένα ορισμένο χαρακτηριστικό. Επομένως μπορούμε να αποφανθούμε για τη σημαντικότητα του συγκεκριμένου χαρακτηριστικού στην ταξινόμηση των δειγμάτων και κατά πόσο επηρεάζεται η τελική απόφαση του μοντέλου μας από αυτό. Για παράδειγμα ένα ICE plot στο οποίο οι καμπύλες παραμένουν σταθερές κατά την μεταβολή των τιμών του χαρακτηριστικού, υποδηλώνει ότι το χαρακτηριστικό δεν κατέχει ιδιαίτερο ρόλο στην τελική απόφαση του μοντέλου. Αντιθέτως, ένα ICE plot στο οποίο οι καμπύλες παρουσιάζουν σημαντικές μεταβολές υποδηλώνει ένα χαρακτηριστικό το οποίο συνιστά βασικό καθοριστικό παράγοντα για την τελική απόφαση του μοντέλου. Συμπεραίνουμε, λοιπόν, ότι η μέθοδος ICE μπορεί να φανεί σημαντικά αποδοτική στην ερμηνευσιμότητα των black box μοντέλων και ιδιαίτερα σε περιπτώσεις που θέλουμε να επικεντρωθούμε στην επιρροή που παρουσιάζουν συγκεκριμένα χαρακτηριστικά στο τελικό αποτέλεσμα του μοντέλου. Αντίθετα, σε περιπτώσεις που το dataset είναι ιδιαίτερα μεγάλο, τα διαγράμματα της μεθόδου ICE δεν θα είναι το ίδιο ευανάγνωστα με αποτέλεσμα η μέθοδος να μην είναι το ίδιο αποδοτική.

Shapley Values

Η τελευταία μέθοδος ερμηνευσιμότητας με την οποία επιλέξαμε να ασχοληθούμε είναι η Shapley Values [11]. Στη μέθοδο αυτή θεωρούμε πως κάθε χαρακτηριστικό είναι παίκτης σε ένα παιχνίδι και η τελική πρόβλεψη του μοντέλου είναι το τελικό "payout" (κέρδος) του παιχνιδιού το οποίο πρέπει να μοιράσουμε δίκαια στους παίκτες ανάλογα με την συμβολή τους στο παιχνίδι. Η μέθοδος Shapley Values ερμηνεύει τη συνεισφορά όλων των χαρακτηριστικών για ένα επιλεγμένο δείγμα. Πιο συγκεκριμένα μας φανερώνει πως τα χαρακτηριστικά του δείγματος επηρέασαν τον αλγόριθμο μηχανικής μάθησης για να λάβει την τελική του απόφαση. Συνοψίζοντας, λοιπόν, μπορούμε να πούμε ότι οι Shapley Values αντιπροσωπεύουν τη συμβολή κάθε χαρακτηριστικού στην απομάκρυνση της τελικής πρόβλεψης από την αναμενόμενη τιμή. Ως αναμενόμενη τιμή θεωρούμε τη μέση τιμή των προβλέψεων του μοντέλου.

Τα Shapley Values του εκάστοτε μοντέλου προκύπτουν από τον παρακάτω μαθηματικό τύπο:

$$\phi_j(val) = \sum_{S \subseteq \{x_1, \dots, x_p\} \setminus \{x_j\}} \frac{|S|!(p - |S| - 1)!}{p!} (val(S \cup \{x_j\}) - val(S))$$

όπου S ένα σύνολο χαρακτηριστικών (features) το οποίο δεν περιέχει το feature που εξετάζουμε, x_j το feature που έχουμε επιλέξει να εξετάσουμε και p ο αριθμός των χαρακτηριστικών. Τέλος με $val(S)$ συμβολίζουμε την πρόβλεψη του μοντέλου με features τα στοιχεία του συνόλου S . Επομένως με τη μαθηματική παράσταση $val(S \cup \{x_j\}) - val(S)$ υπολογίζουμε τη διαφορά που προκύπτει στην τελική πρόβλεψη του μοντέλου όταν στα feature των δειγμάτων συμπεριλαμβάνεται το x_j και όταν αυτό απουσιάζει.

Βασιζόμενοι στα παραπάνω συμπεραίνουμε πως η συγκεκριμένη μέθοδος είναι ιδιαίτερα αποδοτική σε οποιοδήποτε μοντέλο εφαρμοστεί καθώς τα αποτελέσματα ερμηνευσιμότητας που μας αποφέρει στηρίζονται σε μαθηματικά αποδεδειγμένες προτάσεις. Παρ' όλα αυτά, ο υπολογισμός των Shapley Values μπορεί σε αρκετές περιπτώσεις να γίνει δύσχερστος έως και απαγορευτικός λόγω του μεγάλου όγκου πράξεων που πρέπει να υλοποιηθούν.

3.3 Εφαρμογές της ερμηνευσιμότητας στον κλάδο της βιοϊατρικής επιστήμης

Τα εγγενώς ερμηνεύσιμα μοντέλα ήταν τα πρώτα μοντέλα μηχανικής μάθησης που χρησιμοποιήθηκαν στον κλάδο της βιοϊατρικής επιστήμης. Πιο συγκεκριμένα, το 1994 στην μελέτη [36] χρησιμοποιήθηκαν δέντρα απόφασης για την δημιουργία ενός καρδιολογικού συστήματος το οποίο είχε ως σκοπό τον προσδιορισμό της πρόπτωσης της μητροειδούς βαλβίδας.

Μία ακόμη μελέτη που χρησιμοποίησε τα δέντρα απόφασης για να εξυπηρετήσει σκοπούς της ιατρικής επιστήμης, δημιουργώντας ερμηνεύσιμα μοντέλα απόφασης, ήταν η [37] κατά την οποία κατασκευάστηκαν διαφορετικοί συλλογικοί αλγόριθμοι βασιζόμενοι σε δέντρα απόφασης με σκοπό την παρακολούθηση της καρδιαγγειακής αυτόνομης νευροπάθειας σε διαβητικούς ασθενείς.

Σε μία πιο πρόσφατη έρευνα ο T. Brito-Sarracino και άλλοι [38] χρησιμοποίησαν δέντρα απόφασης για τον διαχωρισμό όγκων του στήθους σε καλοήθεις και κακοήθεις πετυχαίνοντας ποσοστό accuracy ίσο με 96%, ερμηνεύοντας παράλληλα την εσωτερική δομή των μοντέλων τους.

Παράλληλα, εγγενώς ερμηνεύσιμα μοντέλα όπως ο ταξινομητής των k κοντινότερων γειτόνων (k -NNN) και ο Naive Bayes, χρησιμοποιήθηκαν για τη δημιουργία μοντέλων

μηχανικής μάθησης με σκοπό την επίλυση βιοϊατρικών προβλημάτων. Πιο συγκεκριμένα ο k-NN αλγόριθμος χρησιμοποιήθηκε στην μελέτη [39] με σκοπό τη δημιουργία ενός ταξινομητή ικανού να συμβάλει στην διάγνωση ασθενών που πάσχουν από καρδιακές παθήσεις. Επίσης, ο k-NN εφαρμόστηκε για την ανίχνευση του καρκίνου στον πνεύμονα [40] αλλά και για την ανίχνευση χρόνιων ασθενιών όπως η νόσος Alzheimer [41]. Η ίδια μελέτη χρησιμοποίησε τον Naive Bayes ταξινομητή για την κατηγοριοποίηση των ασθενών με Alzheimer.

Το υψηλό επίπεδο ερμηνευσιμότητας αλλά και η εύκολη εφαρμογή των εγγενώς ερμηνεύσιμων μοντέλων τα έκαναν ιδιαίτερα δημοφιλή και διέυρυναν την χρήση τους στον κλάδο της βιοιατρικής επιστήμης. Παρ' όλα αυτά η μικρή πολυπλοκότητά τους τα οδηγεί στο να μην αποδίδουν καλά σε δυσκολότερα σύνολα δεδομένων και πιο περίπλοκα προβλήματα σε αντίθεση με τα μοντέλα βαθιάς μάθησης. Μέχρι πρόσφατα, η ερμηνευσιμότητα των μοντέλων βαθιάς μάθησης ήταν πολύ δύσκολο μέχρι και αδύνατο να επιτευχθεί. Η αδυναμία αυτή έχει καταπολεμηθεί από πιο ανεπτυγμένες μεθόδους ερμηνευσιμότητας, οι οποίες χρησιμοποιούνται όλο και περισσότερο σε νέες μελέτες.

Πιο συγκεκριμένα, στη μελέτη [14] δημιουργήθηκαν τρία διαφορετικά μοντέλα μηχανικής μάθησης με σκοπό την κατάταξη όγκων του στήθους σε καλοήθεις και κακοήθεις τα οποία στη συνέχεια ερμηνεύτηκαν μέσω των μεθόδων ερμηνευσιμότητας Global Surrogate, Individual Conditional Expectation και Shapley Values.

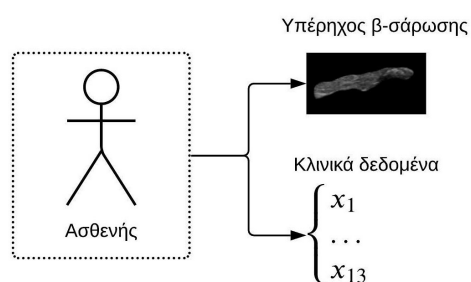
Η μέθοδος Shapley Values χρησιμοποιήθηκε, επιπλέον, από τους Scott M. Lundberg et al. [42] οι οποίοι ανέπτυξαν ένα μοντέλο μηχανικής μάθησης το οποίο προβλέπει τον κίνδυνο υποξαιμίας και παρέχει εξηγήσεις σχετικά με τους παράγοντες κινδύνου σε πραγματικό χρόνο κατά τη διάρκεια της γενικής αναισθησίας. Ακόμα, στην μελέτη [43] χρησιμοποιήθηκαν έξι διαφορετικές τεχνικές μηχανικής μάθησης οι οποίες εκπαιδεύτηκαν έτσι ώστε να προβλέπουν τον κίνδυνο ανάπτυξης υπέρτασης χρησιμοποιώντας δεδομένα καρδιοαναπνευστικής ικανότητας των ασθενών. Στη συνέχεια τα μοντέλα αυτά ερμηνεύτηκαν μέσω επτά διαφορετικών μεθόδων ερμηνευσιμότητας παρέχοντας στο κλινικό προσωπικό την δυνατότητα καλύτερης κατανόησης των αποτελεσμάτων και μεγαλύτερης εμπιστοσύνης σε αυτά.

Κεφάλαιο 4

Υλικό και Μέθοδοι

4.1 Σύνολο Δεδομένων

Στην παρούσα εργασία, αξιοποιήθηκαν δεδομένα 51 ασθενών με καρωτιδική νόσο. Για κάθε έναν από τους ασθενείς έχουμε στη διάθεση μας τις υπερηχογραφικές απεικονιστικές καταγραφές β-σάρωσης των καρωτίδων αρτηριών σε διαμήκη τομή καθώς και ένα διάλυσμα δεκατριών διαφορετικών κλινικών χαρακτηριστικών όπως το φύλο, η ηλικία και ένα σύνολο αιματολογικών εξετάσεων, Σχήμα. 4.1. Πιο συγκεκριμένα, όσον αφορά τις απεικονιστικές υπερηχογραφικές καταγραφές, αυτές προέκυψαν ως στιγμιότυπα (frames) από υπερηχογραφικά βίντεο των καρωτίδων του κάθε ασθενούς. Στην παρούσα εργασία χρησιμοποιήσαμε δύο frames από κάθε βίντεο, από τα οποία το ένα αντιστοιχεί στην συστολή της αρτηρίας ενώ το δεύτερο στη διαστολή. Επιπλέον χρησιμοποιήσαμε μόνο τα μέρη των εικόνων που αντιστοιχούσαν στα αθηρώματα που εμφανίζονταν σε κάθε μία από αυτές. Για αυτό τον λόγο σε αρκετούς ασθενείς οι οποίοι εμφάνιζαν παραπάνω από μία αθηρωματικές πλάκες στα τοιχώματα των καρωτίδων τους, αντιστοιχούν περισσότερες από μία εικόνες.



Σχήμα 4.1: Κάθε ασθενής αντιπροσωπεύεται από μία ή περισσότερες υπερηχογραφικές απεικονιστικές καταγραφές και από ένα διάλυσμα δεκατριών κλινικών χαρακτηριστικών

Ακόμα, για κάθε ασθενή γνωρίζουμε αν εμφάνισε κάποιο ΑΕΕ ή ΠΠΕ τους τελευταίους 6 μήνες, αν δηλαδή θεωρείται συμπτωματικός ή ασυμπτωματικός αλλά και τον βαθμό στένωσης των καρωτίδων αρτηριών του. Με βάση αυτά τα δύο στοιχεία διαχωρίσαμε τους ασθενείς ανάλογα με την μέθοδο θεραπείας που θεωρείται κατάλληλη για κάθε έναν ξεχωριστά. Η ταξινόμησή τους έγινε σε δύο κλάσεις:

1. Αυτούς που θα ακολουθήσουν φαρμακευτική αγωγή σε συνδυασμό με τακτική παρακολούθηση από τον κλινικό ιατρό και χαρακτηρίζονται ως **Ασθενείς Χαμηλού Κινδύνου, Low Risk**.

2. Αυτούς που θα ακολουθήσουν επεμβατική θεραπεία (αγγειοπλαστική με τοποθέτηση stent ή ενδαρτηρεκτομή) και χαρακτηρίζονται ως **Ασθενείς Υψηλού Κινδύνου, High Risk**

Πιο συγκεκριμένα ο διαχωρισμός των ασθενών έγινε με βάση την τρέχουσα κλινική πρακτική [20] [44] για την αξιολόγηση της επικινδυνότητας μίας αθηρωματικής πλάκας κατά την οποία οι ασυμπτωματικοί ασθενείς με στένωση της τάξεως 70% και άνω καθώς και οι συμπτωματικοί ασθενείς με στένωση της τάξεως 50% και άνω κατατάσσονται στην κλάση των ασθενών Υψηλού κινδύνου ενώ σε κάθε άλλη περίπτωση κατατάσσονται στην κλάση των ασθενών Χαμηλού κινδύνου, Πίνακας. 4.1. Στο σύνολο δεδομένων που έχουμε στη διάθεσή μας οι ασθενείς υψηλού κινδύνου είναι 39 και οι ασθενείς χαμηλού κινδύνου 12.

•	BΣ < 50%	50% < BΣ < 70%	70% < BΣ
Συμπτωματικός	Θ1	Θ2	Θ2
Ασυμπτωματικός	-	Θ1	Θ2

Πίνακας 4.1: Τρέχουσα κλινική πρακτική για την επιλογή της θεραπείας ενός ασθενή με καρωτιδική αθηρωμάτωση. Με Θ1 συμβολίζουμε την συντηρητική αγωγή, με Θ2 την επεμβατική μέθοδο θεραπείας και με BΣ τον βαθμό στένωσης της καρωτίδας.

Όπως αναφέραμε και προηγουμένως σε κάθε ασθενή αντιστοιχούν τουλάχιστον δύο εικόνες, αν εμφανίζει μόνο μία αθηρωματική πλάκα αλλά μπορούν να αντιστοιχούν και περισσότερες αν εμφανίζει πάνω από ένα αθηρώματα στις αρτηρίες του. Επομένως συνολικά έχουμε στη διάθεσή μας 106 εικόνες υπερήχων που αντιστοιχούν σε ασθενείς υψηλού κινδύνου και 38 που αντιστοιχούν σε ασθενείς χαμηλού κινδύνου. Αντίστοιχα για τα κλινικά δεδομένα κάθε ασθενής αντιπροσωπεύεται από ένα μοναδικό διάλυσμα δεκατριών χαρακτηριστικών. Επομένως έχουμε 39 διαλύσματα που αντιστοιχούν σε ασθενείς υψηλού κινδύνου και 12 που αντιστοιχούν σε ασθενείς χαμηλού κινδύνου.

Τέλος αξίζει να αναφερθεί πως το υλικό που χρησιμοποιήθηκε στην εργασία αντλήθηκε από τη βάση δεδομένων του έργου με κωδικό 09ΣΥΝ-12-1054 και τίτλο "Η αθηρωματική πλάκα στην καρωτίδα: μία πολυ-επιστημονική προσέγγιση με στόχο τη βέλτιστη διαχείριση συμπτωματικών και ασυμπτωματικών ασθενών", υπαγόμενο στα επιχειρησιακά προγράμματα "Ανταγωνιστικότητα και Επιχειρηματικότητα", ΕΣΠΑ 2007-2013. Οι εξετάσεις πραγματοποιήθηκαν στο αγγειοχειρουργικό τμήμα του Πανεπιστημιακού Γενικού Νοσοκομείου "ΑΤΤΙΚΟΝ", κατά τα έτη 2012-2013.

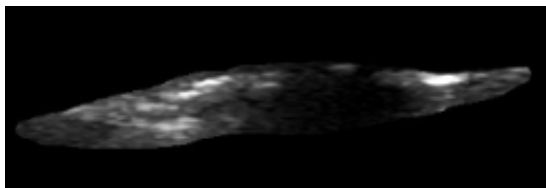
4.1.1 Περιγραφή υπερηχογραφικών απεικονίσεων β-σάρωσης των καρωτιδικών αρτηριών

Για την εξαγωγή των εικόνων πραγματοποιήθηκε μία σειρά από υπερηχογραφικές απεικονιστικές καταγραφές β-σάρωσης της καρωτίδας σε διαμήκη τομή, σε 51 ασθενείς, ακολουθώντας ένα τυποποιημένο πρωτόκολλο, ικανό να διασφαλίσει την αξιόπιστη καταγραφή της κίνησης του αρτηριακού τοιχώματος και της αθηρωματικής πλάκας της καρωτίδας. Ειδικότερα με βάση αυτό το πρωτόκολλο καθορίζονται η θέση και η ειδική προετοιμασία του ασθενούς πριν την εξέταση, η θερμοκρασία του δωματίου αλλά και οι ρυθμίσεις της κεφαλής υπερήχων, οι οποίες αναλύονται στον Πίνακα. 4.2 [20].

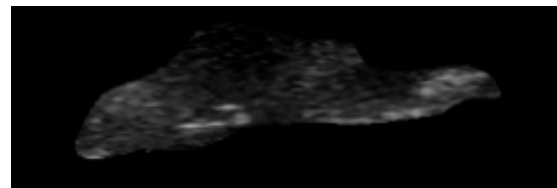
Παράμετρος	Ρύθμιση
Κεφαλή	Γραμμική 3-12 MHz, General Electric
Είδος εικόνων	β-σάρωση
Βάθος	3.5-4cm
Persistence	0
Κέρδος	60dB ή 75Db
Χρονική ανάλυση	≥ 25 καρέ/δευτερόλεπτο

Πίνακας 4.2: Ρυθμίσεις της διάταξης υπερήχων για την καταγραφή πραγματικών χρονικών ακολουθιών εικόνων υπερήχων β-σάρωσης της καρωτίδας σε διαμήκη τομή

Από τους ασθενείς που υποβλήθηκαν στην υπερηχογραφική εξέταση, οι 12 ήταν ασθενείς χαμηλού κινδύνου (Σχήμα. 4.2a) ενώ οι υπόλοιποι 39 ήταν ασθενείς υψηλού κινδύνου (Σχήμα. 4.2b). Παρόλα αυτά οι καταγραφές που έχουμε στην διάθεση μας είναι περισσότερες καθώς πλήθος ασθενών εμφάνιζε περισσότερες από μία αθηρωματικές πλάκες είτε στην ίδια αρτηρία είτε και στις δύο αρτηρίες. Πιο συγκεκριμένα έχουμε στη διάθεση μας 106 υπερήχους που ανήκουν σε ασθενείς υψηλού κινδύνου και 38 που ανήκουν σε ασθενείς χαμηλού κινδύνου.



(a)



(b)

Σχήμα 4.2: Υπέρηχογραφικές απεικονίσεις β-σάρωσης των αθηρωμάτων που υπάρχουν στις καρωτιδικές αρτηρίες οι οποίες ανήκουν σε: (α) Ασθενή υψηλού κινδύνου (b) Ασθενή χαμηλού κινδύνου

4.1.2 Περιγραφή αριθμητικών κλινικών δεδομένων

Εκτός από τους υπερήχους β-σάρωσης έχουμε στη διάθεση μας 13 κλινικά χαρακτηριστικά τα οποία αντιπροσωπεύουν κάθε ασθενή καθώς και τον βαθμό στένωσης των αρτηριών του. Τα χαρακτηριστικά αυτά αναλύονται στη συνέχεια.

Βαθμός στένωσης της καρωτίδας, για τον υπολογισμό του διακρίνονται δύο τεχνικές [20]. Στην πρώτη ο βαθμός στένωσης υπολογίζεται απευθείας ενώ στην δεύτερη προσδιορίζεται μέσω των διαταραχών στις ταχύτητες ροής που προκαλεί η στένωση. Η απευθείας μέτρηση μπορεί να πραγματοποιηθεί σε οποιοδήποτε βαθμό στένωσης αλλά επηρεάζεται πολλές φορές από την ποιότητα της εικόνας. Ο έμμεσος τρόπος μέτρησης αφορά συνήθως στενώσεις της τάξεως άνω του 50% καθώς σε μικρότερες στενώσεις οι αιμοδυναμικές μεταβολές που προκαλούνται δεν είναι αρκετά έντονες. Τη βάση για την εκτίμηση των στενώσεων στην έσω καρωτιδική αρτηρία αποτελούν το φάσμα ταχυτήτων που λαμβάνεται στη στενωτική περιοχή ή λίγο μετά από αυτή. Οι παράμετροι που μετρούνται είναι τέσσερις: 1) η κορυφαία συστολική ταχύτητα (peak systolic velocity, PSV), 2) η τελοδιαστολική ταχύτητα (peak enddiastolic velocity, EDV), 3) ο λόγος των συστολικών ταχυτήτων και 4) ο λόγος των τελοδιαστολικών ταχυτήτων. Με βάση το πρότυπο που χρησιμοποιείται στην

μελέτη ACSRS (Asymptomatic, Carotid Stenosis and Risk of Stroke) και τις παραπάνω τιμές, ο βαθμός στένωσης της καρωτίδα υπολογίζεται όπως περιγράφεται στον Πίνακα. 4.3.

Βαθμός στένωσης	PSV	EDV	PSV _{ICA} /PSV _{CCA}
0% – 29%	< 100	< 40	
30% – 49%	110 < PSV < 130	< 40	
50% – 59%	> 130	< 40	< 3.2
60% – 69%	> 130	40 < EDV < 110	3.2 < PSVratio < 4
70% – 79%	> 210	110 < EDV < 140	> 4
80% – 95%	> 210	> 140	
96% – 99%			
100%			

Πίνακας 4.3: Υπολογισμός του βαθμού στένωσης της καρωτίδας με βάση το πρότυπο που ακολουθείται στη μελέτη ACSRS. Όπου με PSV αναφερόμαστε στη Μέγιστη συστολική ταχύτητα σε cm/sec, με EDV στην Τελοδιαστολική ταχύτητα σε cm/sec, με ICA στη έσω κοινή αρτηρία και με CCA στην κοινή καρωτιδική αρτηρία

Ηλικία - Age, η ηλικία του κάθε ασθενή. Όπως αναφέρθηκε και προηγουμένως [19] η καρωτιδική νόσος εμφανίζεται κυρίως σε ηλικιωμένους ασθενείς. Πιο συγκεκριμένα μόνο το 1% των ασθενών ηλικίας 50-59 ετών εμφανίζουν στένωση σοβαρού βαθμού, ενώ το ποσοστό αυτό είναι 10% για τις ηλικίες 80-89 ετών.

Φύλλο - Sex, το φύλλο του ασθενούς.

