



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΥΛΙΚΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΒΙΟΙΑΤΡΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ

**Η εξηγησιμότητα μοντέλων μηχανικής μάθησης στη
βιοϊατρική μηχανική και η εφαρμογή της σε
συστήματα ψευδοτυχαιωμάτων επινεφριδίων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Γεώργιου Δ. Κοτζιά

Επιβλέπων: Ματσόπουλος Γεώργιος
Καθηγητής ΗΜΜΥ Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2025



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΥΛΙΚΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΒΙΟΙΑΤΡΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ

Η εξηγησιμότητα μοντέλων μηχανικής μάθησης στη
βιοϊατρική μηχανική και η εφαρμογή της σε
συστήματα ψευδοτυχαιωμάτων επινεφριδίων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Γεώργιου Δ. Κοτζιά

Επιβλέπων: Ματσόπουλος Γεώργιος
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 18-02-2025.

.....
Γ. Ματσόπουλος Καθ. Ε.Μ.Π.

.....
Π. Τσανάκας Καθ. Ε.Μ.Π.

.....
Ο. Πετροπούλου Δρ. Ε.ΔΙ.Π. Α'

Αθήνα, Φεβρουάριος 2025

.....

Γεώργιος Δ. Κοτζιάς

Διπλωματούχος Ηλεκτρολόγων Μηχανικών & Μηχανικών
Υπολογιστών Ε.Μ.Π.

© Γεώργιος Δ. Κοτζιάς (2025).

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

Η παρούσα διπλωματική εργασία διερευνά την εφαρμογή μεθόδων Εξηγήσιμης Τεχνητής Νοημοσύνης (ETN) στην ταξινόμηση και διάγνωση τυχαιωμάτων επινεφριδίων, ένα κρίσιμο πεδίο της σύγχρονης ιατρικής. Στο πλαίσιο της έρευνας, αναλύθηκε ένα σύνολο δεδομένων 262 ασθενών, το οποίο περιλάμβανε πλήθος κλινικών, εργαστηριακών και απεικονιστικών παραμέτρων. Η μελέτη επικεντρώθηκε στην ανάπτυξη ενός αξιόπιστου και ερμηνεύσιμου μοντέλου για την υποστήριξη κλινικών αποφάσεων, με ιδιαίτερη έμφαση στην ικανότητα του συστήματος να παρέχει κατανοητές εξηγήσεις για τις προβλέψεις του. Στο πλαίσιο της μεθοδολογικής προσέγγισης, εφαρμόστηκαν και αξιολογήθηκαν συστηματικά διάφοροι αλγόριθμοι μηχανικής μάθησης. Μέσα από μια ενδεδειγμένη διαδικασία συγκριτικής αξιολόγησης, ο XGBoost αναδείχθηκε ως το βέλτιστο μοντέλο, επιτυγχάνοντας εξαιρετική ακρίβεια στη διασταυρούμενη επικύρωση, με αξιοσημείωτη σταθερότητα στις προβλέψεις του. Ιδιαίτερη έμφαση δόθηκε στην ερμηνευσιμότητα των αποτελεσμάτων μέσω των προηγμένων μεθόδων SHAP (SHapley Additive exPlanations) και LIME (Local Interpretable Model-agnostic Explanations), οι οποίες επέτρεψαν τη λεπτομερή κατανόηση των παραγόντων που επηρεάζουν τις προβλέψεις του μοντέλου. Η ανάλυση των αποτελεσμάτων ανέδειξε συγκεκριμένα κλινικά χαρακτηριστικά που διαδραματίζουν καθοριστικό ρόλο στη διάγνωση των τυχαιωμάτων επινεφριδίων, παρέχοντας πολύτιμες πληροφορίες για τους κλινικούς ιατρούς. Τα αποτελέσματα υποδεικνύουν ότι η ενσωμάτωση τεχνικών ETN μπορεί να συμβάλει σημαντικά στην αξιοπιστία και αποδοχή των συστημάτων τεχνητής νοημοσύνης στην κλινική πράξη, προσφέροντας όχι μόνο υψηλή ακρίβεια αλλά και τη δυνατότητα κατανόησης της συλλογιστικής πίσω από κάθε πρόβλεψη. Η έρευνα αυτή ευελπιστούμε να αποτελέσει ένα σημαντικό βήμα προς την κατεύθυνση της ανάπτυξης αξιόπιστων και διαφανών συστημάτων υποστήριξης κλινικών αποφάσεων, τα οποία μπορούν να ενσωματωθούν αποτελεσματικά στην καθημερινή κλινική πρακτική, βελτιώνοντας την ποιότητα της παρεχόμενης φροντίδας υγείας.

Λέξεις Κλειδιά: Εξηγήσιμη Τεχνητή Νοημοσύνη, Τυχαιώματα Επινεφριδίων, Συστήματα Υποστήριξης Κλινικών Αποφάσεων

Abstract

This thesis explores the application of Explainable Artificial Intelligence (AI) methods in the classification and diagnosis of adrenal gland tumors, a critical field of modern medicine. As part of the study, a dataset of 262 patients was analyzed, which included a multitude of clinical, laboratory and imaging parameters. The study focused on developing a reliable and interpretable model to support clinical decisions, with a particular emphasis on the ability of the system to provide understandable explanations for its predictions. As part of the methodological approach, several machine learning algorithms were systematically applied and evaluated. Through a thorough benchmarking process, the XGBoost classifier emerged as the optimal model, achieving excellent accuracy in cross-validation, with remarkable stability in its predictions. Particular emphasis was placed on the interpretability of the results through the advanced SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) methods, which allowed a detailed understanding of the factors influencing the model's predictions. Analysis of the results highlighted specific clinical features that play a key role in the diagnosis of adrenal tychemia, providing valuable information for clinicians. The results suggest that the integration of XAI techniques can significantly contribute to the reliability and acceptance of AI systems in clinical practice, offering not only high accuracy but also the ability to understand the reasoning behind each prediction. We hope this research will prove to be an important step towards the development of reliable and transparent clinical decision support systems that can be effectively integrated into daily clinical practice, improving the quality of health care provided.

Keywords: Explainable Artificial Intelligence, Adrenal Incidentalomas, Clinical Decision Support Systems

Ευχαριστίες

Αρχικώς, θα ήθελα να ευχαριστήσω θερμά τον καθηγητή κ. Γεώργιο Μαστόπουλο για την εμπιστοσύνη την οποία μου έδειξε και την ανάθεση του ομολογουμένως πολύ ενδιαφέροντος θέματος. Παράλληλα, θα ήθελα να ευχαριστήσω τον κ. Μπίνα Δημήτριο, για τη στήριξη και καθοδήγηση, καθ' όλη τη διάρκεια εκπόνησης του παρόντος.

Κοτζιάς Γεώργιος,
Φεβρουάριος 2025

Περιεχόμενα

Κεφάλαιο 1	Εισαγωγή	Σελ. 15
1.1	Σκοπός της εργασίας	15
1.2	Δομή της εργασίας	17
Κεφάλαιο 2	Ιατρικό υπόβαθρο	Σελ. 18
2.1	Επινεφρίδια	18
2.1.1	Ανατομία	18
2.1.2	Φυσιολογία	18
2.1.3	Λειτουργίες	19
2.2	Τυχαιώματα επινεφριδίων	19
2.2.1	Ορισμός	19
2.2.2	Παθοφυσιολογία	19
2.2.3	Τύποι τυχαιωμάτων	20
2.2.4	Κλινική σημασία της έγκαιρης διάγνωσης κακοήθειας	20
Κεφάλαιο 3	Τεχνητή νοημοσύνη και ιατρική	Σελ. 21
3.1	Εισαγωγή στην τεχνητή νοημοσύνη	21
3.1.1	Βασικές έννοιες	21
3.1.2	Μηχανική Μάθηση	21
3.1.3	Βαθιά Μάθηση	22
3.2	Συστήματα Υποστήριξης Κλινικών Αποφάσεων (ΣΥΚΑ)	22
3.2.1	Ορισμός	22
3.2.2	Αρχιτεκτονική ΣΥΚΑ	22
3.2.3	Τύποι ΣΥΚΑ	22
3.2.4	Πλεονεκτήματα και περιορισμοί	23
3.3	Σημασία ερμηνευσιμότητας στην ιατρική	23
3.3.1	Νομικές απαιτήσεις	23
3.3.2	Ηθικά ζητήματα	23
3.3.3	Σχέση ιατρού - ασθενούς	24
3.3.4	Κλινική ευθύνη - λογοδοσία	24

3.4	Ιστορική εξέλιξη	25
3.4.1	Σημαντικοί σταθμοί στην εξέλιξη.....	25

Κεφάλαιο 4	Εξηγήσιμη τεχνητή νοημοσύνη (ETN)	Σελ. 27
4.1	Ανάγκη εξηγησιμότητας	27
4.2	Τύποι εξηγήσεων	27
4.3	Μέθοδοι ETN	28
4.3.1	Local Interpretable Model-agnostic Explanations (LIME)	28
4.3.2	SHapley Additive exPlanations (SHAP)	28
4.3.3	Integrated Gradients	29
4.3.4	Counterfactual Explanations.....	29
4.3.5	Layer-wise Relevance Propagation	30

Κεφάλαιο 5	Βιβλιογραφική ανασκόπηση	Σελ. 31
5.1	Μεθοδολογία έρευνας	31
5.1.1	Κριτήρια αναζήτησης	31
5.1.2	Βάσεις δεδομένων	32
5.2	Σημαντικά επιστημονικά περιοδικά και συνέδρια στο πεδίο	32
5.3	ETN και διάγνωση τυχαιωμάτων	33
5.3.1	Υπάρχοντα συστήματα.....	33
5.3.2	Μέθοδοι εξήγησης.....	33
5.3.3	Προκλήσεις και περιορισμοί.....	34

Κεφάλαιο 6	Σύνολο δεδομένων, επιλογή μοντέλου	Σελ. 35
6.1	Προετοιμασία συνόλου δεδομένων	35
6.2	Ανάπτυξη και Αξιολόγηση Μοντέλων	37
6.2.1	Προεπεξεργασία δεδομένων και μετρικές αξιολόγησης	37
6.2.2	Θεμελιώδεις μετρικές ταξινόμησης	37
6.2.3	Μοντέλα Μηχανικής Μάθησης και Υπερπαραμέτροι.....	38
6.2.4	Αποτελέσματα Αξιολόγησης	40
6.3	Περιορισμοί της Μελέτης	45
6.4	Μελλοντικές Προοπτικές	46

Κεφάλαιο 7	Επεξήγηση μοντέλου επιλογής	Σελ. 47
7.1	Μέθοδοι SHAP και LIME.....	47
7.2	Υποθέσεις ταξινόμησης ψευδοτυχαιωμάτων	52
7.3	Διαγράμματα για SHAP ανάλυση.....	54
7.3.1	Διάγραμμα Σημαντικότερων Χαρακτηριστικών (Top Features) .	54
7.3.2	Διάγραμμα Δυνάμεων (Force Plot)	55

7.3.3	Διάγραμμα Καταράκτη	57
7.3.4	Συνδυαστικές τιμές	59
7.4	Διαγράμματα για LIME ανάλυση	59
7.5	Ανάλυση Εσφαλμένης Ταξινόμησης.	61

Κεφάλαιο 8 Συμπεράσματα - μελλοντικές προεκτάσεις Σελ. 63

BIBΛΙΟΓΡΑΦΙΑ Σελ. 64

Κατάλογος πινάκων

Βασικοί όροι αναζήτησης.....	31
Κατηγορίες συστημάτων TN για τυχαιώματα.....	33
Ομαδοποίηση δεδομένων (χαρακτηριστικών) ασθενών	36
Μοντέλα μηχανικής μάθησης που αξιολογήθηκαν	38
Υπερπαραμέτροι που επελέγησαν για κάθε μοντέλο.....	39
Επίδοση Μοντέλων με τεχνική SMOTE	40
Προγνωστικές Αξίες και Μετρικές Απόδοσης ανά Μοντέλο	44
Πίνακας αντιστοίχισης Θεωρίας Παιγνίων - Μηχανικής Μάθησης	50
Οριακή συνεισφορά και τιμές Shapley για τα 3 συμπτώματα.....	50
(ψευδό)-Παραδείγματα διαταραχών LIME και τα αντίστοιχα βάρη τους	51
Σύγκριση SHAP και LIME.....	51
Υποθέσεις ταξινόμησης ψευδοτυχαιωμάτων	52

Κατάλογος σχημάτων

Ανατομία επινεφριδίου	20
Πίνακες σύγχυσης για τα εννέα μοντέλα μηχανικής μάθησης που εξετάστηκαν	41
Ανάλυση σταθερότητας μοντέλων	41
Καμπύλες απόδοσης	42
Κατανομή επιδόσεων ανά κλάση	43
Σύγκριση μεθόδων ερμηνείας μοντέλου: SHAP και LIME	47
Διαδικασία ανάπτυξης μοντέλου μηχανικής μάθησης	48
Ροή επεξήγησης XAI	53
Σημαντικότερα χαρακτηριστικά	54
Διάγραμμα δυνάμεων για κάθε κλάση	55
Περίπτωση εσφαλμένης ταξινόμησης	57
Διαγράμματα Καταρράκτη (2) ανά κλάση	58
Διάγραμμα lime ανά κλάση	60
Περίπτωση εσφαλμένης ταξινόμησης	61

Εισαγωγή

1.1 Σκοπός της εργασίας

Η παρούσα διπλωματική εργασία επικεντρώνεται στη μελέτη και εφαρμογή μεθόδων Εξηγήσιμης Τεχνητής Νοημοσύνης (ETN) (Explainable Artificial Intelligence - ΧΑΙ) στον τομέα της βιοϊατρικής μηχανικής, με ιδιαίτερη έμφαση στην ταξινόμηση τυχαιωμάτων επινεφριδίων. Η σύγχρονη έρευνα γύρω από την ένταξη της τεχνητής νοημοσύνης στην ιατρική πράξη και ιδιαίτερα της εξηγήσιμης υποδηλώνει την τάση την οποία διακατέχει το συγκεκριμένο πεδίο [1]. Στη σύγχρονη ιατρική πραγματικότητα, η ενσωμάτωση συστημάτων υποστήριξης αποφάσεων βασισμένων στην τεχνητή νοημοσύνη αποτελεί μια αναδυόμενη τάση με σημαντικές προοπτικές [2]. Τα συστήματα αυτά έχουν επιδείξει αξιοσημείωτη ακρίβεια στη διάγνωση και πρόγνωση διαφόρων παθολογικών καταστάσεων, συμπεριλαμβανομένων των τυχαιωμάτων [3]. Ωστόσο, η αποτελεσματική εφαρμογή τέτοιων συστημάτων προϋποθέτει την κατανόηση και ερμηνεία των αποφάσεών τους από την ιατρική κοινότητα [4]. Τα τυχαιώματα επινεφριδίων αποτελούν ένα σημαντικό κλινικό ζήτημα, με την εμφάνισή τους στο γενικό πληθυσμό να κυμαίνεται μεταξύ 1-8.7% και να αυξάνεται με την ηλικία [5]. Η έγκαιρη και ακριβής διάγνωση της κακοήθειας σε αυτές τις περιπτώσεις είναι κρίσιμη για την πορεία εξέλιξης του ευρήματος, καθώς το πρωτοπαθές επινεφριδιακό καρκίνωμα, παρότι σπάνιο, χαρακτηρίζεται από ιδιαίτερα επιθετική συμπεριφορά με σημαντικά μειωμένα ποσοστά επιβίωσης στην περίπτωση καθυστερημένης διάγνωσης [6]. Η μεθοδολογική προσέγγιση που ακολουθείται στην παρούσα εργασία διακρίνεται για την ανάλυση των δεδομένων και την ολιστική αντιμετώπιση του προβλήματος της διάγνωσης. Συγκεκριμένα, η διαδικασία ξεκινά από τη λεπτομερή ανάλυση των κλινικών δεδομένων, προχωρά στην εφαρμογή προηγμένων τεχνικών μηχανικής μάθησης, και καταλήγει στην ερμηνεία των αποτελεσμάτων με τρόπο που είναι ευελπιστούμε να είναι άμεσα αξιοποιήσιμος από τους επαγγελματίες υγείας. Η προσέγγιση αυτή ενσωματώνει τις βέλτιστες πρακτικές που έχουν αναδειχθεί στη βιβλιογραφία, ενώ παράλληλα εισάγει στοιχεία στον τρόπο συνδυασμού των διαφορετικών μεθοδολογικών εργαλείων. Η εργασία αξιοποιεί ένα μοναδικό σύνολο δεδομένων που συνδυάζει παραδοσιακές κλινικές παραμέτρους με σύγχρονους βιοδείκτες και απεικονιστικά χαρακτηριστικά. Αυτή η προσέγγιση εναρμονίζεται με τις σύγχρονες τάσεις στη βιοϊατρική έρευνα [7], όπου η ποιότητα των δεδομένων αποτελεί θεμελιώδη προϋπόθεση για την αξιοπιστία

των συμπερασμάτων και την πρακτική εφαρμογή των αποτελεσμάτων.

Η μεθοδολογική προσέγγιση περιλαμβάνει την ανάπτυξη και αξιολόγηση διαφόρων αλγορίθμων μηχανικής μάθησης, καθώς και της ερμηνείας αυτών, με ιδιαίτερη έμφαση στις μεθόδους SHAP (SHapley Additive exPlanations) [8] και LIME (Local Interpretable Model - agnostic Explanations) [9].

Οι κύριοι στόχοι της εργασίας περιλαμβάνουν:

1. Την ανάπτυξη ενός αξιόπιστου μοντέλου πρόβλεψης για την ταξινόμηση τυχαιωμάτων επινεφριδίων
2. Τη συγκριτική αξιολόγηση των μεθόδων SHAP και LIME στην ερμηνεία των προβλέψεων
3. Την αναγνώριση των σημαντικότερων κλινικών παραμέτρων που επηρεάζουν τη διάγνωση της κακοήθειας

Η πρωτοτυπία της εργασίας έγκειται στη συνδυαστική προσέγγιση της πρόβλεψης και εξήγησης στο συγκεκριμένο κλινικό πλαίσιο των τυχαιωμάτων επινεφριδίων. Η συνεισφορά της εργασίας στο πεδίο είναι πολυεπίπεδη:

- Προσφέρει ένα καινοτόμο εργαλείο υποστήριξης κλινικών αποφάσεων για την αξιολόγηση τυχαιωμάτων επινεφριδίων
- Προάγει την κατανόηση των δυνατοτήτων και περιορισμών των μεθόδων ETN στην ιατρική διάγνωση
- Συμβάλλει στην ευρύτερη συζήτηση για την ενσωμάτωση εξηγήσιμων συστημάτων τεχνητής νοημοσύνης στην κλινική πράξη

Η επιλογή του συγκεκριμένου θέματος υπαγορεύεται από την επιτακτική ανάγκη γεφύρωσης του χάσματος μεταξύ των προηγμένων συστημάτων τεχνητής νοημοσύνης και της καθημερινής κλινικής πρακτικής. Η διάγνωση των τυχαιωμάτων επινεφριδίων αποτελεί ένα ιδανικό πεδίο εφαρμογής τέτοιων συστημάτων, καθώς συνδυάζει την πολυπλοκότητα των κλινικών δεδομένων με την κρισιμότητα της έγκαιρης διάγνωσης. Επιπλέον, η εξηγησιμότητα των αλγοριθμικών αποφάσεων δεν αποτελεί απλώς μια τεχνική πρόκληση, αλλά μια θεμελιώδη προϋπόθεση για την οικοδόμηση εμπιστοσύνης μεταξύ των συστημάτων τεχνητής νοημοσύνης και των επαγγελματιών υγείας [7]. Σε μια εποχή όπου η ιατρική περίθαλψη γίνεται ολοένα και πιο εξαρτώμενη από την τεχνολογία, η ανάπτυξη εξηγήσιμων συστημάτων υποστήριξης αποφάσεων μπορεί να συμβάλει καθοριστικά στη βελτίωση της ποιότητας της υγειονομικής περίθαλψης και στην ασφάλεια των ασθενών [4].

1.2 Δομή της εργασίας

Η παρούσα διπλωματική εργασία διαρθρώνεται σε κύρια κεφάλαια, τα οποία οργανώνονται ως εξής:

Στο [Κεφάλαιο 1](#) παρουσιάζεται η εισαγωγή της εργασίας, όπου αναλύεται ο σκοπός και η σημασία της μελέτης των τυχαιωμάτων επινεφριδίων με τη χρήση εξηγήσιμης τεχνητής νοημοσύνης.

Το [Κεφάλαιο 2](#) εστιάζει στο ιατρικό υπόβαθρο, παρέχοντας μια εκτενή ανάλυση των επινεφριδίων, της ανατομίας τους, της φυσιολογίας τους και των τυχαιωμάτων που μπορεί να εμφανιστούν σε αυτά.

Στο [Κεφάλαιο 3](#) εξετάζεται η σχέση μεταξύ τεχνητής νοημοσύνης και ιατρικής, με έμφαση στα Συστήματα Υποστήριξης Κλινικών Αποφάσεων (ΣΥΚΑ) και τη σημασία της ερμηνευσιμότητας στην ιατρική πράξη.

Το [Κεφάλαιο 4](#) εμβαθύνει στην Εξηγήσιμη Τεχνητή Νοημοσύνη (ETN), παρουσιάζοντας τις βασικές μεθόδους και τεχνικές που χρησιμοποιούνται για την ερμηνεία των αποφάσεων των μοντέλων μηχανικής μάθησης.

Στο [Κεφάλαιο 5](#) παρατίθεται μια εκτενής βιβλιογραφική ανασκόπηση του πεδίου, αναλύοντας τις τρέχουσες εξελίξεις και τάσεις στην εφαρμογή της ETN στη διάγνωση τυχαιωμάτων.

Στο [Κεφάλαιο 6](#) παρουσιάζει το σύνολο δεδομένων που χρησιμοποιήθηκε, την επιλογή και αξιολόγηση των μοντέλων, καθώς και την εξήγηση των αποτελεσμάτων τους με τη χρήση τεχνικών ETN.

Στο [Κεφάλαιο 7](#) αναπτύσσονται οι μέθοδοι SHAP και LIME για το επιλεγέν μοντέλο μηχανικής μάθησης.

Τέλος, στο [Κεφάλαιο 8](#) συγκεντρώνονται τα συμπεράσματα της παρούσας έρευνας καθώς και ορισμένες προτεινόμενες μελλοντικές προεκτάσεις.

Ιατρικό υπόβαθρο

2.1 Επινεφρίδια

Τα επινεφρίδια αποτελούν ενδοκρινείς αδένες ζωτικής σημασίας για τον ανθρώπινο οργανισμό [10]. Εντοπίζονται στον οπισθοπεριτοναϊκό χώρο, ένα σε κάθε πλευρά, στους άνω πόλους των νεφρών [11].

