



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΙΩΑΝΝΙΝΩΝ

ΣΧΟΛΗ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ Πληροφορικής και Τηλεπικοινωνιών

ΠΜΣ Πληροφορικής και Δικτύων

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**ΠΡΟΗΓΜΕΝΕΣ ΤΕΧΝΙΚΕΣ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΓΙΑ ΤΗΝ ΑΝΙΧΝΕΥΣΗ ΤΗΣ ΝΟΣΟΥ
ΤΟΥ ΠΑΡΚΙΝΣΟΝ**

Αιμίλιος Ψαθάς

Επιβλέπων: Νικόλαος Γιαννακέας

Αναπληρωτής Καθηγητής

Άρτα, Απρίλιος, 2025

**ADVANCED DATA ANALYSIS AND MACHINE LEARNING
TECHNIQUES FOR THE DETECTION OF PARKINSON'S
DISEASE**

Εγκρίθηκε από τριμελή εξεταστική επιτροπή

Άρτα, 8-4-2025

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Επιβλέπων καθηγητής

Νικόλαος Γιαννακέας,

Αναπληρωτής Καθηγητής

2. Μέλος επιτροπής

Αλέξανδρος Τζάλλας,

Αναπληρωτής Καθηγητής

3. Μέλος επιτροπής

Ιωάννης Τσούλος,

Καθηγητής

© Ψαθάς, Αιμίλιος, 2025.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Δήλωση μη λογοκλοπής

Δηλώνω υπεύθυνα και γνωρίζοντας τις κυρώσεις του Ν. 2121/1993 περί Πνευματικής Ιδιοκτησίας, ότι η παρούσα μεταπτυχιακή εργασία είναι εκ ολοκλήρου αποτέλεσμα δικής μου ερευνητικής εργασίας, δεν αποτελεί προϊόν αντιγραφής ούτε προέρχεται από ανάθεση σε τρίτους. Όλες οι πηγές που χρησιμοποιήθηκαν (κάθε είδους, μορφής και προέλευσης) για τη συγγραφή της περιλαμβάνονται στη βιβλιογραφία.

Ψαθάς, Αιμίλιος

Υπογραφή

A handwritten signature in black ink, appearing to read 'Aemilios Psathas', written over a horizontal line.

ΕΥΧΑΡΙΣΤΙΕΣ

Εκφράζω θερμή ευγνωμοσύνη στον επιβλέποντα καθηγητή μου, κ. Γιαννακέας, για την πολύτιμη καθοδήγηση και υποστήριξη, καθώς και στην οικογένειά μου για την ατέρμονη αγάπη και ενθάρρυνσή τους. Επίσης, ευχαριστώ τους φίλους και συμφοιτητές μου για τη συντροφιά και τις εποικοδομητικές συζητήσεις.

ΠΕΡΙΛΗΨΗ

Η Νόσος του Πάρκινσον (PD) είναι μια προοδευτική νευροεκφυλιστική διαταραχή, που χαρακτηρίζεται από κινητικά συμπτώματα όπως τρόμος, βραδυκίνησία και ακαμψία. Η έγκαιρη και ακριβής διάγνωση αποτελεί πρόκληση λόγω της υποκειμενικότητας των κλινικών αξιολογήσεων. Η παρούσα μελέτη διερευνά τη συνδυασμένη χρήση τεχνολογιών Διαδικτύου των Αντικειμένων (IoT) και τεχνικών Μηχανικής Μάθησης (ML) για τη βελτίωση της ανίχνευσης και παρακολούθησης της PD. Σχεδιάστηκε ένα έξυπνο γάντι (SmartGlove), εφοδιασμένο με επιταχυνσιόμετρα, γυροσκόπια και αισθητήρες κάμψης, για τη συλλογή δεδομένων κίνησης σε πραγματικό χρόνο από ασθενείς. Το γάντι καταγράφει τα μοτίβα κίνησης κατά τη διάρκεια τυποποιημένων ασκήσεων, επιτρέποντας την ποσοτική ανάλυση των συμπτωμάτων.

Η προεπεξεργασία των δεδομένων περιελάμβανε εξαγωγή χαρακτηριστικών, όπως στατιστικά μεγέθη, χαρακτηριστικά συχνотήτων και δυναμικές παραμέτρους, ακολουθούμενη από αυστηρή επιλογή με t-tests, Random Forest και Ανάλυση Κύριων Συνιστωσών (PCA). Βασικά χαρακτηριστικά, όπως η φασματική εντροπία, το εύρος διατεταρτημορίου (IQR) και οι εκθέτες Lyapunov, αναδείχθηκαν ως σημαντικοί διαχωριστικοί παράγοντες μεταξύ ασθενών και υγιών ατόμων. Μοντέλα μηχανικής μάθησης (Λογιστική Παλινδρόμηση, Μηχανές Διανυσμάτων Υποστήριξης (SVM)), Random Forest και Gradient Boosting, εκπαιδεύτηκαν και αξιολογήθηκαν, με το Random Forest να επιτυγχάνει την υψηλότερη ακρίβεια στην ταξινόμηση των κινητικών διαταραχών.

Τα αποτελέσματα επιβεβαιώνουν ότι οι συσκευές IoT, σε συνδυασμό με ML, μπορούν να ποσοτικοποιήσουν αντικειμενικά τα συμπτώματα της PD, προσφέροντας ένα κλιμακούμενο εργαλείο για απομακρυσμένη παρακολούθηση και έγκαιρη διάγνωση. Μελλοντικές έρευνες θα εστιάσουν στην επικύρωση του συστήματος σε μεγαλύτερες ομάδες ασθενών και στην ενσωμάτωση πολυτροπικών δεδομένων για τη βελτίωση της προβλεπτικής απόδοσης. Η προσέγγιση αυτή διατηρεί υπόσχεση για εξατομικευμένη θεραπεία και βελτιωμένα κλινικά αποτελέσματα.

Λέξεις-κλειδιά: Νόσος του Πάρκινσον, Ανάλυση δεδομένων, Μηχανική μάθηση,
Επιλογή χαρακτηριστικών, Διαδίκτυο των Αντικειμένων

ABSTRACT

Parkinson's Disease (PD) is a progressive neurodegenerative disorder characterized by motor symptoms such as tremors, bradykinesia, and rigidity. Early and accurate diagnosis remains challenging due to the subjective nature of clinical assessments. This study explores the integration of Internet of Things (IoT) technologies and machine learning (ML) techniques to improve PD detection and monitoring. A custom-designed SmartGlove, equipped with accelerometers, gyroscopes, and flex sensors, was developed to collect real-time kinematic data from patients. The glove captures movement patterns during standardized exercises, enabling quantitative analysis of motor symptoms.

Data preprocessing involved feature extraction, including statistical, frequency-domain, and dynamic characteristics, followed by rigorous selection using t-tests, Random Forest, and Principal Component Analysis (PCA). Key features such as spectral entropy, interquartile range (IQR), and Lyapunov exponents were identified as significant discriminators between PD and healthy states. Machine learning models (Logistic Regression, Support Vector Machines (SVM), Random Forest, and Gradient Boosting) were trained and evaluated, with Random Forest achieving the highest accuracy in classifying PD, related motor impairments.

The results demonstrate that IoT-enabled devices, combined with ML, can objectively quantify PD symptoms, offering a scalable tool for remote monitoring and early diagnosis. Future work will focus on validating the system in larger cohorts and integrating multimodal data to enhance predictive performance. This approach holds promise for personalized therapy and improved clinical outcomes in neurodegenerative disease management.