Ουρία αίματος - Urea, είναι προϊόν του μεταβολισμού και αποτελεί δείκτη της καλής λειτουργικότητας των νεφρών και του ήπατος. Οι φυσιολογικές τιμές της στο αίμα κυμαίνονται από 7 έως 20 mg/dl. Όταν το όριο αυτό ξεπεραστεί, σημαίνει ότι η λειτουργία των νεφρών δεν είναι σωστή. Η ουρία διαταράσσεται κυρίως όταν υπάρχουν συμπτώματα ή παθήσεις όπως η καρδιακή ανεπάρκεια, το έμφραγμα του μυοκαρδίου, ο σακχαρώδης διαβήτης, η χίρρωση του ήπατος και άλλα. Σε μελέτες που έγιναν [45], [46] βρέθηκε ότι οι υψηλές τιμές της ουρίας στο αίμα συνυπάρχουν με την ανάπτυξη της αθηρωματικής πλάκας και κατ'επέκταση της καρωτιδικής νόσου.

Κρεατινίνη - Creatinine, η κρεατινίνη είναι και αυτή προϊόν του μεταβολισμού. Πιο συγκεκριμένα είναι μία αζωτούχος ένωση που παράγεται κατά την διάσπαση της φωσφορικής κρεατίνης των σκελετικών μυών και η απομάκρυνσή της από τον οργανισμό μας προϋποθέτει την σωστή λειτουργία των νεφρών. Οι φυσιολογικές τιμές κρεατινίνης αίματος συνήθως είναι 0.6-1.3 mg/dL, όταν ξεπερνούν αυτό το όριο δηλώνουν τη μη φυσιολογική λειτουργία των νεφρών. Η ίδια μελέτη που αναφέραμε προηγουμένως [45] σχετικά με την επίδραση της αυξημένης ουρίας στην αθηρωματική πλάκα, έδειξε ότι η υψηλές τιμές κρεατινίνης παρουσιάζουν παρόμοια συμπεριφορά όσον αφορά την εμφάνιση των αθηρωμάτων στις καρωτιδικές αρτηρίες.

Οξαλοξική Τρανσαμινάση - SGOT, είναι ένα ένζυμο που βρίσκεται κυρίως στην καρδιά, το συκώτι και τους μύες. Απελευθερώνεται στην κυκλοφορία μετά από τραυματισμό ή θάνατο των κυττάρων και τα επίπεδά του αυξάνονται εντός 12 ωρών από τον τραυματισμό και παραμένουν αυξημένα για 5 ημέρες. Ο έλεγχος της SGOT γίνεται συχνά μετά από έμφραγμα του μυοκαρδίου ή για τον έλεγχο ηπατικής βλάβης. Η αυξημένη τιμή της μπορεί να

οφείλεται σε οξεία νεφρική νόσο, εγκεφαλικό τραύμα, απόφραξη χοληφόρων και άλλα. Οι φυσιολογικές τιμές είναι 10-40 IU/L για τους άντρες και 9-32 IU/L για τις γυναίκες.

Πυροσταφυλική Τρανσαμινάση - SGPT, όπως και η SGOT είναι και αυτή ηπατικό ένζυμο που αυξάνεται όταν υπάρχει καταστροφή των κυττάρων. Οι φυσιολογικές τιμές είναι 10-55 IU/L για τους άντρες και 7-30 IU/L για τις γυναίκες.

γ-Γλουταμυλοτρανσφεράση - GGT, η μέτρηση της γ-γλουταμυλοτρανσφεράσης χρησιμοποιείται στη διάγνωση και την παρακολούθηση των ηπατοχολικών νοσημάτων και αποτελεί μέχρι και σήμερα τον πιο ευαίσθητο ενζυμικό δείκτη για τα ηπατικά νοσήματα. Διαθέσιμες ενδείξεις υποστηρίζουν έναν παθογενετικό ρόλο της δραστηριότητας της GGT όσον αφορά στην εξέλιξη και την αστάθεια των αθηροσκληρωτικών πλακών [47] αλλά και καρδιαγγειακών παθήσεων γενικότερα [48]. Οι φυσιολογικές τιμές της κυμαίνονται από 0 έως 30 IU/L.

Αλκαλική Φωσφατάση - ALP, η αλκαλική φωσφατάση είναι ένα ένζυμο το οποίο παράγεται στο ήπαρ, στα οστά και στον πλακούντα. Τα επίπεδα της αλκαλικής φωσφατάσης στον ανθρώπινο οργανισμό είναι πολύ σημαντικά για την διάγνωση φυσιολογικών ή παθολογικών καταστάσεων άμεσα σχετιζόμενων με το σκελετό, το ηπατοχοληφόρο σύστημα και τον πλακούντα. Αρκετές κλινικές μελέτες [49], [50] έχουν δείξει ότι αυξημένα επίπεδα ALP στον οργανισμό αποτελούν προγνωστικό παράγοντα για μελλοντικά καρδιαγγειακά επεισόδια και συνδέονται με αυξημένο κίνδυνο θνησιμότητας σε ασθενείς με AEE [51], [52]. Πιο συγκεκριμένα μέσω της μελέτης [49] αποδείχθηκε ότι αυξημένες τιμές ALP στον οργανισμό συσχετίζονται άμεσα με την ανάπτυξη αθηρωματικών αγγειακών παθήσεων της καρδιάς και του εγκεφάλου, όπως η ΚΝ. Οι φυσιολογικές τιμές κυμαίνονται από 40-129 U/L στους άντρες και 35-104 U/L στις γυναίκες

LDL Χοληστερόλη - LDL, η λιποπρωτεΐνη χαμηλής πυκνότητας (LDL), συσσωρεύεται στα τοιχώματα των αρτηριών και μπορεί να προκαλέσει καρδιοπάθεια και άλλες σοβαρές επιπλοκές. Εάν αποκοπεί από το τοίχωμα της αρτηρίας, μια μάζα LDL χοληστερόλης μπορεί να προκαλέσει θρόμβωση. Η LDL δεν πρέπει να ξεπερνά τα 100 χιλιοστόγραμμα ανά δεκατόλιτρο αίματος (mg/dL). Η τιμή μεταξύ 100-129 mg/dL κρίνεται ανεκτή, η τιμή 130-159 mg/dL οριακά υψηλή, η τιμή 160-189 mg/dL υψηλή, ενώ η τιμή 190 mg/dL και άνω επικίνδυνα υψηλή.

HDL Χοληστερόλη - HDL, η λιποπρωτεΐνη υψηλής πυκνότητας (HDL), βοηθά στο να απομακρυνθεί η LDL χοληστερόλη από το κυκλοφορικό σύστημα ώστε να μειωθεί ο κίνδυνος καρδιοπάθειας. Η φυσιολογική τιμή της HDL είναι 60mg/dL και άνω. Αν η τιμή της είναι κάτω από 40 mg/dl, τότε θεωρείται πολύ χαμηλή.

Τριγλυκερίδια - TG, αποτελούν ένα είδος λιπιδίων, όπως και η χοληστερόλη, που υπάρχουν στο αίμα. Οι θερμίδες που δεν χρησιμοποιούνται από τον οργανισμό όταν τρώμε μετατρέπονται σε τριγλυκερίδια, τα οποία στη συνέχεια αποθηκεύονται στα λιπώδη κύτταρα. Η διαφορά μεταξύ χοληστερόλης και τριγλυκερίδιων είναι ότι τα τελευταία τροφοδοτούν τον οργανισμό με ενέργεια, ενώ η χοληστερόλη χρησιμοποιείται στη δομή των κυττάρων και για την παραγωγή ορισμένων ορμονών. Οι φυσιολογικές τιμές των τριγλυκερίδιων δεν πρέπει να ξεπερνούν τα 150 mg/dl.

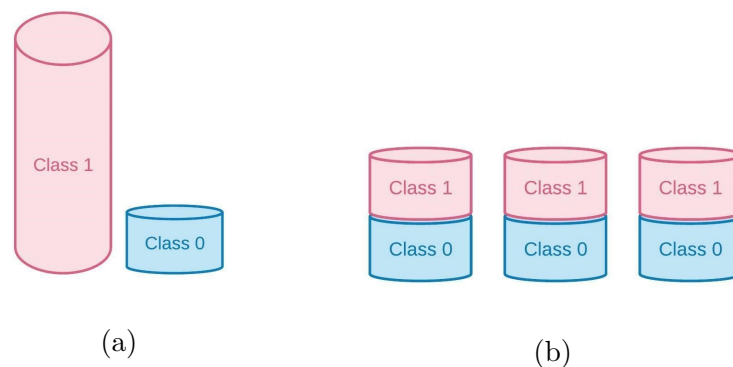
Ολική Χοληστερόλη - Cholesterol, υπολογίζεται προσθέτοντας την HDL, την LDL και 20% των τριγλυκεριδίων. Η τιμή της δεν πρέπει να ξεπερνά τα 200 mg/dL.

Γλυκόζη αίματος - Blood Glucose, το σάκχαρο ή γλυκόζη είναι ένας μονοσακχαρίτης που βρίσκεται φυσικά σε φρούτα. Σχηματίζεται επίσης από την πέψη των υδατανθράκων και τη μετατροπή του γλυκογόνου από το ήπαρ και είναι η κύρια πηγή κυτταρικής ενέργειας στον οργανισμό. Η γλυκόζη είναι απαραίτητη για τη λειτουργία του εγκεφάλου και των ερυθροκυττάρων. Η περίσσεια γλυκόζης αποθηκεύεται ως γλυκογόνο στο ήπαρ και τα μυϊκά κύτταρα. Το αγγειακό εγκεφαλικό επεισόδιο ανήκει στις πιθανές ερμηνείες παθολογικών τιμών της γλυκόζης. Οι φυσιολογικές τιμές της δεν πρέπει να υπερβαίνουν τα 100 mg/dL σε κατάσταση νηστείας και τα 140 mg/dL 2 ώρες μετά το φαγητό. Ακόμα μελέτες όπως οι [53], [54] παρουσιάζουν την άμεση συσχέτιση ανάμεσα σε ασθενής με αυξημένη γλυκόζη αίματος και καρωτιδική νόσο και τονίζουν την συνύπαρξη των δύο αυτών ασθενειών.

4.1.3 Ιδιαιτερότητες του συνόλου δεδομένων

Τα δύο μεγαλύτερα προβλήματα που χρειάστηκε να αντιμετωπίσουμε κατά την δημιουργία των μοντέλων ήταν η μικρή έκταση του συνόλου δεδομένων καθώς και η τεράστια ανισοροπία ανάμεσα στις δύο κλάσεις στις οποίες είναι διαχωρισμένο. Πιο συγκεκριμένα, όπως αναφέραμε και προηγουμένως, είχαμε στη διάθεση μας δεδομένα από 51 ασθενείς. Από αυτούς οι 39 (69.24% του συνόλου των δεδομένων) ανήκουν στην κλάση High Risk και οι 12 (30.76% του συνόλου των δεδομένων) στην κλάση Low Risk. Η ιδιαιτερότητα αυτή κάνει αυτομάτως το σύνολο των δεδομένων μας ανισόροπο (imbalanced) και καλούμαστε να το αντιμετωπίσουμε καταλλήλως για τη σωστή εκπαίδευση των μοντέλων μας.

Ο τρόπος αντιμετώπισης που επιλέξαμε είναι η χρήση Συλλογικών μοντέλων (Ensemble Neural Networks - ENN) των οποίων τα επιμέρους μοντέλα θα εκπαιδεύονται με διαφορετικά ισάριθμα υποσύνολα του αρχικού συνόλου δεδομένων [55]. Κάθε ένα από τα υποσύνολα αυτά αποτελείται από όλα τα δείγματα που ανήκουν στην κλάση της μειοψηφίας, δηλαδή τους ασθενείς χαμηλού κινδύνου και από μέρος των δειγμάτων που ανήκουν στην κλάση της πλειοψηφίας δηλαδή τους ασθενείς υψηλού κινδύνου. Στο παρακάτω σχήμα φαίνεται μία αναπαράσταση του διαχωρισμού που περιγράψαμε προηγουμένως.



Σχήμα 4.3: (a) Αρχικό ανισόροπο σύνολο δεδομένων, όπου με Class 1 συμβολίζουμε την κλάση των ασθενών Υψηλού κινδύνου (High risk) και με Class 0 την κλάση των ασθενών Χαμηλού κινδύνου (Low risk) (b) Διαχωρισμός του αρχικού ανισόροπου συνόλου δεδομένων σε 3 διαφορετικά ισοροπημένα υποσύνολα, χωρίζοντας την Class 1 σε τρία ανεξάρτητα υποσύνολα.

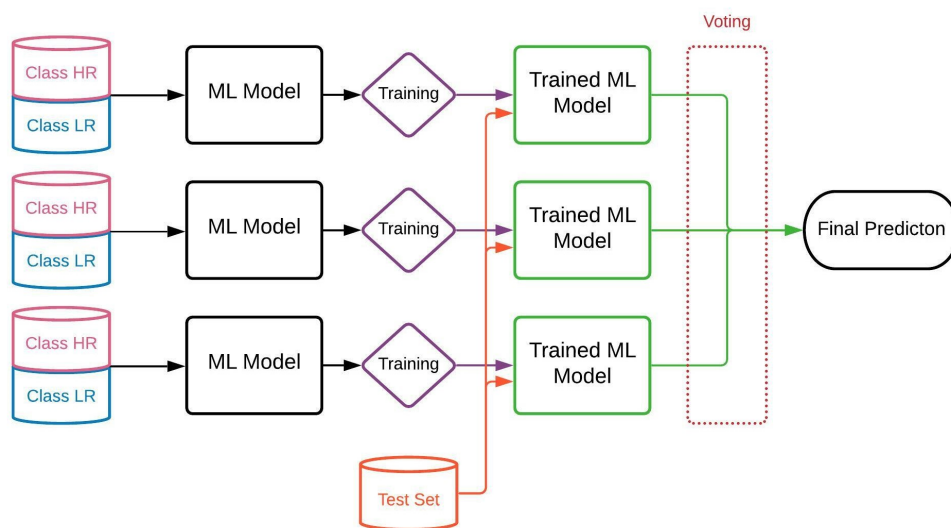
4.2 Μεθοδολογία κατασκευής αλγορίθμων μηχανικής μάθησης

Για την κατασκευή όλων των μοντέλων μηχανικής μάθησης που θα αναλυθούν στη συνέχεια χρησιμοποιήσαμε τη γλώσσα προγραμματισμού Python και τη βιβλιοθήκη της PyTorch.

4.2.1 Σχήμα Συλλογικής Μάθησης

Για την αντιμετώπιση της ανισορροπίας των δεδομένων χρησιμοποιήσαμε την τεχνική της Συλλογικής μάθησης (Ensemble Learning) στην κατασκευή των μοντέλων μας. Η τεχνική που θα αναλύσουμε στην συνέχεια χρησιμοποιήθηκε για την κατασκευή και των δύο μοντέλων (Απλό Συλλογικό μοντέλο και Συνελικτικό Συλλογικό μοντέλο) αλλάζοντας κάθε φορά το είδος και την αρχιτεκτονική των επιμέρους μοντέλων.

Στην παρούσα εργασία κάθε μοντέλο Συλλογικής μάθησης (Ensemble Neural Network - ENN) απαρτίζεται από τρία ίδια μοντέλα μηχανικής μάθησης τα οποία εκπαιδεύονται ξεχωριστά με ένα από τα 3 ισορροπημένα υποσύνολα του αρχικού συνόλου εκπαίδευσης. Στη συνέχεια η πρόβλεψη του τελικού αποτελέσματος προκύπτει μέσω των επιμέρους προβλέψεων των τριών μοντέλων. Στο Σχήμα. 4.4 φαίνεται η αρχιτεκτονική του ENN που χρησιμοποιήσαμε για την ανάπτυξη των μοντέλων μας στη συνέχεια.



Σχήμα 4.4: Αρχιτεκτονική Συλλογικού μοντέλου μηχανικής μάθησης. Κάθε ισορροπημένο υποσύνολο του αρχικού συνόλου δεδομένων αποτελεί είσοδο σε κάθε ένα από τα επιμέρους μοντέλα τα οποία στη συνέχεια εκπαιδεύονται ξεχωριστά. Η τελική απόφαση του Συλλογικού μοντέλου λαμβάνεται μέσω του συμφηρισμού των προβλέψεων των τριών μοντέλων (επίπεδο voting).

Περισσότερες λεπτομέρειες για τον τρόπο που επιλέξαμε να γίνει ο συμφηρισμός των αποτελεσμάτων των επιμέρους μοντέλων θα δοθούν στην συνέχεια, ξεχωριστά για κάθε μοντέλο μηχανικής μάθησης.

4.2.2 Συλλογικό Συνελικτικό Νευρωνικό Δίκτυο

Το μοντέλο που επιλέξαμε να δημιουργήσουμε για την κατηγοριοποίηση των υπερήχων β-σάρωσης στις κλάσεις Υψηλού και Χαμηλού κινδύνου βασίζεται στην τεχνική της Συλλογικής

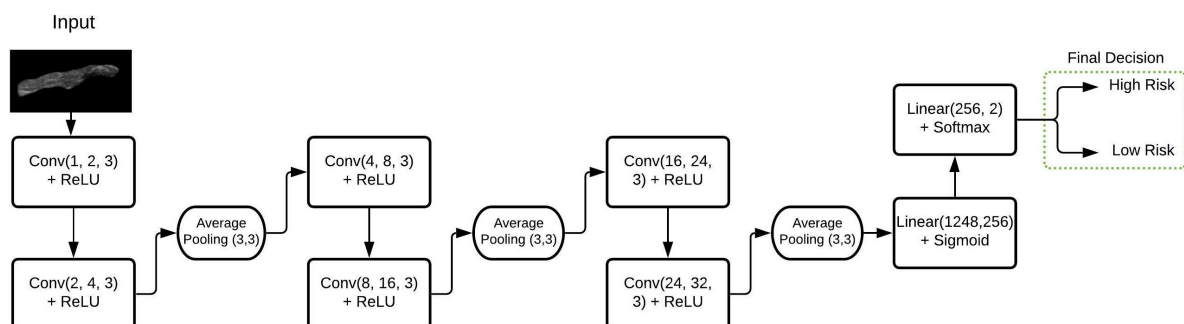
μάθησης. Πιο συγκεκριμένα για τη κατασκευή του χρησιμοποιήσαμε 3 ίδια συνελικτικά νευρωνικά δίκτυα (CNN) τα οποία εκπαιδεύτηκαν ξεχωριστά με ένα από τα 3 ισορροπημένα υποσύνολα του αρχικού συνόλου δεδομένων, τα οποία είναι διαχωρισμένα με τον τρόπο που περιγράψαμε στην Ενότητα 4.1.3. Τέλος, η πρόβλεψη του τελικού συλλογικού μοντέλου προέκυψε μέσω του συμφηρισμού των επιμέρους προβλέψεων των τριών μοντέλων.

Αρχιτεκτονική επιμέρους CNNs

Στην Ενότητα 2.3 περιγράψαμε τα βασικά δομικά στοιχεία ενός CNN μοντέλου. Με βάση αυτά, λοιπόν, η αρχιτεκτονική που αποφασίσαμε να ακολουθήσουμε για την κατασκευή των επιμέρους συνελικτικών δικτύων του Συλλογικού μοντέλου μηχανικής μάθησης (Σχήμα 4.5)

¹ είναι η εξής:

- Το κάθε επιμέρους μοντέλο αποτελείται από 3 ζεύγη συνελικτικών επιπέδων, στα οποία χρησιμοποιείται πυρήνας συνέλιξης με διαστάσεις (3, 3) και εφαρμόζεται συνάρτηση ενεργοποίησης **Relu**.
- Το κάθε ζεύγος συνελικτικών επιπέδων ακολουθείται από επίπεδο **Average Pooling** με διαστάσεις (3, 3).
- Έπειτα από τα συνελικτικά επίπεδα, ακολουθεί το πρώτο πλήρως συνδεδεμένο επίπεδο για τη δημιουργία μονοδιάστατης αναπαράστασης με είσοδο 1248 νευρώνων και έξοδο 256 νευρώνων με **Sigmoid** συνάρτηση ενεργοποίησης.
- Το δεύτερο πλήρως συνδεδεμένο επίπεδο έχει είσοδο 256 νευρώνων και έξοδο 2 νευρώνων με συνάρτηση ενεργοποίησης **Softmax**, μέσω της οποίας προκύπτουν οι τελικές έξοδοι του μοντέλου οι οποίες αντιστοιχούν στις τιμές πιθανότητας που έχει προβλέψει το μοντέλο για κάθε μία από τις δύο κλάσεις (High risk και Low risk).



Σχήμα 4.5: Αρχιτεκτονική CNN μοντέλων. Στα συνελικτικά επίπεδα (Conv) ο πρώτος αριθμός μέσα στην παρένθεση αντιπροσωπεύει τον αριθμό των εικόνων που λαμβάνονται ως είσοδος στο εκάστοτε επίπεδο, ο δεύτερος τον αριθμό των εικόνων που προκύπτουν ως έξοδος από το επίπεδο και ο τρίτος την διάσταση του συνελικτικού πυρήνα που χρησιμοποιείται.

Εκπαίδευση επιμέρους CNNs

Για την εκπαίδευση των μοντέλων μας επιλέξαμε να ακολουθήσουμε τις παρακάτω επιλογές.

¹Η αρχιτεκτονική αυτή είναι κοινή και για τα 3 μοντέλα

- Επιλέξαμε ως συνάρτηση κόστους (Loss Function) την **Binary Cross Entropy Loss** καθώς αυτή είναι κατάλληλη για προβλήματα δυαδικής ταξινόμησης και έδωσε τα καλύτερα αποτελέσματα στο παρόν μοντέλο.
- Ως **βελτιστοποιητή** δοκιμάσαμε τον Adam και τον SGD optimizer, αλλά καταλήξαμε στην επιλογή του **Adam** καθώς έδωσε τα βέλτιστα αποτελέσματα στο πρόβλημά μας.
- Καταλήξαμε πως **10 εποχές** ήταν αρκετές έτσι ώστε τα επιμέρους CNN μοντέλα να εκπαιδευτούν επαρκώς αποφεύγοντας το φαινόμενο της υπερεκπαίδευσης (overfitting).
- Ο ρυθμός εκπαίδευσης, **learning rate**, που χρησιμοποιήσαμε ήταν 0.0005 καθώς για κάθε τιμή μεγαλύτερης ή μικρότερης τάξεως το μοντέλο αδυνατούσε να εκπαιδευτεί επαρκώς.
- Τέλος το **Batch Size** που χρησιμοποιήσαμε είναι ίσο με 8.

Συλλογή αποτελεσμάτων και εξαγωγή τελικής πρόβλεψης

Μετά την εκπαίδευση των επιμέρους μοντέλων το σύνολο αξιολόγησης (test set) δίνεται ως είσοδος σε κάθε ένα από αυτά. Για κάθε δείγμα του test set κάθε ένα από τα μοντέλα προβλέπει την πιθανότητα με την οποία ανήκει σε κάθε μία από τις δύο κλάσεις. Στη συνέχεια οι τρεις προβλέψεις συμψηφίζονται και προκύπτει η τελική πρόβλεψη του συλλογικού μοντέλου. Στο συγκεκριμένο μοντέλο επιλέξαμε να κάνουμε τον συμψηφισμό των τελικών προβλέψεων υπολογίζοντας το μέσο όρο των πιθανοτήτων που εξαγάγουν τα επιμέρους μοντέλα για κάθε μία από τις δύο κλάσεις, με αποτέλεσμα η κλάση με τον μεγαλύτερο μέσο όρο πιθανότητας να αποτελέσει και την τελική απόφαση του συλλογικού μοντέλου.