2.1.1 Ανατομία

Κάθε επινεφρίδιο έχει μάζα περίπου 4-6 γραμμαρίων στους ενήλικες και αποτελείται από δύο διακριτά τμήματα με διαφορετική εμβρυολογική προέλευση [12]. Το μεγαλύτερο τμήμα του αδένα, που αντιπροσωπεύει το 80-90% της συνολικής μάζας του, είναι ο επινεφριδιακός φλοιός (adrenal cortex). Ο φλοιός διαιρείται σε τρεις διακριτές ζώνες: τη σπειροειδή ζώνη (zona glomerulosa), τη στηλιδωτή ζώνη (zona fasciculata) και τη δικτυωτή ζώνη (zona reticularis). Το εσωτερικό τμήμα του αδένα, ο επινεφριδιακός μυελός (adrenal medulla), καταλαμβάνει το υπόλοιπο 10-20% της μάζας [13]. Η αγγείωση των επινεφριδίων είναι ιδιαίτερα πλούσια, με τρεις κύριες αρτηρίες να τροφοδοτούν κάθε αδένα: την άνω, τη μέση και την κάτω επινεφριδιακή αρτηρία. Αυτή η εκτεταμένη αγγείωση είναι απαραίτητη για την επαρκή τροφοδοσία του αδένα με αίμα, δεδομένου του σημαντικού μεταβολικού του ρόλου [14].

2.1.2 Φυσιολογία

Η φυσιολογία των επινεφριδίων χαρακτηρίζεται από την παραγωγή διαφόρων ορμονών, με κάθε ζώνη του φλοιού να εξειδικεύεται στην παραγωγή συγκεκριμένων ορμονών [15]. Η σπειροειδής ζώνη παράγει αλατοκορτικοειδή, με κύριο εκπρόσωπο την αλδοστερόνη. Η στηλιδωτή ζώνη είναι υπεύθυνη για την παραγωγή γλυκοκορτικοειδών, με σημαντικότερη την κορτιζόλη. Η δικτυωτή ζώνη παράγει ανδρογόνα, κυρίως δεϋδροεπιανδροστερόνη (DHEA) και ανδροστενδιόνη. Ο μυελός των επινεφριδίων παράγει κατεχολαμίνες, συγκεκριμένα επινεφρίνη και νορεπινεφρίνη [16].

2.1.3 Λειτουργίες

Οι ορμόνες των επινεφριδίων επιτελούν πολλαπλές και κρίσιμες λειτουργίες στον οργανισμό [17]. Η αλδοστερόνη διαδραματίζει κεντρικό ρόλο στη ρύθμιση του ισοζυγίου νατρίου-καλίου, στον έλεγχο της αρτηριακής πίεσης και στη διατήρηση του όγκου του πλάσματος. Η κορτιζόλη είναι απαραίτητη για το μεταβολισμό των υδατανθράκων, των πρωτεϊνών και των λιπών, ενώ παράλληλα έχει σημαντική ανοσορρυθμιστική και αντιφλεγμονώδη δράση. Επιπλέον, διαδραματίζει κεντρικό ρόλο στην απόκριση του οργανισμού στο στρες [18]. Τα ανδρογόνα των επινεφριδίων συμβάλλουν στην ανάπτυξη των δευτερογενών χαρακτηριστικών του φύλου και στην ανάπτυξη των οστών. Οι κατεχολαμίνες είναι υπεύθυνες για την απόκριση "μάχης ή φυγής", ρυθμίζουν την καρδιακή συχνότητα και την αρτηριακή πίεση, ενώ συμμετέχουν ενεργά στη μεταβολική προσαρμογή του οργανισμού στο στρες. Η διαταραχή της λειτουργίας των επινεφριδίων μπορεί να οδηγήσει σε σοβαρές παθολογικές καταστάσεις, όπως η νόσος Addison (ανεπάρκεια) ή το σύνδρομο Cushing (υπερλειτουργία) [19].

2.2 Τυχαιώματα επινεφριδίων

2.2.1 Ορισμός

Τα τυχαιώματα επινεφριδίων (adrenal incidentalomas) αποτελούν μάζες των επινεφριδίων που ανακαλύπτονται τυχαία κατά τη διάρκεια απεικονιστικού ελέγχου ο οποίος διενεργείται για άλλη αιτία [20]. Ο όρος εισήχθη για πρώτη φορά στην ιατρική βιβλιογραφία τη δεκαετία του 1980 και έκτοτε η συχνότητα ανεύρεσής τους έχει αυξηθεί σημαντικά, κυρίως λόγω της ευρείας χρήσης σύγχρονων απεικονιστικών μεθόδων [5]. Η επίπτωσή τους στο γενικό πληθυσμό κυμαίνεται μεταξύ 1-8.7%, αυξανόμενη με την ηλικία [21, 22, 23].

2.2.2 Παθοφυσιολογία

Η παθοφυσιολογία των τυχαιωμάτων επινεφριδίων είναι πολύπλοκη και περιλαμβάνει διάφορους μηχανισμούς ανάπτυξης ενώ η μη φυσιολογική ανάπτυξη των επινεφριδίων μπορεί να επιβαρύνει πολλαπλά τον ανθρώπινο οργανισμό [24]. Σε μοριακό επίπεδο, η ανάπτυξή τους σχετίζεται με διαταραχές στη ρύθμιση του κυτταρικού πολλαπλασιασμού και της απόπτωσης. Γενετικές μεταλλάξεις σε συγκεκριμένα γονίδια, όπως τα TP53, CTNNB1, και PRKAR1A, έχουν συσχετιστεί με την ανάπτυξη διαφόρων τύπων τυχαιωμάτων [25]. Επιπλέον, διαταραχές στη λειτουργία του άξονα υποθαλάμου-υπόφυσης-επινεφριδίων μπορούν να οδηγήσουν στην ανάπτυξη ορμονοπαραγωγών τυχαιωμάτων [26].

2.2.3 Τύποι τυχαιωμάτων

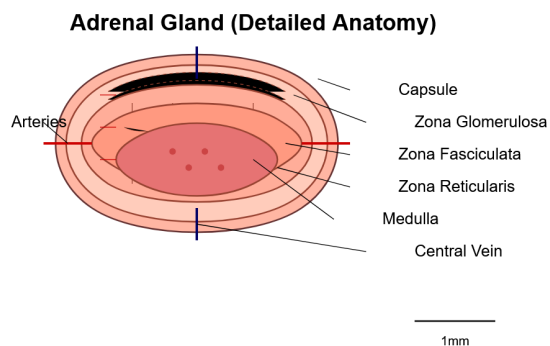
Τα τυχαιώματα επινεφριδίων κατηγοριοποιούνται με βάση τα λειτουργικά και μορφολογικά τους χαρακτηριστικά [27]. Από λειτουργική άποψη, διακρίνονται σε μη λειτουργικά και λειτουργικά. Τα λειτουργικά τυχαιώματα περιλαμβάνουν φαιοχρωμοκυτώματα, αδενώματα που παράγουν κορτιζόλη (σύνδρομο Cushing) και αλδοστερόνη (πρωτοπαθής αλδοστερονισμός). Από μορφολογική άποψη, διακρίνονται σε αδενώματα, μυελολιπώματα, κύστεις, αιμορραγίες και πρωτοπαθή ή μεταστατικά καρκινώματα [28].

2.2.4 Κλινική σημασία της έγκαιρης διάγνωσης κακοήθειας

Η έγκαιρη διάγνωση της κακοήθειας στα τυχαιώματα επινεφριδίων είναι καθοριστικής σημασίας για την πρόγνωση των ασθενών [29]. Το πρωτοπαθές επινεφριδιακό καρκίνωμα, αν και σπάνιο, έχει ιδιαίτερα επιθετική συμπεριφορά. Η πενταετής επιβίωση για τοπικά περιορισμένη νόσο είναι περίπου 60%, ενώ μειώνεται δραματικά σε 10-20% σε περίπτωση μεταστατικής νόσου [30].

Η διαφορική μεταξύ καλοήθων και κακοήθων τυχαιωμάτων βασίζεται σε πολλαπλά κριτήρια, συμπεριλαμβανομένων των απεικονιστικών χαρακτηριστικών, του μεγέθους, της ορμονικής δραστηριότητας και του ρυθμού ανάπτυξης [31]. Η χρήση σύγχρονων απεικονιστικών τεχνικών, όπως η αξονική τομογραφία με πρωτόκολλο αδενώματος και η μαγνητική τομογραφία με ειδικές ακολουθίες, σε συνδυασμό με τον ενδεδειγμένο ορμονικό έλεγχο, επιτρέπει την ακριβέστερη διάγνωση και την έγκαιρη θεραπευτική παρέμβαση [32].

Η έγκαιρη αναγνώριση της κακοήθειας είναι επίσης κρίσιμη για τη βελτιστοποίηση της χειρουργικής προσέγγισης και την αποφυγή διεγχειρητικής ρήξης του όγκου, η οποία σχετίζεται με σημαντικά χειρότερη πρόγνωση. Επιπλέον, η έγκαιρη διάγνωση επιτρέπει την έγκαιρη έναρξη συστηματικής θεραπείας σε περιπτώσεις όπου ενδείκνυται, βελτιώνοντας τη συνολική έκβαση [5].



Σχήμα 2.1: Ανατομία επινεφριδίου

Τεχνητή νοημοσύνη και ιατρική

3.1 Εισαγωγή στην τεχνητή νοημοσύνη

Η τεχνητή νοημοσύνη (TN) αποτελεί έναν από τους ταχύτερα αναπτυσσόμενους τομείς της επιστήμης των υπολογιστών, με εφαρμογές που επεκτείνονται σε όλο το φάσμα της ανθρώπινης δραστηριότητας [33]. Ορίζεται ως ο κλάδος της επιστήμης των υπολογιστών που ασχολείται με τη σχεδίαση και ανάπτυξη ευφύων συστημάτων, ικανών να μιμηθούν ανθρώπινες γνωστικές λειτουργίες, όπως η μάθηση, η προσαρμογή και η επίλυση προβλημάτων [34].

3.1.1 Βασικές έννοιες

Η θεμελιώδης αρχή της τεχνητής νοημοσύνης βασίζεται στην έννοια της υπολογιστικής νοημοσύνης, η οποία περιλαμβάνει την ικανότητα ενός συστήματος να αντιλαμβάνεται, να μαθαίνει και να προσαρμόζεται στο περιβάλλον του [35]. Τα συστήματα TN χαρακτηρίζονται από την ικανότητά τους να επεξεργάζονται μεγάλο όγκο δεδομένων, να αναγνωρίζουν πρότυπα και να εξάγουν συμπεράσματα βάσει των διαθέσιμων πληροφοριών [36]. Η ανάπτυξη της TN στηρίζεται σε διάφορα επιστημονικά πεδία, συμπεριλαμβανομένων των μαθηματικών, της στατιστικής, της νευροεπιστήμης και της γνωσιακής ψυχολογίας [37].

3.1.2 Μηχανική Μάθηση

Η μηχανική μάθηση αποτελεί υποπεδίο της τεχνητής νοημοσύνης που επικεντρώνεται στην ανάπτυξη αλγορίθμων και στατιστικών μοντέλων [38]. Επιτρέπει στους υπολογιστές να βελτιώνουν προοδευτικά την απόδοσή τους σε συγκεκριμένες εργασίες μέσω της εμπειρίας, χωρίς να έχουν ρητά προγραμματιστεί για αυτές [39]. Διακρίνεται σε τρεις βασικές κατηγορίες: την επιβλεπόμενη μάθηση, όπου το σύστημα εκπαιδεύεται με επισημειωμένα δεδομένα, τη μη επιβλεπόμενη μάθηση, που επικεντρώνεται στην ανακάλυψη κρυφών δομών σε μη επισημειωμένα δεδομένα, και την ενισχυτική μάθηση, που βασίζεται στην αλληλεπίδραση με το περιβάλλον [40].

3.1.3 Βαθιά Μάθηση

Η βαθιά μάθηση αποτελεί εξειδικευμένο πεδίο της μηχανικής μάθησης, βασιζόμενη στη χρήση τεχνητών νευρωνικών δικτύων με πολλαπλά επίπεδα επεξεργασίας [40]. Τα νευρωνικά δίκτυα βαθιάς μάθησης αποτελούνται από διαδοχικά επίπεδα διασυνδεδεμένων κόμβων, όπου κάθε επίπεδο μετασχηματίζει την είσοδό του σε πιο αφηρημένες αναπαραστάσεις [41]. Η εκπαίδευση πραγματοποιείται μέσω του αλγορίθμου οπισθοδιάδοσης και της στοχαστικής καθόδου κλίσης, απαιτώντας σημαντική υπολογιστική ισχύ και μεγάλο όγκο δεδομένων [41]. Στον ιατρικό τομέα, η βαθιά μάθηση έχει επιδείξει εξαιρετικά αποτελέσματα στην ανάλυση ιατρικών εικόνων και τη διάγνωση ασθενειών, παρά τις προκλήσεις στην ερμηνευσιμότητα των αποφάσεών της [42].

3.2 Συστήματα Υποστήριξης Κλινικών Αποφάσεων (ΣΥΚΑ)

Η εξέλιξη της τεχνολογίας στον τομέα της υγείας έχει οδηγήσει στην ανάπτυξη εξελιγμένων συστημάτων υποστήριξης κλινικών αποφάσεων, τα οποία έχουν μετασχηματίσει τον τρόπο με τον οποίο λαμβάνονται οι ιατρικές αποφάσεις [43].

3.2.1 Ορισμός

Τα Συστήματα Υποστήριξης Κλινικών Αποφάσεων (ΣΥΚΑ) είναι εξειδικευμένα λογισμικά σχεδιασμένα να παρέχουν στους επαγγελματίες υγείας εξατομικευμένη κλινική υποστήριξη στη λήψη αποφάσεων [44]. Αξιοποιώντας σύγχρονες τεχνολογίες πληροφορικής και τεχνητής νοημοσύνης, τα ΣΥΚΑ ενσωματώνουν κλινικές γνώσεις, ιατρικά δεδομένα και αλγορίθμους για την παροχή έγκαιρων και τεκμηριωμένων συστάσεων στο σημείο φροντίδας [45].

3.2.2 Αρχιτεκτονική ΣΥΚΑ

Η αρχιτεκτονική των ΣΥΚΑ περιλαμβάνει τρία βασικά συστατικά στοιχεία [46]. Η βάση γνώσης περιέχει κωδικοποιημένη ιατρική γνώση, κλινικές κατευθυντήριες οδηγίες και κανόνες συμπερασμού. Η μηχανή συμπερασμού αποτελεί τον πυρήνα του συστήματος, εφαρμόζοντας λογική και αλγορίθμους για την επεξεργασία των δεδομένων. Η διεπαφή χρήστη επιτρέπει την αλληλεπίδραση με το σύστημα και την παρουσίαση των αποτελεσμάτων με κατανοητό τρόπο [47].

3.2.3 Τύποι ΣΥΚΑ

Τα ΣΥΚΑ κατηγοριοποιούνται με βάση διάφορα κριτήρια, συμπεριλαμβανομένης της λειτουργικότητας και του τρόπου παροχής συστάσεων [48]. Τα συστήματα βασισμένα σε κανόνες χρησιμοποιούν προκαθορισμένους κλινικούς

κανόνες για την παραγωγή συστάσεων. Τα συστήματα βασισμένα σε μηχανική μάθηση αξιοποιούν τεχνικές τεχνητής νοημοσύνης για την ανάλυση μεγάλου όγκου δεδομένων και την εξαγωγή προτύπων. Υπάρχουν επίσης υβριδικά συστήματα που συνδυάζουν διαφορετικές προσεγγίσεις για βελτιστοποιημένη απόδοση [49].

3.2.4 Πλεονεκτήματα και περιορισμοί

Τα ΣΥΚΑ προσφέρουν σημαντικά πλεονεκτήματα στην κλινική πράξη, συμπεριλαμβανομένης της βελτίωσης της ποιότητας φροντίδας, της μείωσης ιατρικών σφαλμάτων και της τυποποίησης της κλινικής πρακτικής [50]. Συμβάλλουν στην ταχύτερη λήψη αποφάσεων και στην καλύτερη τεκμηρίωση της κλινικής πρακτικής. Ωστόσο, υπάρχουν και σημαντικοί περιορισμοί [51]. Η ποιότητα των συστάσεων εξαρτάται από την ακρίβεια και την πληρότητα των δεδομένων εισόδου. Η ενσωμάτωση των ΣΥΚΑ στην καθημερινή κλινική πρακτική μπορεί να αντιμετωπίσει αντίσταση από τους επαγγελματίες υγείας. Επιπλέον, η υπερβολική εξάρτηση από τα συστήματα αυτά μπορεί να οδηγήσει σε μείωση της κλινικής κρίσης [43].

3.3 Σημασία ερμηνευσιμότητας στην ιατρική

Στο πλαίσιο της αυξανόμενης χρήσης συστημάτων τεχνητής νοημοσύνης στην ιατρική, η ερμηνευσιμότητα των αποφάσεων αναδεικνύεται ως θεμελιώδης απαίτηση για την ασφαλή και αποτελεσματική εφαρμογή τους[7]. Η σημασία της ερμηνείας φαίνεται από τη σημασία του πώς απαντάμε στο ερώτημα: "Γιατί ο αλγόριθμος έβγαλε αυτό το αποτέλεσμα;".

3.3.1 Νομικές απαιτήσεις

Το νομικό πλαίσιο που διέπει τη χρήση συστημάτων τεχνητής νοημοσύνης στην ιατρική θέτει αυστηρές απαιτήσεις σχετικά με τη διαφάνεια και την ερμηνευσιμότητα των αποφάσεων [52]. Ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR) της Ευρωπαϊκής Ένωσης απαιτεί ρητά το "δικαίωμα στην εξήγηση" για αποφάσεις που λαμβάνονται από αυτοματοποιημένα συστήματα [53]. Στις ΗΠΑ, ο FDA έχει θεσπίσει κατευθυντήριες γραμμές για την αξιολόγηση και έγκριση ιατρικών συσκευών που ενσωματώνουν τεχνητή νοημοσύνη, με έμφαση στην ανάγκη για διαφανή και ερμηνεύσιμα συστήματα [54]. Αντίστοιχη νομοθετική πρακτική υπάρχει και σε άλλες χώρες όπως η Ινδία (DPDP).

3.3.2 Ηθικά ζητήματα

Η ηθική διάσταση της ερμηνευσιμότητας στην ιατρική είναι πολυεπίπεδη και άπτεται θεμελιωδών αρχών της βιοηθικής [55]. Η αυτονομία του ασθενούς προϋποθέτει την κατανόηση των παραγόντων που επηρεάζουν τις ιατρικές αποφάσεις. Η δικαιοσύνη απαιτεί διαφάνεια στον τρόπο με τον οποίο τα

συστήματα τεχνητής νοημοσύνης καταλήγουν σε συγκεκριμένες αποφάσεις, ιδιαίτερα όταν αυτές μπορεί να έχουν διαφορετικό αντίκτυπο σε διαφορετικές πληθυσμιακές ομάδες [56]. Η αρχή της μη βλάβης επιβάλλει την κατανόηση των πιθανών κινδύνων και περιορισμών των αλγοριθμικών συστημάτων [55]. Η ευεργεσία, ως θεμελιώδης αρχή της ιατρικής ηθικής, απαιτεί τη βέλτιστη χρήση της τεχνολογίας προς όφελος του ασθενούς, κάτι που προϋποθέτει την πλήρη κατανόηση της λειτουργίας των συστημάτων από τους επαγγελματίες υγείας [56]. Επιπλέον, η κοινωνική δικαιοσύνη στην υγειονομική περίθαλψη απαιτεί την αξιολόγηση και αντιμετώπιση πιθανών προκαταλήψεων στα αλγοριθμικά συστήματα, κάτι που είναι εφικτό μόνο μέσω της ερμηνευσιμότητας [55]. Η διαφάνεια στη λήψη αποφάσεων είναι επίσης καίρια για την οικοδόμηση εμπιστοσύνης μεταξύ του συστήματος υγείας και των διαφορετικών κοινοτήτων που εξυπηρετεί [56].

3.3.3 Σχέση ιατρού - ασθενούς

Η ερμηνευσιμότητα των ιατρικών αποφάσεων αποτελεί θεμελιώδες στοιχείο της σχέσης ιατρού-ασθενούς [57]. Η εμπιστοσύνη στη θεραπευτική σχέση ενισχύεται όταν οι ασθενείς κατανοούν το σκεπτικό πίσω από τις ιατρικές αποφάσεις. Οι ιατροί πρέπει να είναι σε θέση να εξηγούν με κατανοητό τρόπο τις συστάσεις των συστημάτων τεχνητής νοημοσύνης και να τις ενσωματώνουν στον κλινικό διάλογο [58]. Η έλλειψη ερμηνευσιμότητας μπορεί να υπονομεύσει την εμπιστοσύνη των ασθενών και να επηρεάσει αρνητικά τη συμμόρφωση στη θεραπεία [59]. Αυτό είναι ιδιαίτερα σημαντικό σε περιπτώσεις χρόνιων παθήσεων, όπου η μακροχρόνια συμμόρφωση στη θεραπεία είναι κρίσιμη για την έκβαση [57]. Η αποτελεσματική επικοινωνία των αλγοριθμικών προτάσεων απαιτεί από τους ιατρούς να μπορούν να μεταφράζουν τεχνικές έννοιες σε κατανοητή γλώσσα για τους ασθενείς [58]. Επιπλέον, η διαφάνεια στη λήψη ιατρικών αποφάσεων ενισχύει την αυτονομία του ασθενούς και την ενεργό συμμετοχή του στη διαχείριση της υγείας του [59]. Η ερμηνευσιμότητα διευκολύνει επίσης τον εποικοδομητικό διάλογο σχετικά με εναλλακτικές θεραπευτικές επιλογές και επιτρέπει στους ασθενείς να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με τη φροντίδα τους [57].

3.3.4 Κλινική ευθύνη - λογοδοσία

Η κατανομή της κλινικής ευθύνης και η λογοδοσία σε περιπτώσεις ιατρικών σφαλμάτων αποτελούν κρίσιμα ζητήματα στην εποχή της τεχνητής νοημοσύνης [60]. Η ερμηνευσιμότητα των αλγοριθμικών αποφάσεων είναι απαραίτητη για τον προσδιορισμό της ευθύνης σε περιπτώσεις δυσμενών συμβάντων. Οι ιατροί πρέπει να μπορούν να κατανοούν και να αξιολογούν κριτικά τις προτάσεις των συστημάτων τεχνητής νοημοσύνης, καθώς διατηρούν την τελική ευθύνη για τις κλινικές αποφάσεις [4]. Τα συστήματα καταγραφής και ελέγχου των αλγοριθμικών αποφάσεων είναι απαραίτητα για τη διασφάλιση της λογοδοσίας και την προστασία τόσο των ασθενών όσο και των επαγγελματιών υγείας [61]. Σύμφωνα με το άρθρο 22 του GDPR, οι ασθενείς έχουν το δικαίωμα να μην υπόκεινται σε αποφάσεις που βασίζονται αποκλειστικά σε αυ-

τοματοποιημένη επεξεργασία, συμπεριλαμβανομένης της κατάρτισης προφίλ. Επιπλέον, το άρθρο 13 του GDPR απαιτεί διαφάνεια στην επεξεργασία δεδομένων, συμπεριλαμβανομένης της ύπαρξης αυτοματοποιημένης λήψης αποφάσεων και της λογικής που χρησιμοποιείται [4]. Οι οργανισμοί υγειονομικής περίθαλψης οφείλουν να διενεργούν εκτιμήσεις αντικτύπου (DPIA) για συστήματα τεχνητής νοημοσύνης που επεξεργάζονται ευαίσθητα δεδομένα υγείας, σύμφωνα με την αρχή της λογοδοσίας του GDPR (άρθρο 5.2) [60].