Keywords: Parkinson's Disease, Data Analysis, Machine Learning, Feature Selection, IoT

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΕΥΧΑΡΙΣΤΙΕΣ	iv
ΠΕΡΙΛΗΨΗ	v
ABSTRACT	vii
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ	viii
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	x
ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ/ΕΙΚΟΝΩΝ	xi
ΠΙΝΑΚΑΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ	xii
ΕΙΣΑΓΩΓΗ.....	xiii
1. Ο συνδιασμός του IoT και της μηχανικής μάθησης στη σύγχρονη υγειονομική περίθαλψη.....	1
1.1 IoT σε ιατρικές εφαρμογές	2
1.2 Μηχανική μάθηση σε ιατρικές εφαρμογές	5
1.3 Συνεργασία του IoT και του ML στην υγειονομική περίθαλψη.....	7
1.4 Μελλοντικές τάσεις	11
1.5 Σύνοψη	13
2. Συλλογή δεδομένων και Μεθοδολογία απόκτησης δεδομένων	16
2.1 Σχεδιασμός του SmartGlove	17
2.2 Πλατφόρμα mobPark και Διαχείριση Δεδομένων.....	18
2.3 Πιλοτική εφαρμογή και αρχικά αποτελέσματα	19
2.4 Ασκήσεις με το γάντι και συλλογή δεδομένων	20
3. Επιλογή χαρακτηριστικών και διερευνητική στατιστική	24
3.1 Τεχνικές επιλογής χαρακτηριστικών	25
3.2 Ανάλυση των επιδράσεων της πρόσληψης χαπιών στα εξαγόμενα χαρακτηριστικά....	34
3.3 Δέντρα αποφάσεων και ο ρόλος τους στην ανάλυση της νόσου του Πάρκινσον	38

3.4 Ανάλυση Κύριων Συνιστωσών (PCA) και η Σημασία της στην Ανάλυση της Νόσου του Πάρκινσον.....	43
4. Ανάπτυξη μοντέλου: Εκπαίδευση του μοντέλου μηχανικής μάθησης.....	50
4.1 Εισαγωγή στη μηχανική μάθηση για την ανίχνευση της νόσου του Πάρκινσον	50
4.2 Προετοιμασία δεδομένων.....	52
4.3 Επιλογή μοντέλων	54
4.4 Εκπαίδευση μοντέλων και ρύθμιση υπερπαραμέτρων	57
4.5 Αξιολόγηση Μοντέλου.....	60
4.6 Συζήτηση	68
ΠΑΡΑΡΤΗΜΑ Α	72
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	90

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 3.1 Τα 10 % καλύτερα χαρακτηριστικά που επιλέχθηκαν.....	49
--	----

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ/ΕΙΚΟΝΩΝ

Εικόνα 1. SmartGlove.....	17
Εικόνα 2. Παράδειγμα της 1 ^{ης} άσκησης	20
Εικόνα 3. Παράδειγμα της 2 ^{ης} άσκησης	21
Εικόνα 4. Παράδειγμα της 3 ^{ης} άσκησης	21
Εικόνα 5. Παράδειγμα της 4 ^{ης} άσκησης	22
Εικόνα 1 Ακρίβεια μοντέλων για την Άσκηση 1.....	61
Εικόνα 2 AUC μοντέλων για την Άσκηση 1.....	62
Εικόνα 8 Ακρίβεια μοντέλων για την Άσκηση 2.....	63
Εικόνα 3 AUC μοντέλων για την Άσκηση 2.....	63
Εικόνα 10 Ακρίβεια μοντέλων για την Άσκηση 3.....	64
Εικόνα 4 AUC μοντέλων για την Άσκηση 3.....	65
Εικόνα 12 Ακρίβεια μοντέλων για την Άσκηση 4.....	66
Εικόνα 5 AUC μοντέλων για την Άσκηση 4.....	66

ΠΙΝΑΚΑΣ ΣΥΝΤΟΜΟΓΡΑΦΙΩΝ

PD.....	Νόσος του Πάρκινσον (Parkinson's Disease)
IoT.....	Διαδίκτυο των Αντικειμένων (Internet of Things)
ML	Μηχανική Μάθηση (Machine Learning)
PCA	Ανάλυση Κύριων Συνιστωσών(Principal Component Analysis)
UPDRS	Unified Parkinson's Disease Rating Scale
RMSSD	Root Mean Square of Successive Differences
SVM	Μηχανές Διανυσμάτων Υποστήριξης(Support Vector Machines)
IoMT.....	Διαδίκτυο των Ιατρικών Αντικειμένων (Internet of Medical Things)
RPM.....	Απομακρυσμένη παρακολούθηση ασθενών (Remote Patient Monitoring)
DL.....	Βαθιά Μάθηση (Deep Learning)
RL.....	Ενισχυτική Μάθηση(Reinforcement Learning)
SoC.....	System on Chip

ΕΙΣΑΓΩΓΗ

Η νόσος του Πάρκινσον (PD) είναι μια προοδευτική νευροεκφυλιστική διαταραχή που χαρακτηρίζεται από κινητικά συμπτώματα όπως τρέμουλο , βραδυκινησία, ακαμψία και αστάθεια στάσης, παράλληλα με μη κινητικά συμπτώματα όπως οι διαταραχές της διάθεσης και οι διαταραχές του ύπνου. Παρά την πολύπλοκη συμπτωματολογία της, η διάγνωση της PD παραμένει σε μεγάλο βαθμό κλινική, καθώς δεν υπάρχουν οριστικές βιολογικές εξετάσεις για την έγκαιρη ανίχνευση. Η έγκαιρη διάγνωση είναι κρίσιμη, αλλά η σταδιακή εμφάνιση των συμπτωμάτων οδηγεί συχνά σε λανθασμένη διάγνωση ή καθυστερημένη θεραπεία. Αυτές οι καθυστερήσεις εμποδίζουν την αποτελεσματική διαχείριση και επιδεινώνουν τα αποτελέσματα των ασθενών και την βελτίωση της ποιότητας ζωής τους (Jankovic, 2008).

Οι πρόσφατες τεχνολογικές εξελίξεις, ιδίως στους τομείς του Διαδικτύου των Αντικειμένων (IoT) και της Μηχανικής Μάθησης (ML), προσφέρουν νέες ελπιδοφόρες οδούς για την έγκαιρη ανίχνευση και παρακολούθηση της PD. Οι συσκευές IoT, όπως τα έξυπνα ρολόγια, οι φορητοί αισθητήρες και άλλα εργαλεία παρακολούθησης της υγείας του ανθρώπου, επιτρέπουν τη συνεχή, μη επεμβατική παρακολούθηση των κινητικών λειτουργιών, των συχνотήτων τρέμουλου και των ανωμαλιών βάδισης σε πραγματικό χρόνο. Αυτές οι συσκευές καταγράφουν τεράστιες ποσότητες δεδομένων εκτός των παραδοσιακών κλινικών περιβαλλόντων, καθιστώντας δυνατή την αξιολόγηση της κατάστασης των ασθενών στην καθημερινή ζωή. Με την εφαρμογή τεχνικών σε αυτά τα δεδομένα, μπορούν να ανιχνευθούν με αυξημένη ακρίβεια πρώιμα σημάδια της νόσου του Πάρκινσον, οδηγώντας σε έγκαιρες παρεμβάσεις (Jankovic, 2008).

Προσεγγίσεις μηχανικής μάθησης για την ανίχνευση της νόσου του Πάρκινσον

Η Μηχανική Μάθηση εξελίσσεται ταχύτατα σε ένα παιχνίδι που αλλάζει τα δεδομένα στην ανίχνευση της PD. Οι αλγόριθμοι της ML είναι ικανοί να επεξεργάζονται μεγάλα σύνολα δεδομένων από συσκευές IoT και να εξάγουν σημαντικές πληροφορίες. Αυτά τα μοντέλα μπορούν να αναλύσουν δεδομένα ομιλίας, κινητικές λειτουργίες και άλλα μη κινητικά συμπτώματα για τον

εντοπισμό λεπτών μοτίβων που μπορεί να προηγηθούν πριν την πλήρως ανάπτυξη της νόσου (Pradyumna et al., 2024).