4.2.3 Συλλογικό Απλό Νευρωνικό Δίκτυο

Για την κατασκευή του μοντέλου που επεξεργάζεται και παράγει προβλέψεις με βάση τα κλινικά δεδομένα κάθε ασθενή, χρησιμοποιήσαμε και πάλι την τεχνική της συλλογικής μάθησης. Όπως και προηγουμένως δημιουργήσαμε ένα συλλογικό μοντέλο αποτελούμενο από τρία, ίδιας αρχιτεκτονικής, απλά νευρωνικά δίκτυα (Simple Neural Networks - SNN). Κάθε ένα από αυτά εκπαιδεύτηκε με ένα υποσύνολο του αρχικού συνόλου δεδομένων το οποίο είναι ισορροπημένο, δηλαδή εμπεριέχει ίδιο ποσοστό δεδομένων και από τις δύο κλάσεις. Στη συνέχεια τα εκπαιδευμένα πλέον μοντέλα συνεργάζονται (επίπεδο voting) εξαγάγοντας με αυτόν τον τρόπο την τελική πρόβλεψη του συλλογικού μοντέλου. Στη συνέχεια θα περιγράψουμε αναλυτικά τα τρία στάδια της κατασκευής του τελικού συλλογικού απλού νευρωνικού δικτύου.

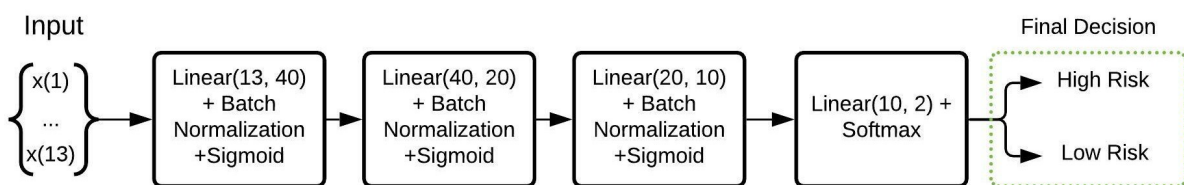
Αρχιτεκτονική επιμέρους SNNs

Μετά από δοκιμές διάφορων αρχιτεκτονικών, καταλήξαμε στην παρακάτω λόγω της υψηλότερης απόδοσής της². Το κάθε layer όπου χρησιμοποιήσαμε είναι πλήρως συνδεδεμένο με το προηγούμενό του και τα βάρη της σύνδεσης καθορίζονται με μεθόδους gradient descent [56]. Το επόμενο μέρος του layer αποτελείται από το Batch Normalization, που χρησιμοποιήθηκε για να αντιμετωπίσουμε το πρόβλημα του internal covariance shift [57]. Το

²Η αρχιτεκτονική αυτή είναι κοινή και για τα 3 μοντέλα

Batch Normalization εφαρμόζει κανονικοποίηση των δεδομένων που εξάγονται από τον linear μετασχηματισμό, επιτυγχάνοντας αριθμητική σταθερότητα. Ως εκ τούτου, μπορούμε να χρησιμοποιήσουμε μεγαλύτερο learning rate (ρυθμό μάθησης) και να πετύχουμε ταχύτερη και πιο αποτελεσματική εκπαίδευση. Το τελευταίο κομμάτι του layer αποτελείται από την συνάρτηση ενεργοποίησης Sigmoid.

Τα τρία παραπάνω μέρη, δηλαδή το γραμμικό μέρος σύνδεσης με το προηγούμενο layer, το Batch Normalization και η συνάρτηση ενεργοποίησης, με τη σειρά αυτή δημιουργούν ένα πιο σύνθετο Layer. Το layer αυτό τοποθετήθηκε τρεις φορές στη σειρά για την κατασκευή του απλού νευρωνικού δικτύου. Η αναλυτική αρχιτεκτονική του μοντέλου μας φαίνεται στο Σχήμα. 4.6. Αξίζει να σημειώσουμε ότι στο τελευταίο layer έχουμε ένα linear μετασχηματισμό με έξοδο 2 στοιχεία τα οποία ακολουθούνται από Softmax ενεργοποίηση και καθορίζουν τις τελικές πιθανότητες High και Low Risk. Επίσης, στην είσοδο του μοντέλου παρατηρούμε τα 13 διαφορετικά χαρακτηριστικά (features) των κλινικών δεδομένων.



Σχήμα 4.6: Αρχιτεκτονική των επιμέρους SNN μοντέλων. Στα γραμμικά μέρη σύνδεσης (Linear) ο πρώτος αριθμός μέσα στην παρένθεση αναφέρεται στον αριθμό των νευρώνων της εισόδου του εκάστοτε layer ενώ ο δεύτερος στον αριθμό νευρώνων της εξόδου.

Εκπαίδευση επιμέρους SNNs

Για την εκπαίδευση των SNN επιλέξαμε να χρησιμοποιήσουμε τις εξής παραμέτρους:

- Όπως και προηγουμένως, στην εκπαίδευση των CNN, έτσι και τώρα επιλέξαμε ως συνάρτηση κόστους την **Binary Cross Entropy Loss**, λόγω της δυαδικής φύσης του προβλήματός μας.
- Ως βελτιστοποιητή επιλέξαμε τον **Adam** καθώς μέσω αυτού λάβαμε τα βέλτιστα αποτελέσματα.
- Οι **εποχές** εκπαίδευσης του κάθε μοντέλου είναι **20**.
- Ο ρυθμός εκπαίδευσης, **learning rate**, τέθηκε ίσος με **0.055**
- Τέλος, το **batch size** που επιλέξαμε είναι ίσο με **8** όπως και προηγουμένως καθώς το πλήθος των δεδομένων που έχουμε να διαχειριστούμε είναι το ίδιο.

Συλλογή αποτελεσμάτων και εξαγωγή τελικής πρόβλεψης

Για την συλλογή των αποτελεσμάτων επιλέξαμε να χρησιμοποιήσουμε την μέθοδο του μέσου όρου των πιθανοτήτων των τριών μοντέλων, όπως και στο Συνελικτικό συλλογικό μοντέλο. Η επιλογή αυτής της τεχνικής έγινε με βάση τα αποτελέσματα που έδωσε, τα οποία ήταν τα βέλτιστα σε σχέση με άλλες τεχνικές που δοκιμάστηκαν.

4.3 Μεθοδολογία κατασκευής μεθόδων ερμηνευσιμότητας

Τα μοντέλα μηχανικής μάθησης που υλοποιήσαμε δεν είναι ερμηνεύσιμα, δηλαδή αρκούνται στο να δώσουν μία απλή "απάντηση" για την κλάση στην οποία προβλέπουν πως ανήκει το κάθε δείγμα, χωρίς να επεξηγήσουν τις αιτίες που τα οδήγησαν να λάβουν αυτή την απόφαση. Αυτό έχει ως αποτέλεσμα να χάνουν την αξιοπιστία τους και να μην μπορούν να χρησιμοποιηθούν σε εφαρμογές που διαχειρίζονται ευαίσθητα ζητήματα γύρω από την υγεία, όπως και το πρόβλημα που καλούμαστε να αντιμετωπίσουμε στην παρούσα εργασία. Επιλέξαμε λοιπόν να χρησιμοποιήσουμε μεθόδους ερμηνευσιμότητας τις οποίες θα εφαρμόσουμε ξεχωριστά σε κάθε μοντέλο με σκοπό να βγάλουμε συμπεράσματα για την λειτουργία και τον τρόπο με τον οποίο λαμβάνουν τις αποφάσεις τους. Για την ερμηνευσιμότητα του απλού συλλογικού μοντέλου (SNN) θα χρησιμοποιήσουμε 3 διαφορετικές μεθόδους: την μέθοδο **Global Surrogate**, τα **ICE plots** και την μέθοδο **Shaplay Values**. Αντίστοιχα για να ερμηνεύσουμε το συνελικτικό συλλογικό μοντέλο (CNN) επιλέξαμε να χρησιμοποιήσουμε την μέθοδο **Shaplay Values**.

4.3.1 Global Surrogate

Η πρώτη μέθοδος που εφαρμόσαμε είναι η Global Surrogate [11]. Όπως αναφέραμε και στην Ενότητα 3.2.2 με τη μέθοδο αυτή δημιουργούμε ένα νέο ερμηνεύσιμο μοντέλο το οποίο εκπαιδεύεται στις προβλέψεις του μοντέλου που θέλουμε να ερμηνεύσουμε έτσι ώστε να προσεγγίζει τις προβλέψεις του. Στην παρούσα εργασία επιλέξαμε να χρησιμοποιήσουμε ένα δέντρο απόφασης, decision tree - DT, για την εξαγωγή των τελικών συμπερασμάτων καθώς μπορεί να οπτικοποιηθεί και να ερμηνευτεί με πολύ εύκολο και ευανάγνωστο τρόπο. Επίσης όπως αναφέραμε και στην Ενότητα 2.7 τα δέντρα απόφασης χρησιμοποιήθηκαν σε πολλές εργασίες μέχρι και σήμερα σχετικές με την λήψη κλινικών αποφάσεων αλλά και άλλων βιοϊατρικών προβλημάτων.

Τα βήματα που ακολουθήσαμε για την δημιουργία του Surrogate μοντέλου αλλά και για την εξαγωγή των συμπερασμάτων ερμηνευσιμότητας μέσω αυτού, περιγράφονται στην συνέχεια.

- Αφού εκπαιδεύσαμε και αξιολογήσαμε το SNN μοντέλο μας σε κάθε ένα από τα 5 folds της μεθόδου 5 Cross Validation διατηρήσαμε τις προβλέψεις που έδωσε για κάθε test set από κάθε fold.
- Στη συνέχεια, δημιουργήσαμε έναν νέο σύνολο δεδομένων το οποίο αποτελούνταν από τα δείγματα που ανήκαν στα σύνολα αξιολόγησης (test sets), κάθε fold, αλλάζοντας τους τα αρχικά τους label, δηλαδή τις τιμές που υποδηλώνουν σε ποια κατηγορία (high ή low risk) ανήκει το κάθε δείγμα, με τις προβλέψεις που προέκυψαν από το SNN μοντέλο.
- Με αυτό το νέο σύνολο δεδομένων εκπαιδεύσαμε και αξιολογήσαμε το υποκατάστατο μοντέλο, δηλαδή το δέντρο απόφασης. Για την δημιουργία του χρησιμοποιήσαμε τη βιβλιοθήκη της Python, scikit-learn [58] και το κριτήριο διαχωρισμού που επιλέξαμε ήταν το Gini Criterion.
- Στη συνέχεια, για να λάβουμε τα αποτελέσματα της ερμηνευσιμότητας οπτικοποιήσαμε το δέντρο απόφασης, έτσι ώστε να παρατηρήσουμε τα χαρακτηριστικά στα οποία βασίζεται ώστε να λάβει την τελική του απόφαση.

4.3.2 Individual Conditional Expectation

Η δεύτερη μέθοδος ερμηνευσιμότητας που επιλέξαμε να χρησιμοποιήσουμε είναι η μέθοδος Individual Conditional Expectation ή για συντομία τα ICE διαγράμματα (plots) [11]. Μέσω των ICE plots μπορούμε να παρατηρήσουμε τη συμπεριφορά που εμφανίζει το μοντέλο μηχανικής μάθησης, στην περίπτωση μας το απλό συλλογικό μοντέλο, στην έξοδο του, ανάλογα με τις αλλαγές σε ένα συγκεκριμένο χαρακτηριστικό (feature) ενώ όλα τα υπόλοιπα παραμένουν σταθερά. Στην παρούσα εργασία επιλέξαμε να ελέγξουμε 100 διαφορετικές τιμές από κάθε χαρακτηριστικό οι οποίες βρίσκονται στο εύρος της μέγιστης και της ελάχιστης τιμής αυτού. Το τελικό διάγραμμα θα αποτελείται από καμπύλες που παρουσιάζουν την συμπεριφορά της εξόδου του μοντέλου για κάθε ασθενή - δείγμα ξεχωριστά.

4.3.3 Shapley Values

Τέλος, χρησιμοποιήσουμε τη μέθοδο Shapley Values [11] και για την υλοποίησή της χρησιμοποιήσαμε τη βιβλιοθήκη SHAP της Python και πιο συγκεκριμένα τον DeepExplainer. Με την μέθοδο αυτή ερμηνεύσαμε και το απλό συλλογικό μοντέλο αλλά και το συνελικτικό συλλογικό μοντέλο. Επιλέξαμε να ερμηνεύσουμε κάθε επιμέρους μοντέλο των συλλογικών μοντέλων ξεχωριστά και στη συνέχεια να συγκεντρώσουμε τα χαρακτηριστικά με την μεγαλύτερη επιρροή καταλήγοντας σε ένα τελικό συμπέρασμα. Η επιλογή αυτή έγινε με βάση τη μέθοδο Shapley Values η οποία αδυνατεί να ερμηνεύσει με κάποιον άλλο τρόπο ένα Ensemble μοντέλο λόγω της αρχιτεκτονικής του.

Κεφάλαιο 5

Ανάλυση Αποτελεσμάτων

5.1 Αποτελέσματα μοντέλων μηχανικής μάθησης

Στον πίνακα 5.1, παρατίθενται τα αποτελέσματα του συνελικτικού συλλογικού μοντέλου.

Accuracy	Specificity	Sensitivity	F1-score	Roc/AUC
67.4%	76%	63.5%	75%	0.7

Πίνακας 5.1: Αποτελέσματα συνελικτικού συλλογικού μοντέλου

Τα αποτελέσματα ενώ δεν είναι ιδιαίτερα υψηλά, θεωρούνται ικανοποιητικά για το πλήθος και την ποιότητα των δεδομένων μας. Στη συνέχεια στον πίνακα 5.2 παρατίθενται τα αποτελέσματα του απλού συλλογικού μοντέλου.

Accuracy	Specificity	Sensitivity	F1-score	Roc/AUC
75%	80%	60%	70%	0.7

Πίνακας 5.2: Αποτελέσματα απλού συλλογικού μοντέλου

Παρατηρούμε πως τα αποτελέσματα του απλού συλλογικού μοντέλου κυμαίνονται στις ίδιες περίπου τιμές με του συνελικτικού μοντέλου. Όπως και στο συνελικτικό μοντέλο τα αποτελέσματα δεν είναι ιδιαίτερα υψηλά όμως είναι αρκετά ικανοποιητικά για τον πλήθος των δεδομένων που έχουμε στη διάθεσή μας. Πιο συγκεκριμένα παρατηρούμε πως η τιμή της μετρικής Sensitivity μπορεί να φτάσει και το ποσοστό 80%. Ενώ και για τα δύο μοντέλα οι τιμές της μετρικής F1-score η οποία χρησιμοποιείται κυρίως σε unbalanced δεδομένα δίνοντας τα πιο αντιπροσωπευτικά αποτελέσματα σε σχέση με τις άλλες μετρικές, είναι αρκούτως υψηλές.

5.2 Αποτελέσματα μεθόδων ερμηνευσιμότητας

Στη συνέχεια θα αναλύσουμε τα αποτελέσματα που εξαγάγαμε από τις μεθόδους ερμηνευσιμότητας που εφαρμόσαμε στα μοντέλα μηχανικής μάθησης. Αρχικά θα αναλύσουμε τα αποτελέσματα που λάβαμε από την ερμηνευσιμότητα του απλού συλλογικού μοντέλου και στη συνέχεια του συνελικτικού συλλογικού μοντέλου.

5.2.1 Ερμηνευσιμότητα Απλού Συλλογικού μοντέλου

Global Surrogate

Όπως αναφέραμε και στην Ενότητα 4.3.1 για την κατασκευή της Global Surrogate μεθόδου χρησιμοποιήσαμε ένα εγγενώς ερμηνεύσιμο μοντέλο, και πιο συγκεκριμένα ένα δέντρο απόφασης, decision tree - DT, με απώτερο σκοπό να λειτουργήσει ως υποκατάστατο για το μη ερμηνεύσιμο μοντέλο μας. Στον Πίνακα 5.3 παρουσιάζονται τα αποτελέσματα του "υποκατάστατου" μοντέλου, έτσι όπως αυτό εκπαιδεύτηκε και αξιολογήθηκε στο νέο σύνολο δεδομένων, καθώς και η τιμή της μετρικής R-square, η οποία υποδηλώνει πόσο καλά αντιπροσωπεύει το προς ερμηνεία μοντέλο μας.

Accuracy	Specificity	Sensitivity	Roc/AUC	R-Square
68.42%	70%	66%	0.68	0.2

Πίνακας 5.3: Αποτελέσματα της αξιολόγησης του Δέντρου απόφασης κατά τη λειτουργία του ως global surrogate μοντέλο

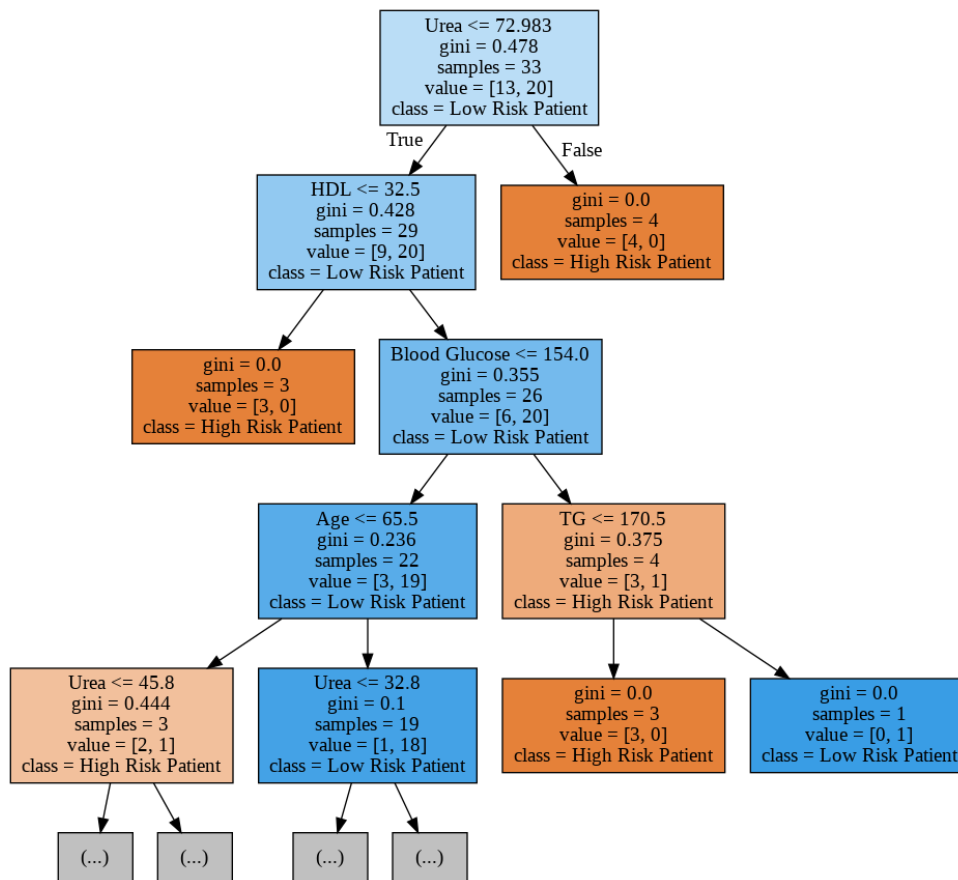
Παρατηρούμε πως τα αποτελέσματα του DT είναι αρκετά καλά, συμπεραίνοντας έτσι ότι υποκαθιστά με επιτυχία το μη ερμηνεύσιμο μοντέλο μας. Η τιμή της μετρικής R-square δεν είναι κοντά στην μονάδα όμως αυτό οφείλεται στο πολύ μικρό αριθμό δεδομένων που έχουμε στη διάθεσή μας και όχι στη μη σωστή λειτουργία του υποκατάστατου μοντέλου.

Στην συνέχεια παρουσιάζουμε το οπτικοποιημένο DT, Σχήμα 5.1, έτσι όπως προέκυψε μετά από την διαδικασία της εκπαίδευσής του, με σκοπό να το ερμηνεύσουμε ποιοτικά. Στην κορυφή του δέντρου (ρίζα) γίνονται οι διαχωρισμοί που είναι βασικότερης σημασίας στην κατάταξη των δειγμάτων ενώ όσο οι κόμβοι καταλήγουν προς τα φύλλα οι διαχωρισμοί είναι μικρότερης σημαντικότητας. Σε κάθε κόμβο εμφανίζεται το χαρακτηριστικό και η αντίστοιχη τιμή του, με βάση το οποίο γίνεται ο διαχωρισμός των δειγμάτων, η τιμή του κριτηρίου του διαχωρισμού στην περίπτωση μας του Gini κριτηρίου, ο αριθμός των δειγμάτων με βάση τα οποία εκπαιδεύτηκε το DT καθώς και ο ακριβής αριθμός δειγμάτων κάθε κλάσης (High και Low risk) πριν το διαχωρισμό. Τέλος σημειώνεται και η κλάση των δειγμάτων της πλειοψηφίας.

Πιο συγκεκριμένα, στην περίπτωση μας βλέπουμε ότι ο πιο σημαντικός διαχωρισμός είναι αυτός που προέκυψε με βάση τις τιμές του χαρακτηριστικού της Ουρίας (Urea), δηλαδή 4 ασθενείς κατατάχθηκαν, χωρίς κανέναν άλλο έλεγχο, στην κλάση των ασθενών υψηλού κινδύνου καθώς είχαν πολύ υψηλή τιμή ουρίας. Το συμπέρασμα αυτό συμβαδίζει με κλινικές μελέτες [45], [46] που έχουν δείξει πως οι υψηλές τιμές ουρίας είναι πιθανόν να εμφανιστούν σε ασθενείς που πάσχουν από καρωτιδική νόσο. Ο αμέσως επόμενος έλεγχος που γίνεται από το δέντρο απόφασης αφορά την τιμή της HDL χοληστερόλης. Παρατηρούμε πως οι ασθενείς με χαμηλές τιμές HDL χοληστερόλης κατατάσσονται ως ασθενείς υψηλού κινδύνου. Όπως και προηγουμένως, το συμπέρασμα που λαμβάνουμε και από αυτόν τον διαχωρισμό είναι αρκετά ενθαρρυντικό ως προς την εγγύτητα του μοντέλου μας καθώς οι χαμηλές τιμές HDL - καλής χοληστερόλης σχετίζονται με την ανάπτυξη των αθηρωμάτων [59], [60] επειδή συντελούν στην διατήρηση της LDL - κακής χοληστερόλης η οποία είναι ένας από τους βασικότερους παράγοντες για την ανάπτυξη των αθηρωμάτων [61].

Στη συνέχεια ο αμέσως επόμενος σε σημαντικότητα διαχωρισμός είναι αυτός που γίνεται με βάση την τιμή της γλυκόζης του αίματος (Blood Glucose), κατά τον οποίο οι ασθενείς με υψηλές τιμές κατατάσσονται στην κλάση του υψηλού κινδύνου. Και αυτή η παρατήρηση

συμβαδίζει με ότι είναι κλινικά αποδεδειγμένο καθώς στις μελέτες [53], [54] παρατηρήθηκε πως οι υψηλές τιμές γλυκόζης στο αίμα είναι ένα κοινό χαρακτηριστικό στους ασθενείς με αθηροσκλήρωση των καρωτιδικών αρτηριών.



Σχήμα 5.1: Ερμηνευσιμότητα Απλού Συλλογικού μοντέλου μέσω ενός Δέντρου Απόφασης (Decision Tree) που λειτουργεί ως υποκατάστατο μοντέλο. Στο διάγραμμα οι κόμβοι που είναι χρωματισμένοι με αποχρώσεις του πορτοκαλί δηλώνουν ότι τα περισσότερα δείγματα, πριν να πραγματοποιηθεί ο διαχωρισμός, ανήκουν στην κλάση των ασθενών Υψηλού κινδύνου (High risk) ενώ με αποχρώσεις του μπλε σημειώνονται οι κόμβοι που υποδηλώνουν το αντίθετο, δηλαδή την κλάση Χαμηλού κινδύνου (Low risk). Οι πιο σκούρες αποχρώσεις υποδηλώνουν πως ο κόμβος είναι τερματικός (φύλλο).