3.4 Ιστορική εξέλιξη

Η εξέλιξη της υπολογιστικής ιατρικής και της βιοϊατρικής μηχανικής αντικατοπτρίζει μια συναρπαστική διαδρομή τεχνολογικής προόδου η οποία εκτείνεται σε περισσότερες από έξι δεκαετίες, από την χρήση υπολογιστικών συστημάτων [62] τη δεκαετία του 1960 έως τις σύγχρονες εφαρμογές της τεχνητής νοημοσύνης [63, 64, 65]. Από τις πρώιμες προσπάθειες αυτοματοποίησης απλών ιατρικών διαδικασιών μέχρι τα σύγχρονα συστήματα τεχνητής νοημοσύνης, ο τομέας έχει μεταμορφώσει ριζικά την παροχή υγειονομικής περίθαλψης.

Στις αρχές της δεκαετίας του 1960, οι υπολογιστές στην ιατρική είχαν περιορισμένο ρόλο, εστιάζοντας κυρίως στην αποθήκευση και ανάκτηση δεδομένων ασθενών. Ωστόσο, η εποχή αυτή έθεσε τα θεμέλια για την μετέπειτα επανάσταση στην ιατρική πληροφορική. Η σταδιακή ανάπτυξη της υπολογιστικής ισχύος, σε συνδυασμό με την εξέλιξη των μαθηματικών μοντέλων και των αλγορίθμων, άνοιξε νέους ορίζοντες στην κατανόηση και αντιμετώπιση των ασθενειών.

3.4.1 Σημαντικοί σταθμοί στην εξέλιξη

Η δεκαετία του 1970 σηματοδότησε μια κρίσιμη καμπή με τη μετάβαση από τις παραδοσιακές στατιστικές μεθόδους στα πρώτα συστήματα τεχνητής νοημοσύνης. Το σύστημα MYCIN [66], που αναπτύχθηκε στο Πανεπιστήμιο Stanford, αποτέλεσε ορόσημο στην ιστορία των έμπειρων συστημάτων. Χρησιμοποιώντας μια βάση κανόνων που προσομοίωνε την ιατρική σκέψη, το MYCIN μπορούσε να προτείνει θεραπείες για σοβαρές λοιμώξεις με ακρίβεια συγκρίσιμη με εκείνη των ειδικών ιατρών της εποχής [67]. Η δεκαετία του 1980 έφερε την επανάσταση των νευρωνικών δικτύων στην ανάλυση ιατρικών εικόνων. Τα πρώτα συστήματα, αν και πρωτόγονα σε σύγκριση με τα σημερινά πρότυπα, έδειξαν την τεράστια δυνατότητα της αυτοματοποιημένης ανάλυσης ιατρικών απεικονίσεων [68, 69, 70]. Παράλληλα, η δεκαετία του 1990 σηματοδότησε την είσοδο της μηχανικής μάθησης στη γονιδιωματική, επιταχύνοντας δραματικά την ανάλυση γονιδιακών δεδομένων και ανοίγοντας νέους δρόμους στην εξατομικευμένη ιατρική [71, 72].

Η δεκαετία του 2010 έφερε την επανάσταση της βαθιάς μάθησης στην ιατρική διάγνωση. Τα συστήματα βαθιάς μάθησης άρχισαν να επιδεικνύουν εντυπωσιακές επιδόσεις σε ποικίλες διαγνωστικές εργασίες, συχνά ξεπερνώντας την ακρίβεια των έμπειρων κλινικών ιατρών σε συγκεκριμένους τομείς

[73, 74, 75]. Το σύστημα Isabel, για παράδειγμα, αποτέλεσε σημαντικό βήμα στην υποστήριξη διαφορικής διάγνωσης, αξιοποιώντας μια εκτεταμένη βάση γνώσεων και προηγμένους αλγόριθμους επεξεργασίας φυσικής γλώσσας [76]. Παράλληλα, η ανάπτυξη του Watson for Oncology της IBM [77] αποτέλεσε ορόσημο στην εφαρμογή της τεχνητής νοημοσύνης στην ογκολογία, παρά τις προκλήσεις που αντιμετώπισε στην κλινική εφαρμογή.

Ιδιαίτερα σημαντική ήταν η εξέλιξη των συστημάτων αναγνώρισης προτύπων στην ανάλυση ιατρικών εικόνων. Οι αλγόριθμοι Support Vector Machines (SVM) και Random Forests άρχισαν να εφαρμόζονται ευρέως στην ανάλυση ακτινολογικών εικόνων [78], ενώ η εισαγωγή των Convolutional Neural Networks (CNNs) στην ιατρική απεικόνιση [38] έθεσε τα θεμέλια για την επανάσταση της βαθιάς μάθησης που ακολούθησε. Η περίοδος 2015-2020 χαρακτηρίστηκε από την εμφάνιση ακόμη πιο εξελιγμένων αρχιτεκτονικών βαθιάς μάθησης, όπως τα Residual Networks [79] και τα Transformer models [80], που έφεραν επανάσταση στην ανάλυση ιατρικών δεδομένων. Ιδιαίτερα σημαντική ήταν η συνεισφορά των Generative Adversarial Networks (GANs) [81] στη σύνθεση και επαύξηση ιατρικών δεδομένων, αντιμετωπίζοντας το χρόνιο πρόβλημα της περιορισμένης διαθεσιμότητας δεδομένων εκπαίδευσης.

Ωστόσο, η αυξανόμενη πολυπλοκότητα των μοντέλων τεχνητής νοημοσύνης έφερε στο προσκήνιο το ζήτημα της εξηγησιμότητας. Η ανάγκη για διαφάνεια και κατανόηση των αποφάσεων των αλγορίθμων οδήγησε στην ανάπτυξη νέων μεθοδολογιών όπως η LIME [9] και το SHAP [8], που επιτρέπουν την ερμηνεία των προβλέψεων των μοντέλων μηχανικής μάθησης. Παράλληλα, η έρευνα στράφηκε προς την ανάπτυξη εγγενώς ερμηνεύσιμων μοντέλων [82], αναγνωρίζοντας ότι η εξηγησιμότητα δεν πρέπει να είναι μεταγενέστερη προσθήκη αλλά εγγενές χαρακτηριστικό των συστημάτων τεχνητής νοημοσύνης στην ιατρική. Σήμερα, η έμφαση δίνεται στην ανάπτυξη υβριδικών συστημάτων που συνδυάζουν την υψηλή ακρίβεια των μοντέλων βαθιάς μάθησης με την εξηγησιμότητα παραδοσιακών προσεγγίσεων [4]. Η τάση αυτή αντανάκλα την αυξανόμενη αναγνώριση ότι η επιτυχής ενσωμάτωση της τεχνητής νοημοσύνης στην κλινική πράξη απαιτεί όχι μόνο υψηλή ακρίβεια αλλά και την εμπιστοσύνη της ιατρικής κοινότητας [57].

Εξηγήσιμη τεχνητή νοημοσύνη (ETN)

Η Εξηγήσιμη Τεχνητή Νοημοσύνη (Explainable Artificial Intelligence - XAI) αναδύεται ως ένα κρίσιμο πεδίο έρευνας που στοχεύει στην αντιμετώπιση του προβλήματος της "μαύρου κουτιού" στα συστήματα τεχνητής νοημοσύνης [83].

4.1 Ανάγκη εξηγησιμότητας

Η αυξανόμενη πολυπλοκότητα των μοντέλων μηχανικής μάθησης, ιδιαίτερα στον τομέα της βαθιάς μάθησης, έχει δημιουργήσει την επιτακτική ανάγκη για κατανόηση και ερμηνεία των αποφάσεών τους [82]. Η εξηγησιμότητα είναι ιδιαίτερα σημαντική σε τομείς υψηλού ρίσκου όπως η ιατρική, όπου οι αποφάσεις μπορεί να έχουν σημαντικές επιπτώσεις στην ανθρώπινη ζωή [4].

4.2 Τύποι εξηγήσεων

Οι εξηγήσεις στην ETN μπορούν να κατηγοριοποιηθούν σε διάφορους τύπους [84]:

- Οι καθολικές εξηγήσεις (global explanations) στοχεύουν στην κατανόηση της συνολικής συμπεριφοράς του μοντέλου.
- Οι τοπικές εξηγήσεις (local explanations) επικεντρώνονται στην ερμηνεία συγκεκριμένων προβλέψεων.
- Οι εξηγήσεις βάσει παραδειγμάτων (example-based explanations) χρησιμοποιούν αντιπροσωπευτικές περιπτώσεις για την επεξήγηση των αποφάσεων του μοντέλου [85].

4.3 Μέθοδοι ETN

4.3.1 Local Interpretable Model-agnostic Explanations (LIME)

Η μέθοδος LIME παρέχει τοπικές εξηγήσεις για τις προβλέψεις οποιουδήποτε μοντέλου μηχανικής μάθησης [9]. Λειτουργεί δημιουργώντας ένα απλό, ερμηνεύσιμο μοντέλο που προσεγγίζει τη συμπεριφορά του πολύπλοκου μοντέλου σε μια συγκεκριμένη πρόβλεψη. Μια εφαρμογή είναι στην ιατρική απεικόνιση, δηλαδή ο εντοπισμός σημείων μιας εικόνας που επηρέασαν περισσότερο την απόφαση του μοντέλου [8].

Η διαδικασία LIME περιλαμβάνει τη δημιουργία διαταραγμένων εκδόσεων των δεδομένων εισόδου και την παρατήρηση του τρόπου με τον οποίο αλλάζουν οι προβλέψεις του μοντέλου. Αυτή η προσέγγιση επιτρέπει την κατανόηση της τοπικής συμπεριφοράς του μοντέλου γύρω από μια συγκεκριμένη πρόβλεψη.

Η LIME λειτουργεί δημιουργώντας ένα τοπικά πιστό γραμμικό μοντέλο γύρω από μια συγκεκριμένη πρόβλεψη μέσω της ακόλουθης διαδικασίας:

Δειγματοληψία: Δημιουργεί διαταραγμένα δείγματα γύρω από το στιγμιότυπο x αντλώντας μη μηδενικά στοιχεία από μια κανονική κατανομή $N(0, \sigma^2)$.

Μέτρηση Απόστασης: Υπολογίζει αποστάσεις μεταξύ των διαταραγμένων δειγμάτων και του αρχικού στιγμιότυπου χρησιμοποιώντας μια συνάρτηση πυρήνα:

$$\pi_x(z) = \exp(-D(x, z)^2 / \sigma^2) \quad (4.1)$$

όπου D είναι μια συνάρτηση απόστασης (συνήθως συνημιτονοειδής απόσταση για κείμενο ή $L2$ για δεδομένα πίνακα).

Τοπική Προσαρμογή Μοντέλου: Εκπαιδεύει ένα σταθμισμένο γραμμικό μοντέλο ελαχιστοποιώντας:

$$\xi(g) = \sum \pi_x(z)(f(z) - g(z))^2 + \Omega(g) \quad (4.2)$$

όπου f είναι το αρχικό μοντέλο, g είναι το ερμηνεύσιμο μοντέλο, και $\Omega(g)$ είναι ποινή πολυπλοκότητας.

4.3.2 SHapley Additive exPlanations (SHAP)

Η μέθοδος SHAP βασίζεται στη θεωρία παιγνίων και συγκεκριμένα στις τιμές Shapley [8]. Παρέχει συνεπείς και ακριβείς εκτιμήσεις της συνεισφοράς κάθε χαρακτηριστικού στην πρόβλεψη του μοντέλου. Η μέθοδος εγγυάται ορισμένες επιθυμητές ιδιότητες, όπως η τοπική ακρίβεια και η δικαιοσύνη [86].

Οι τιμές SHAP υπολογίζονται εξετάζοντας όλους τους πιθανούς συνδυασμούς χαρακτηριστικών και μετρώντας τη μέση οριακή συνεισφορά κάθε χαρακτηριστικού. Αυτή η προσέγγιση εξασφαλίζει ότι η κατανομή της συνεισφοράς είναι δίκαιη και αντικατοπτρίζει την πραγματική επίδραση κάθε χαρακτηριστικού στην πρόβλεψη. Η μέθοδος είναι ιδιαίτερα χρήσιμη σε περιπτώσεις όπου απαιτείται λεπτομερής κατανόηση της σημασίας των διαφόρων παραγόντων που επηρεάζουν την απόφαση του μοντέλου.

Οι τιμές SHAP υπολογίζονται χρησιμοποιώντας το ακόλουθο μαθηματικό πλαίσιο:

Ορισμός Χαρακτηριστικής Συνάρτησης:

$$v(S) = E[f(x)|x_S] - E[f(x)] \quad (4.3)$$

όπου S είναι ένα υποσύνολο χαρακτηριστικών και x_S αντιπροσωπεύει τις τιμές των χαρακτηριστικών στο υποσύνολο S .

Υπολογισμός Τιμής Shapley για το χαρακτηριστικό i :

$$\phi_i = \sum_{S \subseteq N \setminus i} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [v(S \cup i) - v(S)] \quad (4.4)$$

όπου N είναι το σύνολο όλων των χαρακτηριστικών.

4.3.3 Integrated Gradients

Η μέθοδος Integrated Gradients προσφέρει μια προσέγγιση βασισμένη στις κλίσεις για την απόδοση σημαντικότητας στα χαρακτηριστικά εισόδου [87]. Είναι ιδιαίτερα αποτελεσματική σε βαθιά νευρωνικά δίκτυα και παρέχει εξηγήσεις που ικανοποιούν σημαντικές αξιωματικές ιδιότητες [88].

Η τεχνική αυτή υπολογίζει την ολοκλήρωση των κλίσεων κατά μήκος μιας ευθείας γραμμής από ένα σημείο αναφοράς μέχρι την είσοδο. Αυτό επιτρέπει την ακριβή εκτίμηση της συνεισφοράς κάθε χαρακτηριστικού στην τελική πρόβλεψη.

Η μέθοδος υπολογίζει τις αποδόσεις χαρακτηριστικών χρησιμοποιώντας:

Διατύπωση Ολοκληρώματος Διαδρομής:

$$IG_i(x) = (x_i - x'_i) \times \int_0^1 \frac{\partial F}{\partial x_i}(x' + \alpha(x - x')) d\alpha \quad (4.5)$$

όπου x' είναι η βασική είσοδος και α είναι η παράμετρος διαδρομής.

4.3.4 Counterfactual Explanations

Οι αντιπαραδειγματικές εξηγήσεις παρέχουν πληροφορίες για το πώς θα έπρεπε να αλλάξουν τα δεδομένα εισόδου ώστε να επιτευχθεί ένα διαφορετικό αποτέλεσμα [89]. Αυτή η προσέγγιση είναι ιδιαίτερα χρήσιμη στην κλινική πράξη, καθώς παρέχει πρακτική καθοδήγηση για πιθανές παρεμβάσεις [90].

Η δημιουργία αντιπαραδειγματικών εξηγήσεων περιλαμβάνει την αναζήτηση των ελάχιστων αλλαγών που απαιτούνται στα δεδομένα εισόδου για την επίτευξη του επιθυμητού αποτελέσματος. Αυτές οι εξηγήσεις είναι ιδιαίτερα χρήσιμες σε περιπτώσεις όπου απαιτείται η λήψη αποφάσεων βάσει των προβλέψεων του μοντέλου, καθώς παρέχουν συγκεκριμένες οδηγίες για δράση.

Το πρόβλημα βελτιστοποίησης για την εύρεση αντιπαραδειγμάτων διατυπώνεται ως:

Συνάρτηση Στόχου:

$$\min_{x'} \lambda_1 L_1(f(x'), y') + \lambda_2 L_2(x, x') + \lambda_3 L_3(x') \quad (4.6)$$

όπου:

- L_1 : Απώλεια που εξασφαλίζει την επιθυμητή πρόβλεψη
- L_2 : Μετρική απόστασης
- L_3 : Περιορισμός πολλαπλότητας δεδομένων
- λ_i : Παράμετροι στάθμισης

4.3.5 Layer-wise Relevance Propagation

Η μέθοδος Layer-wise Relevance Propagation (LRP) είναι σχεδιασμένη ειδικά για βαθιά νευρωνικά δίκτυα [91]. Λειτουργεί διαδίδοντας την πρόβλεψη προς τα πίσω μέσω του δικτύου και κατανέμοντας τη συνεισφορά σε κάθε χαρακτηριστικό εισόδου. Η μέθοδος είναι ιδιαίτερα αποτελεσματική στην ανάλυση ιατρικών εικόνων και στην αναγνώριση προτύπων [92]. Η LRP χρησιμοποιεί έναν κανόνα διάδοσης που διατηρεί τη συνολική συνεισφορά σε κάθε επίπεδο του δικτύου. Αυτό εξασφαλίζει ότι η συνολική σημαντικότητα που αποδίδεται στα χαρακτηριστικά εισόδου ισούται με την τιμή της πρόβλεψης. Η μέθοδος είναι ιδιαίτερα χρήσιμη για την κατανόηση των εσωτερικών μηχανισμών των βαθιών νευρωνικών δικτύων και την αξιολόγηση της αξιοπιστίας των προβλέψεών τους. Η LRP διαδίδει βαθμολογίες σχετικότητας R_j μέσω του δικτύου χρησιμοποιώντας:

Ιδιότητα Διατήρησης:

$$\sum_j R_j^{(l)} = \sum_i R_i^{(l+1)} \quad (4.7)$$

όπου l υποδηλώνει τον αριθμό στρώματος.

Κανόνας Διάδοσης (ε-κανόνας):

$$R_i = \sum_j \frac{a_i w_{ij}}{\sum_i a_i w_{ij} + \varepsilon} \times R_j \quad (4.8)$$

Στην παρούσα ανάλυση, η επιλογή είναι να επικεντρωθούμε στις μεθόδους LIME (Local Interpretable Model-agnostic Explanations), SHAP (SHapley Additive exPlanations). Αυτές οι τεχνικές είναι ιδιαίτερα κατάλληλες για την ερμηνεία μοντέλων μηχανικής μάθησης, καθώς προσφέρουν model-agnostic προσεγγίσεις που μπορούν να εφαρμοστούν αποτελεσματικά σε οποιοδήποτε black-box μοντέλο. Δεδομένου ότι δεν υλοποιήσαμε νευρωνικά δίκτυα βαθιάς μάθησης, οι δύο τελευταίες προαναφερθείσες τεχνικές εξηγησιμότητας (Layer-Wise Relevance Propagation και Integrated Gradients) θεωρήθηκαν ως μη ταιριαστές.

Βιβλιογραφική ανασκόπηση

5.1 Μεθοδολογία έρευνας

Η βιβλιογραφική ανασκόπηση πραγματοποιήθηκε με συστηματικό τρόπο για την κάλυψη του ερευνητικού πεδίου της εξηγήσιμης τεχνητής νοημοσύνης στη διάγνωση τυχαιωμάτων επινεφριδίων. Η μεθοδολογία ακολούθησε τις κατευθυντήριες γραμμές PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) για τη διασφάλιση της ποιότητας και της αναπαραγωγιμότητας της ανασκόπησης. Η ανασκόπηση επικεντρώθηκε κυρίως σε δημοσιεύσεις της τελευταίας πενταετίας (2019-2024), με έμφαση στις πιο πρόσφατες εξελίξεις στον τομέα της εξηγήσιμης τεχνητής νοημοσύνης. Ωστόσο, συμπεριλήφθηκαν και παλαιότερες σημαντικές δημοσιεύσεις που θεωρούνται θεμελιώδεις για το πεδίο, όπως οι αρχικές δημοσιεύσεις για τις μεθόδους LIME [9] και SHAP [93].

5.1.1 Κριτήρια αναζήτησης

Τα κριτήρια αναζήτησης διαμορφώθηκαν με βάση το ερευνητικό ερώτημα και περιελάμβαναν τους ακόλουθους όρους και συνδυασμούς τους:

Πρωτεύοντες όροι	Δευτερεύοντες όροι
Explainable AI, XAI, Interpretable ML Machine Learning, Deep Learning Clinical Decision Support LIME, SHAP, Integrated Gradients ACC Diagnosis, Adrenal Carcinoma	Adrenal incidentaloma, Adrenal tumor Medical imaging, Diagnosis Healthcare, Medical AI Feature importance, Model interpretation Pheochromocytoma, Adrenal Glands

Πίνακας 5.1: Βασικοί όροι αναζήτησης

5.1.2 Βάσεις δεδομένων

Η αναζήτηση πραγματοποιήθηκε σε πολλαπλές βάσεις δεδομένων για την εξασφάλιση ευρείας κάλυψης της βιβλιογραφίας:

- PubMed Central (PMC) και MEDLINE για ιατρική βιβλιογραφία
- IEEE Xplore για τεχνολογική βιβλιογραφία
- ACM Digital Library για επιστήμη υπολογιστών
- Google Scholar για διεπιστημονική κάλυψη
- arXiv για προδημοσιεύσεις σχετικές με την τεχνητή νοημοσύνη

5.2 Σημαντικά επιστημονικά περιοδικά και συνέδρια στο πεδίο

Η έρευνα στο πεδίο της εξηγήσιμης τεχνητής νοημοσύνης στη βιοϊατρική δημοσιεύεται κυρίως σε κορυφαία επιστημονικά περιοδικά και παρουσιάζεται σε διεθνή συνέδρια υψηλού κύρους. Έγινε προσπάθεια να ενταχθούν πηγές από τα σημαντικότερα περιοδικά καθώς και διεθνή συνέδρια:

- | | |
|--|---|
| • <i>Nature Medicine Intelligence</i> | • <i>Artificial Intelligence in Medicine</i> (Elsevier) |
| • <i>Nature Biomedical Engineering</i> | |
| • <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> | • <i>Journal of Biomedical Informatics</i> (Elsevier) |
| • <i>IEEE Transactions on Medical Imaging</i> | • <i>NPJ Digital Medicine</i> (Nature Partner Journal) |
| • <i>IEEE Transactions on Biomedical Engineering</i> | • <i>JAMA</i> |
| | • <i>The Lancet Digital Health</i> |
| • <i>Medical Image Analysis</i> (Elsevier) | • <i>Radiology: Artificial Intelligence</i> |
| • NeurIPS (Neural Information Processing Systems) | • EMBC (IEEE Engineering in Medicine and Biology Conference) |
| • ICML (International Conference on Machine Learning) | • AIME (Artificial Intelligence in Medicine) |
| • MICCAI (Medical Image Computing and Computer Assisted Intervention) | • MLHC (Machine Learning for Healthcare) |
| • AAAI (Association for the Advancement of Artificial Intelligence) | • SPIE Medical Imaging |
| • CHIL (ACM Conference on Health, Inference, and Learning) | • MIDL (Medical Imaging with Deep Learning) |

5.3 ETN και διάγνωση τυχαιωμάτων

5.3.1 Υπάρχοντα συστήματα

Η εφαρμογή της τεχνητής νοημοσύνης στη διάγνωση τυχαιωμάτων επανεφριδίων έχει σημειώσει σημαντική πρόοδο τα τελευταία χρόνια. Τα υπάρχοντα συστήματα μπορούν να κατηγοριοποιηθούν ως εξής:

Τύπος Συστήματος	Χαρακτηριστικά
Συστήματα Ανάλυσης Εικόνων	Αυτοματοποιημένη ανάλυση CT/MRI για χαρακτηριστικά όγκου
Συστήματα Κατηγοριοποίησης	Διάκριση μεταξύ καλοήθων και κακοήθων όγκων
Συστήματα Πρόγνωσης	Πρόβλεψη εξέλιξης και επικινδυνότητας

Πίνακας 5.2: Κατηγορίες συστημάτων TN για τυχαιώματα

5.3.2 Μέθοδοι εξήγησης

Οι κύριες μέθοδοι εξήγησης που χρησιμοποιούνται στη διάγνωση τυχαιωμάτων περιλαμβάνουν:

Τοπικές Εξηγήσεις (LIME): Παρέχουν εξηγήσεις για μεμονωμένες προβλέψεις, αναδεικνύοντας τα χαρακτηριστικά που επηρέασαν περισσότερο την απόφαση του μοντέλου [9].