Για παράδειγμα, μια μελέτη των (Bukhari and Ogudo, 2024) πρότεινε ένα μοντέλο ML που χρησιμοποιεί σήματα ομιλίας για την ανίχνευση της πρώιμης PD. Η προσέγγισή τους σε σύνολο, βασισμένη στον ταξινομητή AdaBoost, ανέλυσε χαρακτηριστικά της φωνής, όπως οι συντελεστές Μελ-frequency cepstrum (Abdul and Al-Talabani, 2022) και οι μετασχηματισμοί Wavelet (Antoine, 2003), για να διαφοροποιήσει τους ασθενείς με PD από τα υγιή άτομα. Το μοντέλο επέδειξε υψηλή ακρίβεια, ακρίβεια και ανθεκτικότητα μέσω διασταυρούμενης επικύρωσης, καθιστώντας το ένα πολλά υποσχόμενο εργαλείο για την έγκαιρη ανίχνευση της νόσου (Javaid et al., 2022; Li et al., 2024).

Συσκευές IoT και παρακολούθηση σε πραγματικό χρόνο

Οι συσκευές IoT βελτιώνουν τις παραδοσιακές κλινικές αξιολογήσεις προσφέροντας συνεχή παρακολούθηση, παρέχοντας πιο λεπτομερείς πληροφορίες για την κατάσταση ενός ασθενούς με την πάροδο του χρόνου. Αυτές οι συσκευές παρακολουθούν βασικές κινητικές λειτουργίες, όπως η ταχύτητα βάρδιας, η ένταση του τρόμου και η ισορροπία, οι οποίες μπορούν να επεξεργαστούν με αλγόριθμους ML για την ανίχνευση αποκλίσεων από τα φυσιολογικά πρότυπα που είναι ενδεικτικά της εξέλιξης της νόσου PD. Αυτά τα δεδομένα σε πραγματικό χρόνο είναι ανεκτίμητα για τη δημιουργία μιας πιο ολιστικής κατανόησης του τρόπου με τον οποίο η PD εκδηλώνεται σε μεμονωμένους ασθενείς, γεγονός που μπορεί να οδηγήσει σε πιο έγκαιρες και πιο εξατομικευμένες παρεμβάσεις (Javaid et al., 2022; Li et al., 2024).

Μη κινητικά συμπτώματα και έγκαιρη ανίχνευση

Πέρα από τα κινητικά συμπτώματα, τα μη κινητικά χαρακτηριστικά, όπως οι γνωστικές διαταραχές, οι διαταραχές του ύπνου και οι αλλαγές στη διάθεση, διαδραματίζουν κρίσιμο ρόλο στη διάγνωση της PD. Οι τεχνικές ML χρησιμοποιούνται πλέον για την ανάλυση αυτών των μη κινητικών συμπτωμάτων, ιδίως στα αρχικά τους στάδια. Για παράδειγμα, η διαταραχή της συμπεριφοράς του ύπνου REM (Peever and Fuller, 2017) και η απώλεια της αίσθησης της όσφρησης είναι γνωστά πρόδρομα συμπτώματα της νόσου, και τα μοντέλα ML που

εκπαιδεύονται σε αυτά τα δεδομένα μπορούν να βοηθήσουν στον εντοπισμό ατόμων υψηλού κινδύνου χρόνια πριν την εμφάνιση κινητικών συμπτωμάτων παρεμβάσεις (Javaid et al., 2022; Li et al., 2024). Με την ενσωμάτωση διαφορετικών συνόλων δεδομένων, τα μοντέλα ML προσφέρουν μια πιο ολοκληρωμένη εικόνα της PD, επιτρέποντας την έγκαιρη διάγνωση σε σχέση με τις παραδοσιακές μεθόδους παρεμβάσεις (Javaid et al., 2022; Li et al., 2024).

Προκλήσεις και μελλοντικές κατευθύνσεις

Ενώ η ενσωμάτωση του IoT και της ML υπόσχεται πολλά για τη διαχείριση της PD, υπάρχουν ακόμη προκλήσεις που πρέπει να αντιμετωπιστούν. Ένα σημαντικό ζήτημα είναι η μεταβλητότητα των συμπτωμάτων μεταξύ των ασθενών, η οποία μπορεί να καταστήσει δύσκολη τη δημιουργία γενικευμένων μοντέλων που λειτουργούν σε διαφορετικούς πληθυσμούς. Υψηλής ποιότητας, αντιπροσωπευτικά σύνολα δεδομένων είναι απαραίτητα για την εκπαίδευση αποτελεσματικών μοντέλων ML, ωστόσο η πρόσβαση σε τέτοια δεδομένα δέχεται περιορισμένες παρεμβάσεις (Javaid et al., 2022; Li et al., 2024).

Μια άλλη πρόκληση είναι η ερμηνευσιμότητα των μοντέλων. Οι επιβλέπων γιατροί πρέπει να κατανοούν πώς τα μοντέλα ML καταλήγουν στα συμπεράσματά τους για να λαμβάνουν τεκμηριωμένες αποφάσεις. Η διασφάλιση ότι τα μοντέλα αυτά είναι διαφανή και κατανοητά είναι ζωτικής σημασίας για την ευρεία υιοθέτησή τους στην κλινική πρακτική των παρεμβάσεων (Javaid et al., 2022; Li et al., 2024).

Προχωρώντας προς τα εμπρός, ο συνδυασμός δεδομένων ομιλίας, κινητικών δεδομένων από συσκευές IoT και η ανάλυση μη κινητικών συμπτωμάτων σε πολυτροπικά μοντέλα ML ενέχει τεράστιες δυνατότητες για την προώθηση της ανίχνευσης της PD. Αυτές οι καινοτομίες θα μπορούσαν να ανοίξουν το δρόμο για εξατομικευμένα σχέδια θεραπείας, βελτιστοποιημένη παρακολούθηση της νόσου και καλύτερα αποτελέσματα για τους ασθενείς παρεμβάσεις (Javaid et al., 2022; Li et al., 2024).

Επιρροή τεχνολογιών IoT και ML

Η ανάπτυξη των τεχνολογιών IoT και ML έχει φέρει επανάσταση στην υγειονομική περίθαλψη, προσφέροντας νέες δυνατότητες για την έγκαιρη διάγνωση και διαχείριση της PD. Οι φορητές συσκευές όπως τα έξυπνα ρολόγια, οι αισθητήρες κίνησης και οι συσκευές παρακολούθησης ύπνου συλλέγουν βιομετρικά δεδομένα που τροφοδοτούνται σε αλγορίθμους ML. Αυτές οι τεχνολογίες επιτρέπουν τη συνεχή παρακολούθηση των ασθενών, βοηθώντας στην ανίχνευση της εξέλιξης της νόσου σε πραγματικό χρόνο (Javaid et al., 2022; Li et al., 2024).