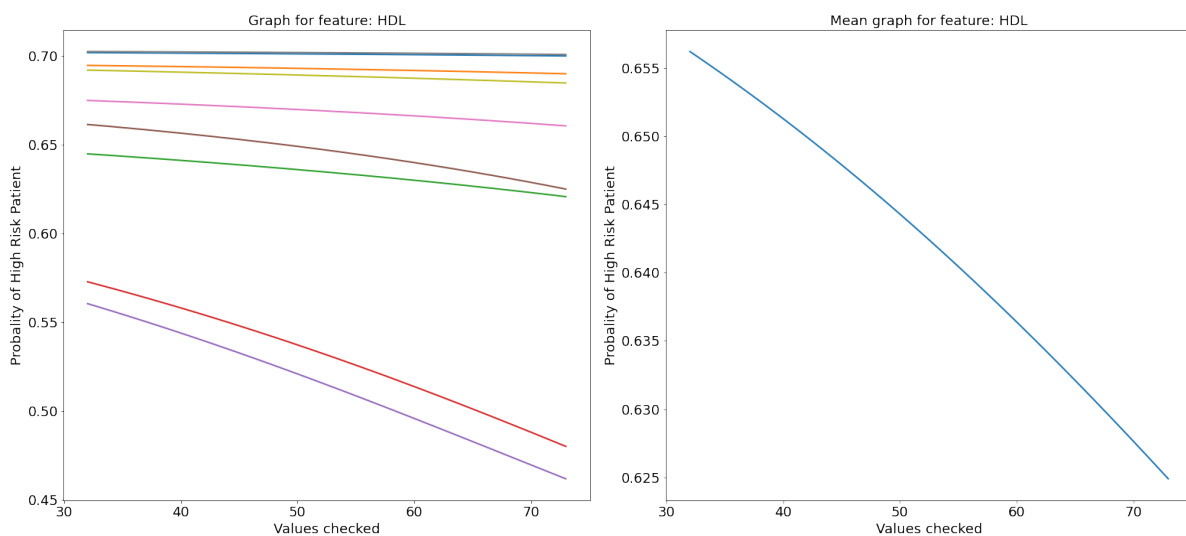
Individual Conditional Expectation (ICE)

Η δεύτερη μέθοδος ερμηνευσιμότητας που επιλέξαμε να εφαρμόσουμε είναι η μέθοδος Individual Conditional Expectation ή για συντομία ICE. Στη μέθοδο αυτή δημιουργούμε διαγράμματα τα οποία αντιπροσωπεύουν την συμπεριφορά που εμφανίζει το μοντέλο όταν ένα συγκεκριμένο χαρακτηριστικό αλλάζει την τιμή του ενώ όλα τα υπόλοιπα διατηρούνται σταθερά. Τα διαγράμματα αυτά ονομάζονται ICE plots. Όπως αναφέραμε και στην Ενότητα 4.3.2 στην παρούσα εργασία επιλέξαμε να εξετάσουμε για κάθε χαρακτηριστικό 100 διαφορετικές τιμές οι οποίες λήφθηκαν από το διάστημα μεταξύ της μέγιστης και της ελάχιστης τιμής του, με σταθερό βήμα. Έτσι λοιπόν, κάθε ICE διάγραμμα αποτελείται από καμπύλες, μία για κάθε εξεταζόμενο ασθενή οι οποίες αντιπροσωπεύουν τις μεταβολές του μοντέλου. Για να γίνουν πιο ευανάγνωστα τα διαγράμματα επιλέξαμε να δημιουργήσουμε ένα επιπλέον διάγραμμα το οποίο αποτελείται από μία μόνο καμπύλη η οποία προκύπτει από τον μέσο όρο

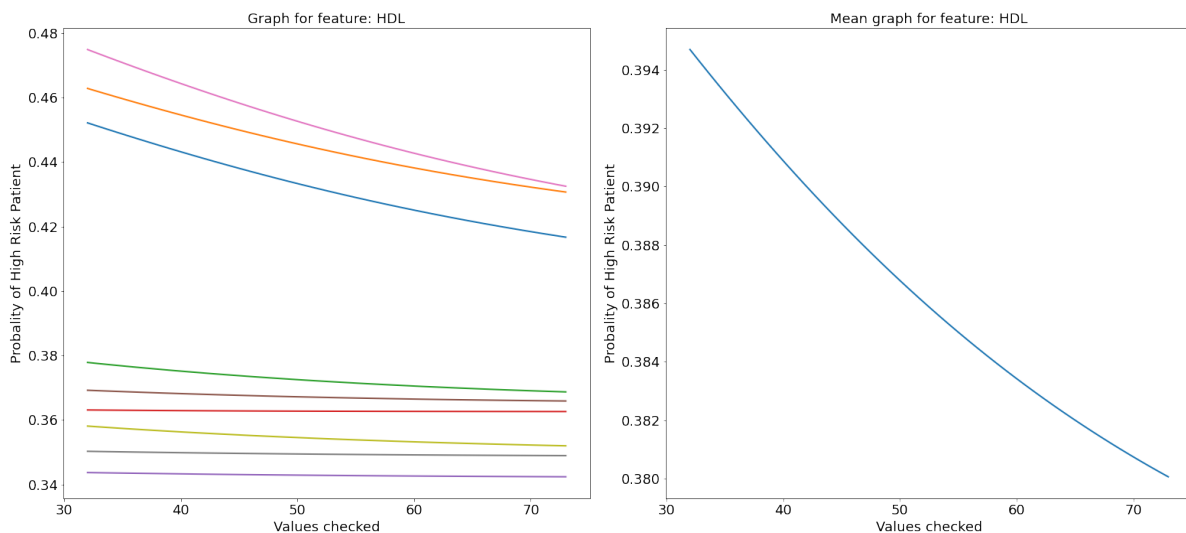
των συνολικών παρατηρήσεων.

Στον x-άξονα των διαγραμμάτων εμφανίζονται οι 100 διαφορετικές τιμές του εξεταζόμενου χαρακτηριστικού ενώ στον y-άξονα οι αντίστοιχες πιθανότητες που εξαγάγει το μοντέλο με βάση την αλλαγή αυτή. Οι τιμές πιθανότητας που εμφανίζονται στον y-άξονα αντιστοιχούν στην πιθανότητα ενός ασθενή να ταξινομηθεί από το μοντέλο στην κλάση υψηλού κινδύνου.

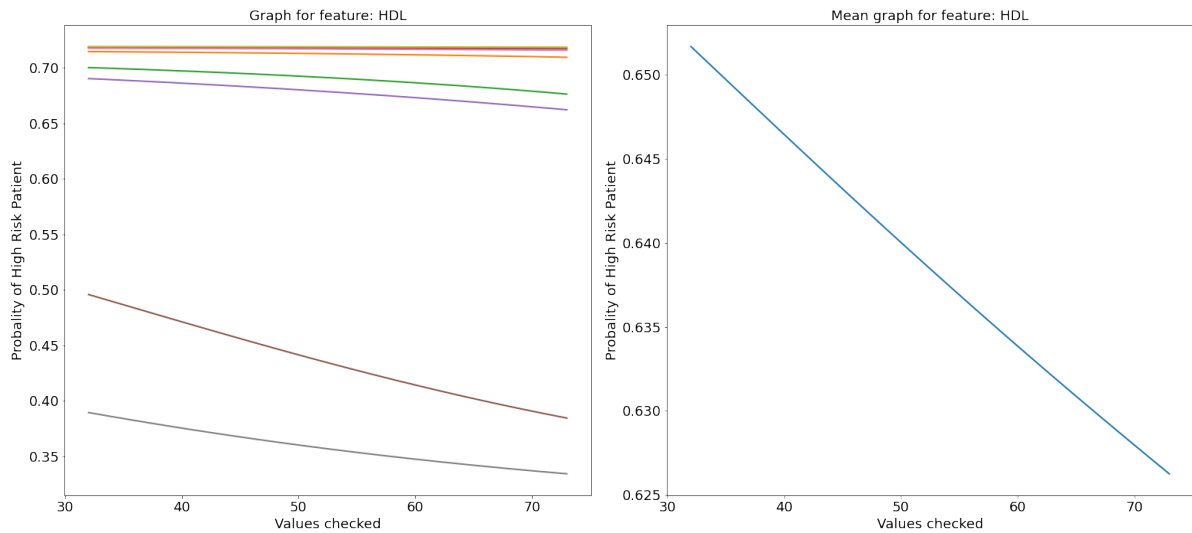
Αρχικά εφαρμόσαμε την ICE μέθοδο σε κάθε ένα από τα επιμέρους μοντέλα ξεχωριστά και στη συνέχεια εμφανίζουμε τα αποτελέσματά του για τα χαρακτηριστικά: HDL χοληστερόλη (HDL), LDL χοληστερόλη (LDL), Γλυκόζη αίματος (Blood Glucose) και Τριγλυκερίδια (TG)



Σχήμα 5.2: ICE plot 1ου επιμέρους μοντέλου για το χαρακτηριστικό της HDL χοληστερόλης. Το αριστερό διάγραμμα αποτελείται από καμπύλες που η κάθε μία αντιπροσωπεύει έναν ξεχωριστό ασθενή. Το δεξί διάγραμμα αποτελεί τον μέσο όρο αυτών των καμπυλών

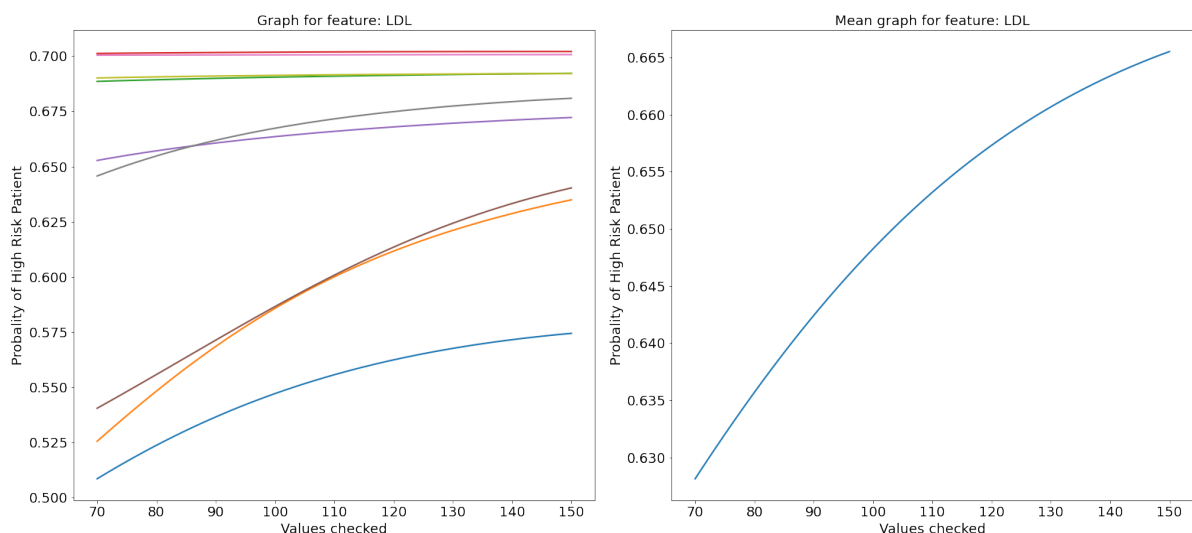


Σχήμα 5.3: ICE plot 2ου επιμέρους μοντέλου για το χαρακτηριστικό της HDL χοληστερόλης



Σχήμα 5.4: ICE plot 3ου επιμέρους μοντέλου για το χαρακτηριστικό της HDL χοληστερόλης

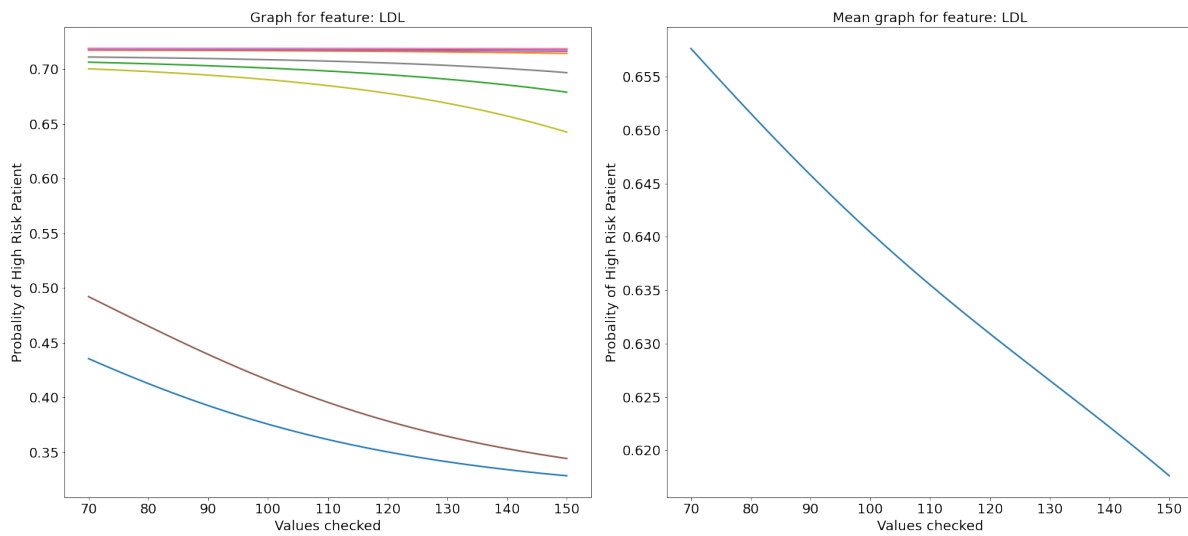
Παρατηρούμε πως τα ICE διαγράμματα για το χαρακτηριστικό της HDL χοληστερόλης παρουσιάζουν την επιθυμητή συμπεριφορά και για τα τρία επιμέρους μοντέλα, Σχήματα 5.2, 5.3, 5.4. Πιο συγκεκριμένα για μικρές τιμές της HDL χοληστερόλης η πιθανότητα ενός ασθενή να ταξινομηθεί στην κλάση υψηλού κινδύνου είναι μεγαλύτερη από ότι για μεγαλύτερες τιμές της. Σε μελέτες που έχουν πραγματοποιηθεί βρέθηκε πως τα υψηλά επίπεδα HDL χοληστερόλης έχουν συσχετιστεί με μικρότερο κίνδυνο εμφάνισης καρδιαγγειακών νοσημάτων. Αυτό συμβαίνει καθώς η HDL χοληστερόλη αφαιρεί τη χοληστερόλη από τα αφρώδη κύτταρα αναστέλλοντας την οξειδωση της LDL χοληστερόλης. Με αυτόν τον τρόπο καταφέρνει να περιορίσει της φλεγμονώδεις διαδικασίες που αποτελούν την βάση της αθηροσκλήρωσης [62]. Ειδικότερά έχει παρατηρηθεί πως όταν η τιμή της HDL χοληστερόλης ξεπεράσει τα $60mg/dl$ θεωρείται ως "αντίστροφος παράγοντας κινδύνου", δηλαδή μειώνει την πιθανότητα αθηροσκλήρωσης [63].



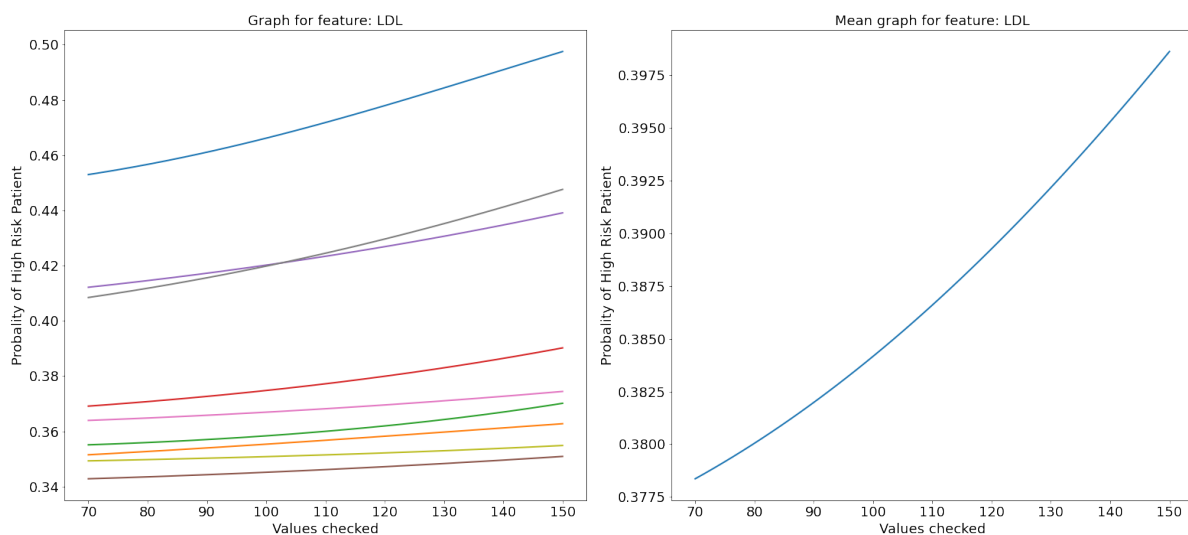
Σχήμα 5.5: ICE plot 1ου επιμέρους μοντέλου για το χαρακτηριστικό της LDL χοληστερόλης

Την αντίστροφη συμπεριφορά αναμένουμε να λάβουμε από το χαρακτηριστικό της LDL χοληστερόλης, Σχήματα. 5.5, 5.6, 5.7. Καθώς οι μεγάλες τιμές της ευνοούν την ανάπτυξη

της αθηρωματικής πλάκας και κατ' επέκταση της αθηρωματικής νόσου [64]. Παρόλα αυτά παρατηρούμε πως ενώ το πρώτο και το τρίτο επιμέρους μοντέλο ανταποκρίνονται, το δεύτερο μοντέλο συμπεριφέρεται διαφορετικά. Πιο συγκεκριμένα, βλέπουμε πως για μεγάλες τιμές της LDL χοληστερόλης η τιμή της πιθανότητας να ταξινομηθεί ο ασθενής στην κλάση υψηλού κινδύνου μειώνεται αντί να αυξάνεται. Η συμπεριφορά αυτή είναι παράδοξη και δεν συνάδει με αυτές των υπολοίπων επιμέρους μοντέλων. Εφόσον τα άλλα δύο επιμέρους μοντέλα δεν εμφανίζουν "πρόβλημα" και η αρχιτεκτονική και των τριών μοντέλων είναι η ίδια καταλήγουμε στο συμπέρασμα πως η αστοχία αυτή πιθανόν να προκύπτει από τον αρχικό διαμοιρασμό των δεδομένων στα μοντέλα, όπως περιγράψαμε στη Ενότητα 4.1.3.



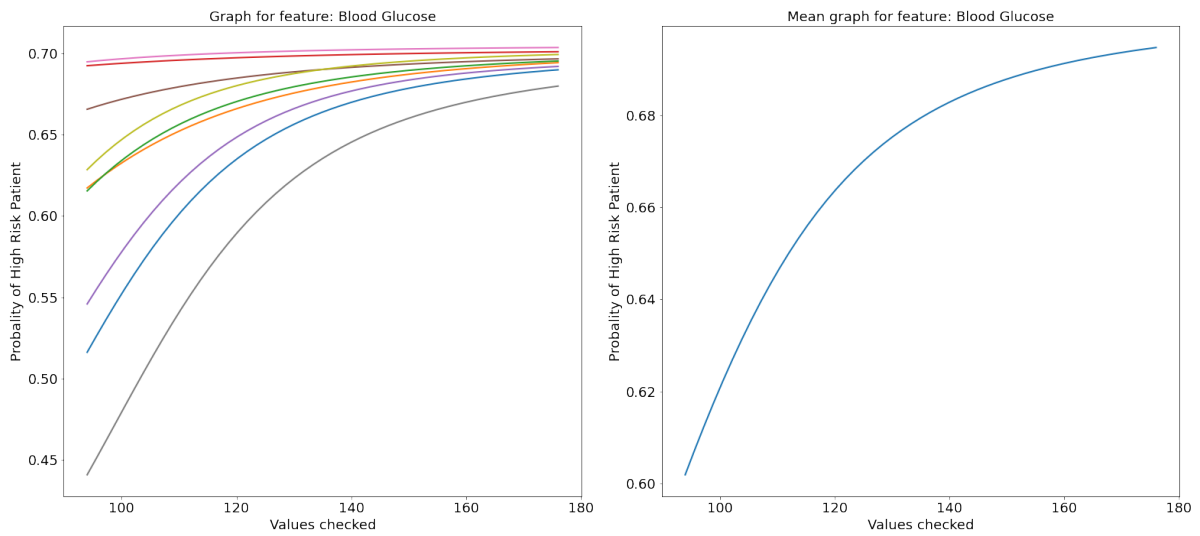
Σχήμα 5.6: ICE plot 2ου επιμέρους μοντέλου για το χαρακτηριστικό LDL χοληστερόλη



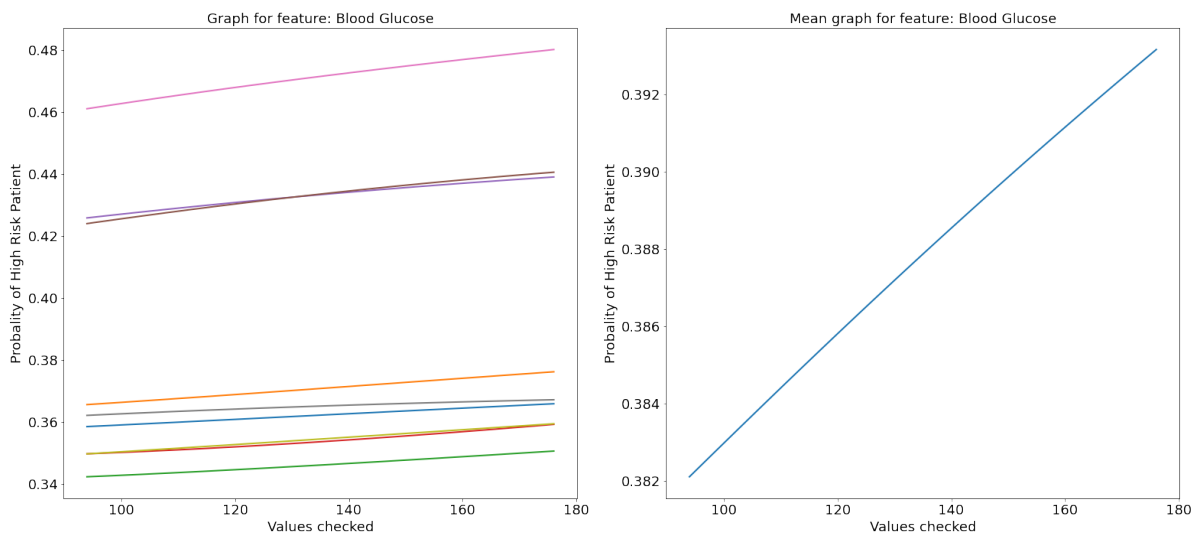
Σχήμα 5.7: ICE plot 3ου επιμέρους μοντέλου για το χαρακτηριστικό LDL χοληστερόλη

Συνεχίζουμε με τα ICE διαγράμματα του χαρακτηριστικού της γλυκόζης αίματος (Blood Glucose), Σχήμα. 5.8, 5.9, 5.10. Παρατηρούμε πως οι καμπύλες των διαγραμμάτων και για τα τρία μοντέλα παρουσιάζουν ανοδική πορεία. Δηλαδή για μεγάλες τιμές της γλυκόζης του

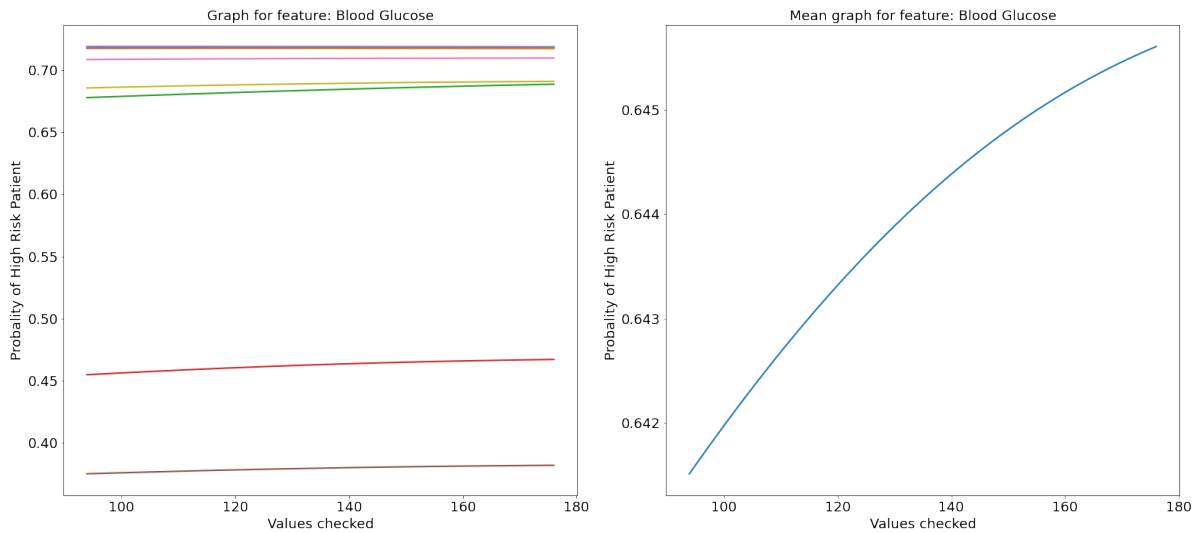
αίματος η πιθανότητα του ασθενούς να ταξινομηθεί από τα μοντέλα στην κλάση του υψηλού κινδύνου αυξάνεται. Η συμπεριφορά αυτή μπορεί να θεωρηθεί επιθυμητή και να συσχετιστεί με ότι είναι κλινικά αποδεδειγμένο καθώς οι διαβητικοί ασθενείς με στένωση ≥ 50 έχουν μεγάλες πιθανότητες να εμφανίσουν κάποιο καρδιαγγειακό νόσημα το οποίο μπορεί να αποβεί θανατηφόρο [53].