SHAP Values: Προσφέρουν συνεπείς και ακριβείς εκτιμήσεις της συνεισφοράς κάθε χαρακτηριστικού στην πρόβλεψη [93].

Layer-wise Relevance Propagation: Ιδιαίτερα χρήσιμη για την ανάλυση απεικονιστικών δεδομένων, επιτρέποντας την οπτικοποίηση των περιοχών που επηρέασαν την απόφαση [91].

5.3.3 Προκλήσεις και περιορισμοί

Οι κύριες προκλήσεις στην εφαρμογή της ETN στη διάγνωση τυχαιωμάτων περιλαμβάνουν:

Ποιότητα Δεδομένων: Η περιορισμένη διαθεσιμότητα μεγάλων, καλά επισημειωμένων συνόλων δεδομένων αποτελεί σημαντικό εμπόδιο στην ανάπτυξη αξιόπιστων μοντέλων [6].

Κλινική Εγκυρότητα: Η ανάγκη για αξιόπιστη επικύρωση των εξηγήσεων από κλινικούς ιατρούς παραμένει κρίσιμη για την αποδοχή των συστημάτων στην κλινική πράξη [4].

Νομικές Απαιτήσεις: Η συμμόρφωση με κανονισμούς όπως ο GDPR που απαιτούν διαφάνεια στη λήψη αποφάσεων επηρεάζει σημαντικά το σχεδιασμό των συστημάτων [53].

Υπολογιστική Πολυπλοκότητα: Η ανάγκη για αποτελεσματική παραγωγή εξηγήσεων σε πραγματικό χρόνο αποτελεί τεχνική πρόκληση που απαιτεί προσεκτική εξισορρόπηση μεταξύ ακρίβειας και ταχύτητας [83].

Διαχείριση Αβεβαιότητας: Η ανάγκη για αξιόπιστη ποσοτικοποίηση και επικοινωνία της αβεβαιότητας στις προβλέψεις του μοντέλου αποτελεί σημαντική πρόκληση, ιδιαίτερα σε περιπτώσεις όπου τα διαθέσιμα δεδομένα είναι ανεπαρκή ή αμφίβολης ποιότητας.

Διαλειτουργικότητα Συστημάτων: Η ενσωμάτωση των συστημάτων ETN στα υπάρχοντα κλινικά πληροφοριακά συστήματα απαιτεί σημαντική προσπάθεια προσαρμογής και τυποποίησης των διεπαφών και των πρωτοκόλλων επικοινωνίας.

Εκπαίδευση Προσωπικού: Η αποτελεσματική χρήση των συστημάτων ETN προϋποθέτει επαρκή εκπαίδευση του ιατρικού και νοσηλευτικού προσωπικού, τόσο στην τεχνική όσο και στην κλινική πτυχή των συστημάτων.

Διαχείριση Ακραίων Περιπτώσεων: Η ικανότητα του συστήματος να αναγνωρίζει και να χειρίζεται κατάλληλα σπάνιες ή ακραίες περιπτώσεις, παρέχοντας αξιόπιστες εξηγήσεις για τις αποφάσεις του, αποτελεί σημαντική πρόκληση για την ασφαλή κλινική εφαρμογή.

Σύνολο δεδομένων, επιλογή μοντέλου

6.1 Προετοιμασία συνόλου δεδομένων

Τα πλήρη και αναλυτικά δεδομένα στα οποία αναπτύχθηκαν διεξοδικά τα ερευνητικά μοντέλα είναι ιατρικής και κλινικής φύσεως και αφορούν συνολικά 262 πραγματικούς ανθρώπους από διάφορες περιοχές. Συγκεκριμένα και αναλυτικά, για κάθε έναν συμμετέχοντα στη μελέτη υπάρχει λεπτομερής καταγραφή ιατρικών δεδομένων τόσο γενικά όσο και ειδικά των όγκων επινεφριδίων σε βάθος χρόνου. Δηλαδή, εντοπίζουμε συστηματική καταγραφή για το εάν πάσχουν από διαβήτη, αν λαμβάνουν συστηματική θεραπεία για αρτηριακή πίεση σε τακτική βάση, καθώς και απώλεια σήματος MRI υψηλής ευκρίνειας, ή σημαντική αλλαγή στη διάμετρο του όγκου το τελευταίο κρίσιμο εξάμηνο πριν την επίσημη καταγραφή. Πέραν αυτών των βασικών στοιχείων, υπάρχουν και άλλα σημαντικά χαρακτηριστικά, όπως ύψος σε εκατοστά, βιολογικό φύλο, σωματικό βάρος, και χρονολογική ηλικία.

Στο αναλυτικό πλαίσιο της παρούσας ερευνητικής εργασίας, αφαιρέθηκαν οι στήλες του AMKA καθώς και η πλήρης ημερομηνία γέννησης και ο μοναδικός αύξων αριθμός ασθενούς, διότι θεωρήθηκαν εύλογα μη σχετιζόμενες πληροφορίες με τη μελέτη. Ενδεχομένως, να υπάρχει μια μικρή στατιστική συσχέτιση με τον ακριβή χρόνο γέννησης, σε περίπτωση που ο συμμετέχων έζησε γεωγραφικά κοντά σε κάποιο σημαντικό γεγονός με ιστορικό χαρακτήρα που επηρεάζει την εμφάνιση καρκίνων διαφόρων τύπων (π.χ. τοπική διαρροή ραδιενέργειας). Βλέπουμε αναλυτικά πως τα συλλεχθέντα δεδομένα είναι επιστημονικά επαρκή και αφορούν πολλαπλές σημαντικές κατηγορίες, από δημογραφικά στοιχεία, σωματικά χαρακτηριστικά, εκτενείς εξετάσεις ούρων/αίματος, πλήρες λιπιδαιμικό προφίλ έως και συγκεκριμένα διαγνωστικά στοιχεία αναφορικά με τον όγκο στα επινεφρίδια. Θεωρούνται κατ' αυτήν την επιστημονική έννοια ολοκληρωμένα και πλήρη, παρά το σχετικά μικρό συνολικό αριθμό δεδομένων. Παρακάτω στον αναλυτικό συγκεντρωτικό πίνακα φαίνονται λεπτομερώς τα διαφορετικά είδη των καταγραφών ανά συμμετέχοντα στην έρευνα:

Κατηγορία	Παράμετρος - Μεταβλητή
Δημογραφικά	Ηλικία, Φύλο
Διαγνωστικά	Λειτουργικότητα, Διάγνωση, Ταξινόμηση Διάγνωσης, Ιστολογική Επιβεβαίωση, Ηλικία Διάγνωσης
Σωματικά Χαρακτηριστικά	Ύψος, Βάρος, Δείκτης Μάζας Σώματος (BMI)
Αναπαραγωγική Υγεία	Εμμηνόπαυση, Ηλικία Εμμηνόπαυσης
Εξαρτήσεις	Κάπνισμα
Καρδιακή Υγεία	Αρτηριακή Πίεση, Θεραπεία Αρτηριακής Πίεσης
Μεταβολική Υγεία	Υποκαλιαιμία, Υπερλιπιδαιμία, Αγωγή για Χοληστερόλη, Σακχαρώδης Διαβήτης, Θεραπεία Σακχαρώδους Διαβήτη
Εξετάσεις Ούρων	(Nor)-μετανεφρίνη
Υγεία Οστών	Οστεοπόρωση
Οικογενειακό Ιστορικό	Ύπαρξη Πρότερου Συγγενούς με Σακχαρώδη Διαβήτη
Χαρακτηριστικά Όγκου	Τοποθεσία, Μέγιστη Διατομή
Ενδοκρινολογικά	Ελεύθερη Κορτιζόλη (πρωί και βράδυ), Αδρενοκορτικοτρόπος Ορμόνη, Δραστητικότητα Ρενίνης Πλάσματος, Άμεση Μέτρηση Ρενίνης Πλάσματος, Αλδοστερόνη, Δ-Αλδοστερόνη, Θεική Δεϋδροεπιανδροστερόνη, 17-Υδροξυπρογεστερόνη, Τεστοστερόνη, Ελεύθερη Κορτιζόλη Ούρων, Overnight fasting cortisol
Διαγνωστικά	Δοκιμασία Καταστολής με Δεξαμεθαζόνη
Μεταβολισμός Γλυκόζης	Σάκχαρο Νηστείας, Αιμογλοβίνη A1c, Δοκιμασία ανοχής γλυκόζης από το στόμα (2 ωρών)
Προφίλ λιπιδίων	Χοληστερόλη, HDL Χοληστερόλη, LDL Χοληστερόλη, Τριγλυκερίδια
Οστική Πυκνότητα	T-σκορ σε μηριαίο οστό, T-σκορ σε συνολική μέση οστική πυκνότητα
Απεικονιστικά	Μονάδες Hounsfield (HU) σε Εικόνες Υπολογιστικής Τομογραφίας, Έκπλυση Σκιαγραφικού στην Αξονική Τομογραφία, Εξασθένηση Σήματος MRI
Εξέλιξη Όγκου	Αύξηση διαμέτρου

Πίνακας 6.1: Ομαδοποίηση δεδομένων (χαρακτηριστικών) ασθενών

Για τον καθαρισμό του συνόλου δεδομένων, αφαιρέθηκαν 4 συμμετέχοντες για τους οποίους δεν υπήρχε ιατρική εκτίμηση για τον τύπο όγκου. Αναφορικά με την έλλειψη δεδομένων, 3455 κελιά εκ του συνόλου δεν είχαν τιμή, είτε συνεχόμενη, είτε διακριτή για κάποια κατηγορία. Αυτό είναι της τάξεως του 25%. Στις συνεχόμενες τιμές, συμπληρώθηκε με την μέση τιμή ως προς την ίδια κατηγοριοποίηση του όγκου, για να αποφευχθούν συμπληρωματικοί με διαφορετικούς τύπους (π.χ. καρκίνωμα, μη παραγωγικό όγκο). Στις κατηγορηματικές τιμές, η έννοια της μέσης τιμής θα αποτύγχανε, οπότε συμπληρώθηκε πάλι ανά ομάδα, η συχνότερη τιμή.

6.2 Ανάπτυξη και Αξιολόγηση Μοντέλων

6.2.1 Προεπεξεργασία δεδομένων και μετρικές αξιολόγησης

Πριν από την εκπαίδευση των μοντέλων, εφαρμόστηκε τυποποίηση των χαρακτηριστικών χρησιμοποιώντας τον StandardScaler σύμφωνα με τον τύπο:

$$z = \frac{x - \mu}{\sigma} \quad (6.1)$$

όπου μ αντιπροσωπεύει τη μέση τιμή και σ την τυπική απόκλιση κάθε χαρακτηριστικού.

6.2.2 Θεμελιώδεις μετρικές ταξινόμησης

Στο πλαίσιο της ταξινόμησης όγκων επινεφριδίων, χρησιμοποιήθηκαν οι ακόλουθες μετρικές αξιολόγησης:

- **Ακρίβεια (Accuracy):** Ο λόγος των σωστών προβλέψεων προς το σύνολο των προβλέψεων:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6.2)$$

- **Ευαισθησία (Sensitivity):** Η ικανότητα του μοντέλου να αναγνωρίζει σωστά τα θετικά δείγματα:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (6.3)$$

- **Ειδικότητα (Specificity):** Η ικανότητα του μοντέλου να αναγνωρίζει σωστά τα αρνητικά δείγματα:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (6.4)$$

- **Θετική Προγνώστική Αξία (PPV):** Η πιθανότητα μια θετική πρόβλεψη να είναι σωστή:

$$\text{PPV} = \frac{TP}{TP + FP} \quad (6.5)$$

- **Αρνητική Προγνωστική Αξία (NPV):** Η πιθανότητα μια αρνητική πρόβλεψη να είναι σωστή:

$$NPV = \frac{TN}{TN + FN} \quad (6.6)$$

- **Διασταυρούμενη Επικύρωση (Cross-validation Score):** Μετρική που υπολογίζεται διαχωρίζοντας το σύνολο δεδομένων σε k υποσύνολα:

$$CV \text{ Score} = \frac{1}{k} \sum_{i=1}^k \text{Accuracy}_i \quad (6.7)$$

όπου:

- TP (True Positives): Σωστά ταξινομημένα θετικά δείγματα
- TN (True Negatives): Σωστά ταξινομημένα αρνητικά δείγματα
- FP (False Positives): Λανθασμένα ταξινομημένα ως θετικά δείγματα
- FN (False Negatives): Λανθασμένα ταξινομημένα ως αρνητικά δείγματα

6.2.3 Μοντέλα Μηχανικής Μάθησης και Υπερπαράμετροι

Στον Πίνακα 6.2 παρουσιάζονται τα μοντέλα μηχανικής μάθησης που εκπαιδεύτηκαν:

Μοντέλο	Περιγραφή
Logistic Regression	Γραμμικός ταξινομητής με ρυθμιζόμενη κανονικοποίηση
KNN	Ταξινομητής k -πλησιέστερων γειτόνων
Random Forest	Σύνθετος ταξινομητής τυχαίου δάσους
Decision Tree	Απλός ταξινομητής δέντρου απόφασης
Bagging	Ταξινομητής συνόλου με τεχνική bagging
AdaBoost	Προσαρμοστικός ταξινομητής ενίσχυσης
XGBoost	Ταξινομητής ακραίας κλίσης gradient boosting
SVM	Μηχανή διανυσμάτων υποστήριξης
Voting Classifier	Σύνθετος ταξινομητής με σταθμισμένη ψηφοφορία

Πίνακας 6.2: Μοντέλα μηχανικής μάθησης που αξιολογήθηκαν

Οι υπερπαράμετροι που επελέγησαν για κάθε μοντέλο παρουσιάζονται συγκεντρωτικά στον Πίνακα 6.3:

Μοντέλο	Υπερπαράμετρος	Επιλεγμένες τιμές
Logistic Regression	Συντελεστής κανονικοποίησης (C)	0.0005
	Ποινή (penalty)	elasticnet
	Αναλογία L1 (l1_ratio)	0.03
	Επιλυτής (solver)	saga
	Μέγιστες επαναλήψεις (max_iter)	100
KNN	Αριθμός γειτόνων (n_neighbors)	12
	Βάρη (weights)	distance
	Μετρική απόστασης (metric)	minkowski
	Παράμετρος Minkowski (p)	3
Random Forest	Αριθμός δέντρων (n_estimators)	50
	Μέγιστο βάθος (max_depth)	1
	Ελάχιστα δείγματα διαχωρισμού (min_samples_split)	3
	Ελάχιστα δείγματα φύλλου (min_samples_leaf)	3
	Μέγιστα χαρακτηριστικά (max_features)	sqrt
	Κριτήριο (criterion)	gini
Decision Tree	Τυχαία κατάσταση (random_state)	42
	Μέγιστο βάθος (max_depth)	8
	Ελάχιστα δείγματα διαχωρισμού (min_samples_split)	5
	Ελάχιστα δείγματα φύλλου (min_samples_leaf)	2
	Κριτήριο (criterion)	entropy
Bagging	Τυχαία κατάσταση (random_state)	42
	Αριθμός εκτιμητών (n_estimators)	80
	Μέγιστα δείγματα (max_samples)	0.98
	Μέγιστα χαρακτηριστικά (max_features)	0.78
AdaBoost	Τυχαία κατάσταση (random_state)	42
	Αριθμός εκτιμητών (n_estimators)	50
	Ρυθμός μάθησης (learning_rate)	0.01
	Αλγόριθμος (algorithm)	SAMME
XGBoost	Τυχαία κατάσταση (random_state)	42
	Αριθμός εκτιμητών (n_estimators)	40
	Μέγιστο βάθος (max_depth)	2
	Ρυθμός μάθησης (learning_rate)	0.01
	Υποδειγματοληψία (subsample)	0.24
	Δειγματοληψία στηλών (colsample_bytree)	0.48
	Ελάχιστο βάρος παιδιού (min_child_weight)	2
	Αντικειμενική συνάρτηση (objective)	multi:softprob
SVM	Αριθμός κλάσεων (num_class)	len(classes)
	Τυχαία κατάσταση (random_state)	42
	Συντελεστής κανονικοποίησης (C)	0.80
	Πυρήνας (kernel)	sigmoid
	Γάμμα (gamma)	auto
	Σχήμα συνάρτησης απόφασης (decision_function_shape)	ovo
	Τυχαία κατάσταση (random_state)	42

Πίνακας 6.3: Υπερπαράμετροι που επελέγησαν για κάθε μοντέλο

6.2.4 Αποτελέσματα Αξιολόγησης

Η συγκριτική ανάλυση των εννέα μοντέλων μηχανικής μάθησης αποκάλυψε σημαντικές διαφορές στην απόδοση και τη σταθερότητά τους. Τα μοντέλα αξιολογήθηκαν χρησιμοποιώντας πολλαπλές μετρικές, με ιδιαίτερη έμφαση στη διασταυρούμενη επικύρωση και τη σταθερότητα των προβλέψεων.

Classifier	Prec.	Recall	Spec.	NPV	F1	Acc.	AUC	CV
XGBoost	0.952	0.992	0.993	0.970	0.970	0.980	0.998	0.989
Bagging	0.952	0.992	0.993	0.970	0.970	0.980	0.999	0.980
SVM	0.917	0.984	0.985	0.944	0.944	0.961	0.994	0.996
Decision Tree	0.936	0.936	0.959	0.959	0.936	0.961	0.948	0.982
Log. Regression	0.886	0.984	0.986	0.944	0.929	0.961	0.999	0.978
Random Forest	0.889	0.936	0.986	0.962	0.899	0.961	0.947	0.976
AdaBoost	0.878	0.936	0.986	0.962	0.903	0.961	0.944	0.980
KNN	0.825	0.920	0.945	0.909	0.861	0.922	0.981	0.987

Πίνακας 6.4: Επίδοση Μοντέλων με τεχνική SMOTE

6.2.4.1 Συγκριτική Ανάλυση Επιδόσεων

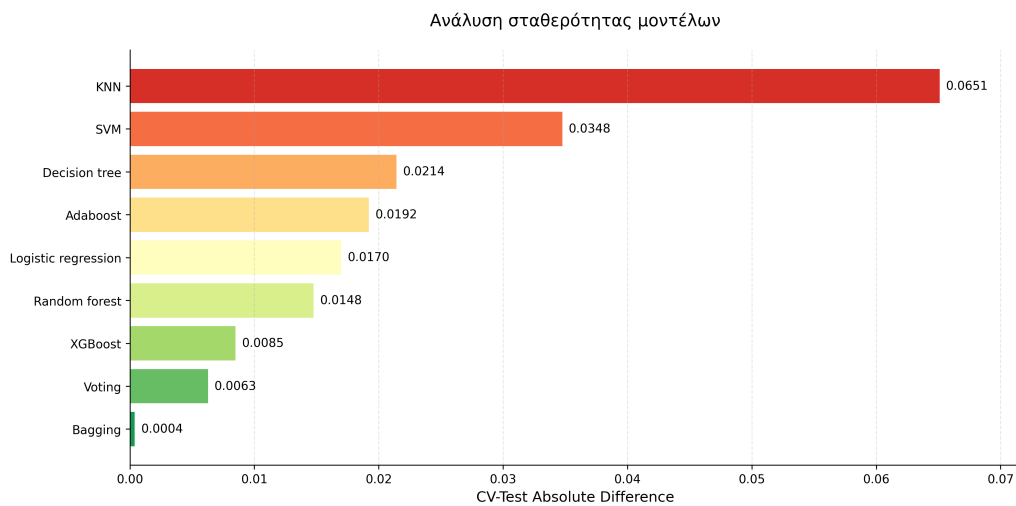
Η ανάλυση των επιδόσεων των μοντέλων (Πίνακας 6.4) αποκαλύπτει μια σαφή διαφοροποίηση μεταξύ των διαφόρων προσεγγίσεων. Το XGBoost και το Bagging επιτυγχάνουν τις υψηλότερες επιδόσεις, με σχεδόν πανομοιότυπες μετρικές στις περισσότερες κατηγορίες (Precision: 0.952, Recall: 0.992, Specificity: 0.993, NPV: 0.970, F1: 0.970, Accuracy: 0.980). Το XGBoost ξεχωρίζει ελαφρώς με υψηλότερη βαθμολογία διασταυρούμενης επικύρωσης (CV Score: 0.989 έναντι 0.980 του Bagging), υποδεικνύοντας καλύτερη γενίκευση σε άγνωστα δεδομένα.

Αξιοσημείωτο είναι ότι τα ensemble μοντέλα (XGBoost, Bagging, SVM) γενικά επιδεικνύουν καλύτερη απόδοση σε σύγκριση με τα μεμονωμένα μοντέλα, επιβεβαιώνοντας τη σημασία του συνδυασμού πολλαπλών ταξινομητών για τη βελτίωση της ακρίβειας πρόβλεψης. Η υπεροχή των ensemble μεθόδων, και ιδιαίτερα του XGBoost, μπορεί να αποδοθεί στην προηγμένη διαχείριση της μεροληψίας-διακύμανσης και στην αποτελεσματική αντιμετώπιση της υπερπροσαρμογής μέσω τεχνικών όπως η στοχαστική κλίση και η L1/L2 κανονικοποίηση.

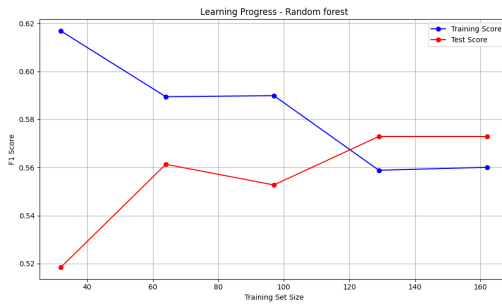
Επιπλέον, η εξαιρετική απόδοση στις μετρικές Specificity (0.993) και NPV (0.970) υποδηλώνει την ιδιαίτερη ικανότητα του μοντέλου στην αναγνώριση και των τριών κατηγοριών ψευδοτυχαιωμάτων, με ελάχιστα ψευδώς θετικά αποτελέσματα. Η υψηλή τιμή AUC (0.998) επιβεβαιώνει περαιτέρω την ισχυρή διακριτική ικανότητα του μοντέλου μεταξύ των διαφορετικών κλάσεων.