Συμπέρασμα

Η εφαρμογή προηγμένων τεχνικών ανάλυσης δεδομένων και ML, σε συνδυασμό με το IoT, φέρνει επανάσταση στον τρόπο διάγνωσης και διαχείρισης της PD. Αξιοποιώντας συσκευές IoT για τη συλλογή δεδομένων σε πραγματικό χρόνο σχετικά με κινητικά και μη κινητικά συμπτώματα και εφαρμόζοντας μοντέλα ML για την ανάλυση αυτών των συνόλων δεδομένων, οι κλινικοί γιατροί αποκτούν νέα εργαλεία για την έγκαιρη ανίχνευση και την εξατομικευμένη θεραπεία. Αυτή η τεχνολογία όχι μόνο βελτιώνει τη διαγνωστική ακρίβεια, αλλά δίνει επίσης τη δυνατότητα προληπτικής διαχείρισης της νόσου PD, βελτιώνοντας τελικά την ποιότητα ζωής όσων πάσχουν από αυτή την εξουθενωτική ασθένεια παρεμβάσεις (Javaid et al., 2022; Li et al., 2024).

1. Ο συνδιασμός του IoT και της μηχανικής μάθησης στη σύγχρονη υγειονομική περίθαλψη

Ο ψηφιακός μετασχηματισμός της σύγχρονης υγειονομικής περίθαλψης καθοδηγείται από τις ραγδαίες τεχνολογικές εξελίξεις, με το IoT και τη ML να διαδραματίζουν καθοριστικό ρόλο. Το IoT αναφέρεται σε συνδεδεμένα συστήματα φυσικών ενσωματωμένων συσκευών με αισθητήρες, λογισμικό και συνδεσιμότητα στο δίκτυο, που τους επιτρέπουν να συλλέγουν, να μεταδίδουν και να ανταλλάσσουν δεδομένα (Atzori et al., 2010).

Στην υγειονομική περίθαλψη, αυτό μεταφράζεται ως ένα δίκτυο φορητών αισθητήρων, εργαλείων απομακρυσμένης παρακολούθησης, έξυπνων εμφυτευμάτων και άλλων συνδεδεμένων συσκευών που παράγουν δεδομένα σε πραγματικό χρόνο, δίνοντας τη δυνατότητα λήψης αποφάσεων βάσει δεδομένων (Giannakopoulou et al., 2022).

Συμπληρώνοντας το IoT, η ML χρησιμοποιεί αλγορίθμους που μαθαίνουν από τα δεδομένα χωρίς ρητό προγραμματισμό, εντοπίζοντας μοτίβα, κάνοντας προβλέψεις και βελτιώνοντας την απόδοσή τους με την πάροδο του χρόνου. Η σύγκλιση του IoT και της ML δημιουργεί ένα ισχυρό οικοσύστημα που μετατοπίζει την υγειονομική περίθαλψη από ένα αντιδραστικό, εστιασμένο στη θεραπεία μοντέλο σε ένα προληπτικό, προγνωστικό, προληπτικό και εξατομικευμένο παράδειγμα (Mitchell, 1997).

Αυτή η συνέργεια είναι ιδιαίτερα σημαντική για τη διαχείριση χρόνιων ασθενειών όπως η PD, προσφέροντας δυνατότητες για έγκαιρη ανίχνευση, ακριβή παρακολούθηση, προσαρμοσμένες παρεμβάσεις και βελτιωμένα αποτελέσματα για τους ασθενείς. Το παρόν κεφάλαιο διερευνά τις εφαρμογές του IoT και της ML στην υγειονομική περίθαλψη, με ιδιαίτερη έμφαση στη συνδυασμένη επίδρασή τους στη διαχείριση της PD (Giannakopoulou et al., 2022).

1.1 IoT σε ιατρικές εφαρμογές

1.1.1 Ορισμός και πεδίο εφαρμογής του IoT στην υγειονομική περίθαλψη

Η εφαρμογή του IoT στην υγειονομική περίθαλψη, που συχνά αποκαλείται Διαδίκτυο των Ιατρικών Αντικειμένων (IoMT), περιλαμβάνει ένα δίκτυο διασυνδεδεμένων ιατρικών συσκευών, ασθενών και παρόχων υγειονομικής περίθαλψης που διευκολύνεται από προηγμένα δίκτυα επικοινωνίας (Singh et al., 2017).

Το δίκτυο αυτό περιλαμβάνει φορητές συσκευές, διαγνωστικά εργαλεία, πλατφόρμες ενσωμάτωσης δεδομένων και άλλα συνδεδεμένα συστήματα που αλληλεπιδρούν απρόσκοπτα για να επιτρέπουν την παρακολούθηση της υγείας σε πραγματικό χρόνο και τη λήψη τεκμηριωμένων αποφάσεων. Το IoMT στοχεύει στη μείωση του ανθρώπινου σφάλματος, στην αυτοματοποίηση επαναλαμβανόμενων εργασιών, στη βελτίωση των αποτελεσμάτων των ασθενών και στη βελτίωση της αποτελεσματικότητας της παροχής υγειονομικής περίθαλψης.

Οι συσκευές IoT συλλέγουν ένα ευρύ φάσμα μετρήσεων υγείας, από βασικούς δείκτες όπως ο καρδιακός ρυθμός και ο κορεσμός οξυγόνου έως πιο σύνθετες παραμέτρους όπως τα πρότυπα βάδισης, η ποιότητα του ύπνου και οι νευρολογικοί δείκτες, οι οποίοι είναι ιδιαίτερα σημαντικοί για καταστάσεις όπως η PD (Ray, 2018). Ο συνεχής βρόχος ανατροφοδότησης που επιτρέπει το IoT διευκολύνει τις δυναμικές προσαρμογές των σχεδίων φροντίδας με βάση τις ατομικές ανάγκες των ασθενών.

1.1.2 Βασικές εφαρμογές του IoT στην υγειονομική περίθαλψη

Οι βασικές εφαρμογές του IoT στην υγειονομική περίθαλψη είναι οι εξής:

- **Απομακρυσμένη παρακολούθηση ασθενών (RPM):** Η RPM χρησιμοποιεί φορητούς αισθητήρες, βιοαισθητήρες και εξοπλισμό παρακολούθησης στο σπίτι για τη συνεχή συλλογή και μετάδοση δεδομένων υγείας του ασθενούς, όπως τα επίπεδα δραστηριότητας, η μεταβλητότητα του καρδιακού ρυθμού, τα

πρότυπα ύπνου και τα ζωτικά σημεία (Arora and Meumeu Yomsi, 2019). Στο πλαίσιο της νόσου PD, τα συστήματα RPM μπορούν να παρακολουθούν μετρήσεις κινητικής λειτουργίας, συχνότητα τρόμου, παραμέτρους βάδισης και καθημερινές δραστηριότητες, παρέχοντας στους κλινικούς γιατρούς πολύτιμες πληροφορίες για την εξέλιξη της νόσου και επιτρέποντας προληπτικές παρεμβάσεις. Μελέτες έχουν δείξει ότι οι φορητές τεχνολογίες μπορούν να καταγράψουν λεπτές αλλαγές στα κινητικά πρότυπα που μπορεί να μην είναι άμεσα παρατηρήσιμες κατά τη διάρκεια κλινικών επισκέψεων ρουτίνας, διευκολύνοντας την έγκαιρη διάγνωση και τα βελτιστοποιημένα σχέδια θεραπείας (Liu et al., 2016).