Σχήμα 5.8: ICE plot 1ου επιμέρους μοντέλου για το χαρακτηριστικό της γλυκόζης αίματος, Blood Glucose

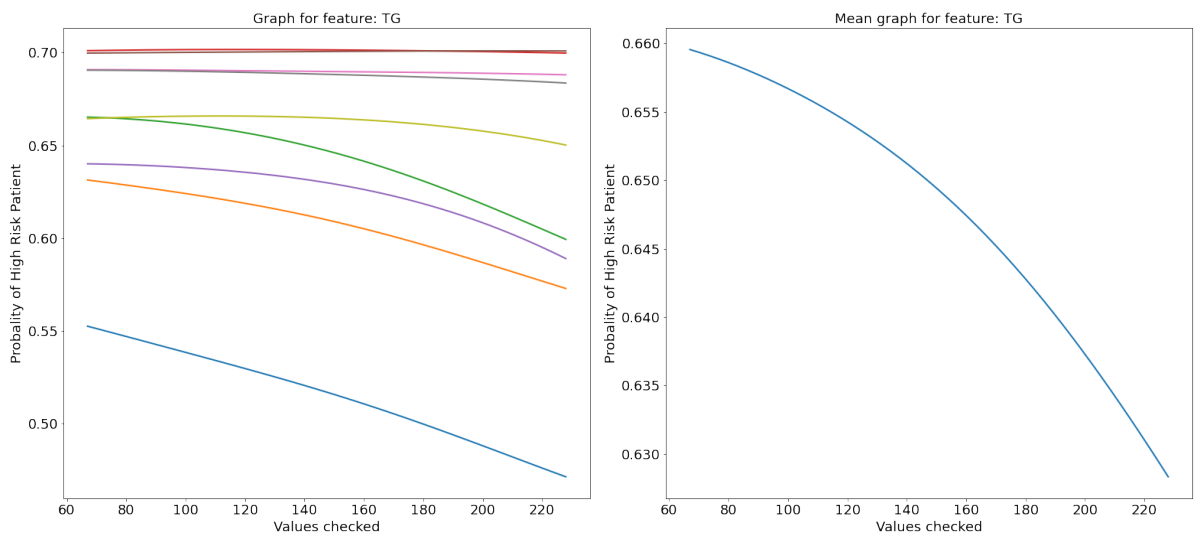


Σχήμα 5.9: ICE plot 2ου επιμέρους μοντέλου για το χαρακτηριστικό της γλυκόζης αίματος, Blood Glucose

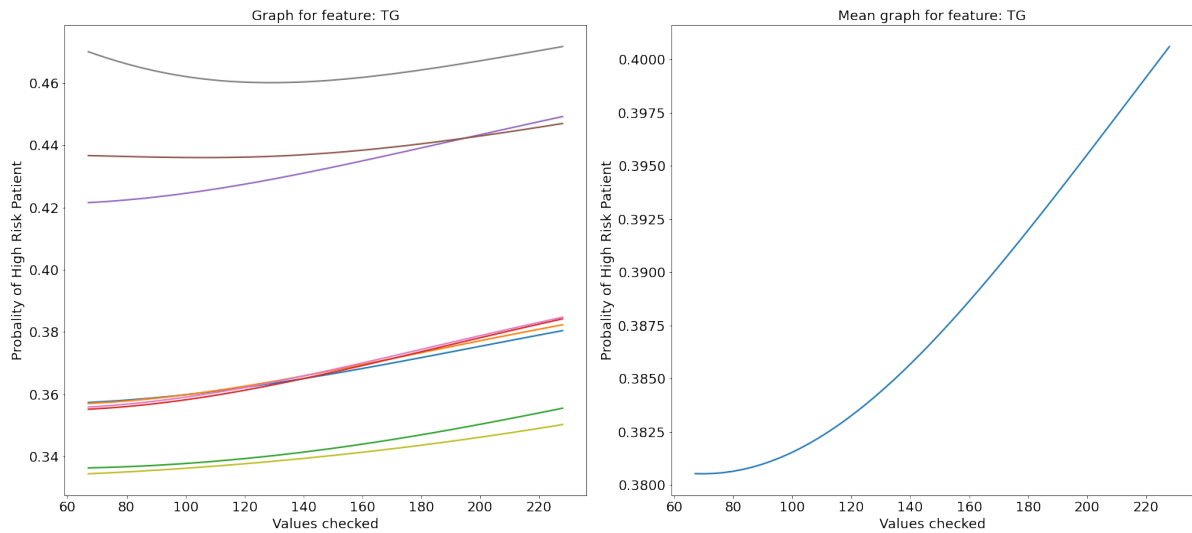


Σχήμα 5.10: ICE plot 3ου επιμέρους μοντέλου για το χαρακτηριστικό της γλυκόζης αίματος, Blood Glucose

Παρακάτω, παρουσιάζουμε τα ICE plots για το χαρακτηριστικό των τριγλυκεριδίων, Σχήμα. 5.11, 5.12, 5.13. Τα υψηλά επίπεδα των τριγλυκεριδίων (TG) αποτελούν δείκτες για αρκετούς τύπους αθηρογόνων λιποπρωτεϊνών και οι ασθενείς που έχουν ανεβασμένες τιμές τριγλυκεριδίων έχουν αυξημένες πιθανότητες να εμφανίσουν κάποια καρδιαγγειακή νόσο, ακόμα και αν τα επίπεδα της LDL χοληστερόλης τους είναι φυσιολογικά [65]. Επομένως οι υψηλές τιμές των τριγλυκεριδίων συσχετίζονται άμεσα με την αύξηση της πιθανότητας ενός ασθενούς να ταξινομηθεί στην κλάση υψηλού κινδύνου.

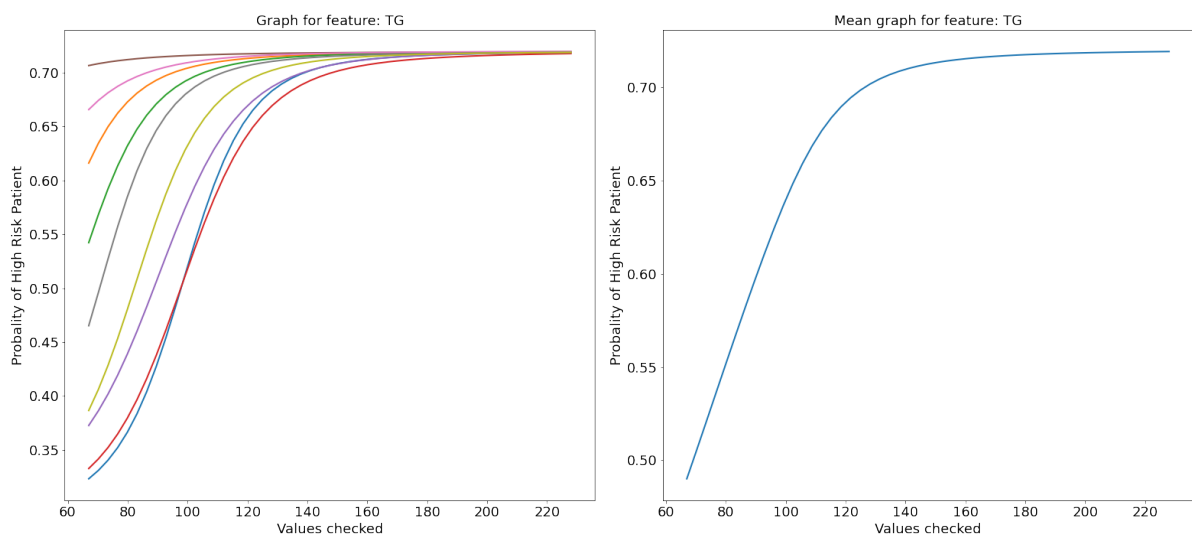


Σχήμα 5.11: ICE plot 1ου επιμέρους μοντέλου για το χαρακτηριστικό των τριγλυκεριδίων, TG



Σχήμα 5.12: ICE plot 2ου επιμέρους μοντέλου για το χαρακτηριστικό των τριγλυκεριδίων, TG

Παρόλα αυτά, από τα διαγράμματα μπορούμε να παρατηρήσουμε πως ενώ για το δεύτερο και το τρίτο επιμέρους μοντέλο παρουσιάζουν την αναμενόμενη συμπεριφορά το ICE plot του πρώτου μοντέλου εμφανίζει την αντίστροφη συμπεριφορά. Η αστοχία αυτή, όπως αναφέραμε και παραπάνω για το χαρακτηριστικό της LDL χοληστερόλης, μπορεί να οφείλεται είτε σε αδυναμία του μοντέλου να εκπαιδευτεί σωστά καθώς το πλήθος των δεδομένων είναι μικρό, είτε στον αρχικό διαμοιρασμό του συνόλου δεδομένων, στα σύνολα εκπαίδευσης και αξιολόγησης του κάθε επιμέρους μοντέλου.

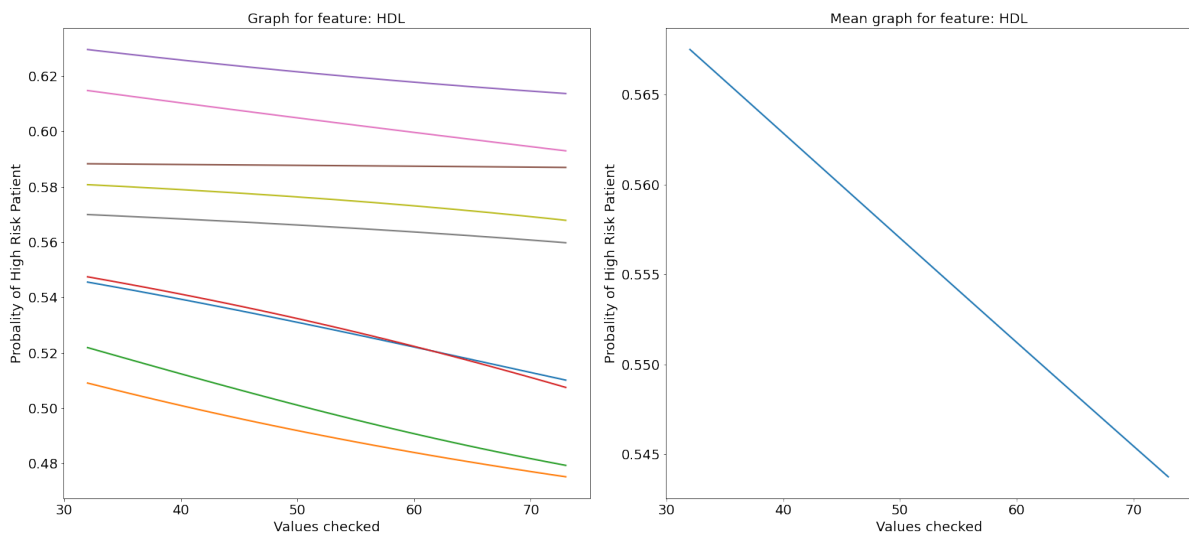


Σχήμα 5.13: ICE plot 3ου επιμέρους μοντέλου για το χαρακτηριστικό των τριγλυκεριδίων, TG

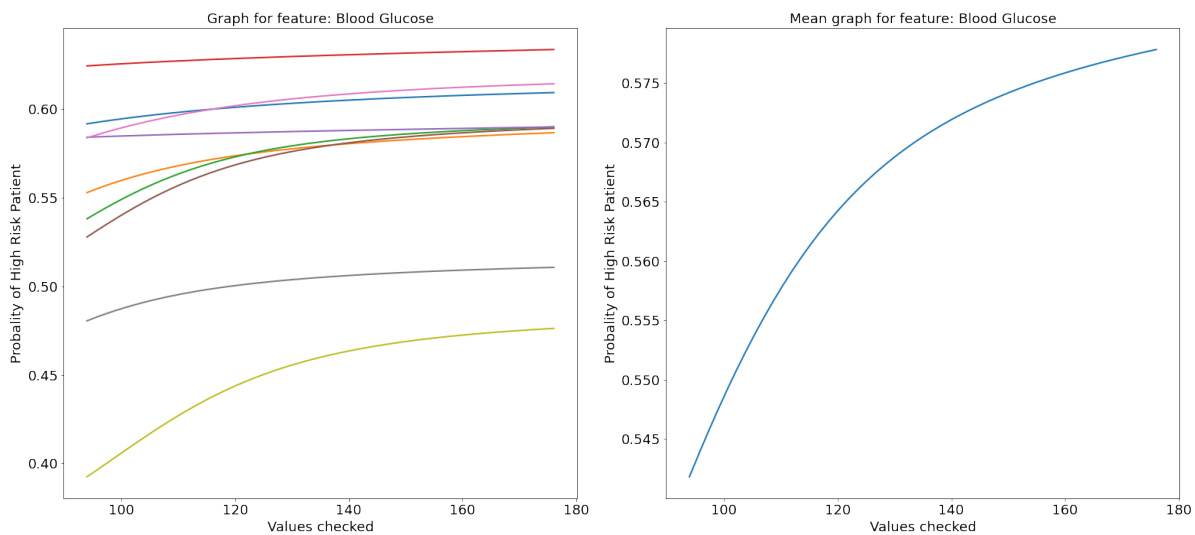
Τέλος αποφασίσαμε να εξαγάγουμε τα ICE plots των χαρακτηριστικών που αναφέραμε προηγουμένως για την ερμηνευσιμότητα του απλού συλλογικού μοντέλου, λαμβάνοντας έτσι και ένα πιο καθολικό αποτέλεσμα ερμηνευσιμότητας, Σχήμα. 5.14, 5.16, 5.15, 5.17.

Παρατηρούμε πως όλα τα ICE plots του απλού συλλογικού μοντέλου παρουσιάζουν την αναμενόμενη συμπεριφορά με βάση τις παρατηρήσεις που κάναμε προηγουμένως. Πιο συγκεκριμένα για τα χαρακτηριστικά της LDL χοληστερόλης, της γλυκόζης αίματος (Blood

Glucose) και των τριγλυκεριδίων (TG) η συμπεριφορά των καμπυλών στα διαγράμματα είναι αύξουσα ενώ για το χαρακτηριστικό της HDL χοληστερόλης η συμπεριφορά είναι φθίνουσα.

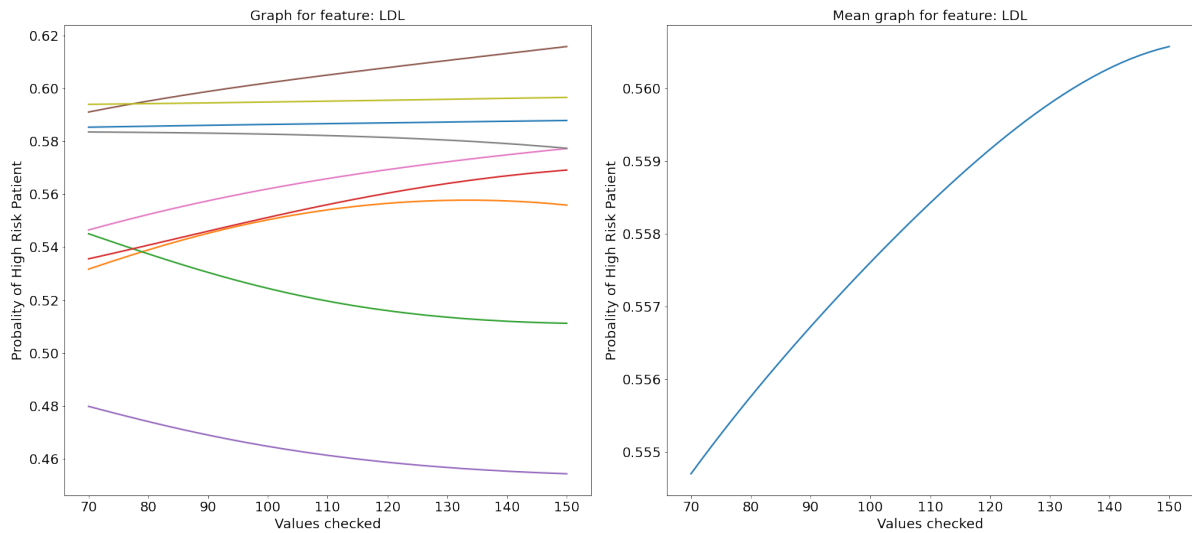


Σχήμα 5.14: ICE plot απλού συλλογικού μοντέλου για το χαρακτηριστικό της HDL χοληστερόλης

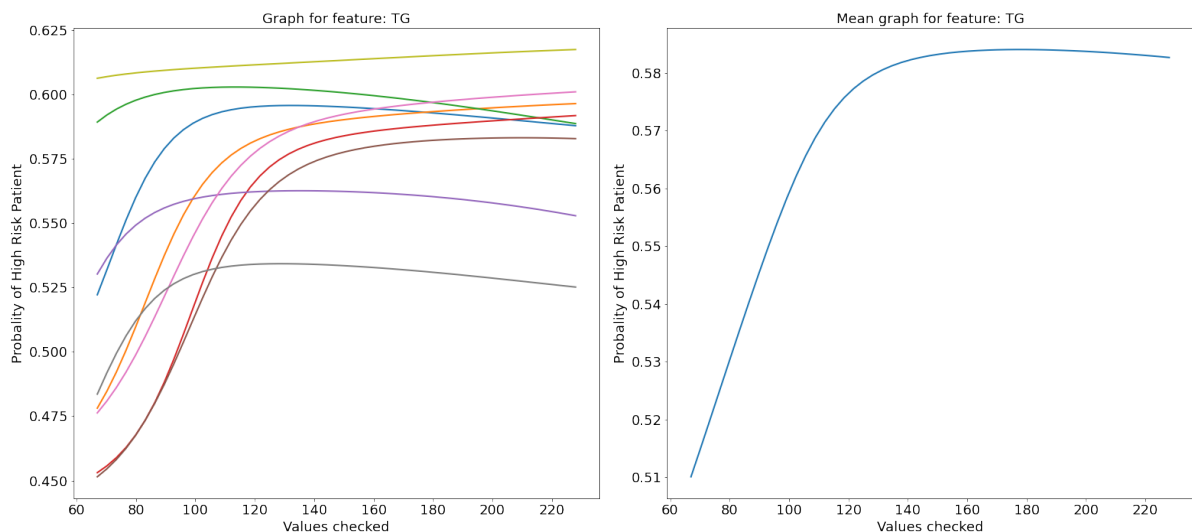


Σχήμα 5.15: ICE plot απλού συλλογικού μοντέλου για το χαρακτηριστικό της γλυκόζης αίματος, Blood Glucose

Τα χαρακτηριστικά της HDL χοληστερόλης (Σχήμα 5.14) καθώς και της γλυκόζης αίματος (Σχήμα 5.15), παρουσίασαν την επιθυμητή συμπεριφορά και για τα τρία επιμέρους μοντέλα, όπως είδαμε προηγουμένως. Επομένως ήταν αναμενόμενο να λάβουμε την αντίστοιχα επιθυμητή συμπεριφορά για τα ICE plots του απλού συλλογικού μοντέλου του οποίου οι αποφάσεις προκύπτουν από το μέσο όρο των αποφάσεων των τριών επιμέρους μοντέλων.



Σχήμα 5.16: ICE plot απλού συλλογικού μοντέλου για το χαρακτηριστικό της LDL χοληστερόλης



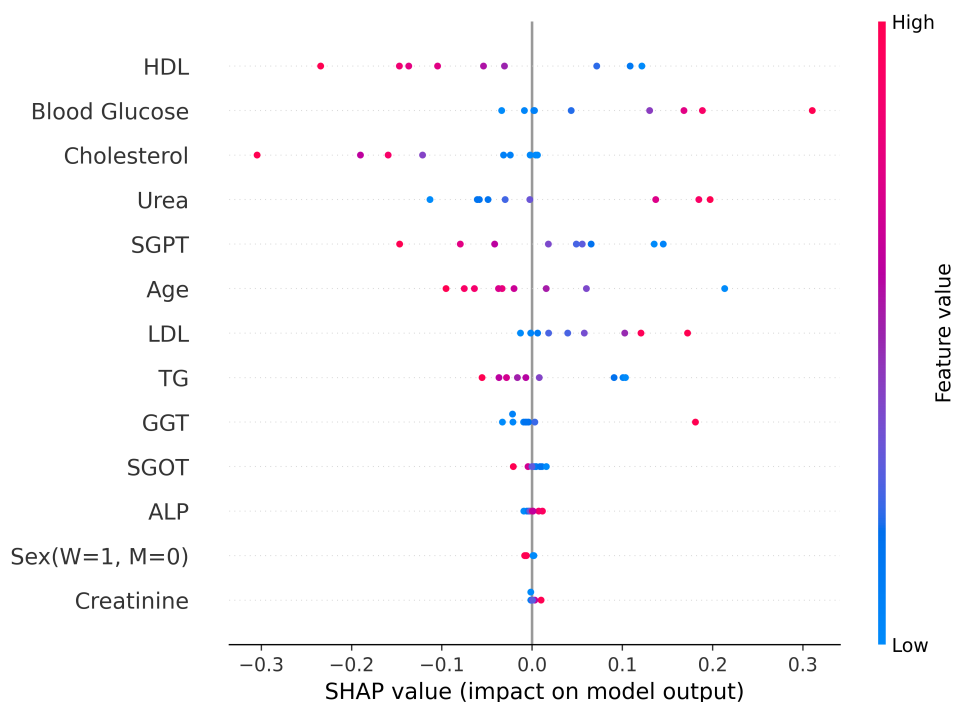
Σχήμα 5.17: ICE plot απλού συλλογικού μοντέλου για το χαρακτηριστικό των τριγλυκεριδίων, TG

Για το χαρακτηριστικό της LDL χοληστερόλης (Σχήμα. 5.16) και των τριγλυκεριδίων (Σχήμα. 5.17) ένα από τα τρία επιμέρους μοντέλα παρουσίασε μη αναμενόμενη συμπεριφορά ενώ τα υπόλοιπα εμφάνισαν την επιθυμητή. Πιο συγκεκριμένα όπως είδαμε και προηγουμένως, για το χαρακτηριστικό της LDL χοληστερόλης το ICE plot για το δεύτερο επιμέρους μοντέλο αντιστοιχούσε τις μεγαλύτερες τιμές της LDL με μικρότερη πιθανότητα του ασθενούς να ταξινομηθεί στην κλάση του υψηλού κινδύνου. Αντίστοιχη συμπεριφορά παρουσίασε και το ICE plot του πρώτου επιμέρους μοντέλου για το χαρακτηριστικό των τριγλυκεριδίων. Παρόλα αυτά το συλλογικό μοντέλο δεν φαίνεται να επηρεάζεται από τις αστοχίες αυτές καθώς και για τα δύο χαρακτηριστικά τα τελικά ICE plots που προκύπτουν έχουν αύξουσα συμπεριφορά.