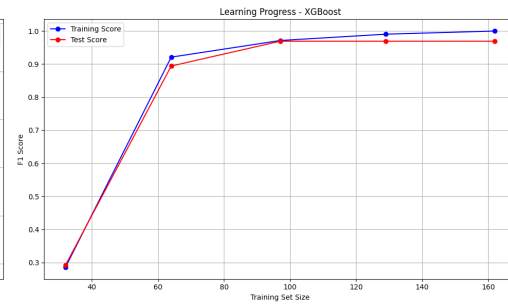
Σχήμα 6.1: Πίνακες σύγχυσης για τα εννέα μοντέλα μηχανικής μάθησης που εξετάστηκαν



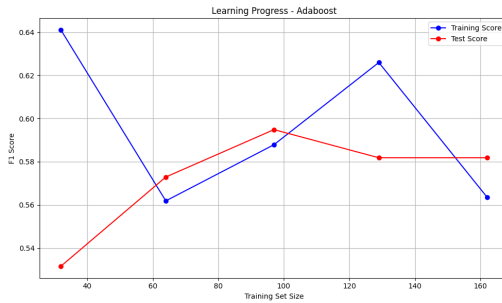
Σχήμα 6.2: Ανάλυση σταθερότητας μοντέλων



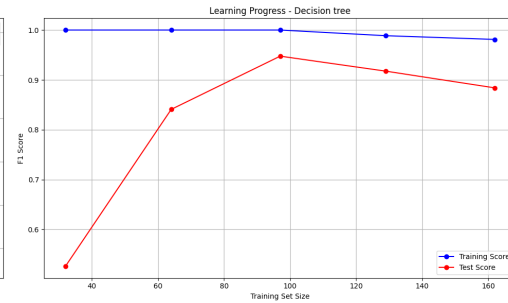
Random Forest



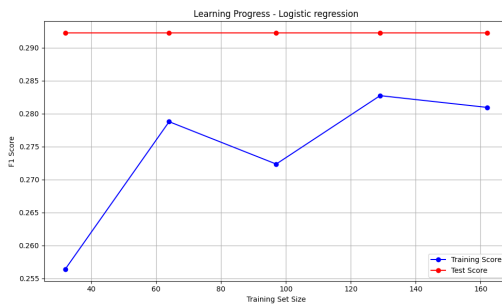
XGBoost



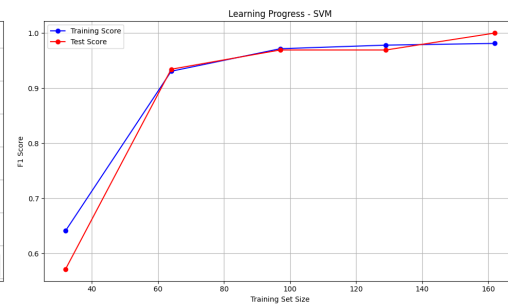
AdaBoost



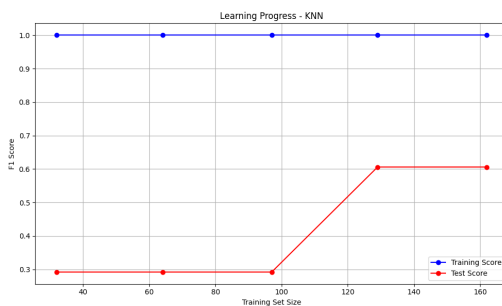
Decision Tree



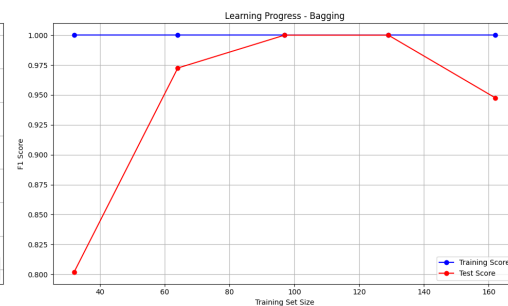
Logistic Regression



SVM



KNN



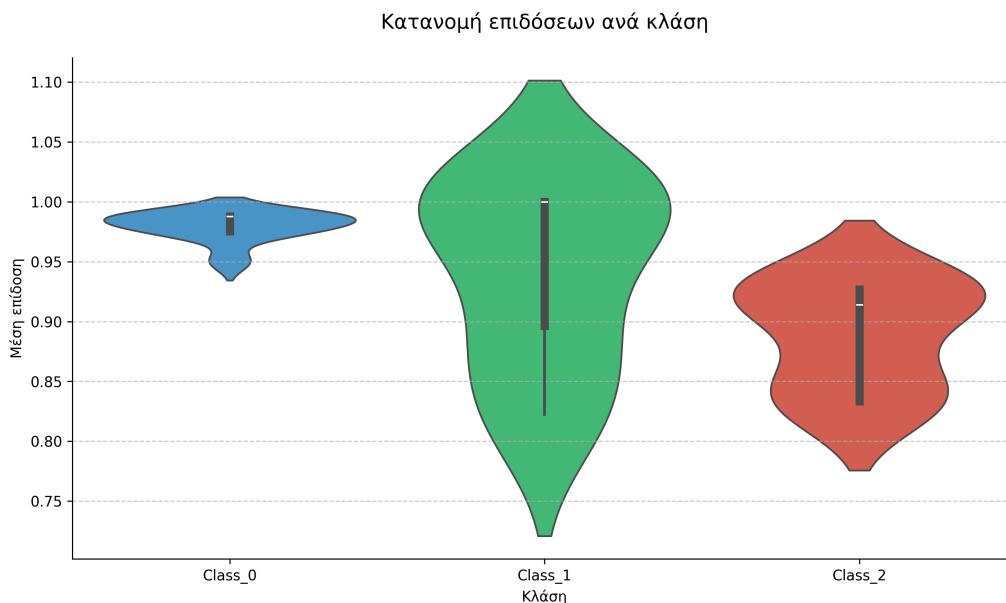
Bagging

Σχήμα 6.3: Καμπύλες απόδοσης

6.2.4.2 Ανάλυση Σταθερότητας

Η σταθερότητα των μοντέλων αξιολογήθηκε μέσω της σύγκρισης των επιδόσεων μεταξύ της διασταυρούμενης επικύρωσης και του συνόλου ελέγχου (Εικόνα 6.2). Παρατηρούνται τα εξής:

- Το μοντέλο Bagging επιδεικνύει εξαιρετική σταθερότητα με ελάχιστη διαφορά (0.0004) μεταξύ CV και test scores, υποδεικνύοντας αξιοσημείωτη ικανότητα γενίκευσης.
- Τα ensemble μοντέλα (Bagging, Voting, XGBoost) καταλαμβάνουν τις τρεις πρώτες θέσεις στη σταθερότητα (0.0004, 0.0063 και 0.0085 αντίστοιχα), επιβεβαιώνοντας το θεωρητικό πλεονέκτημα των συνδυαστικών προσεγγίσεων στη μείωση της διακύμανσης.
- Τα μοντέλα KNN και SVM εμφανίζουν τη μεγαλύτερη διακύμανση (0.0651 και 0.0348 αντίστοιχα), γεγονός που υποδηλώνει σημαντική ευαισθησία στις διαφοροποιήσεις μεταξύ των συνόλων εκπαίδευσης και ελέγχου.



Σχήμα 6.4: Κατανομή επιδόσεων ανά κλάση

6.2.4.3 Ανάλυση Προγνωστικών Αξιών

Η ανάλυση των θετικών και αρνητικών προγνωστικών αξιών παρέχει σημαντική πληροφορία για την αξιοπιστία των προβλέψεων κάθε μοντέλου. Οι τιμές αυτές είναι ιδιαίτερα σημαντικές στο κλινικό πλαίσιο, καθώς υποδεικνύουν την αξιοπιστία των θετικών και αρνητικών προβλέψεων. Από την ανάλυση των προγνωστικών αξιών και μετρικών απόδοσης (Πίνακας 6.5) προκύπτουν τα εξής σημαντικά συμπεράσματα:

Μοντέλο	Precision	Recall	Specificity	NPV
Voting	0.952	0.992	0.993	0.970
Bagging	0.952	0.992	0.993	0.970
XGBoost	0.952	0.992	0.993	0.970
SVM	0.917	0.984	0.985	0.944
Decision Tree	0.936	0.936	0.959	0.959
Random Forest	0.889	0.936	0.986	0.962
AdaBoost	0.878	0.936	0.986	0.962
Logistic Regression	0.886	0.984	0.986	0.944
KNN	0.825	0.920	0.945	0.909

Πίνακας 6.5: Προγνωστικές Αξίες και Μετρικές Απόδοσης ανά Μοντέλο

- **Ισορροπία Μετρικών:**
 - Οι αλγόριθμοι Voting, Bagging και XGBoost επιτυγχάνουν την καλύτερη ισορροπία μεταξύ όλων των μετρικών, με Precision 0.952, Recall 0.992, και εξαιρετική Specificity 0.993
 - Το NPV 0.970 για αυτά τα μοντέλα υποδεικνύει υψηλή αξιοπιστία στην αναγνώριση αρνητικών περιπτώσεων
- **Ειδικότητα και Ευαισθησία:**
 - Παρατηρείται εξαιρετικά υψηλή ειδικότητα (Specificity > 0.98) στα περισσότερα μοντέλα
 - Ο αλγόριθμος SVM επιτυγχάνει πολύ καλή ισορροπία μεταξύ Recall (0.984) και Specificity (0.985)
- **Προγνωστικές Αξίες:**
 - Τα κορυφαία μοντέλα (Voting, Bagging, XGBoost) παρουσιάζουν εξαιρετική ισορροπία μεταξύ Precision (0.952) και NPV (0.970)
 - Ο αλγόριθμος KNN εμφανίζει τις χαμηλότερες τιμές σε όλες τις μετρικές, υποδεικνύοντας μειωμένη αξιοπιστία συγκριτικά με τα άλλα μοντέλα

6.2.4.4 Ανάλυση Επιδόσεων ανά Κλάση

Η κατανομή των επιδόσεων ανά κλάση (Εικόνα 6.4) αποκαλύπτει σημαντικές διαφοροποιήσεις στην ικανότητα των μοντέλων να αναγνωρίζουν διαφορετικούς τύπους όγκων. Συγκεκριμένα:

- **Καλοήθεις Όγκοι (Class 0):** Παρατηρείται σταθερή και υψηλή απόδοση με μέση επίδοση περίπου 0.98, με μικρή διακύμανση μεταξύ των μοντέλων, υποδεικνύοντας αξιόπιστη και συνεπή αναγνώριση των καλοήθων περιπτώσεων.

- **Ύποπτοι Όγκοι (Class 1):** Εμφανίζεται η υψηλότερη μέση επίδοση με εξαιρετικές τιμές Precision και Recall, υποδηλώνοντας ιδιαίτερα αποτελεσματική αναγνώριση των ύποπτων περιπτώσεων παρά τη σχετική ανισορροπία του αρχικού συνόλου δεδομένων (20 περιπτώσεις πριν το SMOTE).
- **Κακοήθεις Όγκοι (Class 2):** Παρουσιάζεται η χαμηλότερη μέση επίδοση (περίπου 0.92) με σημαντική διακύμανση, γεγονός που πιθανώς οφείλεται στην αρχική ανισορροπία των δεδομένων (33 περιπτώσεις) και στην πολυπλοκότητα αναγνώρισης των κακοήθων χαρακτηριστικών.

Παρατηρώντας τα πρότυπα μάθησης (Εικόνα 6.3) των ξεχωριστών μοντέλων, βλέπουμε διαφορετικά μοτίβα εκπαίδευσης σε κάθε αλγόριθμο. Τα μοντέλα SVM και XGBoost επιδεικνύουν ιδιαίτερα σταθερές καμπύλες μάθησης, με τις επιδόσεις εκπαίδευσης και δοκιμής να συγκλίνουν ομαλά και να διατηρούν υψηλή επίδοση (πάνω από 0.9 F1 score) αφού φτάσουν σε επαρκή δεδομένα εκπαίδευσης. Αυτό υποδεικνύει ότι αυτά τα μοντέλα είναι κατάλληλα για το πρόβλημα. Αντιθέτως, τα μοντέλα Random Forest και Adaboost επιδεικνύουν πιο ασταθή συμπεριφορά, με αξιοσημείωτα κενά μεταξύ των επιδόσεων εκπαίδευσης και δοκιμής, που υποδεικνύει πιθανά ζητήματα υπερπροσαρμογής. Το Random Forest συγκεκριμένα δείχνει μειωμένη επίδοση με περισσότερα δεδομένα, υποδεικνύοντας ότι θα μπορούσε να ωφεληθεί από βελτίωση παραμέτρων.

Τα μοντέλα Logistic Regression και KNN παρουσιάζουν ενδιαφέροντα χαρακτηριστικά στα αντίθετα άκρα. Το Logistic Regression διατηρεί σταθερά υψηλές επιδόσεις δοκιμής κατά τη διάρκεια της διαδικασίας εκπαίδευσης, υποδεικνύοντας ότι συλλαμβάνει καλά τα υποκείμενα μοτίβα παρά την απλότητά του. Όμως, η επίδοση εκπαίδευσής του κυμαίνεται, δείχνοντας κάποια αστάθεια στη διαδικασία μάθησης. Το μοντέλο KNN, ενώ αρχικά δείχνει χαμηλή επίδοση, παρουσιάζει μια απότομη βελτίωση σε ένα συγκεκριμένο όριο δεδομένων, μετά από το οποίο διατηρεί σταθερή επίδοση. Το Decision Tree δείχνει καλή σύγκλιση μεταξύ επιδόσεων εκπαίδευσης και δοκιμής αλλά με ελαφρώς χαμηλότερη συνολική επίδοση σε σύγκριση με τα πιο εξελιγμένα μοντέλα όπως το SVM και το XGBoost, κάτι που είναι τυπικό για απλούστερα δεντρικά μοντέλα.

6.3 Περιορισμοί της Μελέτης

- Το περιορισμένο μέγεθος του συνόλου δεδομένων (262 περιπτώσεις) θέτει προκλήσεις στη γενίκευση των αποτελεσμάτων και υποδεικνύει την ανάγκη για εκτενέστερη επικύρωση σε μεγαλύτερους πληθυσμούς
- Η σημαντική έλλειψη δεδομένων (25%) αντιμετωπίστηκε με στρατηγική συμπλήρωση ανά κατηγορία όγκου, ωστόσο αυτή η προσέγγιση ενδέχεται να εισάγει συστηματικά σφάλματα στην ανάλυση

6.4 Μελλοντικές Προοπτικές

- Η εξαιρετική απόδοση των ensemble μοντέλων υποδεικνύει μια πολλά υποσχόμενη κατεύθυνση για μελλοντική έρευνα, με έμφαση στη βελτιστοποίηση των υπερπαραμέτρων και την ανάπτυξη πιο εξελιγμένων συνδυαστικών τεχνικών
- Η ενσωμάτωση προηγμένων τεχνικών επεξεργασίας εικόνας και η συσχέτισή τους με τα κλινικά δεδομένα θα μπορούσε να βελτιώσει περαιτέρω την ακρίβεια του μοντέλου
- Μελλοντική έρευνα θα μπορούσε να επικεντρωθεί στην ανάπτυξη εξειδικευμένων μοντέλων για συγκεκριμένους υποτύπους τυχαιωμάτων επινεφριδίων

Η συγκριτική ανάλυση των εννέα μοντέλων μηχανικής μάθησης ανέδειξε τα μοντέλα Voting, Bagging και XGBoost ως τα πιο αποτελεσματικά, επιτυγχάνοντας τις υψηλότερες επιδόσεις με ταυτόσημες τιμές στις βασικές μετρικές (Precision: 0.952, Recall: 0.992, Specificity: 0.993, NPV: 0.970). Τα μοντέλα αυτά επέδειξαν εξαιρετική ισορροπία μεταξύ ακρίβειας και ανάκλησης, υποδεικνύοντας την υψηλή τους αποτελεσματικότητα στην ταξινόμηση τόσο των θετικών όσο και των αρνητικών περιπτώσεων.

Αξιοσημείωτη είναι η σταθερότητα των προαναφερθέντων μοντέλων, όπως φαίνεται από τα διαγράμματα σύγχυσης και την ανάλυση σταθερότητας. Ιδιαίτερα το XGBoost παρουσίασε εξαιρετική απόδοση στις επιπρόσθετες μετρικές αξιολόγησης, συμπεριλαμβανομένων των F1 (0.970), Accuracy (0.980), AUC (0.998) και CV (0.989), καταδεικνύοντας την υψηλή του αξιοπιστία και προσαρμοστικότητα στα δεδομένα.

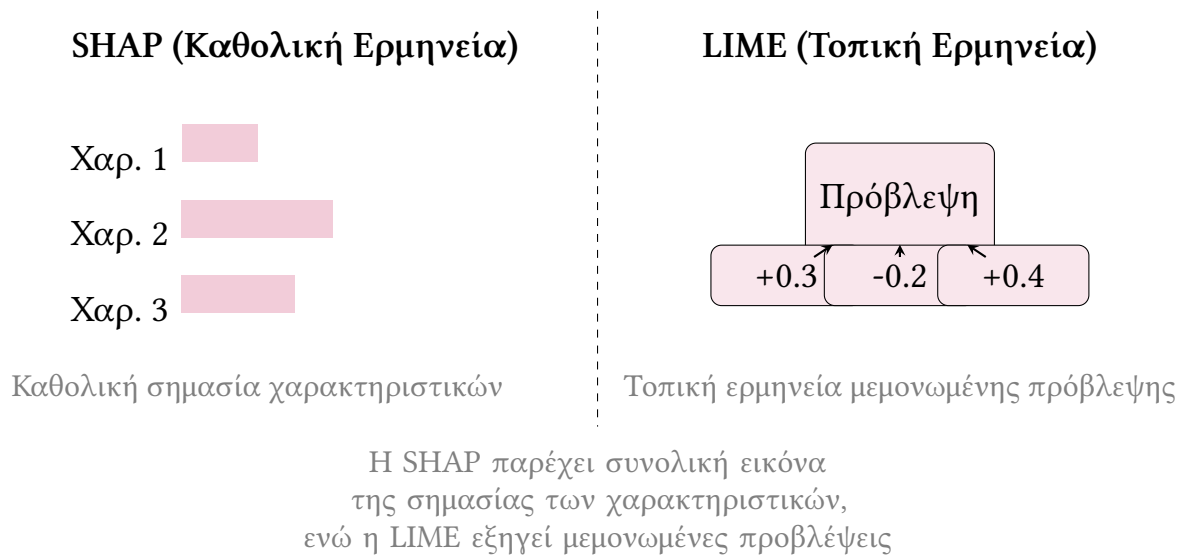
Η υπεροχή των ensemble μεθόδων (Voting, Bagging, XGBoost) έναντι των απλούστερων μοντέλων όπως το KNN και η Λογιστική Παλινδρόμηση είναι εμφανής, με τη διαφορά στην απόδοση να είναι ιδιαίτερα αισθητή στις μετρικές ακρίβειας και ανάκλησης. Τα violin plots επιβεβαιώνουν επίσης την υψηλότερη σταθερότητα των κορυφαίων μοντέλων στις διαφορετικές κλάσεις, υποδεικνύοντας την καταλληλότητά τους για πρακτική εφαρμογή.

Αξίζει να σημειωθεί ότι η τεχνική SMOTE που εφαρμόστηκε για την αντιμετώπιση της ανισορροπίας των κλάσεων συνέβαλε σημαντικά στην βελτιστοποίηση της απόδοσης όλων των μοντέλων, με ιδιαίτερη έμφαση στα ensemble μοντέλα. Η επιτυχής εφαρμογή της τεχνικής αυτής αντικατοπτρίζεται στις υψηλές τιμές των μετρικών Specificity και NPV, υποδεικνύοντας την αποτελεσματική διαχείριση τόσο της πλειοψηφούσας όσο και της μειοψηφούσας κλάσης, ενώ παράλληλα διατηρήθηκε η υψηλή ακρίβεια πρόβλεψης.

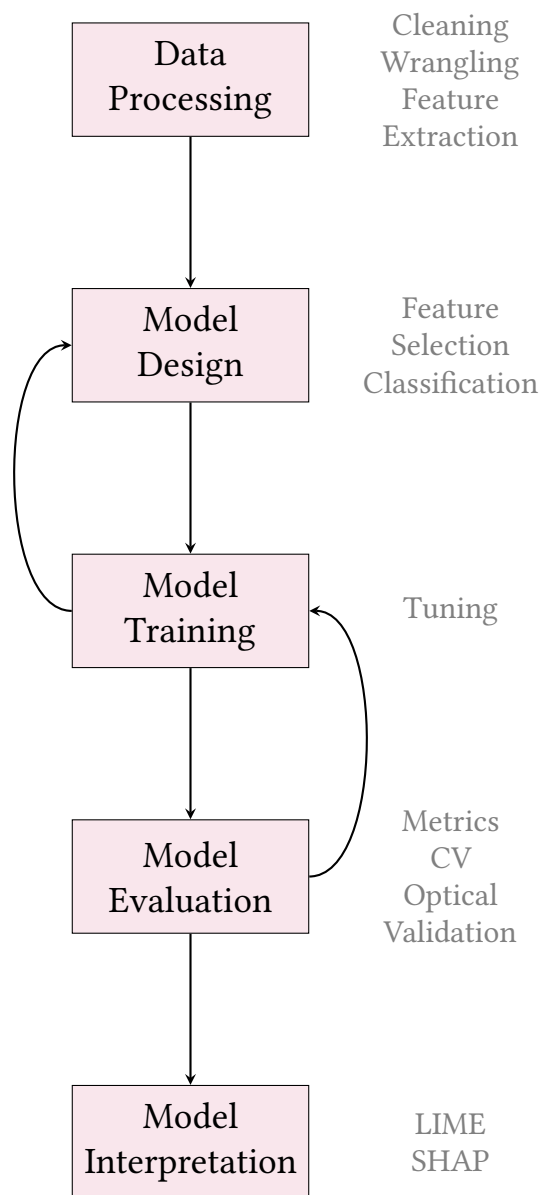
Επεξήγηση μοντέλου επιλογής

7.1 Μέθοδοι SHAP και LIME

Στο τελευταίο στάδιο της διαδικασίας μηχανικής μάθησης (Εικόνα 7.2), επικεντρωνόμαστε στην ερμηνεία του μοντέλου, η οποία είναι ζωτικής σημασίας για την κλινική εφαρμογή. Η χρήση των μεθόδων SHAP (SHapley Additive exPlanations) και LIME (Local Interpretable Model-agnostic Explanations) παρέχει ένα ισχυρό πλαίσιο για την κατανόηση των προβλέψεων του μοντέλου. Αυτές οι μέθοδοι έχουν ιδιαίτερη κλινική σημασία, καθώς επιτρέπουν στους ιατρούς να κατανοήσουν τον τρόπο με τον οποίο το μοντέλο καταλήγει σε συγκεκριμένες προβλέψεις. Η SHAP αναλύει τη συνεισφορά κάθε χαρακτηριστικού στην τελική πρόβλεψη, ενώ η LIME δημιουργεί τοπικά ερμηνεύσιμα μοντέλα για κάθε πρόβλεψη. Αυτή η διπλή προσέγγιση στην ερμηνευσιμότητα είναι κρίσιμη για την εφαρμογή της τεχνητής νοημοσύνης στην κλινική πράξη, καθώς παρέχει την απαραίτητη διαφάνεια και αξιοπιστία που απαιτείται στον τομέα της υγείας.