- **Τηλεϊατρική:** Οι πλατφόρμες τηλεϊατρικής με τη βοήθεια του IoT γεφυρώνουν γεωγραφικά και υλικοτεχνικά εμπόδια, επιτρέποντας εξ αποστάσεως διαβουλεύσεις, διαγνώσεις και εκπαίδευση των ασθενών (Dorsey and Topol, 2016). Αυτές οι πλατφόρμες διευκολύνουν τις αξιολογήσεις σε πραγματικό χρόνο με βάση δεδομένα που συλλέγονται από συσκευές IoT, ωφελώντας ασθενείς σε υποβαθμισμένες περιοχές ή ασθενείς με περιορισμούς κινητικότητας. Στη διαχείριση της PD, η τηλεϊατρική μπορεί να παρέχει συνεχή ανατροφοδότηση σχετικά με τη σοβαρότητα των συμπτωμάτων, την αποτελεσματικότητα της φαρμακευτικής αγωγής και την τήρηση των θεραπευτικών σχεδίων.
- **Έξυπνα συστήματα διαχείρισης νοσοκομείων:** Πέρα από την άμεση φροντίδα των ασθενών, το IoT βελτιστοποιεί τις λειτουργίες των νοσοκομείων μέσω διασυνδεδεμένων συστημάτων που παρακολουθούν τη διαθεσιμότητα κλινών, παρακολουθούν τον ιατρικό εξοπλισμό, διαχειρίζονται τα αποθέματα και ελέγχουν περιβαλλοντικούς παράγοντες όπως η ποιότητα του αέρα και η θερμοκρασία (Gubbi et al., 2013). Για χρόνιες παθήσεις όπως η PD, οι οποίες μπορεί να απαιτούν συχνές νοσηλείες, τα συστήματα IoT μπορούν να βελτιώσουν τη χρήση των πόρων, να μειώσουν τους χρόνους αναμονής και να βελτιώσουν τη συνολική ικανοποίηση των ασθενών.
- **Συμμόρφωση με τη φαρμακευτική αγωγή και αυτοματοποιημένη χορήγηση φαρμάκων:** Η τήρηση της φαρμακευτικής αγωγής είναι ζωτικής σημασίας για τη διαχείριση χρόνιων παθήσεων όπως η PD. Οι διανομείς φαρμάκων με δυνατότητα IoT, τα έξυπνα μπουκάλια χαπιών και τα συστήματα

υπενθύμισης διασφαλίζουν την έγκαιρη λήψη φαρμάκων και παρέχουν δεδομένα σχετικά με τα πρότυπα συμμόρφωσης (Patel et al., 2009). Τα προηγμένα συστήματα μπορούν ακόμη και να προσαρμόζουν τις δοσολογίες με βάση βιομετρική ανατροφοδότηση σε πραγματικό χρόνο, βελτιστοποιώντας τα θεραπευτικά αποτελέσματα και ελαχιστοποιώντας τις ανεπιθύμητες ενέργειες.

- **Επιδημιολογία χρόνιων ασθενειών και ανάλυση πρόβλεψης:** Τα συστήματα IoT παράγουν μεγάλα σύνολα δεδομένων υγείας σε επίπεδο πληθυσμού, παρέχοντας πολύτιμες πληροφορίες σχετικά με τις τάσεις χρόνιων ασθενειών, τους παράγοντες κινδύνου και την αποτελεσματικότητα των παρεμβάσεων δημόσιας υγείας (López-Martínez et al., 2020). Στο πλαίσιο της PD, τα εν λόγω σύνολα δεδομένων μπορούν να συμβάλουν στην κατανόηση των περιβαλλοντικών ή γενετικών παραγόντων που συμβάλλουν στην ασθένεια και να καθοδηγήσουν στοχευμένα προληπτικά μέτρα.

1.1.3 Οφέλη του IoT στην υγειονομική περίθαλψη

Η ενσωμάτωση των τεχνολογιών IoT προσφέρει πολλά βασικά οφέλη:

- **Ενισχυμένη προσβασιμότητα:** Επεκτείνει την πρόσβαση σε υπηρεσίες υγειονομικής περίθαλψης για ασθενείς σε απομακρυσμένες ή αγροτικές περιοχές, μειώνοντας την ανάγκη για συχνές προσωπικές επισκέψεις. Αυτό είναι ιδιαίτερα σημαντικό για πληθυσμούς με περιορισμένη πρόσβαση σε ειδικούς ή για όσους ζουν σε γεωγραφικά απομονωμένες περιοχές. Οι λύσεις τηλεϊατρικής που υποστηρίζονται από συσκευές IoT επιτρέπουν την εξ αποστάσεως διαβούλευση, παρακολούθηση, ακόμη και την εξ αποστάσεως παροχή φροντίδας (Bashshur et al., 2020).
- **Διαπιστώσεις με βάση τα δεδομένα:** Επιτρέπει τη συνεχή παρακολούθηση και συλλογή δεδομένων, παρέχοντας στους κλινικούς γιατρούς πολύτιμες πληροφορίες για τη λήψη τεκμηριωμένων αποφάσεων και εξατομικευμένων σχεδίων θεραπείας. Η συνεχής ροή δεδομένων από τις συσκευές IoT επιτρέπει την ανάλυση της κατάστασης της υγείας των ασθενών σε πραγματικό χρόνο, επιτρέποντας την έγκαιρη ανίχνευση ανωμαλιών, την παρακολούθηση της αποτελεσματικότητας της θεραπείας και την προσαρμογή των σχεδίων

περίθαλψης ανάλογα με τις ανάγκες. Αυτή η προσέγγιση με βάση τα δεδομένα διευκολύνει την πιο προληπτική και εξατομικευμένη φροντίδα (Li et al., 2024).

- **Μείωση του κόστους:** Ελαχιστοποιεί τις επανεισαγωγές στο νοσοκομείο, μειώνει την ανάγκη για προσωπικές διαβουλεύσεις και βελτιστοποιεί τη χρήση των πόρων, συμβάλλοντας στη μείωση του κόστους υγειονομικής περίθαλψης. Επιτρέποντας την απομακρυσμένη παρακολούθηση και την έγκαιρη παρέμβαση, το IoT μπορεί να αποτρέψει δαπανηρές νοσηλείες και επισκέψεις στα επείγοντα περιστατικά. Επιπλέον, τα συστήματα που βασίζονται στο IoT μπορούν να βελτιστοποιήσουν τις νοσοκομειακές λειτουργίες, όπως η διαχείριση των κλινών και ο έλεγχος των αποθεμάτων, οδηγώντας σε περαιτέρω εξοικονόμηση κόστους (Tan et al., 2024).
- **Εξατομικευμένη φροντίδα:** Διευκολύνει την ανάπτυξη εξατομικευμένων θεραπειών και παρεμβάσεων με βάση τις ατομικές ανάγκες και συνθήκες των ασθενών, βελτιώνοντας τη συμμόρφωση με τη θεραπεία και τα αποτελέσματα. Οι συσκευές IoT επιτρέπουν τη συλλογή λεπτομερών δεδομένων για μεμονωμένους ασθενείς, συμπεριλαμβανομένων των φυσιολογικών παραμέτρων, των επιπέδων δραστηριότητας, των προτύπων ύπνου και της συμμόρφωσης με τη φαρμακευτική αγωγή. Τα δεδομένα αυτά μπορούν να χρησιμοποιηθούν για την ανάπτυξη εξατομικευμένων σχεδίων θεραπείας, την παροχή στοχευμένων παρεμβάσεων και την παρακολούθηση της αποτελεσματικότητας της θεραπείας (Swan, 2012).

1.2 Μηχανική μάθηση σε ιατρικές εφαρμογές

1.2.1 Σύντομη επισκόπηση της μηχανικής μάθησης στην ιατρική

Η ML είναι μια μετασχηματιστική τεχνολογία στον τομέα της υγειονομικής περίθαλψης, η οποία χρησιμοποιεί υπολογιστικούς αλγορίθμους για την ανάλυση μεγάλων συνόλων δεδομένων, τον εντοπισμό σύνθετων μοτίβων και τη λήψη προβλέψεων ή αποφάσεων (Jiang et al., 2017).