Shapley Values

Στη συνέχεια θα παρουσιάσουμε τα αποτελέσματα που προέκυψαν μετά την εφαρμογή της μεθόδου Shapley Values. Για την οπτικοποίηση των αποτελεσμάτων της ερμηνευσιμότητας θα χρησιμοποιήσουμε διαγράμματα τα οποία παρουσιάζουν την καθολική επίδραση των χαρακτηριστικών στην τελική πρόβλεψη του μοντέλου. Τα διαγράμματα αυτά ονομάζονται Summary plots και περιγράφονται ως εξής. Το κάθε χαρακτηριστικό κατατάσσεται στο y-άξονα (κατακόρυφο) ανάλογα με την σημαντικότητά του στην τελική απόφαση του μοντέλου και οι κουκκίδες που αντιπροσωπεύουν μία συγκεκριμένη τιμή του χαρακτηριστικού αυτού για ένα συγκεκριμένο δείγμα, τοποθετούνται στην αντίστοιχη νοητή γραμμή του y-άξονα. Στον x-άξονα (οριζόντιο) η κάθε κουκκίδα λαμβάνει την θέση της ανάλογα με την shapley τιμή που της αντιστοιχεί, έτσι όπως αυτή υπολογίστηκε από τις μαθηματικές διαδικασίες που περιγράψαμε στην Ενότητα 3.2.2. Τέλος το χρώμα της κάθε κουκκίδας καθορίζεται από την τιμή του συγκεκριμένου χαρακτηριστικού που αντιπροσωπεύει. Τα χαρακτηριστικά με υψηλή τιμή σημειώνονται με κόκκινο χρώμα ενώ αυτά με χαμηλή τιμή σημειώνονται με μπλε χρώμα.

Αρχικά επιλέξαμε να εφαρμόσουμε τη μέθοδο σε κάθε ένα από τα 3 επιμέρους μοντέλα του απλού συλλογικού μοντέλου και να ερμηνεύσουμε το κάθε ένα ξεχωριστά. Τα αποτελέσματα που λάβαμε παρουσιάζονται στα Σχήματα 5.18, 5.19, 5.20.



Σχήμα 5.18: Summary plot 1ου επιμέρους μοντέλου. Κάθε κουκκίδα αντιπροσωπεύει την τιμή ενός χαρακτηριστικού ενός συγκεκριμένου δείγματος. Οι κόκκινες κουκκίδες αντιστοιχούν σε υψηλές τιμές του χαρακτηριστικού ενώ οι μπλε σε χαμηλές.

Στο Σχήμα. 5.18, παρατηρούμε ότι για το 1ο επιμέρους μοντέλο το χαρακτηριστικό με τη μεγαλύτερη σημαντικότητα είναι η HDL χοληστερόλη. Όπως έχουμε αναφέρει ήδη, η HDL χοληστερόλη μεταφέρει τη χοληστερίνη από τους ιστούς του σώματος στο ήπαρ όπου διασπάται. Έτσι μειώνεται ο κίνδυνος ανάπτυξης αθηρωματικής πλάκας και κατ' επέκταση εμφάνισης καρωτιδικής νόσου μεγάλης επικινδυνότητας. Όπως είναι αναμενόμενο, μπορούμε να παρατηρήσουμε στο summary plot, πως οι χαμηλές τιμές της HDL χοληστερόλης αυξάνουν

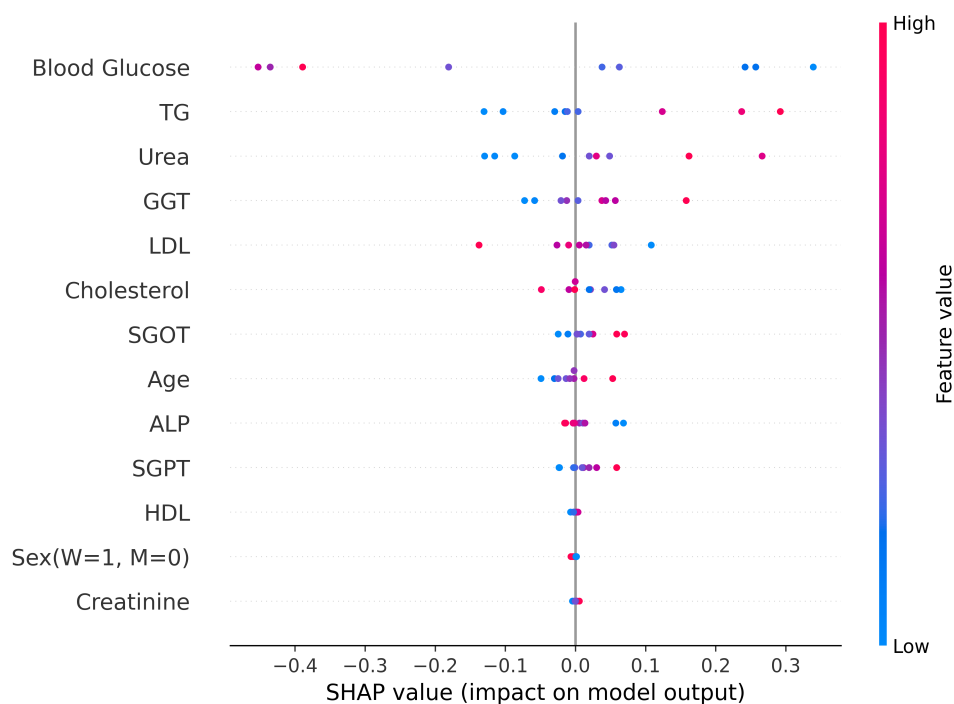
τον κίνδυνο εμφάνισης αθηρωματικής πλάκας ενώ οι μεγαλύτερες τον μειώνουν σημαντικά [61], [59].

Αντίστοιχα για την LDL χοληστερόλη ισχύει το αντίστροφο καθώς αποτελεί τον βασικό παράγοντα για την δημιουργία των αθηρωματικών πλακών [64]. Ειδικότερα, έχει παρατηρηθεί πως η μείωση των επιπέδων της LDL χοληστερόλης κάτω από 70mg/dL σχετίζεται άμεσα με την μείωση των ρυθμών της εξέλιξης της αθηρωσκλήρωσης των καρωτιδικών αρτηριών [66].

Σημαντική επιρροή στην τελική απόφαση του μοντέλου, ακολουθώντας παράλληλα και την αναμενόμενη συμπεριφορά, ασκούν επίσης η γλυκόζη αίματος (Blood Glucose) και η ουρία (Urea). Στην μελέτη [54] αναφέρεται πως οι διαβητικοί ασθενείς με στένωση καρωτίδας της τάξεως 50% και άνω διατρέχουν εξαιρετικά υψηλό κίνδυνο για καρδιαγγειακό θάνατο. Η παρατήρηση αυτή συμπίπτει με την συμπεριφορά του χαρακτηριστικού της γλυκόζης του αίματος στο summary plot (Σχήμα 5.18) στο οποίο οι υψηλές τιμές της γλυκόζης οδηγούν το μοντέλο να λάβει την απόφαση πως ο ασθενής ανήκει στην κατηγορία του υψηλού κινδύνου.

Στο δεύτερο διάγραμμα (Σχήμα. 5.19) παρουσιάζουμε το summary plot που περιγράφει στην ερμηνευσιμότητα του 2ου επιμέρους μοντέλου. Το πρώτο σε σημαντικότητα χαρακτηριστικό είναι η γλυκόζη αίματος (Blood Glucose) η οποία, όπως αναφέραμε και προηγουμένως, θα περιμέναμε να έχει την "ανάποδη" συμπεριφορά από αυτήν που φαίνεται στο διάγραμμα. Παρόλα αυτά, τα χαρακτηριστικά των τριγλυκεριδίων (TG), της ουρίας (Urea) και της γ-Γλουταμυλοτρανσφεράσης (GGT) εμφανίζουν την αναμενόμενη συμπεριφορά.

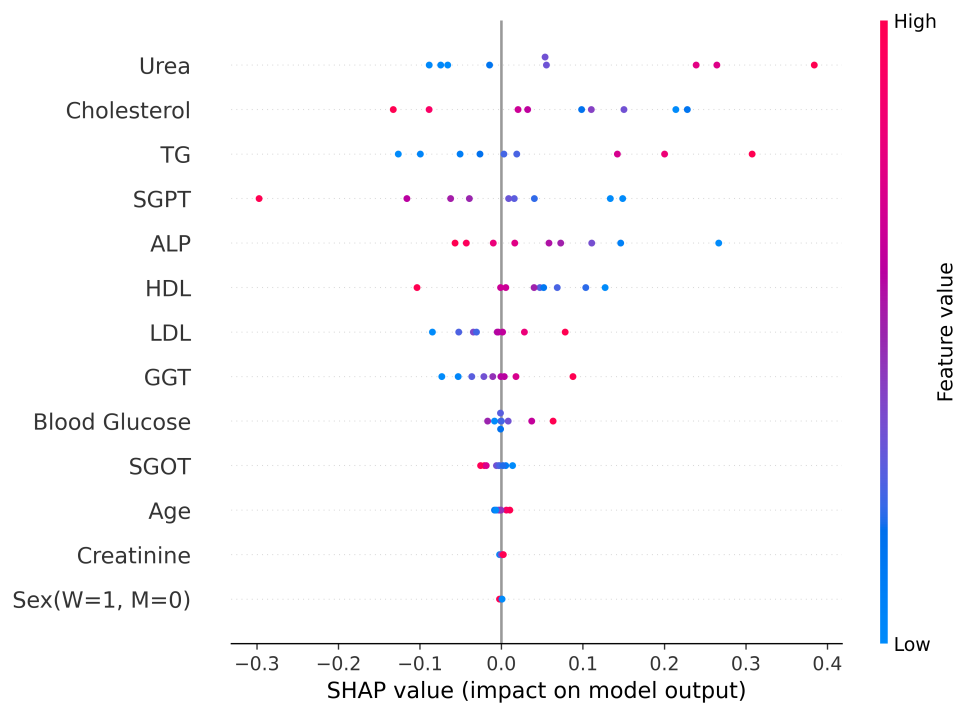
Οι αυξημένες τιμές των τριγλυκεριδίων ευνοούν την δημιουργία υπερλιπιδαιμίας η οποία αποτελεί βασικό παράγοντα για την ανάπτυξη της αρτηριοσκλήρωσης. Συνεπώς τα αυξημένα επίπεδα των τριγλυκεριδίων μπορούν να οδηγήσουν σε σημαντική στένωση των καρωτίδων με αποτέλεσμα ο ασθενής να κατηγοριοποιηθεί στην κλάση του υψηλού κινδύνου. Στο διάγραμμα παρουσιάζεται ακριβώς αυτή η συμπεριφορά. Μεγάλες τιμές των τριγλυκεριδίων ωθούν το μοντέλο να λάβει την απόφαση πως πρόκειται για ασθενή υψηλού κινδύνου ενώ οι μικρές πως πρόκειται για ασθενή χαμηλού κινδύνου.



Σχήμα 5.19: Summary plot 2ου επιμέρους μοντέλου

Τα υψηλά επίπεδα της γ-Γλουταμυλοτρανσφεράσης (GGT) σχετίζονται με πληθώρα παραγόντων καρδιαγγειακών κινδύνων. Πειραματικές μελέτες έχουν αποκαλύψει πως η παρουσία υψηλών GGT τιμών όχι μόνο δρα καταλυτικά στην εμφάνιση αθηρωμάτων αλλά έχει και ενεργό ρόλο στο χαρακτηρισμό της αστάθειας-επικινδυνότητας αυτών [67]. Η συμπεριφορά του χαρακτηριστικού στο διάγραμμα συμπίπτει με την παρατήρηση αυτή καθώς, υψηλές τιμές της GGT επηρεάζουν το μοντέλο θετικά ώστε να λάβει την απόφαση ότι ασθενής ανήκει στην κατηγορία υψηλού κινδύνου ενώ οι χαμηλές συντελούν για το αντίθετο.

Τέλος αξίζει να παρατηρήσουμε πως η συμπεριφορά της LDL χοληστερόλης δεν είναι η αναμενόμενη καθώς μικρές τιμές αυτής ωθούν το μοντέλο να κατατάξει τον ασθενή στην κλάση του υψηλού κινδύνου.



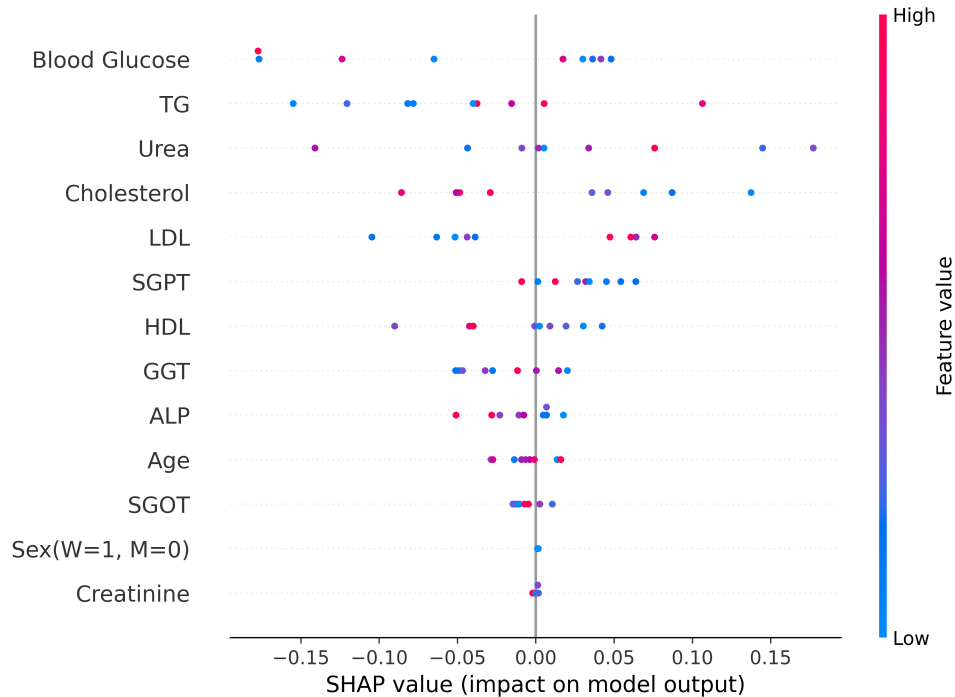
Σχήμα 5.20: Summary plot 3ου επιμέρους μοντέλου

Στο Σχήμα. 5.20 μπορούμε να παρατηρήσουμε το summary plot για την ερμηνευσιμότητα του 3ου επιμέρους μοντέλου. Το χαρακτηριστικό με τη μεγαλύτερη σημαντικότητα είναι η ουρία, η οποία όπως και στα προηγούμενα δύο διαγράμματα έτσι και σε αυτό παρουσιάζει την αναμενόμενη συμπεριφορά. Δηλαδή οι υψηλές τιμές αυξάνουν την πιθανότητα του ασθενούς να ταξινομηθεί στην κατηγορία του υψηλού κινδύνου.

Η συμπεριφορά των χαρακτηριστικών HDL και LDL χοληστερόλης είναι η επιθυμητή όπως και στο πρώτο επιμέρους μοντέλο. Αντίστοιχα, τη συμπεριφορά που αναμένουμε με βάση τις μελέτες που αναφέραμε προηγουμένως ([67], [68]) παρουσιάζουν και τα χαρακτηριστικά των τριγλυκεριδίων (TG) και της γ-Γλουταμυλοτρανσφεράσης (GGT).

Λόγω της φύσης του μοντέλου μας επιλέξαμε να εφαρμόσουμε τη μέθοδο Shapley Values και με έναν διαφορετικό τρόπο, έτσι ώστε να μην ερμηνεύσουμε το κάθε επιμέρους μοντέλο ξεχωριστά αλλά να επικεντρωθούμε σε ένα πιο καθολικό αποτέλεσμα ερμηνευσιμότητας που θα αφορά τις τελικές αποφάσεις του απλού συλλογικού μοντέλου. Πιο συγκεκριμένα το απλό συλλογικό μοντέλο αποτελείται από τρία επιμέρους μοντέλα και η τελική του απόφαση λαμβάνεται με βάση τον μέσο όρο των πιθανοτήτων αυτών των τριών μοντέλων. Ακολουθώντας αυτήν την λογική αποφασίσαμε να εφαρμόσουμε με παρόμοιο τρόπο και την μέθοδο

ερμηνευσιμότητας Shapley Values. Έτσι αφού εξαγάγαμε ξεχωριστά για κάθε μοντέλο τις shapley τιμές του, το τελικό summary plot προέκυψε από τους μέσους όρους των shapley τιμών των τριών μοντέλων. Το συνολικό summary plot παρουσιάζεται στη συνέχεια, Σχήμα 5.21.



Σχήμα 5.21: Summary plot απλού συλλογικού μοντέλου. Οι shapley values τιμές για το απλό συλλογικό μοντέλο προέκυψαν από το μέσο όρο των shapley value τιμών των τριών επιμέρους μοντέλων

Παρατηρούμε πως η γλυκόζη αίματος και η ουρία παρόλο που έχουν μεγάλη σημαντικότητα, δεν ακολουθούν κάποιο συγκεκριμένο - λογικό μοτίβο στην συμπεριφορά τους. Για παράδειγμα μεγάλες τιμές των χαρακτηριστικών αυτών μπορούν να ωθήσουν το μοντέλο είτε να λάβει την απόφαση πως πρόκειται για ασθενή χαμηλού κινδύνου είτε υψηλού. Σε αντίθεση, το χαρακτηριστικό των τριγλυκεριδίων το οποίο έχει μεγάλη επιρροή στην τελική απόφαση του μοντέλου, καθώς είναι δεύτερο σε σειρά σημαντικότητας, και παράλληλα παρουσιάζει την επιθυμητή συμπεριφορά. Όπως βλέπουμε στο διάγραμμα οι χαμηλές τιμές των τριγλυκεριδίων μειώνουν την πιθανότητα του ασθενούς να ταξινομηθεί στην κλάση του υψηλού κινδύνου.

Αντίστοιχα επιθυμητή συμπεριφορά παρουσιάζουν και τα χαρακτηριστικά της HDL και της LDL χοληστερόλης. Στο διάγραμμα φαίνεται πως οι υψηλές τιμές της LDL χοληστερόλης αυξάνουν την πιθανότητα ώστε ο ασθενής να χαρακτηριστεί ως ασθενής υψηλού κινδύνου και αντίστοιχα οι χαμηλές τιμές για την HDL χοληστερόλη. Όπως αναφέραμε και προηγουμένως η συμπεριφορά αυτή είναι σύμφωνη με ότι έχει διατυπωθεί στις μελέτες [60], [66] στις οποίες αναφέρεται πως οι υψηλές τιμές της LDL χοληστερόλης ευνοούν την δημιουργία αθηρωμάτων και αντίστοιχα οι μειωμένες τιμές της HDL χοληστερόλης δεν επαρκούν ώστε να καταπολεμηθούν τα υψηλά επίπεδα της χοληστερόλης από τον οργανισμό με αποτέλεσμα ο κίνδυνος για εμφάνιση καρδιαγγειακών προβλημάτων να αυξάνεται.

Τέλος, αξίζει να παρατηρήσουμε πως το συλλογικό μοντέλο δεν επηρεάζεται από τις αστοχίες που εμφάνισαν τα χαρακτηριστικά της LDL χοληστερόλης και των τριγλυκεριδίων

στο δεύτερο και το πρώτο επιμέρους μοντέλο αντίστοιχα.

Σχολιασμός των αποτελεσμάτων

Στις παραπάνω ενότητες καταφέραμε να ερμηνεύσουμε το απλό συλλογικό μοντέλο αλλά και τα επιμέρους μοντέλα που το απαρτίζουν, εφαρμόζοντας τρεις διαφορετικές μεθόδους ερμηνευσιμότητας. Τα αποτελέσματα που λάβαμε και από τις τρεις μεθόδους ήταν αρκετά ενθαρρυντικά για την σωστή λειτουργία του μοντέλου μας καθώς έδειξαν πως το μοντέλο μας στηρίζει τις τελικές του αποφάσεις σε συγκρίσεις και παρατηρήσεις που είναι κλινικά αποδεδειγμένες. Για παράδειγμα και οι τρεις μέθοδοι ερμηνευσιμότητας έδειξαν πως ένα βασικό κριτήριο για να ταξινομηθεί ένας ασθενής στην κατηγορία του υψηλού κινδύνου είναι οι χαμηλές τιμές της HDL χοληστερόλης.

Παρόλα αυτά υπήρξαν και χαρακτηριστικά που παρουσίασαν μη αναμενόμενη συμπεριφορά στην ερμηνευσιμότητα των επιμέρους μοντέλων. Πιο συγκεκριμένα για το χαρακτηριστικό της LDL χοληστερόλης μπορούμε να παρατηρήσουμε πως δεν παρουσιάζει την επιθυμητή συμπεριφορά για το δεύτερο επιμέρους μοντέλο, όπως αυτή προκύπτει από τα διαγράμματα της ICE και της Shapley Value μεθόδου. Εν αντιθέσει στο πρώτο και στο τρίτο επιμέρους μοντέλο η συμπεριφορά της LDL φαίνεται φυσιολογική, δηλαδή οι υψηλές τιμές της αυξάνουν την πιθανότητα του δείγματος να ανήκει στην κατηγορία του υψηλού κινδύνου.

Η ασυνέπεια αυτή μπορεί να οφείλεται σε αδυναμία του συγκεκριμένου επιμέρους μοντέλου, όμως το γεγονός πως τα υπόλοιπα δύο επιμέρους μοντέλα δεν την παρουσίασαν, ενώ ακολουθούν την ίδια αρχιτεκτονική, μας οδηγεί στο συμπέρασμα πως ίσως τα δεδομένα εκπαίδευσης και αξιολόγησης δεν έχουν διαμοιραστεί ισοδύναμα και στα τρία μοντέλα. Με τον όρο ισοδύναμα δεν αναφερόμαστε στο πλήθος αλλά στην ποιότητα των δεδομένων.

Ωστόσο, αξίζει να παρατηρήσουμε πως οι αστοχίες που παρουσιάστηκαν παραπάνω δεν επηρέασαν την συμπεριφορά του απλού συλλογικού μοντέλου. Εν κατακλείδι, συμπεραίνουμε πως και οι τρεις μέθοδοι ερμηνευσιμότητας μας παρέχουν αξιόλογα και χρήσιμα αποτελέσματα τα οποία μας διαβεβαιώνουν ότι το τελικό μοντέλο μας βασίζει τις αποφάσεις του σε λογικές παρατηρήσεις βοηθώντας μας παράλληλα να ανακαλύψουμε ορισμένες αδυναμίες της μεθόδου που ακολουθήσαμε.