Σχήμα 7.1: Σύγκριση μεθόδων ερμηνείας μοντέλου: SHAP και LIME



Σχήμα 7.2: Διαδικασία ανάπτυξης μοντέλου μηχανικής μάθησης

Πιο συγκεκριμένα, εμπνευσμένοι από το παράδειγμα στη δημοσίευση [94], παραθέτουμε το δικό μας: Η ανάλυση SHAP βασίζεται στις τιμές Shapley, μια έννοια από τη συνεργατική θεωρία παιγνίων. Οι τιμές Shapley παρέχουν μια δίκαιη κατανομή της απόδοσης μεταξύ των παικτών σε ένα συνεργατικό παίγνιο όπου οι παίκτες συνεργάζονται για έναν κοινό στόχο, ακόμη και αν οι παίκτες μπορεί να έχουν συνεισφέρει άνισα. Θα χρησιμοποιήσουμε το παράδειγμα της ιατρικής διάγνωσης για να απεικονίσουμε τις έννοιες. Ας υποθέσουμε ότι τρία συμπτώματα (Πυρετός, Βήχας και Κόπωση), όταν είναι όλα παρόντα μαζί, υποδεικνύουν πιθανότητα 90% για διάγνωση γρίπης. Οι μεμονωμένες πιθανότητες των συμπτωμάτων είναι:

- Πυρετός 40%
- Βήχας 45%
- Κόπωση 35%

Ας υποθέσουμε επίσης ότι γνωρίζουμε τις πιθανότητες των συνδυασμών δύο συμπτωμάτων:

- Πυρετός & Βήχας: 65%
- Πυρετός & Κόπωση: 55%
- Βήχας & Κόπωση: 60%

Δεδομένων αυτών των πιθανοτήτων, θέτουμε το ερώτημα: ποια θα ήταν η δίκαιη ατομική συνεισφορά κάθε συμπτώματος στην πιθανότητα 90% όταν είναι όλα παρόντα;

Ο Lloyd Shapley [95] εξήγαγε μια μαθηματική φόρμουλα των τιμών ϕ_j που ικανοποιεί ένα πλήθος ιδιοτήτων και απέδειξε ότι αυτές οι τιμές είναι οι μόνες ικανές να ικανοποιήσουν όλες αυτές τις ιδιότητες. Αυτές οι τιμές είναι οι λεγόμενες τιμές Shapley. Λαμβάνοντας υπόψη μια απόδοση V , σε κάθε παίκτη j αντιστοιχεί μια τιμή Shapley, ϕ_j , που αντιστοιχεί στο δίκαιο μερίδιό του στην απόδοση με βάση την ατομική του συνεισφορά. Η μαθηματική φόρμουλα για το ϕ_j είναι:

$$\phi_j = \sum_{S \subseteq N \setminus j} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [V(S \cup j) - V(S)]$$

όπου S είναι ένας συνασπισμός, N είναι το σύνολο όλων των παικτών, $V(S \cup j) - V(S)$ ποσοτικοποιεί την οριακή συνεισφορά του παίκτη j στον συνασπισμό S , $\frac{|S|!(|N| - |S| - 1)!}{|N|!}$ είναι το βάρος της οριακής συνεισφοράς και $\sum_{S \subseteq N \setminus j}$ αθροίζει όλους τους πιθανούς συνασπισμούς χωρίς το j .

Η σύνδεση μεταξύ των τιμών Shapley και της ανάλυσης SHAP στη μηχανική μάθηση μπορεί να γίνει κατανοητή θεωρώντας τα πιθανά χαρακτηριστικά του μοντέλου (συμπτώματα στην περίπτωσή μας) ως "παίκτες" σε ένα συνεργατικό παίγνιο:

Θεωρία Παιγνίων	Μηχανική Μάθηση
Παίκτες (Συμπτώματα)	Χαρακτηριστικά
Απόδοση (% Διάγνωσης)	Πρόβλεψη για μεμονωμένη παρατήρηση
Οριακή συμπεριφορά	Τιμή του χαρακτηριστικού
Τιμή Shapley του συμπτώματος j	Τιμή shar του χαρακτηριστικού j

Πίνακας 7.1: Πίνακας αντιστοίχισης Θεωρίας Παιγνίων - Μηχανικής Μάθησης

Coalition size of S	Symptom A's marginal contributions	Symptom B's marginal contributions	Symptom C's marginal contributions
0	$V(A) - V(\emptyset) = 40$	$V(B) - V(\emptyset) = 45$	$V(C) - V(\emptyset) = 35$
1	$V(B,A) - V(B) = 65-45 = 20$ $V(C,A) - V(C) = 55-35 = 20$	$V(A,B) - V(A) = 65-40 = 25$ $V(C,B) - V(C) = 60-35 = 25$	$V(A,C) - V(A) = 55-40 = 15$ $V(B,C) - V(B) = 60-45 = 15$
2	$V(B,C,A) - V(B,C) = 90-60 = 30$	$V(A,C,B) - V(A,C) = 90-55 = 35$	$V(A,B,C) - V(A,B) = 90-65 = 25$
Shapley values	$\frac{1}{3}40 + \frac{1}{6}20 + \frac{1}{6}20 + \frac{1}{3}30 = 30\%$	$\frac{1}{3}45 + \frac{1}{6}25 + \frac{1}{6}25 + \frac{1}{3}35 = 35\%$	$\frac{1}{3}35 + \frac{1}{6}15 + \frac{1}{6}15 + \frac{1}{3}25 = 25\%$

Πίνακας 7.2: Οριακή συνεισφορά και τιμές Shapley για τα 3 συμπτώματα

Στην περίπτωση του LIME, η προσέγγιση είναι διαφορετική. Αντί να υπολογίζουμε τις συνεισφορές με βάση όλους τους πιθανούς συνδυασμούς, η LIME: Δημιουργεί τοπικές διαταραχές γύρω από το σημείο ενδιαφέροντος, προσαρμόζει ένα απλό γραμμικό μοντέλο στη γειτονιά του σημείου, χρησιμοποιεί τους συντελεστές του γραμμικού μοντέλου ως ερμηνείες.

Πιο συγκεκριμένα:

$$\xi(x) = \operatorname{argmin}_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g)$$

όπου:

- f είναι το αρχικό μοντέλο (π.χ., το μοντέλο διάγνωσης)
- g είναι το επεξηγήσιμο μοντέλο
- π_x είναι η συνάρτηση πυρήνα που ορίζει την τοπική γειτονιά
- $\Omega(g)$ είναι ένας όρος ποινής για την πολυπλοκότητα του g

Για το παράδειγμα της διάγνωσης γρίπης, θα μπορούσαμε να έχουμε:

Διαταραχή	Πυρετός	Βήχας	Κόπωση	Απόσταση	Βάρος
1	1	1	1	0.0	1.0
2	1	1	0	1.0	0.8
3	1	0	1	1.0	0.8
4	0	1	1	1.0	0.8
5	1	0	0	1.4	0.6
6	0	1	0	1.4	0.6
7	0	0	1	1.4	0.6
8	0	0	0	1.7	0.4

Πίνακας 7.3: (ψευδό)-Παραδείγματα διαταραχών LIME και τα αντίστοιχα βάρη τους

Συνοψίζοντας, έχουμε ότι η μέθοδος shap μπορεί να είναι και καθολική και τοπική, ενώ η μέθοδος lime εξηγεί μόνον τοπικά ένα μοντέλο μηχανικής μάθησης. Οι υπολογισμοί με τη πρώτη μέθοδο είναι ακριβείς, ενώ με την δεύτερη είναι προσεγγιστικοί.

SHAP	LIME
Καθολική ερμηνεία	Τοπική ερμηνεία
Βασίζεται σε συνασπισμούς	Βασίζεται σε διαταραχές
Ακριβείς υπολογισμοί	Προσεγγιστικοί υπολογισμοί
Εγγυημένη συνέπεια	Τοπική συνέπεια

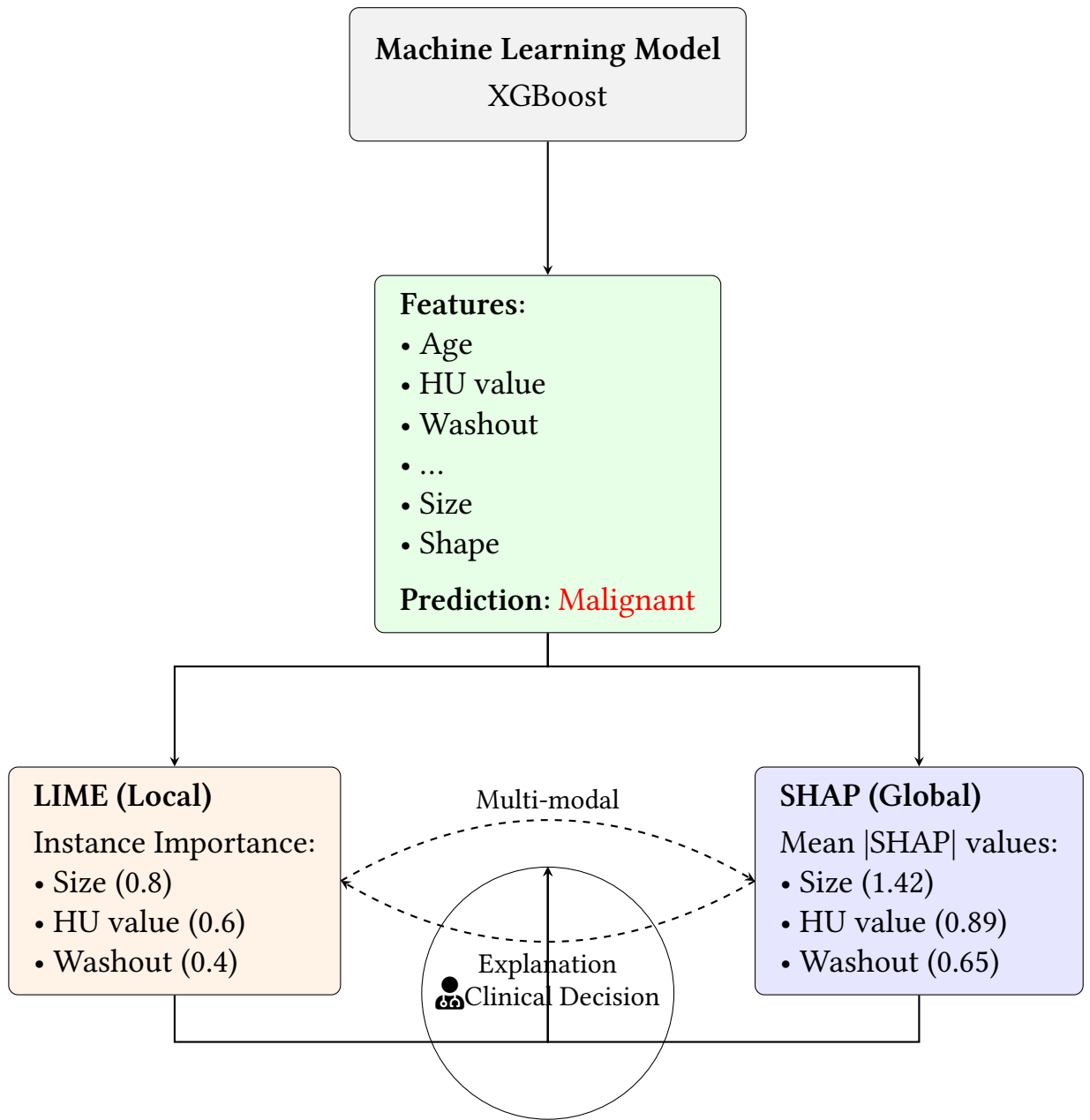
Πίνακας 7.4: Σύγκριση SHAP και LIME

7.2 Υποθέσεις ταξινόμησης ψευδοτυχαιωμάτων

Έχοντας προχωρήσει στην επιλογή του βέλτιστου μοντέλου μηχανικής μάθησης για την ταξινόμηση των ψευδοτυχαιωμάτων, είναι κρίσιμο να εξετάσουμε εάν οι κλινικές παρατηρήσεις που έχουν καταγραφεί στη βιβλιογραφία επιβεβαιώνονται μέσω της ερμηνείας της λειτουργίας του μοντέλου: Η **πρώτη** υπόθεση ($\mathcal{H}_1$) αφορά το μέγεθος του όγκου ως πρωταρχικό δείκτη κακοήθειας. Σύμφωνα με τη βιβλιογραφία, όγκοι μεγαλύτεροι των 4 εκατοστών παρουσιάζουν ευαισθησία 97% και ειδικότητα 52%, ενώ για όγκους άνω των 6 εκατοστών η ευαισθησία είναι 91% με ειδικότητα 80%. Αξιοσημείωτο είναι ότι η μέση διάμετρος των κακοήθων όγκων υπερβαίνει τα 10 εκατοστά. Η **δεύτερη** υπόθεση ($\mathcal{H}_2$) εστιάζει στην ορμονική λειτουργία των όγκων. Περίπου το 60% των κακοήθων όγκων εμφανίζουν ορμονική δραστηριότητα, με το 30% να παρουσιάζει σύνδρομο Cushing, το 20% ανδρογονοποίηση, και ένα 10-20% να εμφανίζει συνδυασμό των δύο. Ιδιαίτερη σημασία έχει η παρατήρηση ότι η ταυτόχρονη έκκριση κορτιζόλης και ανδρογόνων αποτελεί ισχυρή ένδειξη κακοήθειας. Η **τρίτη** υπόθεση ($\mathcal{H}_3$) σχετίζεται με τα απεικονιστικά χαρακτηριστικά των κακοήθων όγκων. Αυτοί τυπικά παρουσιάζουν ετερογένεια και ακανόνιστα όρια, ενώ συχνά εμφανίζουν νέκρωση, αιμορραγία ή ασβεστοποίηση. Σημαντική διαγνωστική αξία έχει η μέση εξασθένηση στην αξονική τομογραφία, με τα αδενώματα να παρουσιάζουν τιμές 8 ± 18 HU και τα καρκινώματα 39 ± 14 HU. Τέλος, η **τέταρτη** υπόθεση ($\mathcal{H}_4$) αφορά τα δημογραφικά χαρακτηριστικά των ασθενών. Παρατηρείται υψηλότερη συχνότητα εμφάνισης στις ηλικίες 50-60 ετών, με τις γυναίκες να εμφανίζουν μεγαλύτερη συχνότητα προσβολής. Αξίζει να σημειωθεί ότι δεν έχει τεκμηριωθεί σαφής συσχέτιση με παράγοντες όπως η κατανάλωση αλκοόλ ή οι διατροφικές συνήθειες.

$\mathcal{H}_1$: Μέγεθος Όγκου [96, 97] • >4 cm: 97% ευαισθησία, 52% ειδικότητα • >6 cm: 91% ευαισθησία, 80% ειδικότητα • >10 cm: μέση διάμετρος κακοήθων
$\mathcal{H}_2$: Ορμονική Λειτουργία [98, 30] • 60% λειτουργικοί όγκοι • 30% σύνδρομο Cushing • 20% ανδρογονοποίηση • 10-20% συνδυασμός
$\mathcal{H}_3$: Απεικονιστικά Χαρακτηριστικά [99] • Αδενώματα: 8 ± 18 HU • Καρκινώματα: 39 ± 14 HU
$\mathcal{H}_4$: Δημογραφικά Χαρακτηριστικά [97, 100, 101] • Ηλικίες: 50-60 ετών • Υψηλότερη συχνότητα στις γυναίκες

Πίνακας 7.5: Υποθέσεις ταξινόμησης ψευδοτυχαιωμάτων



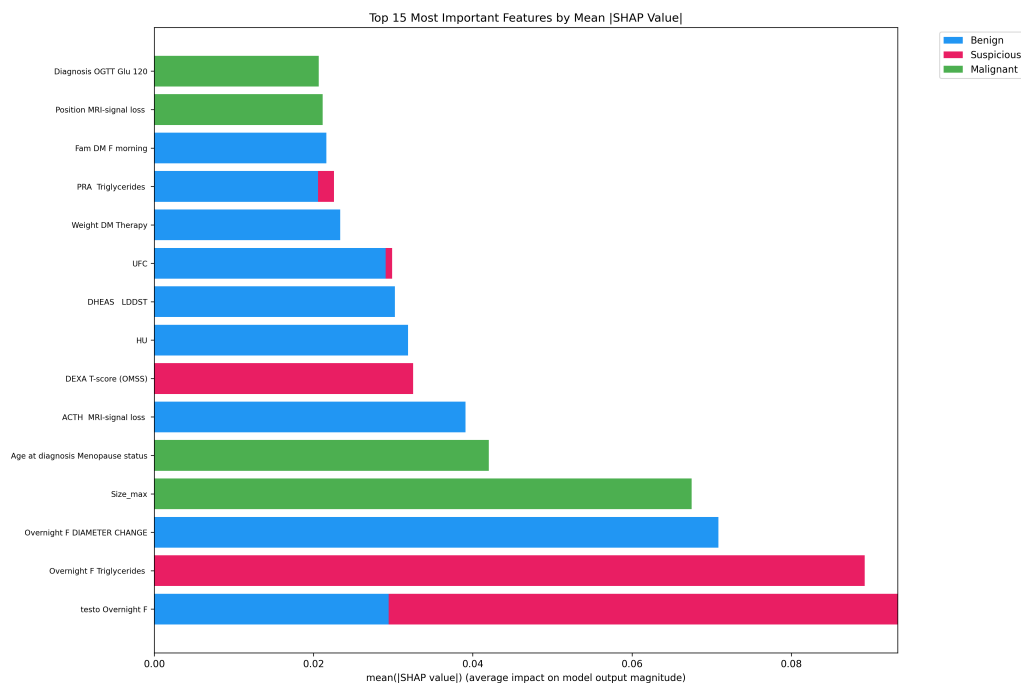
Σχήμα 7.3: Ροή επεξήγησης XAI

Όπως γίνεται προσπάθεια να αποτυπωθεί και στην Εικόνα 7.3, την τελική κρίση για όλα τα περιστατικά την έχει ο ιατρός, ο οποίος αξιολογεί τις ίδιες τις μετρικές, εν είδει ιστορικού ασθενούς, την κατηγοριοποίηση του αλγορίθμου, καθώς και την εξήγηση αυτού, ενδεχομένως με συνδυαστικές μεθόδους. Αυτή η προσέγγιση εξασφαλίζει ότι η τεχνητή νοημοσύνη λειτουργεί ως υποστηρικτικό εργαλείο στη διαγνωστική διαδικασία, ενισχύοντας την κλινική κρίση του ιατρού χωρίς να την υποκαθιστά.

7.3 Διαγράμματα για SHAP ανάλυση

7.3.1 Διάγραμμα Σημαντικότερων Χαρακτηριστικών (Top Features)

Το διάγραμμα αυτό (Εικόνα 7.4) παρουσιάζει τα 15 πιο σημαντικά χαρακτηριστικά του μοντέλου, ταξινομημένα κατά τη συνολική τους επίδραση στις προβλέψεις. Κάθε μπάρα χωρίζεται σε τρία χρώματα που αντιπροσωπεύουν την επίδραση του χαρακτηριστικού σε κάθε κλάση (Καλοήθης, Κακοήθης, Ύποπτος). Το μήκος κάθε τμήματος δείχνει τη μέση απόλυτη τιμή SHAP για κάθε κλάση, υποδεικνύοντας πόσο σημαντικό είναι το χαρακτηριστικό για την αντίστοιχη πρόβλεψη.



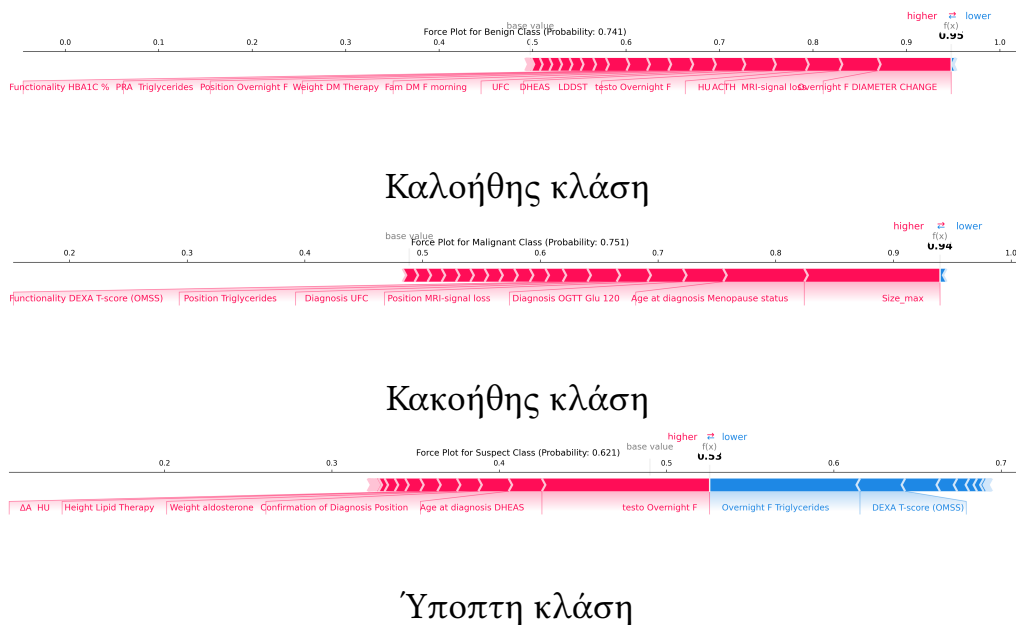
Σχήμα 7.4: Σημαντικότερα χαρακτηριστικά

Η ανάλυση SHAP εντοπίζει ότι οι σημαντικότερες αλληλεπιδράσεις χαρακτηριστικών περιλαμβάνουν τον συνδυασμό "testo Overnight F" και "Overnight F Triglycerides", με ιδιαίτερα υψηλές τιμές SHAP (>0.08). Αυτές οι αλληλεπιδράσεις υποδεικνύουν την κρίσιμη σημασία των ορμονικών δεικτών στην ταξινόμηση των όγκων. Το μέγεθος του όγκου (" $Size_{max}$ ") εμφανίζεται ως ισχυρός προγνωστικός παράγοντας για την κακοήθεια, με τιμή SHAP ~ 0.06 , επιβεβαιώνοντας την κλινική υπόθεση για τη σημασία του μεγέθους στην αξιολόγηση των επινεφριδιακών μαζών. Οι μονάδες Hounsfield (HU) και η απώλεια σήματος MRI εμφανίζουν σημαντικές αλληλεπιδράσεις με άλλους παράγοντες, συμπεριλαμβανομένων των ορμονικών δεικτών (ACTH). Η ηλικία διάγνωσης σε συνδυασμό με την εμμηνοπαυσιακή κατάσταση (τιμή SHAP ~ 0.04) αναδεικνύεται ως σημαντικός παράγοντας για την πρόβλεψη κακοήθειας, υποστηρίζοντας την υπόθεση για τη σημασία των δημογραφικών χαρακτηριστικών. Επιπλέον, η παράμετρος "DEXA T-score (OMSS)" με τιμή SHAP περίπου

0.035 υποδεικνύει μια απρόσμενη συσχέτιση με την οστική πυκνότητα. Η ανάλυση αποκάλυψε πολύπλοκες ορμονικές αλληλεπιδράσεις, με τους συνδυασμούς "DHEAS LDDST" και "UFC" να εμφανίζουν σημαντικές τιμές SHAP (περίπου 0.03). Αυτά τα ευρήματα υποστηρίζουν την υπόθεση για την ύπαρξη λειτουργικών όγκων και την σημασία των συνδυασμένων ορμονικών εκκρίσεων. Ιδιαίτερα ενδιαφέροντα είναι η αλληλεπίδραση μεταξύ "PRA Triglycerides" και "Weight DM Therapy" (τιμές SHAP ~0.02), υποδεικνύοντας μια πιθανή σχέση μεταξύ μεταβολικών παραμέτρων και της φύσης του όγκου.