Μαθαίνοντας από τα δεδομένα, τα μοντέλα ML μπορούν να αντιμετωπίσουν προκλήσεις στη διάγνωση, τη βελτιστοποίηση της θεραπείας, την ανακάλυψη

φαρμάκων και τη διαχείριση ασθενειών. Σε αντίθεση με τον παραδοσιακό προγραμματισμό, όπου ορίζονται ρητοί κανόνες, τα συστήματα ML μαθαίνουν από τα δεδομένα, βελτιώνοντας την ακρίβεια και την απόδοσή τους καθώς επεξεργάζονται περισσότερες πληροφορίες.

1.2.2 Βασικές τεχνικές ML που χρησιμοποιούνται στην υγειονομική περίθαλψη

Οι βασικές τεχνικές της ML είναι:

- **Μάθηση με επίβλεψη:** Τα ζεύγη εισόδου-εξόδου ορίζονται ρητά. Οι συνήθεις αλγόριθμοι περιλαμβάνουν τη γραμμική παλινδρόμηση, τη λογιστική παλινδρόμηση, τα δέντρα αποφάσεων, τα τυχαία δάση, τις μηχανές διανυσμάτων υποστήριξης (SVM) και τα νευρωνικά δίκτυα (Jordan and Mitchell, 2015).
- **Μάθηση χωρίς επίβλεψη:** Προσδιορίζει κρυφά μοτίβα και δομές σε μη επισημειωμένα σύνολα δεδομένων, που χρησιμοποιούνται συχνά για διερευνητική ανάλυση δεδομένων, ομαδοποίηση και μείωση διαστάσεων. Παραδείγματα περιλαμβάνουν την ομαδοποίηση k-means, την ανάλυση κύριων συνιστωσών (PCA) και την εξόρυξη κανόνων συσχέτισης (Bishop, 2006).
- **Βαθιά μάθηση (DL):** Ένα υποπεδίο της ML που χρησιμοποιεί τεχνητά νευρωνικά δίκτυα με πολλαπλά επίπεδα (βαθιά νευρωνικά δίκτυα) για την επεξεργασία σύνθετων δεδομένων, όπως εικόνες, κείμενο και δεδομένα χρονοσειρών. Η DL έχει σημειώσει αξιοσημείωτη επιτυχία στην ανάλυση ιατρικών εικόνων, στην επεξεργασία φυσικής γλώσσας και στη γονιδιωματική (LeCun et al., 2015).
- **Ενισχυτική μάθηση (RL):** Εκπαιδεύει έναν πράκτορα να αλληλοεπιδρά με ένα περιβάλλον και να μαθαίνει τις βέλτιστες ενέργειες μέσω δοκιμής και σφάλματος, καθοδηγούμενος από ανταμοιβές και ποινές. Η RL έχει πιθανές εφαρμογές στον εξατομικευμένο σχεδιασμό θεραπείας, στη ρομποτική χειρουργική και στην ανακάλυψη φαρμάκων (Sutton and Barto, 2018).
- **Μάθηση μεταφοράς:** Αξιοποιεί τη γνώση που αποκτάται από την εκπαίδευση ενός μοντέλου σε μια εργασία για τη βελτίωση της απόδοσης σε μια συναφή

εργασία, ιδιαίτερα χρήσιμη όταν τα επισημασμένα δεδομένα είναι περιορισμένα.

1.2.3 Περιπτώσεις χρήσης της ML στη διάγνωση ασθενειών, την πρόγνωση και την εξατομικευμένη ιατρική

Μερικές περιπτώσεις όπου η χρήση της ML βοηθάει στην ιατρική είναι:

- **Διάγνωση της νόσου:** Τα μοντέλα ML αναλύουν ιατρικά δεδομένα, όπως απεικονιστικά (ακτίνες X, μαγνητικές τομογραφίες, αξονικές τομογραφίες), ηλεκτρονικά αρχεία υγείας, γονιδιωματικά δεδομένα και βιοσήματα, για την ανίχνευση ασθενειών με μεγάλη ακρίβεια και ενδεχομένως σε πρώιμα στάδια (Esteva et al., 2017).
- **Πρόγνωση ασθενειών:** Τα προγνωστικά μοντέλα ML αξιολογούν την εξέλιξη της νόσου, προβλέπουν τις μελλοντικές εκβάσεις και εντοπίζουν ασθενείς με υψηλό κίνδυνο επιπλοκών με βάση διαχρονικά δεδομένα υγείας και παράγοντες κινδύνου (Yu et al., 2018).
- **Εξατομικευμένη ιατρική:** Οι αλγόριθμοι ML προσαρμόζουν τις στρατηγικές θεραπείας στα ατομικά προφίλ των ασθενών με βάση τις γενετικές πληροφορίες, τους παράγοντες του τρόπου ζωής και την ανταπόκριση σε προηγούμενες θεραπείες, βελτιστοποιώντας τη θεραπευτική αποτελεσματικότητα και ελαχιστοποιώντας τις ανεπιθύμητες ενέργειες (Ashley, 2015).
- **Εφαρμογές δημόσιας υγείας:** Η ML υποστηρίζει την παρακολούθηση της υγείας σε επίπεδο πληθυσμού, την επιτήρηση ασθενειών, την πρόβλεψη κρουσμάτων και την αξιολόγηση παρεμβάσεων δημόσιας υγείας (Ashley, 2015).

1.3 Συνεργασία του IoT και του ML στην υγειονομική περίθαλψη

Η πραγματική ισχύς του IoT στην υγειονομική περίθαλψη γίνεται αντιληπτή όταν συνδυάζεται με τις αναλυτικές δυνατότητες της ML. Οι συσκευές IoT λειτουργούν ως το μπροστινό άκρο συλλογής δεδομένων, ενώ το ML παρέχει την ευφυΐα για

την εξαγωγή ουσιαστικών πληροφοριών από τις τεράστιες ποσότητες δεδομένων που παράγονται. Αυτή η συνέργεια επιτρέπει προληπτικές, προγνωστικές και εξατομικευμένες παρεμβάσεις υγειονομικής περίθαλψης.

1.3.1 Πώς το IoT παράγει δεδομένα για μοντέλα ML

Οι συσκευές IoT παράγουν μια συνεχή ροή δεδομένων σε πραγματικό χρόνο που χρησιμεύουν ως βάση για τα μοντέλα ML στην υγειονομική περίθαλψη. Τα δεδομένα αυτά μπορούν να κατηγοριοποιηθούν ως εξής:

- **Δομημένα δεδομένα:** Αυτός ο τύπος δεδομένων είναι οργανωμένος σε μια προκαθορισμένη μορφή, καθιστώντας τα εύκολα αναζητήσιμα και αναλύσιμα. Παραδείγματα περιλαμβάνουν ζωτικές ενδείξεις (καρδιακός ρυθμός, αρτηριακή πίεση, θερμοκρασία), ενδείξεις αισθητήρων από φορητές συσκευές (π.χ. βήματα που πραγματοποιήθηκαν, διάρκεια ύπνου) και εργαστηριακά αποτελέσματα.
- **Μη δομημένα δεδομένα:** Αυτά τα δεδομένα δεν έχουν προκαθορισμένη μορφή και είναι πιο δύσκολο να αναλυθούν. Παραδείγματα περιλαμβάνουν ιατρικές εικόνες (ακτινογραφίες, μαγνητικές τομογραφίες, αξονικές τομογραφίες), κλινικές σημειώσεις σε ηλεκτρονικούς φακέλους υγείας και ηχογραφήσεις αλληλεπιδράσεων μεταξύ ασθενούς και γιατρού.