5.2.2 Ερμηνευσιμότητα CNN μοντέλου

Shapley Values

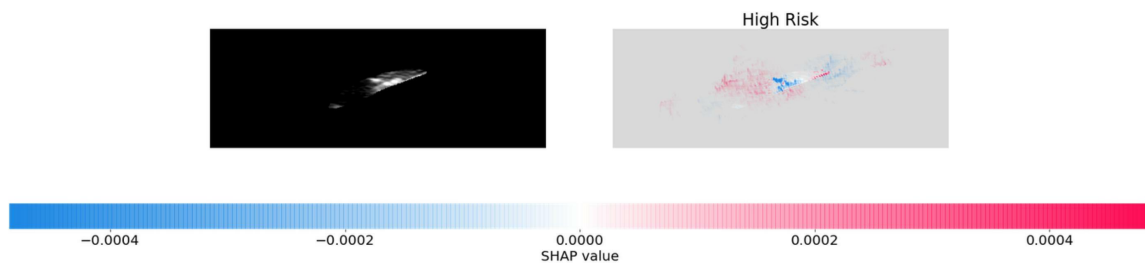
Για να ερμηνεύσουμε το συνελικτικό συλλογικό μοντέλο επιλέξαμε να χρησιμοποιήσουμε τη μέθοδο Shapley Values. Πιο συγκεκριμένα όπως και στο απλό συλλογικό μοντέλο, χρησιμοποιήσαμε τη βιβλιοθήκη Shap της Python και κάθε επιμέρους μοντέλο του συνελικτικού συλλογικού μοντέλου ερμηνεύτηκε ξεχωριστά. Επιλέξαμε να ερμηνεύσουμε τρεις διαφορετικές περιπτώσεις ασθενών και πιο συγκεκριμένα: έναν ασθενή υψηλού κινδύνου ο οποίος ταξινομήθηκε σωστά από τα δύο μοντέλα, έναν ασθενή χαμηλού κινδύνου ο οποίος ταξινομήθηκε σωστά και από τα τρία μοντέλα, και έναν ασθενή υψηλού κινδύνου ο οποίος ταξινομήθηκε εσφαλμένα και από τα τρία μοντέλα. Στη συνέχεια παρουσιάζουμε τα ερμηνεύσιμα πλέον αποτελέσματα έτσι όπως προέκυψαν από κάθε μοντέλο. Σε κάθε εικόνα εμφανίζονται 2 επιμέρους εικόνες. Η πρώτη αντιστοιχεί στο δείγμα που έχουμε επιλέξει να ερμηνεύσουμε και αποτελεί την είσοδο που δέχεται το μοντέλο ώστε να λάβει την τελική του απόφαση. Η δεύτερη αποτελεί το ερμηνεύσιμο αποτέλεσμα του μοντέλου, δηλαδή παρουσιάζει την εικόνα του εξεταζόμενου δείγματος με τη διαφορά ότι πλέον έχουν προσδιοριστεί τα σημεία

του αθηρώματος που οδήγησαν το μοντέλο να λάβει την τελική του απόφαση χρωματίζοντας κάθε εικονοστοιχείο με το κατάλληλο χρώμα. Πιο συγκεκριμένα τα εικονοστοιχεία που ωθούν το μοντέλο να λάβει την απόφαση ότι ο ασθενής πρέπει να ταξινομηθεί στην κατηγορία υψηλού κινδύνου, χρωματίζονται με κόκκινο χρώμα ενώ αντίστοιχα αυτά που το απωθούν από το λάβει αυτή την απόφαση, χρωματίζονται με μπλε.

Ερμηνευσιμότητα 1ου δείγματος, Ασθενής Υψηλού Κινδύνου

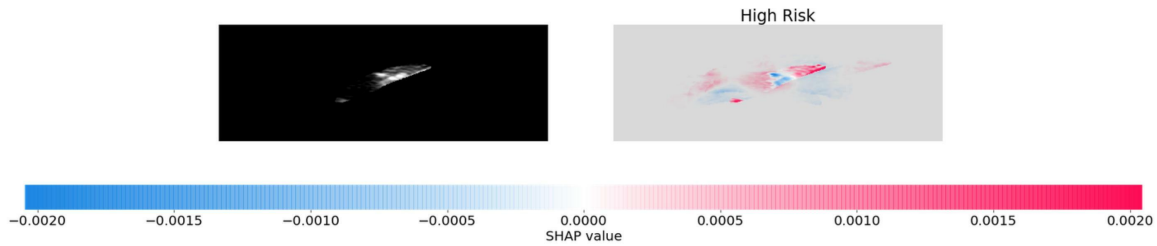
Το πρώτο δείγμα που επιλέξαμε να ερμηνεύσουμε ανήκει σε ασθενή υψηλού κινδύνου. Μέσω των παρακάτω εικόνων μπορούμε να ερμηνεύσουμε τον τρόπο με τον οποίο κάθε ένα από τα επιμέρους μοντέλα λαμβάνει την τελική του απόφαση.

Το πρώτο μοντέλο κατατάσσει το δείγμα εσφαλμένα στην κατηγορία των ασθενών χαμηλού κινδύνου. Όπως μπορούμε να παρατηρήσουμε και από την εικόνα της ερμηνευσιμότητας Σχήμα, 5.22, το πρώτο μοντέλο λαμβάνει την τελική του απόφαση βασιζόμενο κυρίως στην έντονα λευκή περιοχή που υπάρχει στο κέντρο του αθηρώματος, δίνοντας πολύ λιγότερη βαρύτητα στις γύρω περιοχές.

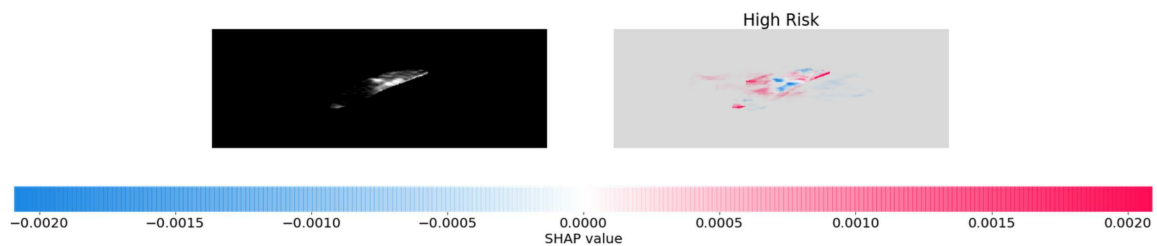


Σχήμα 5.22: Ερμηνευσιμότητα 1ου επιμέρους μοντέλου μέσω της μεθόδου Shapley Values, για δείγμα ασθενούς που ανήκει στην κατηγορία υψηλού κινδύνου. Τα εικονοστοιχεία που χρωματίζονται με κόκκινο χρώμα ωθούν το μοντέλο να λάβει την απόφαση ότι το δείγμα ανήκει στην κατηγορία του υψηλού κινδύνου ενώ τα μπλε το απωθούν. Το εικονιζόμενο δείγμα κατατάσσεται εσφαλμένα από το 1ο επιμέρους μοντέλο στην κατηγορία του χαμηλού κινδύνου καθώς η τελική απόφασή του επηρεάζεται σημαντικά από την έντονα λευκή περιοχή στο κέντρο του αθηρώματος

Εν αντιθέσει με το πρώτο μοντέλο τα δύο επόμενα μοντέλα κατατάσσουν σωστά το προς εξέταση δείγμα στην κατηγορία του υψηλού κινδύνου. Πιο συγκεκριμένα από τις δύο παρακάτω εικόνες ερμηνευσιμότητας (5.23, 5.24) μπορούμε να συμπεράνουμε πως το δεύτερο και το τρίτο επιμέρους μοντέλο παρουσιάζουν παρόμοια συμπεριφορά, καθώς λαμβάνουν την τελική τους απόφαση, ότι το δείγμα δηλαδή ανήκει σε ασθενή υψηλού κινδύνου, δίνοντας μεγάλη βαρύτητα στις δύο πιο σκοτεινές περιοχές γύρω από την φωτεινότερη περιοχή στο κέντρο του αθηρώματος. Οι περιοχές αυτές είναι χρωματισμένες με κόκκινο χρώμα.



Σχήμα 5.23: Ερμηνευσιμότητα 2ου επιμέρους μοντέλου με SV



Σχήμα 5.24: Ερμηνευσιμότητα 3ου επιμέρους μοντέλου με SV

Μια από τις απλούστερες παραμέτρους που μπορούν να χρησιμοποιηθούν για την κατηγοριοποίηση υπερηχογραφικών απεικονίσεων των αθηρωματικών πλακών, είναι η πρώτης τάξης στατιστικά ενδιάμεση τιμή της φωτεινότητας σε κλίμακα του γκρι (GSM) [69], δηλαδή η ενδιάμεση τιμή της φωτεινότητας των εικονοστοιχείων της εικόνας του υπερήχου. Στην μελέτη [70] αναφέρεται πως ο χαμηλός δείκτης GSM για μία αθηρωματική πλάκα συσχετίζεται άμεσα με υψηλότερη συχνότητα εμφάνισης εγκεφαλικού επεισοδίου. Ακόμα ο G.Biasi και άλλοι [71] έδειξαν πως το GSM μπορεί να θεωρηθεί καταλληλότερος δείκτης από τον βαθμό στένωσης της καρωτίδας για την εύρεση πλακών που αποτελούν κίνδυνο για εγκεφαλικά επεισόδια.

Ένας ακόμα δείκτης που χρησιμοποιείται στην μελέτη της επικινδυνότητας των αθηρωματικών πλακών είναι ο δείκτης JBA, ο οποίος προκύπτει από το εμβαδόν της σκούρας περιοχής κοντά στον αυλό (Juxtaluminal Black Area - JBA) και πιο συγκεκριμένα από τα εικονοστοιχεία με φωτεινότητα γκριζας κλίμακας < 25 . Ειδικότερα, έχει παρατηρηθεί πως οι μεγάλες τιμές του δείκτη JBA και ο κίνδυνος εμφάνισης ενός εγκεφαλικού επεισοδίου, συνδέονται με γραμμική σχέση [72].

Συμφωνά με τους παραπάνω δείκτες συμπεραίνουμε πως τα σημεία του αθηρώματος που συνδέονται με την μεγαλύτερη επικινδυνότητα και καθορίζουν σημαντικά την κλάση (χαμηλού ή υψηλού κινδύνου) στην οποία θα ταξινομηθεί ο ασθενής, είναι οι πιο σκούρες περιοχές, δηλαδή οι περιοχές που αποτελούνται από εικονοστοιχεία χαμηλής φωτεινότητας.

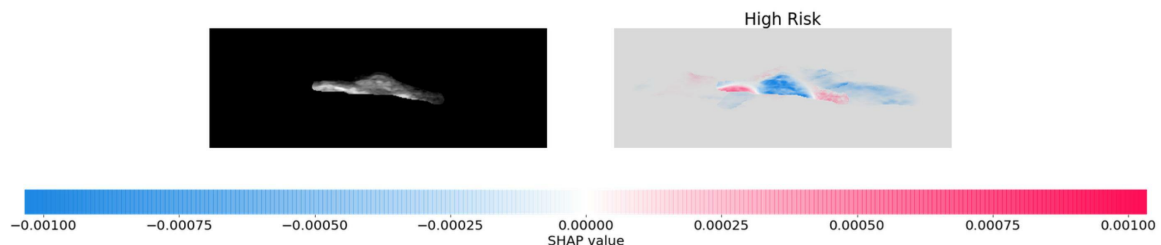
Με βάση τις παραπάνω παρατηρήσεις, μία ένδειξη πως το μοντέλο μας εκπαιδεύεται σωστά και καταλήγει στα συμπεράσματά του βασιζόμενο σε αξιόπιστες παρατηρήσεις θα ήταν η μέθοδος Shapley Values να δίνει μεγαλύτερες shap τιμές στα πιο σκουρόχρωμα pixels (κόκκινο χρώμα). Έτσι, για κάθε εικόνα που ερμηνεύσαμε υπολογίσαμε την ενδιάμεση τιμή (median) των εικονοστοιχείων τα οποία η μέθοδος Shap χρωμάτισε κόκκινα και αντίστοιχα για αυτά που χρωμάτισε μπλε. Στη συνέχεια παρατίθεται ο πίνακας με τον Grayscale median για τα κόκκινα και τα μπλε εικονοστοιχεία, έτσι όπως αυτά υπολογίστηκαν από τη μέθοδο Shap για κάθε μοντέλο.

Επιμέρους Μοντέλα ENN	Grayscale Median	
	Κόκκινα εικονοστοιχεία	Μπλε εικονοστοιχεία
1ο μοντέλο	193	208
2ο μοντέλο	224	241
3ο μοντέλο	190	200

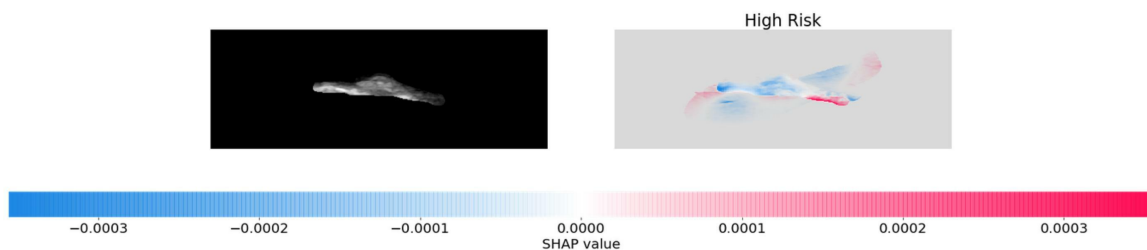
Τα αποτελέσματα που λαμβάνουμε είναι αρκετά ενθαρρυντικά καθώς και στις τρεις περιπτώσεις τα κόκκινα εικονοστοιχεία έχουν μικρότερη ενδιάμεση grayscale τιμή. Αυτό σημαίνει πως αντιστοιχούν στα πιο σκουρόχρωμα εικονοστοιχεία του αθηρώματος, όπως και αναμέναμε.

Ερμηνευσιμότητα 2ου δείγματος, Ασθενής Χαμηλού Κινδύνου

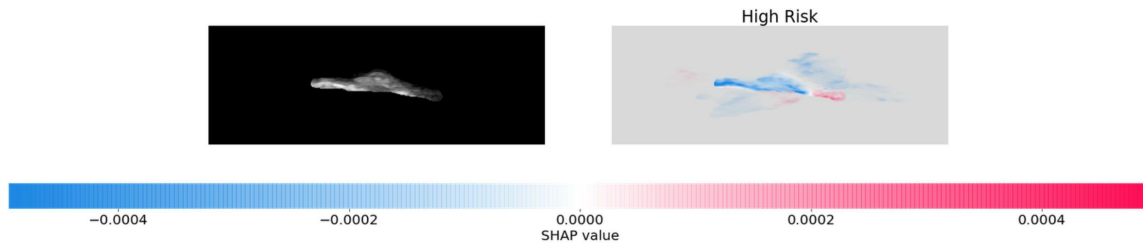
Συνεχίζουμε την ερμηνευσιμότητα του CNN μοντέλου παρουσιάζοντας τον τρόπο που διαχειρίστηκαν τα τρία επιμέρους μοντέλα ένα δείγμα που ανήκει σε ασθενή Χαμηλού κινδύνου. Στη συγκεκριμένη περίπτωση και τα τρία μοντέλα κατατάσσουν σωστά το δείγμα και αυτό συμπεραίνεται και από τα 3 παρακάτω σχήματα ερμηνευσιμότητας. Ειδικότερα μπορούμε να παρατηρήσουμε ότι και τα τρία μοντέλα έχουν την ίδια συμπεριφορά, δηλαδή η περιοχή που ασκεί την μεγαλύτερη επιρροή και στα τρία είναι η μεγάλη σε έκταση φωτεινή κεντρική περιοχή η οποία σημειώνεται με μπλε χρώμα.



Σχήμα 5.25: Ερμηνευσιμότητα 1ου επιμέρους μοντέλου με SV



Σχήμα 5.26: Ερμηνευσιμότητα 2ου επιμέρους μοντέλου με SV



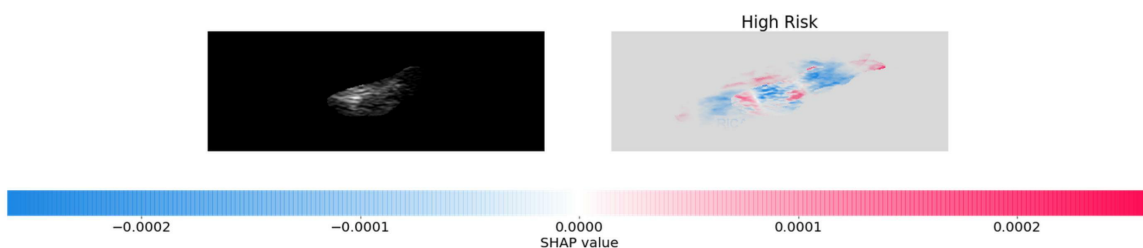
Σχήμα 5.27: Ερμηνευσιμότητα 3ου επιμέρους μοντέλου με SV

Στη συνέχεια παρατίθεται ο πίνακας με την grayscale median τιμή για τα κόκκινα και τα μπλε εικονοστοιχεία.

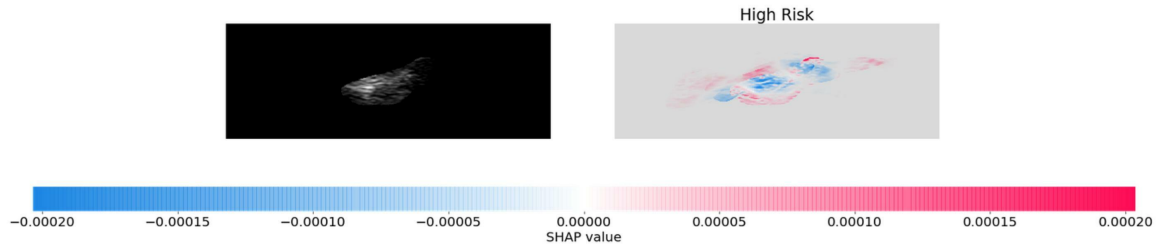
Επιμέρους Μοντέλα ENN	Grayscale median	
	Κόκκινα εικονοστοιχεία	Μπλε εικονοστοιχεία
1ο μοντέλο	147	184
2ο μοντέλο	177	192
3ο μοντέλο	77	165

Ερμηνευσιμότητα 3ου δείγματος, Ασθενής Χαμηλού Κινδύνου

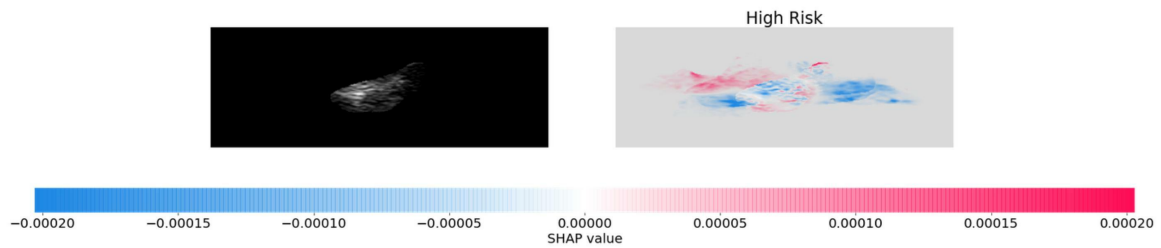
Τέλος θα εξετάσουμε ένα δείγμα το οποίο κατατάσσουν και τα τρία μοντέλα εσφαλμένα. Πρόκειται για ένα δείγμα που ανήκει σε ασθενή υψηλού κινδύνου αλλά και τα τρία μοντέλα το αναγνωρίζουν ως δείγμα που ανήκει στην κατηγορία του χαμηλού κινδύνου. Παρατηρούμε πως και τα τρία μοντέλα απέτυχαν να το κατηγοριοποιήσουν καθώς επηρεάστηκαν από εικονοστοιχεία που δεν αντιστοιχούν καν στα όρια του αθηρώματος που εξετάζουμε, σε αντίθεση με τις περιπτώσεις που εξετάσαμε προηγουμένως όπου τα εικονοστοιχεία που επηρέασαν τις τελικές αποφάσεις των μοντέλων αντιστοιχούσαν σε «σημαντικές» περιοχές του αθηρώματος.



Σχήμα 5.28: Ερμηνευσιμότητα 1ου επιμέρους μοντέλου με SV



Σχήμα 5.29: Ερμηνευσιμότητα 2ου επιμέρους μοντέλου με SV



Σχήμα 5.30: Ερμηνευσιμότητα 3ου επιμέρους μοντέλου με SV

Παρόλα αυτά όπως και προηγουμένως υπολογίσαμε τον grayscale median των κόκκινων και των μπλε εικονοστοιχείων. Τα αποτελέσματα που λάβαμε ήταν τα αναμενόμενα παρόλο που η τελική πρόβλεψη των μοντέλων είναι λάθος. Αυτό λοιπόν μας δείχνει πως η συμπεριφορά των μοντέλων παραμένει σταθερή όσον αφορά τον τρόπο που κατατάσσουν τα δείγματα με βάση τις σκουρόχρωμες και τις ανοιχτόχρωμες περιοχές, απλά στη συγκεκριμένη περίπτωση τα εικονοστοιχεία εκτός του αθηρώματος τα οδηγούν σε λανθασμένες αποφάσεις.

Επιμέρους Μοντέλα ENN	Grayscale median	
	Κόκκινα εικονοστοιχεία	Μπλε εικονοστοιχεία
1ο μοντέλο	66	189
2ο μοντέλο	62	135
3ο μοντέλο	81	166

Κεφάλαιο 6

Συμπεράσματα και μελλοντικές προεκτάσεις

6.1 Συμπεράσματα

Στόχος της παρούσας διπλωματικής εργασίας ήταν η χρήση τεχνικών Βαθιάς μάθησης για την ανάπτυξη μοντέλων ικανών να διαχειριστούν αριθμητικά δεδομένα αλλά και δεδομένα εικόνες με στόχο την κατηγοριοποίηση των ασθενών με καρωτιδική νόσο σε δύο κλάσεις: Ασθενείς Υψηλού και Χαμηλού κινδύνου, δηλαδή σε ασθενείς που θα ακολουθήσουν συντηρητική αγωγή και ασθενείς που θα πρέπει να υποβληθούν σε επεμβατική θεραπεία όπως η ενδαρτηρεκτομή ή η αγγειοπλαστική με τοποθέτηση stent, αντίστοιχα. Ειδικότερα, δημιουργήσαμε δύο μοντέλα συλλογικής μάθησης έτσι ώστε να αντιμετωπίσουμε το πρόβλημα της ανισορροπίας των δεδομένων μας. Τα μοντέλα που κατασκευάσαμε πέτυχαν αξιολογικά αποτελέσματα. Πιο συγκεκριμένα το Απλό Συλλογικό μοντέλο πέτυχε 75% accuracy, 80% specificity, 60% sensitivity, 70% F1-score και 0.7 Roc/AUC και το Συνελικτικό Συλλογικό μοντέλο πέτυχε 67.4% accuracy, 76% specificity, 63.5% Sensitivity, 75% F1-score και 0.7 Roc/AUC.

Στη συνέχεια διαφορετικές τεχνικές ερμηνευσιμότητας, και ειδικότερα οι **Global Surrogate**, **Individual Conditional Expectation plots** και **Shapley values** μέθοδοι, χρησιμοποιήθηκαν με σκοπό να ερμηνεύσουν τις τελικές προβλέψεις των μοντέλων μηχανικής μάθησης. Μέσω της εφαρμογής των μεθόδων ερμηνευσιμότητας συμπεράναμε πως οι μετρικές των μοντέλων μας δεν είναι "τυχαία" ικανοποιητικές αλλά στηρίζονται σε παρατηρήσεις που επιβεβαιώνονται μέσω ιατρικών μελετών και κλινικών συμπερασμάτων.