7.3.2 Διάγραμμα Δυνάμεων (Force Plot)

Το διάγραμμα δυνάμεων για τις καλοήθεις, κακοήθεις και ύποπτες περιπτώσεις δείχνει πώς κάθε χαρακτηριστικό επηρεάζει την πρόβλεψη του μοντέλου. Το κόκκινο χρώμα υποδεικνύει χαρακτηριστικά που ωθούν την πρόβλεψη προς την κατηγορία πρόβλεψης (αυξάνουν την πιθανότητα), ενώ το μπλε χρώμα δείχνει χαρακτηριστικά που ωθούν την πρόβλεψη μακριά από την κατηγορία πρόβλεψης(μειώνουν την πιθανότητα). Η τιμή $E[f(X)]$ αντιπροσωπεύει τη μέση πρόβλεψη του μοντέλου για όλα τα δεδομένα, ενώ η τελική τιμή $f(x)$ δείχνει την πραγματική πρόβλεψη για τη συγκεκριμένη περίπτωση. Το παραπάνω διάγραμμα μπορεί να προσαρμοστεί σε οριζόντια διάταξη. Παραθέτομε, τρία αντιπροσωπευτικά γραφήματα (Εικόνα 7.5) για κάθε κλάση:



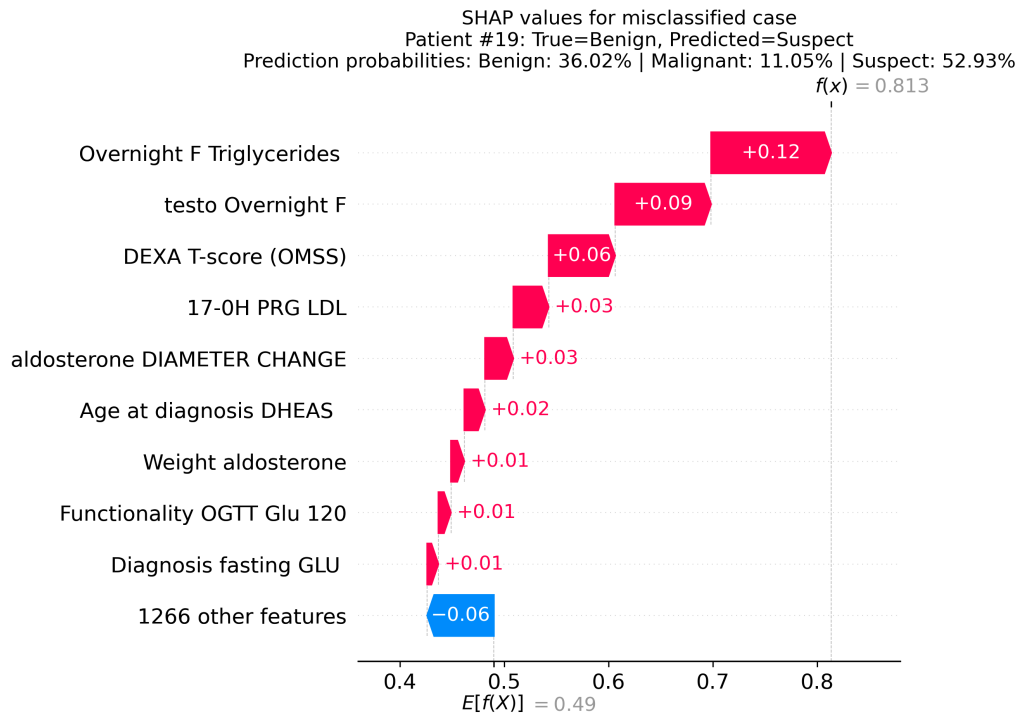
Σχήμα 7.5: Διάγραμμα δυνάμεων για κάθε κλάση

Στο **καλόηθες** περιστατικό, η πρόβλεψη της κατηγορίας "καλόηθες", για τη συγκεκριμένη περίπτωση είναι 0.741, και παρατηρείται ότι παράμετροι όπως το MRI-signal loss, HUACTH, και diameter change συμβάλλουν θετικά στην πρόβλεψη της κατηγορίας αυτής. Αυτές οι παράμετροι φέρνουν την πρόβλεψη πιο κοντά στο καλόηθες, υποδεικνύοντας ότι τα χαρακτηριστικά που σχετίζονται με την απουσία σοβαρών ανωμαλιών (όπως η έλλειψη σήματος στην MRI ή η μικρή αλλαγή στη διάμετρο) είναι δείκτες για λιγότερο επικίνδυνες καταστάσεις.

Στο **κακόηθες** περιστατικό, η πρόβλεψη είναι 0.751, η οποία υποδηλώνει έναν υψηλότερο κίνδυνο για καρκίνο, και αυτός ο κίνδυνος επηρεάζεται κυρίως από το μέγεθος του όγκου ($Size_{max}$). Η συγκεκριμένη παράμετρος είναι η πιο σημαντική για την κατηγοριοποίηση του περιστατικού ως κακόηθες, κάτι που συμφωνεί με την ιατρική πρακτική όπου το μέγεθος του όγκου αποτελεί κλειδί στην αξιολόγηση της σοβαρότητας. Επιπλέον, οι παράμετροι Menopause status και Age at diagnosis παίζουν επίσης σημαντικό ρόλο, υποδεικνύοντας ότι παράγοντες όπως η ηλικία και η εμμηνόπαυση μπορούν να συνδέονται με αυξημένο κίνδυνο για καρκίνο. Τον δεύτερο παράγοντα τον περιμέναμε, βάσει των υποθέσεων που κάναμε.

Το **ύποπτο** περιστατικό με πρόβλεψη 0.621 είναι πιο ασαφές και διαφοροποιημένο, αφού οι παράμετροι εδώ δείχνουν μια ισχυρή διχοτόμηση. Η παρουσία του χαρακτηριστικού test Overnight F, το οποίο έχει θετική επίδραση, αντισταθμίζεται από τη σημαντική αρνητική επίδραση άλλων παραμέτρων, όπως τα Overnight F Triglycerides και το DEXA T-score (OMSS). Αυτό το περιστατικό φαίνεται να έχει χαρακτηριστικά που δεν εντάσσονται καθαρά στις κατηγορίες "καλόηθες" ή "κακόηθες", γεγονός που καθιστά την πρόβλεψη πιο αβέβαιη και υπαινίσσεται την ανάγκη για περαιτέρω ιατρική διερεύνηση. Η διχοτόμηση των χαρακτηριστικών μπορεί να αντικατοπτρίζει την πολυπλοκότητα της ιατρικής διάγνωσης, όπου ορισμένα χαρακτηριστικά μπορεί να είναι αλληλοεπικαλυπτόμενα ή αντιφατικά, απαιτώντας προσεκτική εκτίμηση και επιπλέον διαγνωστικές εξετάσεις για μια ακριβέστερη κατηγοριοποίηση.

Στο σύνολο των περιπτώσεων που αναλύθηκαν, μόνο μία περίπτωση (Ασθενής #19) (Εικόνα 7.6) ταξινομήθηκε λανθασμένα, όπου ένα καλόηθες περιστατικό χαρακτηρίστηκε ως ύποπτο. Αξιοσημείωτο είναι ότι η πρόβλεψη του μοντέλου παρουσιάζει σχετικά μοιρασμένες πιθανότητες μεταξύ του καλοήθους (36.02%) και ύποπτου (52.93%) περιστατικού, υποδεικνύοντας την αβεβαιότητα του μοντέλου στη συγκεκριμένη περίπτωση. Εξετάζοντας τους παράγοντες που οδήγησαν σε αυτή την ταξινόμηση, παρατηρούμε ότι τα Overnight F Triglycerides (+0.12), το test Overnight F (+0.09) και το DEXA T-score (OMSS) (+0.06) είχαν τη μεγαλύτερη θετική επίδραση στην κατηγοριοποίηση ως ύποπτο περιστατικό. Είναι ενδιαφέρον ότι αυτοί οι παράγοντες συχνά εμφανίζονται ως σημαντικοί δείκτες στα ύποπτα περιστατικά, όπως είδαμε και στην προηγούμενη ανάλυση των Force Plots. Η μόνη αρνητική συνεισφορά προέρχεται από τα υπόλοιπα 1266 (βλ. περισσότερα 7.3.4) χαρακτηριστικά (-0.06), υποδηλώνοντας ότι το σύνολο των υπόλοιπων παραμέτρων έτεινε προς την ορθή (καλοήθη) κατηγοριοποίηση, αλλά δεν ήταν αρκετό για να αντισταθμίσει την επίδραση των κύριων χαρακτηριστικών.



Σχήμα 7.6: Περίπτωση εσφαλμένης ταξινόμησης

7.3.3 Διάγραμμα Καταράκτη

Τα Force Plots παρουσιάζουν τη συνεισφορά κάθε χαρακτηριστικού ως μια συνεχή "δύναμη" που ωθεί την πρόβλεψη είτε προς τα πάνω (κόκκινο) είτε προς τα κάτω (μπλε) από τη βασική τιμή (base value), με το πλάτος κάθε τμήματος να αντιπροσωπεύει το μέγεθος της επίδρασης. Από την άλλη πλευρά, τα Waterfall Plots (Εικόνα 7.7) παρουσιάζουν την ίδια πληροφορία σε μορφή "καταρράκτη", όπου κάθε χαρακτηριστικό προσθέτει ή αφαιρεί από την προηγούμενη τιμή, ξεκινώντας από τη βασική τιμή ($E[f(X)]$) και καταλήγοντας στην τελική πρόβλεψη $f(x)$. Αυτή η διαφορετική οπτικοποίηση επιτρέπει την ευκολότερη κατανόηση της σταδιακής συνεισφοράς κάθε χαρακτηριστικού στην τελική πρόβλεψη, ενώ παράλληλα διατηρεί την ίδια ποσοτική πληροφορία με τα Force Plots.

Τα διαγράμματα καταρράκτη SHAP παρουσιάζουν μια λεπτομερή ανάλυση των προβλέψεων του μοντέλου για κάθε κατηγορία (καλοήθη, κακοήθη και ύποπτα περιστατικά), επιβεβαιώνοντας και επεκτείνοντας τα ευρήματα των Force Plots. Συγκεκριμένα, για τα καλοήθη περιστατικά (επάνω), παρατηρούμε μια συνεκτική θετική συνεισφορά των χαρακτηριστικών, με την αλλαγή διαμέτρου, απώλεια σήματος MRI και HU να έχουν τις ισχυρότερες θετικές επιδράσεις (+0.06, +0.04 και +0.04 αντίστοιχα), ενώ τα υπόλοιπα χαρακτηριστικά συνεισφέρουν με μικρότερες τιμές (+0.02 έως +0.03).

Στα κακοήθη περιστατικά (μέση), το $Size_{max}$ αναδεικνύεται ως ο κυριότερος προγνωστικός παράγοντας με σημαντική θετική επίδραση (+0.12), ακολουθούμενο από την εμμηνοπαυσιακή κατάσταση και τη διάγνωση OGTT Glu 120. Τα υπόλοιπα χαρακτηριστικά, όπως η θέση MRI-signal loss και η διάγνωση



Σχήμα 7.7: Διαγράμματα Καταρράκτη (2) ανά κλάση

UFC, έχουν μικρότερες αλλά σταθερά θετικές συνεισφορές, υποδεικνύοντας μια συνεκτική ταξινόμηση των κακοήθων περιστατικών.

Ιδιαίτερο ενδιαφέρον παρουσιάζουν τα ύποπτα περιστατικά (κάτω), όπου παρατηρείται μια πιο σύνθετη εικόνα με τόσο θετικές όσο και αρνητικές συνεισφορές. Το *testo Overnight F* και τα *Overnight F Triglycerides* εμφανίζουν ισχυρές αλλά αντίθετες επιδράσεις (+0.13 και -0.09 αντίστοιχα), ενώ το *DEXA T-score (OMSS)* και το *17-OH PRG LDL* έχουν επίσης σημαντική αρνητική επίδραση. Αυτή η μικτή εικόνα των επιδράσεων υποδηλώνει την πολυπλοκότητα της ταξινόμησης των ύποπτων περιστατικών, όπου το μοντέλο λαμβάνει υπόψη τόσο θετικούς όσο και αρνητικούς παράγοντες για να καταλήξει στην τελική του πρόβλεψη. Αξίζει να σημειωθεί ότι οι πιθανότητες πρόβλεψης για τα ύποπτα περιστατικά είναι γενικά χαμηλότερες (περίπου 48-62%) σε σύγκριση με τις άλλες κατηγορίες, αντανakλώντας την εγγενή αβεβαιότητα στην ταξινόμηση αυτών των περιπτώσεων. Αυτή η ανάλυση επιβεβαιώνει τα ευρήματα των *Force Plots*, παρέχοντας παράλληλα μια πιο λεπτομερή ποσοτική εικόνα της συνεισφοράς κάθε χαρακτηριστικού στην τελική πρόβλεψη του μοντέλου.

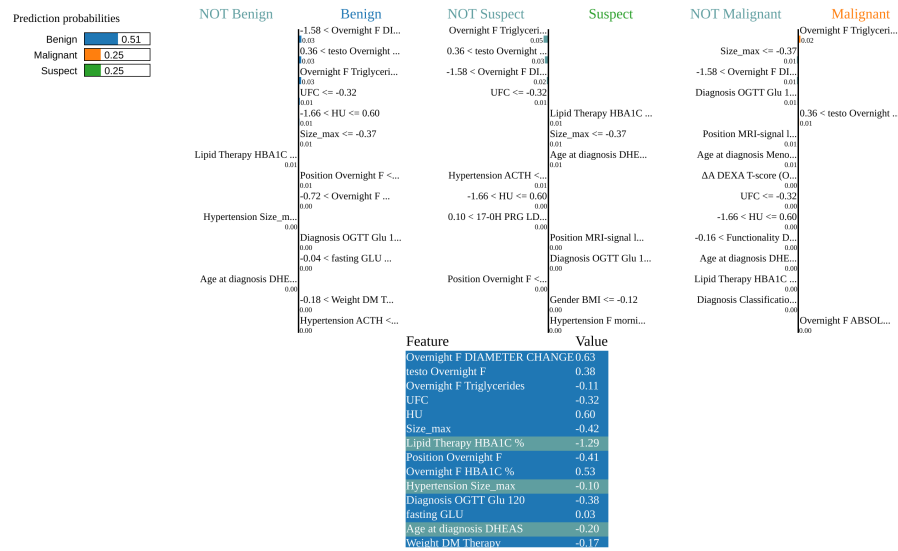
7.3.4 Συνδυαστικές τιμές

Για τα παραπάνω γραφήματα, είναι σημαντικό να διευκρινιστεί ότι τα χαρακτηριστικά που παρουσιάζονται δεν αντιπροσωπεύουν μεμονωμένες μετρήσεις, αλλά πολυωνμικούς συνδυασμούς (*polynomial features*) διαφορετικών παραμέτρων. Για παράδειγμα, όταν βλέπουμε "*testo Overnight F*", αυτό αντιπροσωπεύει τον συνδυασμό των επιπέδων τεστοστερόνης και των τιμών νηστείας κατά τη διάρκεια της νύχτας, λαμβάνοντας υπόψη την πιθανή αλληλεπίδραση μεταξύ αυτών των δύο παραμέτρων.

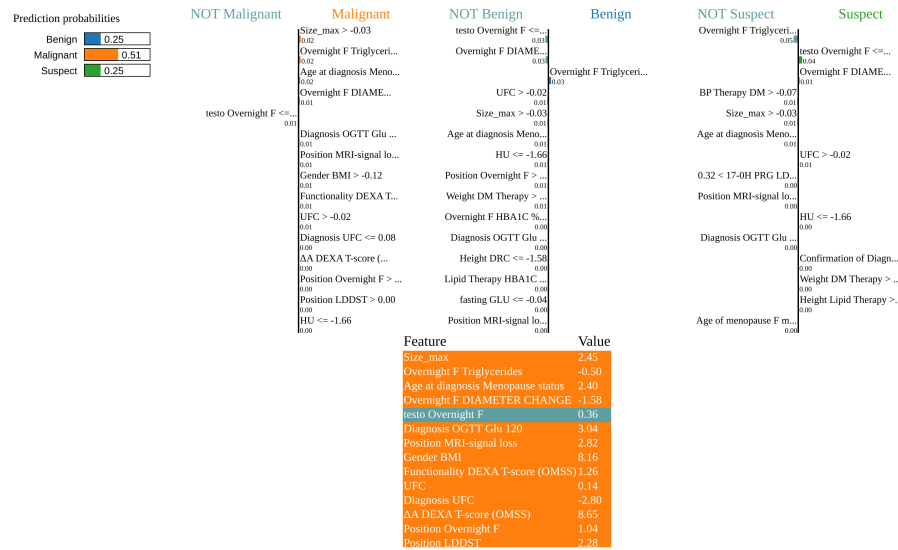
Αυτή η συνδυαστική προσέγγιση επιτρέπει στο μοντέλο να συλλάβει πιο σύνθετες σχέσεις μεταξύ των μεταβλητών, καθώς η διαγνωστική αξία συχνά δεν έγκειται σε μεμονωμένες μετρήσεις αλλά στον τρόπο που διάφορες παράμετροι αλληλεπιδρούν μεταξύ τους. Για παράδειγμα, το "*Overnight F Triglycerides*" αντιπροσωπεύει τη συνδυασμένη επίδραση των επιπέδων τριγλυκεριδίων και των τιμών νηστείας κατά τη διάρκεια της νύχτας, παρέχοντας έτσι μια πιο ολοκληρωμένη εικόνα της μεταβολικής κατάστασης του ασθενούς. Αυτή η πολυωνμική προσέγγιση αποτελεί ένα βασικό χαρακτηριστικό του μοντέλου που ενισχύει την ικανότητά του να αναγνωρίζει πολύπλοκα πρότυπα στα δεδομένα. Το παραπάνω ισχύει και για την δεύτερη μέθοδο.

7.4 Διαγράμματα για LIME ανάλυση

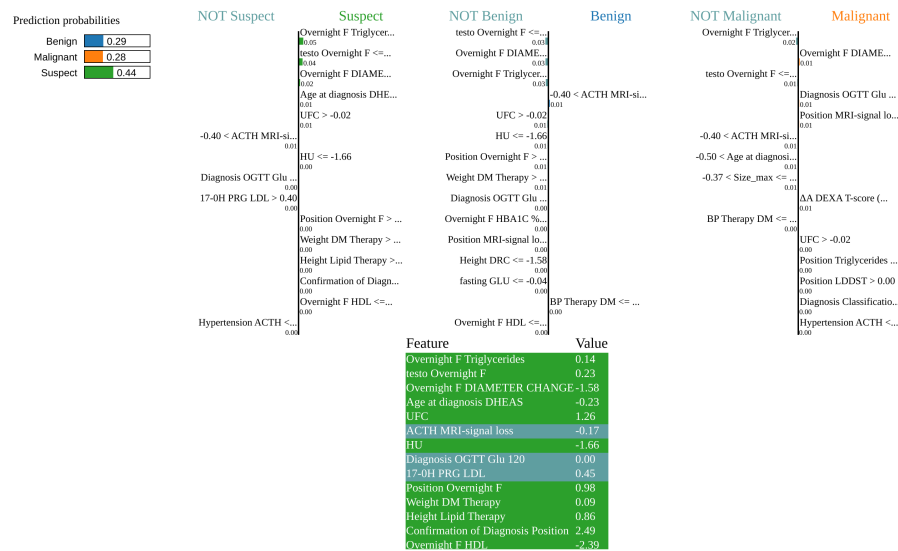
Ομοίως με την προηγούμενη ανάλυση εξετάζουμε τα 15 κορυφαία χαρακτηριστικά ανά κλάση, και πώς επηρεάζουν την τελική επιλογή του μοντέλου. Παρακάτω παραθέτουμε τρία γραφήματα (Εικόνα 7.8), όπου αποτυπώνεται η σχετική πληροφορία.



Καλοήθης κλάση



Κακοήθης κλάση



Υποπτη κλάση

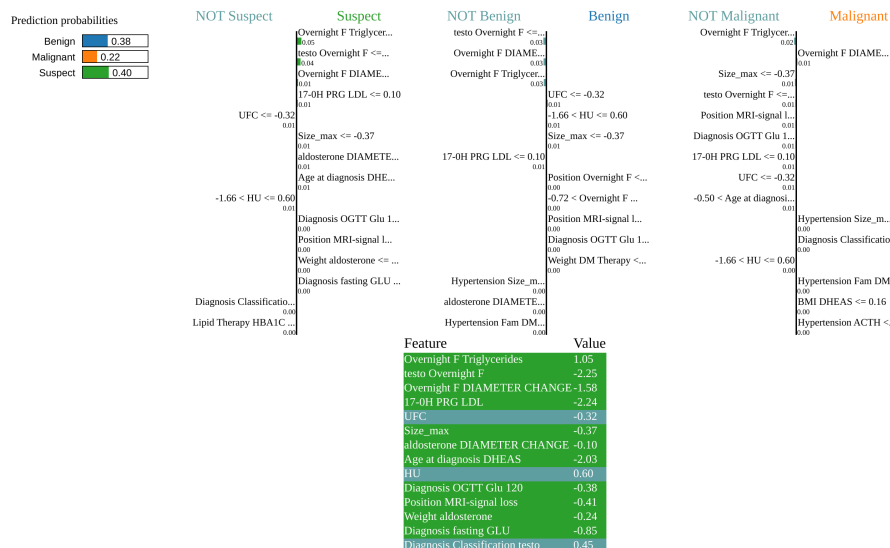
Σχήμα 7.8: Διάγραμμα lime ανά κλάση

Η ανάλυση της επεξηγησιμότητας του μοντέλου αποκάλυψε ότι οι σημαντικότερες παράμετροι για την ταξινόμηση, με βάση τη συνεισφορά τους, είναι:

- *Size_{max}*: Η μέγιστη διάσταση του όγκου αποτελεί τον πρωταρχικό δείκτη με τιμή κατωφλίου -0.37, επιβεβαιώνοντας την υπόθεση $\mathcal{H}_1$.
- **Overnight F Triglycerides & diameter change**: Οι μεταβολές στις τιμές τριγλυκεριδίων κατά τη διάρκεια της νύχτας και της διαμέτρου είναι ισχυροί προγνωστικοί παράγοντες.
- **Diagnosis OGTT Glu 120**: Η τιμή γλυκόζης μετά από δοκιμασία ανοχής εμφανίζει σημαντική προγνωστική αξία.
- **Position MRI-signal loss**: Τα απεικονιστικά χαρακτηριστικά από την μαγνητική τομογραφία συμβάλλουν σημαντικά στην ταξινόμηση.

Οι τιμές Hounsfield Units (HU) από την αξονική τομογραφία επιβεβαιώνουν την υπόθεση $\mathcal{H}_3$, με το κατώφλι $-1.66 < \text{HU} \leq 0.60$ να διαχωρίζει αποτελεσματικά τα αδενώματα από τα καρκινώματα. Η ορμονική λειτουργικότητα ($\mathcal{H}_2$) αντικατοπτρίζεται στις μετρήσεις των τριγλυκεριδίων και της δοκιμασίας OGTT, ενώ τα δημογραφικά χαρακτηριστικά ($\mathcal{H}_4$) επιβεβαιώνονται μέσω των παραμέτρων ηλικίας και φύλου.