Τα δεδομένα αυτά χαρακτηρίζονται από τα «3 Vs» των μεγάλων δεδομένων:

- **Όγκος:** Ο καθαρός όγκος των δεδομένων που παράγονται από τις συσκευές IoT είναι τεράστιος και αυξάνεται συνεχώς.
- **Ταχύτητα:** Τα δεδομένα παράγονται και μεταδίδονται με υψηλή ταχύτητα, απαιτώντας δυνατότητες επεξεργασίας σε πραγματικό χρόνο.
- **Ποικιλία:** Τα δεδομένα έρχονται σε διάφορες μορφές (δομημένες και μη δομημένες), απαιτώντας εξελιγμένες τεχνικές ολοκλήρωσης και ανάλυσης δεδομένων.

Η διαδικασία του IoT που παράγει δεδομένα για μοντέλα ML περιλαμβάνει διάφορα βασικά βήματα:

- **Συλλογή δεδομένων σε πραγματικό χρόνο:** Οι συσκευές IoT συλλέγουν δεδομένα συνεχώς, παρέχοντας δυναμικά σύνολα δεδομένων που αντικατοπτρίζουν τις αλλαγές στην κατάσταση της υγείας των ασθενών σε πραγματικό χρόνο. Για παράδειγμα, οι φορητοί αισθητήρες μπορούν να παρακολουθούν τη μεταβλητότητα του καρδιακού ρυθμού ενός ασθενούς καθ' όλη τη διάρκεια της ημέρας, παρέχοντας μια πιο ολοκληρωμένη εικόνα της καρδιαγγειακής του υγείας από ό,τι μια μεμονωμένη μέτρηση που λαμβάνεται σε μια κλινική (Singh et al., 2017).
- **Μετάδοση και ενσωμάτωση δεδομένων:** Οι συσκευές IoT μεταδίδουν δεδομένα ασύρματα (π.χ. με χρήση Bluetooth, Wi-Fi, κυψελοειδών δικτύων) σε πλατφόρμες cloud ή σε συσκευές υπολογισμού άκρων. Οι πλατφόρμες νέφους προσφέρουν κλιμακούμενες δυνατότητες αποθήκευσης και επεξεργασίας, ενώ η υπολογιστική ακμής επιτρέπει την ταχύτερη ανάλυση των δεδομένων πιο κοντά στην πηγή, μειώνοντας την καθυστέρηση. Στη συνέχεια, τα δεδομένα συγκεντρώνονται, υποβάλλονται σε προεπεξεργασία και ενσωματώνονται με άλλες πηγές δεδομένων, όπως τα EHR, τα γονιδιωμικά δεδομένα και τα δημογραφικά στοιχεία των ασθενών (Meskó et al., 2018).
- **Προκλήσεις στην προετοιμασία δεδομένων:** Τα δεδομένα IoT μπορεί να είναι θορυβώδη (περιέχουν σφάλματα ή άσχετες πληροφορίες), ελλιπή (ελλείπουσες τιμές) ή ασυνεπή (διαφορετικές μορφές ή μονάδες δεδομένων). Αυτό απαιτεί προσεκτική προεπεξεργασία δεδομένων, η οποία περιλαμβάνει βήματα όπως:
 - **Καθαρισμός δεδομένων:** Αφαίρεση σφαλμάτων, ασυνέπειας και ελλιπών τιμών.
 - **Κανονικοποίηση:** Κλιμάκωση των δεδομένων σε ένα τυπικό εύρος για να αποτραπεί η κυριαρχία χαρακτηριστικών με μεγαλύτερες τιμές στο μοντέλο ML.
 - **Εξαγωγή χαρακτηριστικών:** Προσδιορισμός σχετικών χαρακτηριστικών από τα ακατέργαστα δεδομένα που μπορούν να χρησιμοποιηθούν από το μοντέλο ML.
 - **Ενίσχυση δεδομένων:** Δημιουργία συνθετικών δεδομένων για την αύξηση του μεγέθους και της ποικιλομορφίας του συνόλου δεδομένων

εκπαίδευσης, βελτιώνοντας την ανθεκτικότητα του μοντέλου. (Chen et al., 2018)

1.3.2 Παραδείγματα ενσωμάτωσης IoT-ML σε ιατρικά σενάρια πραγματικού κόσμου

Η ενσωμάτωση του IoT και της ML έχει οδηγήσει σε σημαντικές προόδους σε διάφορους τομείς της υγειονομικής περίθαλψης:

- **Διαχείριση χρόνιων ασθενειών:** Συσκευές IoT παρακολουθούν φυσιολογικές παραμέτρους (π.χ. επίπεδα γλυκόζης στο αίμα, αρτηριακή πίεση, καρδιακός ρυθμός), επίπεδα δραστηριότητας, μοτίβα ύπνου και τήρηση της φαρμακευτικής αγωγής, ενώ τα μοντέλα ML αναλύουν αυτά τα δεδομένα για να προβλέπουν τις εξάρσεις, να εξατομικεύουν τα σχέδια θεραπείας και να παρέχουν έγκαιρες παρεμβάσεις.
 - **Διαχείριση διαβήτη:** Οι συνεχείς μετρητές γλυκόζης σε συνδυασμό με αντλίες ινσουλίνης και αλγόριθμους ML δημιουργούν συστήματα «τεχνητού παγκρέατος» που ρυθμίζουν αυτόματα τη χορήγηση ινσουλίνης με βάση τα επίπεδα γλυκόζης σε πραγματικό χρόνο, ελαχιστοποιώντας τον κίνδυνο υπογλυκαιμίας και υπεργλυκαιμίας (Klonoff et al., 2011).
 - **Διαχείριση άσθματος:** Έξυπνοι εισπνευστήρες εξοπλισμένοι με αισθητήρες παρακολουθούν τη χρήση φαρμάκων και περιβαλλοντικούς παράγοντες (π.χ. ατμοσφαιρική ρύπανση, αριθμός γύρης). Τα μοντέλα ML αναλύουν αυτά τα δεδομένα για να προβλέψουν τις κρίσεις άσθματος και να παρέχουν εξατομικευμένες ειδοποιήσεις στους ασθενείς (Wang et al., 2020).
 - **Διαχείριση της PD:** Φορητοί αισθητήρες παρακολουθούν κινητικά συμπτώματα (π.χ. τρόμος, διαταραχές βάδισης, βραδυκινησία) και μη κινητικά συμπτώματα (π.χ. διαταραχές ύπνου, αλλαγές στη διάθεση). Μοντέλα ML αναλύουν αυτά τα δεδομένα για να αξιολογήσουν την εξέλιξη της νόσου, να προβλέψουν πτώσεις, να βελτιστοποιήσουν τις δοσολογίες φαρμάκων και να εξατομικεύσουν προγράμματα αποκατάστασης (Liu et al., 2016).

- **Έγκαιρη ανίχνευση καρδιακών παθήσεων:** Οι ηλεκτροκαρδιογράφοι που βασίζονται στο IoT και άλλοι φορητοί αισθητήρες συλλέγουν συνεχή καρδιακά δεδομένα, τα οποία αναλύουν οι αλγόριθμοι ML για την ανίχνευση αρρυθμιών (ακανόνιστοι καρδιακοί παλμοί), την πρόβλεψη του κινδύνου καρδιακής ανεπάρκειας και την ενεργοποίηση ειδοποιήσεων για έγκαιρη ιατρική φροντίδα. Αυτά τα συστήματα μπορούν να ανιχνεύσουν λεπτές αλλαγές στα μοτίβα του ηλεκτροκαρδιογραφήματος που μπορεί να διαφύγουν από τις παραδοσιακές μεθόδους, επιτρέποντας την έγκαιρη παρέμβαση και την πρόληψη δυνητικά απειλητικών για τη ζωή συμβάντων (Acharya et al., 2017).
- **Απομακρυσμένη παρακολούθηση ασθενών για μετεγχειρητική φροντίδα:** Οι φορητοί αισθητήρες με δυνατότητα IoT παρακολουθούν τα ζωτικά σημεία, τα επίπεδα δραστηριότητας και την πρόοδο επούλωσης των τραυμάτων των ασθενών μετά από χειρουργική επέμβαση. Οι αλγόριθμοι ML αναλύουν αυτά τα δεδομένα για να εντοπίζουν πρώιμα σημάδια επιπλοκών, όπως λοιμώξεις ή καθυστερημένη επούλωση, επιτρέποντας την έγκαιρη παρέμβαση και μειώνοντας την ανάγκη για επανεισαγωγές στο νοσοκομείο (Darkins et al., 2008).