Μέσω της μεθόδου Global Surrogate καταλήξαμε στο συμπέρασμα πως το απλό συλλογικό μοντέλο κατηγοριοποιεί με μεγαλύτερη πιθανότητα τους ασθενείς με υψηλές τιμές ουρίας και χαμηλές τιμές HDL χοληστερόλης στην κλάση του υψηλού κινδύνου. Ειδικότερα η HDL χοληστερόλη παρουσίασε παρόμοια συμπεριφορά και μέσω των μεθόδων ICE και Shapley Values. Παράλληλα μέσω των δύο παραπάνω μεθόδων παρατηρήσαμε πως ένας ακόμα παράγοντας που ωθεί το μοντέλο να κατατάξει έναν ασθενή στην κατηγορία του υψηλού κινδύνου είναι οι χαμηλές τιμές της LDL χοληστερόλης καθώς και οι υψηλές τιμές των τριγλυκεριδίων.

Όσον αφορά το συνελικτικό συλλογικό μοντέλο, η μέθοδος ερμηνευσιμότητας Shapley Values μας οδήγησε στο συμπέρασμα πως τα εικονοστοιχεία με τη μικρότερη φωτεινότητα οδηγούν το μοντέλο να αποφασίσει με μεγαλύτερη πιθανότητα πως ο ασθενής ανήκει στην κατηγορία του υψηλού κινδύνου. Η συμπεριφορά αυτή έρχεται σε συμφωνία με προηγούμενες μελέτες όπου η JBA περιοχή και περιοχές της αθηρωματικής πλάκας με χαμηλό GSM

συνδέονται με την εμφάνιση καρδιαγγειακού συμβάματος.

Εν κατακλείδι, στην παρούσα διπλωματική εργασία καταφέραμε να υλοποιήσουμε δύο ερμηνεύσιμα μοντέλα βαθιάς μηχανικής μάθησης τα οποία διαχειρίζονται αριθμητικά και απεικονιστικά δεδομένα και συμβάλουν στην κατηγοριοποίηση των ασθενών με καρωτιδική νόσο.

6.2 Μελλοντικές Προεκτάσεις

Στην παρούσα εργασία μελλοντικά βήματα είναι τα ακόλουθα:

- Αρχικά θα μπορούσαμε να διαμοιράσουμε τα δεδομένα στα επιμέρους μοντέλα ακολουθώντας κάποιον διαφορετικό αλγόριθμο και όχι με τυχαίο διαμοιρασμό, με σκοπό να καταπολεμήσουμε τις αστοχίες που παρουσιάστηκαν στις μεθόδους ICE και Shapley values για τα χαρακτηριστικά της LDL χοληστερόλης και των τριγλυκεριδίων. Κάποιοι από τους αλγόριθμους διαμοιρασμού που θα μπορούσαμε να ακολουθήσουμε [73], οι οποίοι πιθανόν να βελτιώναν την συμπεριφορά και την απόδοση των μοντέλων είναι:
 - Η μέθοδος **Cadex-Duplex** [74] κατά την οποία ο διαχωρισμός των δεδομένων γίνεται με βάση την ευκλείδεια απόστασή τους
 - Η μέθοδος της **στρωματοποιημένης δειγματοληψίας - stratified sampling** [75], κατά την οποία διερευνάται η εσωτερική δομή του συνόλου δεδομένων, το οποίο στη συνέχεια διαχωρίζεται σε ομοιογενείς ομάδες - clusters και κατά τον διαχωρισμό τα δείγματα επιλέγονται ξεχωριστά από κάθε cluster.
- Επίσης θα ήταν χρήσιμη η εφαρμογή επιπλέον μεθόδων ερμηνευσιμότητας, όπως η μέθοδος LIME για το απλό αλλά και για το συνελικτικό συλλογικό μοντέλο. Ακόμα όσον αφορά το συνελικτικό συλλογικό μοντέλο θα ήταν ενδιαφέρουσα η εφαρμογή μεθόδων ερμηνευσιμότητας που βασίζονται στην εξαγωγή συμπερασμάτων μέσω heatmaps, όπως στην μελέτη [76].
- Ακόμα θα ήταν ενδιαφέρον ο συνδυασμός των κλινικών δεδομένων με τα δεδομένα των υπερήχων και η δημιουργία ενός πολυτροπικού μοντέλου [77] που θα τα αξιοποιούσε ταυτόχρονα.
- Τέλος, θα ήταν ωφέλιμο να εξεταστούν τα συμπεράσματα που εξαγάγαμε μέσω των μεθόδων ερμηνευσιμότητας και από κλινικούς γιατρούς προκειμένου να διερευνηθεί η ορθότητα των αποφάσεων του μοντέλου.

Bibliography

- [1] Véronique L. Roger, Alan S. Go, Donald M. Lloyd-Jones, Robert J. Adams, Jarett D. Berry, et al, *Heart disease and stroke statistics—2021 update: a report from the American Heart Association*. Circulation Journal of the American Heart Association, 2021.
- [2] Bharathikunisetty , Stephen , Senthilnathan , R.Deepak Arjundas , Ayyar , S.S.K and Suresh Kumar., *Prevalence and risk factors of carotid artery stenosis in ischaemic stroke patients a cross sectional study in a tertiary care hospital*. International Journal of Advanced Research (IJAR), 2019.
- [3] K. Prasad, “Pathophysiology and medical treatment of carotid artery stenosis,” *International Journal of Angiology*, vol. 24, no. 03, pp. 158–172, 2015.
- [4] K. Lekadir, A. Galimzianova, A. Betriu, M. Vila, L. Igual, DL. Rubil, et al, “A Convolutional Neural Network for Automatic Characterization of Plaque Composition in Carotid Ultrasound,” *IEEE J Biomed Health Inform*, 2016.
- [5] Abd-Ellah, M. K., Khalaf, A. A. M., Gharieb, R. R., Hassanin, D. A., “Automatic diagnosis of common carotid artery disease using diferent machine learning techniques,” *Journal of Ambient Intelligence and Humanized Computing*, 2021.
- [6] Skandha SS, Gupta SK, Saba L, et al, “3-D optimized classification and characterization artificial intelligence paradigm for cardiovascular/stroke risk stratification using carotid ultrasound-based delineated plaque: Atheromatic™,” *Comput Biol Med*, 2020.
- [7] Wei Ma, Xinyao Cheng, Xiangyang Xu, Furong Wang, Ran Zhou, Aaron Fenster, Mingyue Ding, “Multilevel Strip Pooling-Based Convolutional Neural Network for the Classification of Carotid Plaque Echogenicity,” *Machine Learning and Network Methods for Biology and Medicine*, 2021.
- [8] S. Golemati, A. Gastounioti, and K. S. Nikita, “Ultrasound-image-based cardiovascular tissue motion estimation,” *IEEE Reviews in Biomedical Engineering*, vol. 9, pp. 208–218, 2016.
- [9] F. Y. Rizi, J. Au, H. Yli-Ollila, S. Golemati, M. Makūnaitė, M. Orkisz, N. Navab, M. MacDonald, T. M. Laitinen, H. Behnam, *et al.*, “Carotid wall longitudinal motion in ultrasound imaging: An expert consensus review,” *Ultrasound in Medicine & Biology*, 2020.
- [10] S. Golemati, E. Patelaki, A. Gastounioti, I. Andreadis, C. D. Liapis, and K. S. Nikita, “Motion synchronisation patterns of the carotid atheromatous plaque from b-mode ultrasound,” *Scientific reports*, vol. 10, no. 1, pp. 1–13, 2020.

- [11] C. Molnar, *Interpretable Machine Learning*. 2021 [Online]. Available: <https://christophm.github.io/interpretable-ml-book/>.
- [12] Linardatos P, Papastefanopoulos V, Kotsiantis S, “Explainable AI: A Review of Machine Learning Interpretability Methods,” *Entropy*, 2021.
- [13] T. Brito-Sarracino, M. Santos, E. Antunes, I. Santos, J. Kasmanas, A. Carvalho, “Explainable machine learning for breast cancer diagnosis,” in *Proc. 8th Brazilian Conference on Intelligent Systems*, 2019.
- [14] P. Karatza, K. Dalakleidi, M. Athanasiou, K.S. Nikita, “Interpretability methods of machine learning algorithms with applications in breast cancer diagnosis,” *Engineering in Medicine and Biology Society (EMBC)*, 2021.
- [15] Donkor, Eric S, “Stroke in the century: a snapshot of the burden, epidemiology, and quality of life,” *Stroke research and treatment*, 2018.
- [16] Κωστούλας Ι., *Σύσταση αθηρωματικής πλάκας σε σταθερή και ασταθή στεφανιαία νόσο*. Διδακτορική Διατριβή, Ιατρική Σχολή, Πανεπιστήμιο Θεσσαλίας, 2007.
- [17] Βασίλειος Γ. Σαλεπτοής, *Η σημασία της υπερηχογραφικής διερεύνησης ως εξέταση πρόληψης για την διάγνωση ασυμπτωματικής στένωσης καρωτίδας*. Μεταπτυχιακή Διπλωματική Εργασία, Ιατρική Σχολή, Πανεπιστήμιο Θεσσαλίας, 2017.
- [18] Ι. Κατσιόρης, *Αθηρωματική νόσος του καρωτιδικού διχασμού και περιοδοντική νόσο*. Διδακτορική Διατριβή, Ιατρική Σχολή, Πανεπιστήμιο Αθηνών, 2014.
- [19] Δρ. Ιωάννης Κ. Μπέλλος, Αγγειοχειρουργός Ενδαγγειοχειρουργός, *Καρωτιδική νόσος και εγκεφαλικό επεισόδιο*. <https://aggeiopathia.gr/el/pathiseis/pathiseis-artirion/karotidiki-nosos.html>.
- [20] Γαστουγιώτη Α., «Υπολογιστική ανάλυση μηχανικών χαρακτηριστικών του αρτηριακού τοιχώματος από απεικονιστικές καταγραφές με στόχο την υποβοήθηση της διάγνωσης της καρωτιδικής αθηρωμάτωσης». Διδακτορική Διατριβή, Σχολή Ηλεκτρολόγων Μηχανικών Μηχανικών Υπολογιστών, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, 2014.
- [21] K. S. Nikita, “Atherosclerosis: the evolving role of vascular image analysis,” *Computerized medical imaging and graphics*, vol. 37, no. 1, pp. 1–3, 2013.
- [22] A. Gastouniotti, S. Golemati, J. Stoitsis, and K. Nikita, “Comparison of kalman-filter-based approaches for block matching in arterial wall motion analysis from b-mode ultrasound,” *Measurement Science and Technology*, vol. 22, no. 11, p. 114008, 2011.
- [23] V. Aboyans, J. Ricco, M. Bartelink, M. Björck, M. Brodmann, T. Cohnert, J. Collet, M. Czerny, M. De Carlo, S. Debus, *et al.*, “Esc scientific document group. 2017 esc guidelines on the diagnosis and treatment of peripheral arterial diseases, in collaboration with the european society for vascular surgery (esvs): Document covering atherosclerotic disease of extracranial carotid and vertebral, mesenteric, renal, upper and lower extremity arteries endorsed by: the european stroke organization (eso) the task force for the diagnosis and treatment of peripheral arterial diseases of the european society of cardiology (esc) and of the european society for vascular surgery (esvs),” *Eur Heart J*, vol. 39, no. 9, pp. 763–816, 2018.

- [24] A. S. Panayides, A. Amini, N. D. Filipovic, A. Sharma, S. A. Tsiftaris, A. Young, D. Foran, N. Do, S. Golemati, T. Kurc, *et al.*, “Ai in medical imaging informatics: current challenges and future directions,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 7, pp. 1837–1857, 2020.
- [25] S. Golemati and K. S. Nikita, *Cardiovascular Computing-Methodologies and Clinical Applications*. Springer, 2019.
- [26] I. K. Valavanis, S. G. Mougiakakou, K. A. Grimaldi, and K. S. Nikita, “A multifactorial analysis of obesity as cvd risk factor: use of neural network based methods in a nutrigenetics context,” *BMC bioinformatics*, vol. 11, no. 1, pp. 1–10, 2010.
- [27] M. Skevofilakas, K. Zarkogianni, B. G. Karamanos, and K. S. Nikita, “A hybrid decision support system for the risk assessment of retinopathy development as a long term complication of type 1 diabetes mellitus,” in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 6713–6716, IEEE, 2010.
- [28] P. Refaeilzadeh, L. Tang, and H. Liu, “Cross-validation.,” *Encyclopedia of database systems*, vol. 5, pp. 532–538, 2009.
- [29] D. Opitz and R. Maclin, “Popular ensemble methods: An empirical study,” *Journal of artificial intelligence research*, vol. 11, pp. 169–198, 1999.
- [30] N. N. Tsiaparas, S. Golemati, I. Andreadis, J. S. Stoitsis, I. Valavanis, and K. S. Nikita, “Comparison of multiresolution features for texture classification of carotid atherosclerosis from b-mode ultrasound,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 1, pp. 130–137, 2010.
- [31] S. Golemati, J. S. Stoitsis, A. Gastounioti, A. C. Dimopoulos, V. Koropouli, and K. S. Nikita, “Comparison of block matching and differential methods for motion analysis of the carotid artery wall from ultrasound images,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 5, pp. 852–858, 2012.
- [32] A. Gastounioti, S. Golemati, J. Stoitsis, and K. Nikita, “Carotid artery wall motion analysis from b-mode ultrasound using adaptive block matching: in silico evaluation and in vivo application,” *Physics in Medicine & Biology*, vol. 58, no. 24, p. 8647, 2013.
- [33] S. Golemati, J. Stoitsis, E. G. Sifakis, T. Balkizas, and K. S. Nikita, “Using the hough transform to segment ultrasound images of longitudinal and transverse sections of the carotid artery,” *Ultrasound in medicine & biology*, vol. 33, no. 12, pp. 1918–1932, 2007.
- [34] J. Stoitsis, I. Valavanis, S. G. Mougiakakou, S. Golemati, A. Nikita, and K. S. Nikita, “Computer aided diagnosis based on medical image processing and artificial intelligence methods,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 569, no. 2, pp. 591–595, 2006.

- [35] S. G. Mougiakakou, S. Golemati, I. Gousias, A. N. Nicolaidis, and K. S. Nikita, “Computer-aided diagnosis of carotid atherosclerosis based on ultrasound image statistics, laws’ texture and neural networks,” *Ultrasound in medicine & biology*, vol. 33, no. 1, pp. 26–36, 2007.
- [36] P. Kokol, M. Mernik, J. Završnik, K. Kancler, and I. Malčić, “Decision trees based on automatic learning and their use in cardiology,” *Journal of Medical Systems*, vol. 18, no. 4, pp. 201–206, 1994.
- [37] A. V. Kelarev, A. Stranieri, J. Yearwood, and H. F. Jelinek, “Empirical study of decision trees and ensemble classifiers for monitoring of diabetes patients in pervasive healthcare,” in *2012 15th International Conference on Network-Based Information Systems*, pp. 441–446, IEEE, 2012.
- [38] T. Brito-Sarracino, M. R. dos Santos, E. F. Antunes, I. B. de Andrade Santos, J. C. Kasmanas, A. C. P. de Leon Ferreira, *et al.*, “Explainable machine learning for breast cancer diagnosis,” in *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 681–686, IEEE, 2019.
- [39] M. Shouman, T. Turner, and R. Stocker, “Applying k-nearest neighbour in diagnosing heart disease patients,” *International Journal of Information and Education Technology*, vol. 2, no. 3, pp. 220–223, 2012.
- [40] P. Bhuvaneswari and A. B. Therese, “Detection of cancer in lung with k-nn classification using genetic algorithm,” *Procedia Materials Science*, vol. 10, pp. 433–440, 2015.
- [41] M. Shahbaz, S. Ali, A. Guergachi, A. Niazi, and A. Umer, “Classification of alzheimer’s disease using machine learning techniques,” in *DATA*, pp. 296–303, 2019.
- [42] S. M. Lundberg, B. Nair, M. S. Vavilala, M. Horibe, M. J. Eisses, T. Adams, D. E. Liston, D. K.-W. Low, S.-F. Newman, J. Kim, *et al.*, “Explainable machine-learning predictions for the prevention of hypoxaemia during surgery,” *Nature biomedical engineering*, vol. 2, no. 10, pp. 749–760, 2018.
- [43] R. Elshawi, M. H. Al-Mallah, and S. Sakr, “On the interpretability of machine learning-based model for predicting hypertension,” *BMC medical informatics and decision making*, vol. 19, no. 1, pp. 1–32, 2019.
- [44] A. Gastounioti, V. Koliass, S. Golemati, N. N. Tsiaparas, A. Matsakou, J. S. Stoitsis, N. P. Kadoglou, C. Gkekas, J. D. Kakasis, C. D. Liapis, *et al.*, “Carotid—a web-based platform for optimal personalized management of atherosclerotic patients,” *Computer methods and programs in biomedicine*, vol. 114, no. 2, pp. 183–193, 2014.
- [45] Jaroslav Pelisek, Ines N. Hahntow, Hans-Henning Eckstein, Stefan Ockert, Christian Reeps, MD, a Peter Heider, *et al.*, “Impact of chronic kidney disease on carotid plaque vulnerability,” *Journal of vascular surgery*, 2011.
- [46] Nobuo Kajitani, Haruhito A. Uchida, Isao Suminoe, Yuki Kakio, Masashi Kitagawa, Hajime Sato, Jun Wada, “Chronic kidney disease is associated with carotid atherosclerosis and symptomatic ischaemic stroke,” *Journal of International Medical Research*, 2018.

- [47] N. Βαλλιάνου, Π. Κουταλάς, “Η γ-γλουταμυλτρανσφεράση (γ-GT) του ορού και τα καρδιαγγειακά νοσήματα,” *Αρχαία Ελληνικής Ιατρικής*, 2008.
- [48] Shengyang Jiang, Donglin Jiang, Yijia Tao, “Role of gamma-glutamyltransferase in cardiovascular diseases,” *Experimental and Clinical Cardiology*, 2013.
- [49] Han-Bin Lee, Jinkwon Kim, Sang-Heum Kim, Soonhag Kim, Ok-Joon Kim, Seung-Hun O, “Association between Serum Alkaline Phosphatase Level and Cerebral Small Vessel Disease,” *PLoS ONE* 10, 2015.
- [50] Marcello Tonelli, Gary Curhan, Marc Pfeffer, Frank Sacks, Ravi Thadhani, Michal L Melamed, et al, “Relation between alkaline phosphatase, serum phosphate, and all-cause or cardiovascular mortality,” *Circulation*, 2009.
- [51] Wannamethee SG, Sattar N, Papcosta O, Lennon, Whincup PH, “Alkaline phosphatase, serum phosphate, and incident cardiovascular disease and total mortality in older men,” *Europe PMC*, 2013.
- [52] Jinkwon Kim, Tae-Jin Song, Dongbeom Song, Hye Sun Lee, Chung Mo Nam, Hyo Suk Nam, Young Dae Kim, Ji Hoe Heo, “Serum alkaline phosphatase and phosphate in cerebral atherosclerosis and functional outcomes after cerebral infarction,” *Stroke*, 2013.
- [53] Matthias Hoke, Martin Schillinger, Erich Minar, Georg Goliasch, Christoph J. Binder, Florian J. Mayer, “Carotid ultrasound investigation as a prognostic tool for patients with diabetes mellitus,” *Cardiovascular Diabetology*, 2019.
- [54] Minsu Noh, Hyunwook Kwon, Chang Hee Jung, Sun U. Kwon, Min Seon Kim, Woo Je Lee, “Impact of diabetes duration and degree of carotid artery stenosis on major adverse cardiovascular events: a single-center, retrospective, observational cohort study,” *Cardiovasc Diabetol*, 2017.
- [55] K. Zarkogianni, M. Athanasiou, and A. C. Thanopoulou, “Comparison of Machine Learning Approaches Toward Assessing the Risk of Developing Cardiovascular Disease as a Long-Term Diabetes Complication,” *IEEE J Biomed Health*, vol 22, no. 5, 2018.
- [56] S. Ruder, “An overview of gradient descent optimization algorithms,” *arXiv:1609.04747*, 2016.
- [57] S. Santurkar, D. Tsipras, A. Ilyas, A. Madry, “How Does Batch Normalization Help Optimization?,” *arXiv:1805.11604*, 2018.
- [58] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [59] P.P Toth, “Subclinical atherosclerosis: what it is, what it means and what we can do about it,” 2008.
- [60] Welty, Francine K, “How do elevated triglycerides and low HDL-cholesterol affect inflammation and atherothrombosis?,” 2013.

- [61] “LDL and HDL Cholesterol: ”Bad” and ”Good” Cholesterol,” *Center of disease control and prevention*. https://www.cdc.gov/cholesterol/ldl_hdl.htm.
- [62] P. Barter, “The role of hdl-cholesterol in preventing atherosclerotic disease,” *European heart journal Supplements*, vol. 7, no. suppl_F, pp. F4–F8, 2005.
- [63] M. Bohm and N. Werner, “Hdl cholesterol in the atherosclerotic disease,”
- [64] Brown, M. S., Goldstein, J. L., “How LDL Receptors Influence Cholesterol and Atherosclerosis,” pp. 58–69, 1984.
- [65] B. G. Talayero and F. M. Sacks, “The role of triglycerides in atherosclerosis,” *Current cardiology reports*, vol. 13, no. 6, pp. 544–552, 2011.
- [66] Zhang Q, Liu S, Liu Y, et al, “Achieving low density lipoprotein-cholesterol \leq 70 mg/dL may be associated with a trend of reduced progression of carotid artery atherosclerosis in ischemic stroke patients,” 2017.
- [67] Ndrepepa G, Colleran R, Kastrati A., “Gamma-glutamyl transferase and the risk of atherosclerosis and coronary heart disease,” *Clin Chim Acta*, 2018.
- [68] Talayero, Beatriz G., and Frank M. Sacks., “The Role of Triglycerides in Atherosclerosis,” *Current cardiology reports* 13.6, 2011.
- [69] S. Golemati, T. J. Tegos, A. Sassano, K. S. Nikita, and A. N. Nicolaides, “Echogenicity of b-mode sonographic images of the carotid artery: work in progress,” *Journal of ultrasound in medicine*, vol. 23, no. 5, pp. 659–669, 2004.
- [70] N. El-Barghouty, G. Geroulakos, A. Nicolaides, A. Androulakis, and V. Bahal, “Computer-assisted carotid plaque characterisation,” *European Journal of Vascular and Endovascular Surgery*, vol. 9, no. 4, pp. 389–393, 1995.
- [71] G. Biasi, A. Sampaolo, P. Mingazzini, P. De Amicis, N. El-Barghouty, and A. Nicolaides, “Computer analysis of ultrasonic plaque echolucency in identifying high risk carotid bifurcation lesions,” *European journal of vascular and endovascular surgery*, vol. 17, no. 6, pp. 476–479, 1999.
- [72] S. K. Kakkos, M. B. Griffin, A. N. Nicolaides, E. Kyriacou, M. M. Sabetai, T. Tegos, G. C. Makris, D. J. Thomas, G. Geroulakos, R. of Stroke (ACSRS) Study Group, *et al.*, “The size of juxtaluminal hypoechoic area in ultrasound images of asymptomatic carotid plaques predicts the occurrence of stroke,” *Journal of vascular surgery*, vol. 57, no. 3, pp. 609–618, 2013.
- [73] Z. Reitermanov , “Data Splitting,” 2010.
- [74] Kennard, R. and Stone, L., “Computer aided design of experiments,” 1969.
- [75] May, R. J., Maier, H. R., Dandy, G. C., “Data splitting for artificial neural networks using som-based stratified sampling,” 2010.
- [76] E. Tjoa and G. Cuntai, “Convolutional neural network interpretability with general pattern theory,” *arXiv preprint arXiv:2102.04247*, 2021.

- [77] S.-C. Huang, A. Pareek, S. Seyyedi, I. Banerjee, and M. P. Lungren, “Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines,” *NPJ digital medicine*, vol. 3, no. 1, pp. 1–9, 2020.