7.5 Ανάλυση Εσφαλμένης Ταξινόμησης



Σχήμα 7.9: Περίπτωση εσφαλμένης ταξινόμησης

Στη συγκεκριμένη περίπτωση (Εικόνα 7.9), το μοντέλο παρουσίασε την ακόλουθη κατανομή πιθανοτήτων:

- καλόηθες (Benign): 38%
- κακόηθες (Malignant): 22%
- ύποπτο (Suspect): 40%

Η περίπτωση αυτή είναι ιδιαίτερα ενδιαφέρουσα καθώς το μοντέλο την ταξινομεί ως ύποπτη (αντί για καλοήθη), με τις ακόλουθες κρίσιμες παραμέτρους:

1. Ισχυροί Δείκτες Υποψίας:

- Overnight F Triglycerides: 1.05 (σημαντικά θετική τιμή)
- testo Overnight F: -2.25 (έντονα αρνητική τιμή)
- 17-OH PRG LDL: -2.24 (έντονα αρνητική τιμή)
- Age at diagnosis DHEAS: -2.03 (σημαντικά αρνητική τιμή)

2. Οριακοί Δείκτες:

- $Size_{max}$: -0.37 (ακριβώς στο όριο)
- UFC: -0.32 (οριακή τιμή)
- HU: 0.60 (εντός του διαστήματος αβεβαιότητας)

3. Ορμονικοί Παράγοντες:

- Overnight F diameter change: -1.58 (σημαντικά αρνητική τιμή)
- aldosterone diameter change: -0.10 (ελαφρώς αρνητική τιμή)

Η ταξινόμηση ως ύποπτη περίπτωση (40%) δικαιολογείται από:

- Την παρουσία πολλαπλών ακραίων τιμών στους ορμονικούς δείκτες
- Το μέγεθος του όγκου που βρίσκεται ακριβώς στο όριο
- Την ύπαρξη αντικρουόμενων ενδείξεων μεταξύ απεικονιστικών και βιοχημικών παραμέτρων

Συμπεράσματα - μελλοντικές προεκτάσεις

Το μοντέλο φαίνεται να προβλέπει σωστά, συσχετίζοντας ιατρική πληροφορία, η οποία βάσει βιβλιογραφίας αφορά στο χαρακτηρισμό όγκων επινεφριδίων, όπως η διάμετρος του όγκου, η ηλικία διάγνωσης, καθώς και αιματολογικές εξετάσεις.

Ως προς τις 4 υποθέσεις φαίνεται να έχουν ισχύ όλες.

Ενδιαφέρον είναι το στοιχείο ότι το μοντέλο επιλογής αποτυγχάνει μόνον σε μία περίπτωση ταξινόμησης, κάτι το οποίο θα μπορούσε να μελετηθεί περαιτέρω, εξετάζοντας τυχόν υπερ-προσαρμογή του μοντέλου στα παρόντα δεδομένα.

Αναφορικά με την επέκταση της παρούσας έρευνας, θα μπορούσαν να ληφθούν υπ' όψιν τα εξής:

- Λήψη πιο ισορροπημένου συνόλου δεδομένων
- Παρουσίαση των αποτελεσμάτων στην ιατρική κοινότητα για συνεκτίμηση
 - Δημιουργία ερωτηματολογίου για την συνεκτίμηση
 - * Αξιολόγηση κατανόησης γραφημάτων
 - * Επικύρωση συσχετίσεων χαρακτηριστικών
 - * Εκτίμηση βαρύτητας χαρακτηριστικών ανά κατηγορία

Η συνεκτίμηση είναι απαραίτητη όπως προκύπτει από την βιβλιογραφία. Πέραν όμως της συνεκτίμησης, ένα ερωτηματολόγιο θα μπορούσε να βοηθήσει στην περαιτέρω αξιολόγηση της ίδιας της ερμηνείας της εξήγησης του μοντέλου. Δηλαδή, κατά πόσον τα ίδια τα γραφήματα είναι βοηθητικά στην ερμηνεία, αν είναι "εύκολο" να διαβαστούν από ιατρούς και προφανώς το ίδιο το περιεχόμενο, πέραν της μορφής, δηλαδή αν υπάρχει συμφωνία με τα συμπεράσματα ως προς τη συσχέτιση των χαρακτηριστικών-ιατρικών δεδομένων. Παράλληλα, θα ήταν χρήσιμο να αξιολογηθεί και η βαρύτητα του κάθε χαρακτηριστικού, ανά κατηγορία όγκου.

Βιβλιογραφία

- [1] H. W. Loh, C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari, and U. R. Acharya, ``Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011--2022),'' *Computer Methods and Programs in Biomedicine*, vol. 226, p. 107161, 2022.
- [2] K.-H. Yu, A. L. Beam, and I. S. Kohane, ``Artificial intelligence in healthcare,'' *Nat. Biomed. Eng.*, vol. 2, pp. 719--731, Oct. 2018.
- [3] G. Choy, O. Khalilzadeh, M. Michalski, S. Do, A. E. Samir, O. S. Pianykh, J. R. Geis, P. V. Pandharipande, J. A. Brink, and K. J. Dreyer, ``Current applications and future impact of machine learning in radiology,'' *Radiology*, vol. 288, pp. 318--328, Aug. 2018.
- [4] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, ``Causability and explainability of artificial intelligence in medicine,'' *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 9, p. e1312, July 2019.
- [5] M. Fassnacht, W. Arlt, I. Bancos, H. Dralle, J. Newell-Price, A. Sahdev, A. Tabarin, M. Terzolo, S. Tsagarakis, and O. M. Dekkers, ``Management of adrenal incidentalomas: European society of endocrinology clinical practice guideline in collaboration with the european network for the study of adrenal tumors,'' *Eur. J. Endocrinol.*, vol. 175, pp. G1--G34, Aug. 2016.
- [6] G.-S. Fu, Y. Levin-Schwartz, Q.-H. Lin, and D. Zhang, ``Machine learning for medical imaging,'' *J. Healthc. Eng.*, vol. 2019, p. 9874591, Apr. 2019.
- [7] M. A. Ahmad, C. Eckert, and A. Teredesai, ``Interpretable machine learning in healthcare,'' in *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*, pp. 559--560, 2018.
- [8] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, ``From local explanations to global understanding with explainable ai for trees,'' *Nature machine intelligence*, vol. 2, no. 1, pp. 56--67, 2020.
- [9] M. T. Ribeiro, S. Singh, and C. Guestrin, ``" why should i trust you?" explaining the predictions of any classifier,'' in *Proceedings of the 22nd*

ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1135--1144, 2016.

- [10] W. Pawlina, *Histology: A Text and Atlas: With Correlated Cell and Molecular Biology*, 9e. Lippincott Williams & Wilkins, a Wolters Kluwer business, 01 2024.
- [11] F. H. Netter, *LATIN TERMINOLOGY netter atlas of human anatomy: Classic regional approach with Latin terminology*. Netter Basic Science, Philadelphia, PA: Elsevier - Health Sciences Division, 8 ed., Aug. 2022.
- [12] R. Moore, ``Atlas of head & neck surgery---otolaryngology, by byron j. bailey and karen h. calhoun, lippincott williams & wilkins, philadelphia, 2001, 1018 pp," *Head Neck*, vol. 25, pp. 418--418, May 2003.
- [13] S. Standring, *Gray's anatomy Elsevier eBook on VitalSource (retail access card)*. Gray's Anatomy, London, England: Elsevier Health Sciences, 41 ed., Jan. 2017.
- [14] M. Loukas, R. S. Tubbs, P. H. Abrahams, and S. W. Carmichael, *Gray's Anatomy Review*. Elsevier Health Sciences, 2020.
- [15] J. E. Hall and A. C. Guyton, *Guyton and Hall textbook of medical physiology*. Student consult, Philadelphia, Pa: Elsevier, 13th edition ed., 2016.
- [16] K. E. Barrett, *Ganong's Review of Medical Physiology*. McGraw-Hill Medical, July 2009.
- [17] S. Khonsary, ``Guyton and hall: Textbook of medical physiology," *Surg. Neurol. Int.*, vol. 8, no. 1, p. 275, 2017.
- [18] W. F. Boron and E. L. Boulpaep, *Medical Physiology*. Philadelphia, PA: Elsevier - Health Sciences Division, 3 ed., May 2016.
- [19] V. Kumar, A. K. Abbas, J. C. Aster, and A. T. Deyrup, *Robbins and Kumar basic pathology*. Elsevier, 11 ed., 2023.
- [20] G. Arnaldi and M. Boscaro, ``Adrenal incidentaloma," *Best Practice & Research Clinical Endocrinology & Metabolism*, vol. 26, no. 4, pp. 405--419, 2012. Subclinical Endocrine Disorders.
- [21] G. Mansmann, J. Lau, E. Balk, M. Rothberg, Y. Miyachi, and S. R. Bornstein, ``The clinically inapparent adrenal mass: Update in diagnosis and management," *Endocrine Reviews*, vol. 25, pp. 309--340, 04 2004.
- [22] NIH State-of-the-Science Conference Panel, ``Management of the clinically inapparent adrenal mass (incidentaloma)," *Annals of Internal Medicine*, vol. 138, no. 5, pp. 424--429, 2003.

- [23] M. Sherlock, A. Scarsbrook, A. Abbas, S. Fraser, P. Limumpornpetch, R. Dineen, and P. M. Stewart, "Adrenal incidentaloma," *Endocr. Rev.*, vol. 41, pp. 775--820, Dec. 2020.
- [24] C. Gauss, D. Rowland, and B. Ergun-Longmire, "An update on genetics of adrenal gland and associated disorders," *Endocrines*, vol. 3, pp. 187--197, Apr. 2022.
- [25] S. Zheng, A. D. Cherniack, N. Dewal, R. A. Moffitt, L. Danilova, B. A. Murray, A. M. Lerario, T. Else, T. A. Knijnenburg, G. Ciriello, S. Kim, G. Assie, O. Morozova, R. Akbani, J. Shih, K. A. Hoadley, T. K. Choueiri, J. Waldmann, O. Mete, A. G. Robertson, H.-T. Wu, B. J. Raphael, L. Shao, M. Meyerson, M. J. Demeure, F. Beuschlein, A. J. Gill, S. B. Sidhu, M. Q. Almeida, M. C. B. V. Fragoso, L. M. Cope, E. Kebebew, M. A. Habra, T. G. Whitsett, K. J. Bussey, W. E. Rainey, S. L. Asa, J. Bertherat, M. Fassnacht, D. A. Wheeler, Cancer Genome Atlas Research Network, G. D. Hammer, T. J. Giordano, and R. G. W. Verhaak, "Comprehensive pan-genomic characterization of adrenocortical carcinoma," *Cancer Cell*, vol. 29, pp. 723--736, May 2016.
- [26] A. Lacroix, R. A. Feelders, C. A. Stratakis, and L. K. Nieman, "Cushing's syndrome," *Lancet*, vol. 386, pp. 913--927, Aug. 2015.
- [27] M. Fassnacht, S. Tsagarakis, M. Terzolo, A. Tabarin, A. Sahdev, J. Newell-Price, I. Pelsma, L. Marina, K. Lorenz, I. Bancos, W. Arlt, and O. M. Dekkers, "European society of endocrinology clinical practice guidelines on the management of adrenal incidentalomas, in collaboration with the european network for the study of adrenal tumors," *European Journal of Endocrinology*, vol. 189, pp. G1--G42, 06 2023.
- [28] F. Mantero, M. Terzolo, G. Arnaldi, G. Osella, A. M. Masini, A. Ali, M. Giovagnetti, G. Opocher, and A. Angeli, "A survey on adrenal incidentaloma in Italy. study group on adrenal tumors of the Italian society of endocrinology," *J. Clin. Endocrinol. Metab.*, vol. 85, pp. 637--644, Feb. 2000.
- [29] T. Else, A. C. Kim, A. Sabolch, V. M. Raymond, A. Kandathil, E. M. Caoili, S. Jolly, B. S. Miller, T. J. Giordano, and G. D. Hammer, "Adrenocortical carcinoma," *Endocrine Reviews*, vol. 35, pp. 282--326, 04 2014.
- [30] R. Libé, "Adrenocortical carcinoma (acc): diagnosis, prognosis, and treatment," *Frontiers in cell and developmental biology*, vol. 3, p. 45, 2015.
- [31] L. K. Nieman, "Approach to the patient with an adrenal incidentaloma," *J. Clin. Endocrinol. Metab.*, vol. 95, pp. 4106--4113, Sept. 2010.
- [32] W. W. Mayo-Smith, J. H. Song, G. L. Boland, I. R. Francis, G. M. Israel, P. J. Mazzaglia, L. L. Berland, and P. V. Pandharipande, "Management of incidental adrenal masses: A white paper of the ACR incidental findings committee," *J. Am. Coll. Radiol.*, vol. 14, pp. 1038--1044, Aug. 2017.

- [33] S. Russell and P. Norvig, *Artificial intelligence: A modern approach, global edition*. London, England: Pearson Education, 4 ed., May 2021.
- [34] N. J. Nilsson, *The Quest for Artificial Intelligence*. Cambridge University Press, 2009.
- [35] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [36] M. Mitchell, *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux, 2019.
- [37] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, pp. 255--260, July 2015.
- [38] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436--444, 2015.
- [39] I. Goodfellow, "Deep learning," 2016.
- [40] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [41] Y. Bengio, Y. Lecun, and G. Hinton, "Deep learning for ai," *Communications of the ACM*, vol. 64, no. 7, pp. 58--65, 2021.
- [42] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," *Nature medicine*, vol. 25, no. 1, pp. 24--29, 2019.
- [43] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, "An overview of clinical decision support systems: benefits, risks, and strategies for success," *NPJ digital medicine*, vol. 3, no. 1, p. 17, 2020.
- [44] E. S. Berner, *Clinical decision support systems*, vol. 233. Springer, 2007.
- [45] M. A. Musen, B. Middleton, and R. A. Greenes, "Clinical decision-support systems," in *Biomedical informatics: computer applications in health care and biomedicine*, pp. 795--840, Springer, 2021.
- [46] A. Wright, A. Ai, J. Ash, J. F. Wiesen, T.-T. T. Hickman, S. Aaron, D. McEvoy, S. Borkowsky, P. I. Dissanayake, P. Embi, *et al.*, "Clinical decision support alert malfunctions: analysis and empirically derived taxonomy," *Journal of the American Medical Informatics Association*, vol. 25, no. 5, pp. 496--506, 2018.
- [47] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success," *Bmj*, vol. 330, no. 7494, p. 765, 2005.

- [48] J. A. Osheroff, J. Teich, D. Levick, L. Saldana, F. Velasco, D. Sittig, K. Rogers, and R. Jenders, *Improving outcomes with clinical decision support: an implementer's guide*. Himss Publishing, 2012.
- [49] R. Greenes, *Clinical decision support: the road to broad adoption*. Academic Press, 2014.
- [50] B. Middleton, D. Sittig, and A. Wright, "Clinical decision support: a 25 year retrospective and a 25 year vision," *Yearbook of medical informatics*, vol. 25, no. S 01, pp. S103--S116, 2016.
- [51] A. M. Antoniadis, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, and C. Mooney, "Current challenges and future opportunities for xai in machine learning-based clinical decision support systems: a systematic review," *Applied Sciences*, vol. 11, no. 11, p. 5088, 2021.
- [52] W. Price and I. Nicholson, "Medical ai and contextual bias," *Harv. JL & Tech.*, vol. 33, p. 65, 2019.
- [53] A. Selbst and J. Powles, "'meaningful information' and the right to explanation," in *conference on fairness, accountability and transparency*, pp. 48--48, PMLR, 2018.
- [54] G. Joshi, A. Jain, S. R. Araveeti, S. Adhikari, H. Garg, and M. Bhandari, "Fda-approved artificial intelligence and machine learning (ai/ml)-enabled medical devices: an updated landscape," *Electronics*, vol. 13, no. 3, p. 498, 2024.
- [55] B. Mittelstadt, "Principles alone cannot guarantee ethical ai," *Nature machine intelligence*, vol. 1, no. 11, pp. 501--507, 2019.
- [56] D. S. Char, M. D. Abramoff, and C. Feudtner, "Identifying ethical considerations for machine learning healthcare applications," *The American Journal of Bioethics*, vol. 20, no. 11, pp. 7--17, 2020.
- [57] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature medicine*, vol. 25, no. 1, pp. 44--56, 2019.
- [58] A. Verghese, N. H. Shah, and R. A. Harrington, "What this computer needs is a physician: humanism and artificial intelligence," *Jama*, vol. 319, no. 1, pp. 19--20, 2018.
- [59] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347--1358, 2019.
- [60] I. Price and W. Nicholson, "Medical malpractice and black-box medicine," *Big Data, Health Law, and Bioethics (Cambridge University Press, 2018), U of Michigan Public Law Research Paper*, no. 536, 2017.
- [61] E. H. Shortliffe and M. J. Sepúlveda, "Clinical decision support in the era of artificial intelligence," *Jama*, vol. 320, no. 21, pp. 2199--2200, 2018.

- [62] R. S. Ledley and J. B. Wilson, *Use of computers in biology and medicine*, vol. 1965. McGraw-Hill New York, 1965.
- [63] R. J. Chen, J. J. Wang, D. F. Williamson, T. Y. Chen, J. Lipkova, M. Y. Lu, S. Sahai, and F. Mahmood, "Algorithmic fairness in artificial intelligence for medicine and healthcare," *Nature biomedical engineering*, vol. 7, no. 6, pp. 719--742, 2023.
- [64] K. Cheng, Q. Guo, Y. He, Y. Lu, S. Gu, and H. Wu, "Exploring the potential of gpt-4 in biomedical engineering: the dawn of a new era," *Annals of Biomedical Engineering*, vol. 51, no. 8, pp. 1645--1653, 2023.
- [65] C. Park, C. C. Took, and J.-K. Seong, "Machine learning in biomedical engineering," *Biomedical Engineering Letters*, vol. 8, pp. 1--3, 2018.
- [66] E. H. Shortliffe, "Mycin: A knowledge-based computer program applied to infectious diseases," in *Proceedings of the Annual Symposium on Computer Application in Medical Care*, p. 66, American Medical Informatics Association, 1977.
- [67] L. Y. Victor, L. M. Fagan, S. W. Bennett, W. J. Clancey, A. C. Scott, J. F. Hannigan, R. L. Blum, B. G. Buchanan, and S. N. Cohen, "An evaluation of mycin's advice," *Buchanan and Shortliffe [16]*, pp. 589--596, 1984.
- [68] A. S. Miller, *The application of neural networks to imaging and signal processing in astronomy and medicine*. PhD thesis, University of Southampton, 1993.
- [69] N. Asada, K. Doi, H. MacMahon, S. Montner, M. Giger, C. Abe, and Y. Wu, "Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung diseases: pilot study," *Radiology*, vol. 177, no. 3, pp. 857--860, 1990.
- [70] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological cybernetics*, vol. 36, no. 4, pp. 193--202, 1980.
- [71] P. Baldi and S. Brunak, *Bioinformatics: the machine learning approach*. MIT press, 2001.
- [72] S. Salzberg, "Locating protein coding regions in human dna using a decision tree algorithm," *Journal of Computational Biology*, vol. 2, no. 3, pp. 473--485, 1995.
- [73] X. Liu, L. Faes, A. U. Kale, S. K. Wagner, D. J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern, *et al.*, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *The lancet digital health*, vol. 1, no. 6, pp. e271--e297, 2019.

- [74] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115--118, 2017.
- [75] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *jama*, vol. 316, no. 22, pp. 2402--2410, 2016.
- [76] P. Ramnarayan, A. Tomlinson, G. Kulkarni, A. Rao, and J. Britto, "A novel diagnostic aid (isabel): development and preliminary evaluation of clinical performance," in *MEDINFO 2004*, pp. 1091--1095, IOS Press, 2004.
- [77] S. Somashekhar, M.-J. Sepúlveda, S. Puglielli, A. Norden, E. H. Shortliffe, C. R. Kumar, A. Rauthan, N. A. Kumar, P. Patil, K. Rhee, *et al.*, "Watson for oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board," *Annals of Oncology*, vol. 29, no. 2, pp. 418--423, 2018.
- [78] H. Wang and D. Hu, "Comparison of svm and ls-svm for regression," in *2005 International conference on neural networks and brain*, vol. 1, pp. 279--283, IEEE, 2005.
- [79] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770--778, 2016.
- [80] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.
- [81] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139--144, 2020.
- [82] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature machine intelligence*, vol. 1, no. 5, pp. 206--215, 2019.
- [83] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information fusion*, vol. 58, pp. 82--115, 2020.
- [84] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM computing surveys (CSUR)*, vol. 51, no. 5, pp. 1--42, 2018.
- [85] C. Molnar, *Interpretable machine learning*. Lulu. com, 2020.

- [86] M. Sundararajan and A. Najmi, ``The many shapley values for model explanation," in *International conference on machine learning*, pp. 9269--9278, PMLR, 2020.
- [87] M. Sundararajan, A. Taly, and Q. Yan, ``Axiomatic attribution for deep networks," in *International conference on machine learning*, pp. 3319--3328, PMLR, 2017.
- [88] M. Ancona, E. Ceolini, C. Öztireli, and M. Gross, ``Gradient-based attribution methods," *Explainable AI: Interpreting, explaining and visualizing deep learning*, pp. 169--191, 2019.
- [89] S. Wachter, B. Mittelstadt, and C. Russell, ``Counterfactual explanations without opening the black box: Automated decisions and the gdpr," *Harv. JL & Tech.*, vol. 31, p. 841, 2017.
- [90] S. Verma, J. Dickerson, and K. Hines, ``Counterfactual explanations for machine learning: A review," *arXiv preprint arXiv:2010.10596*, vol. 2, p. 1, 2020.
- [91] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, ``On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PloS one*, vol. 10, no. 7, p. e0130140, 2015.
- [92] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K.-R. Müller, ``Layer-wise relevance propagation: an overview," *Explainable AI: interpreting, explaining and visualizing deep learning*, pp. 193--209, 2019.
- [93] S. M. Lundberg and S.-I. Lee, ``A unified approach to interpreting model predictions," in *Neural Information Processing Systems*, 2017.
- [94] A. V. Ponce-Bobadilla, V. Schmitt, C. S. Maier, S. Mensing, and S. Stodtmann, ``Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development," *Clin. Transl. Sci.*, vol. 17, p. e70056, Nov. 2024.
- [95] L. S. Shapley, ``A value for n-person games," *Contribution to the Theory of Games*, vol. 2, 1953.
- [96] C. Sturgeon, W. T. Shen, O. H. Clark, Q.-Y. Duh, and E. Kebebew, ``Risk assessment in 457 adrenal cortical carcinomas: how much does tumor size predict the likelihood of malignancy?," *Journal of the American College of Surgeons*, vol. 202, no. 3, pp. 423--430, 2006.
- [97] M. Terzolo, A. Ali, G. Osella, and E. Mazza, ``Prevalence of adrenal carcinoma among incidentally discovered adrenal masses: a retrospective study from 1989 to 1994," *Archives of surgery*, vol. 132, no. 8, pp. 914--919, 1997.

- [98] S. Leboulleux, C. Dromain, G. Bonniaud, A. Aupérin, B. Caillou, J. Lumbroso, R. Sigal, E. Baudin, and M. Schlumberger, ``Diagnostic and prognostic value of 18-fluorodeoxyglucose positron emission tomography in adrenocortical carcinoma: a prospective comparison with computed tomography," *The Journal of Clinical Endocrinology & Metabolism*, vol. 91, no. 3, pp. 920--925, 2006.
- [99] D. H. Szolar, M. Korobkin, P. Reittner, A. Berghold, T. Bauernhofer, H. Trummer, H. Schoellnast, K. W. Preidler, and H. Samonigg, ``Adrenocortical carcinomas and adrenal pheochromocytomas: mass and enhancement loss evaluation at delayed contrast-enhanced ct," *Radiology*, vol. 234, no. 2, pp. 479--485, 2005.
- [100] F. Audenet, A. Méjean, E. Chartier-Kastler, and M. Rouprêt, ``Adrenal tumours are more predominant in females regardless of their histological subtype: a review," *World journal of urology*, vol. 31, pp. 1037--1043, 2013.
- [101] E. Sharma, S. Dahal, P. Sharma, A. Bhandari, V. Gupta, B. Amgai, and S. Dahal, ``The characteristics and trends in adrenocortical carcinoma: a united states population based study," *Journal of clinical medicine research*, vol. 10, no. 8, p. 636, 2018.