1.4 Μελλοντικές τάσεις

Η σύγκλιση του IoT και της ML στην υγειονομική περίθαλψη είναι ένας ταχέως εξελισσόμενος τομέας, με τη συνεχή έρευνα και ανάπτυξη να διευρύνει τα όρια του εφικτού. Αρκετές βασικές τάσεις διαμορφώνουν το μέλλον αυτού του τομέα.

1.4.1 Αναδυόμενες συσκευές IoT για προηγμένες ιατρικές εφαρμογές

- **Μικροσκοπικοί και εμφυτεύσιμοι αισθητήρες:** Οι εξελίξεις στη μικροηλεκτρονική και τη νανοτεχνολογία οδηγούν στην ανάπτυξη μικρότερων, λιγότερο επεμβατικών, ακόμη και εμφυτεύσιμων αισθητήρων που μπορούν να παρακολουθούν συνεχώς ένα ευρύ φάσμα φυσιολογικών παραμέτρων, συμπεριλαμβανομένων των επιπέδων γλυκόζης, της αρτηριακής πίεσης, της ενδοκρανιακής πίεσης και των βιοχημικών δεικτών (Khan and Kim, 2024).

- **Έξυπνα κλωστοϋφαντουργικά προϊόντα και φορητά προϊόντα:** Η ενσωμάτωση αισθητήρων σε υφάσματα και ενδύματα δημιουργεί άνετες και διακριτικές φορητές συσκευές για συνεχή παρακολούθηση της υγείας, παρακολούθηση της δραστηριότητας και αποκατάστασης (Islam et al., 2020).
- **Αισθητήρες που καταπίνονται:** Αυτοί οι αισθητήρες μπορούν να καταποθούν όπως τα χάπια για την παρακολούθηση της γαστρεντερικής υγείας, της συμμόρφωσης με τη φαρμακευτική αγωγή και άλλων εσωτερικών φυσιολογικών διεργασιών (Kalantar-Zadeh et al., 2017).
- **Βιοαισθητήρες για την ανίχνευση βιοδεικτών:** Ανάπτυξη εξαιρετικά ευαίσθητων βιοαισθητήρων ικανών να ανιχνεύουν συγκεκριμένους βιοδείκτες σε σωματικά υγρά (π.χ. αίμα, σάλιο, ιδρώτας) για την έγκαιρη ανίχνευση ασθενειών και την εξατομικευμένη διάγνωση (Turner, 2013).

1.4.2 Καινοτομίες στους αλγορίθμους ML για βελτιωμένη διάγνωση και παρακολούθηση

- **Εξηγήσιμη τεχνητή νοημοσύνη:** Αυξανόμενη εστίαση στην ανάπτυξη μοντέλων ML που είναι διαφανή και ερμηνεύσιμα, επιτρέποντας στους κλινικούς γιατρούς να κατανοήσουν το σκεπτικό πίσω από τις προβλέψεις και να οικοδομήσουν εμπιστοσύνη στην υγειονομική περίθαλψη με βάση την ML (Khan and Kim, 2024).
- **Ομοσπονδιακή μάθηση:** Επιτρέπει τη συνεργατική εκπαίδευση μοντέλων ML σε αποκεντρωμένα σύνολα δεδομένων (π.χ. δεδομένα από πολλά νοσοκομεία) χωρίς να μοιράζονται άμεσα ευαίσθητες πληροφορίες ασθενών, αντιμετωπίζοντας τις ανησυχίες για την προστασία της ιδιωτικής ζωής (Yang et al., 2019).
- **Σύντηξη πολυτροπικών δεδομένων:** Ανάπτυξη αλγορίθμων ML που μπορούν να ενσωματώσουν και να αναλύσουν αποτελεσματικά διαφορετικές πηγές δεδομένων (π.χ. απεικόνιση, γονιδιωματική, δεδομένα αισθητήρων) για την παροχή μιας πιο ολιστικής άποψης της υγείας του ασθενούς και τη βελτίωση της διαγνωστικής ακρίβειας (Krones et al., 2024).
- **Ανακάλυψη και ανάπτυξη φαρμάκων με βάση την τεχνητή νοημοσύνη:** Αξιοποίηση της ML για την επιτάχυνση της ανακάλυψης φαρμάκων, τον

εντοπισμό πιθανών στόχων φαρμάκων, την πρόβλεψη της αποτελεσματικότητας και της τοξικότητας των φαρμάκων και την εξατομίκευση των φαρμακευτικών θεραπειών (Vamathevan et al., 2019).

1.4.3 Δυνατότητες ενσωμάτωσης IoT-ML στην προληπτική ιατρική

- **Προβλεπτική ανάλυση για τη διαστρωμάτωση κινδύνου:** Συνδυασμός δεδομένων που παράγονται από το IoT με αλγορίθμους ML για τον εντοπισμό ατόμων με υψηλό κίνδυνο εμφάνισης συγκεκριμένων ασθενειών, επιτρέποντας προληπτικές παρεμβάσεις και προληπτικές στρατηγικές (Obermeyer et al., 2019).
- **Εξατομικευμένες παρεμβάσεις στον τρόπο ζωής:** Χρήση συσκευών IoT για την παρακολούθηση παραγόντων του τρόπου ζωής (π.χ. διατροφή, άσκηση, ύπνος) και παροχή εξατομικευμένων συστάσεων για τη βελτίωση της υγείας και την πρόληψη ασθενειών (Chiang et al., 2021).
- **Έγκαιρη ανίχνευση της εκδήλωσης ασθενειών:** Αξιοποίηση δεδομένων IoT και ML για την παρακολούθηση των τάσεων των ασθενειών, την έγκαιρη ανίχνευση εστιών και την εφαρμογή έγκαιρων παρεμβάσεων δημόσιας υγείας (Alaa et al., 2019).
- **Απομακρυσμένη παρακολούθηση για μετεγχειρητική φροντίδα και αποκατάσταση:** Χρήση συσκευών IoT και ML για την εξ αποστάσεως παρακολούθηση ασθενών μετά από χειρουργική επέμβαση ή κατά τη διάρκεια της αποκατάστασης, επιτρέποντας την έγκαιρη ανίχνευση επιπλοκών και εξατομικευμένα προγράμματα αποκατάστασης (Rajendran et al., 2025).

1.5 Σύνοψη

Η σύγκλιση του IoT και του ML φέρνει επανάσταση στην υγειονομική περίθαλψη, προσφέροντας πρωτοφανείς ευκαιρίες για τη βελτίωση της διαχείρισης των ασθενειών, την ενίσχυση των αποτελεσμάτων των ασθενών και τον μετασχηματισμό της παροχής υγειονομικής περίθαλψης. Επιτρέποντας τη συνεχή παρακολούθηση, τη λήψη αποφάσεων βάσει δεδομένων, τις εξατομικευμένες

παρεμβάσεις και την προληπτική διαχείριση ασθενειών, το IoT και το ML ανοίγουν το δρόμο για ένα πιο αποδοτικό, αποτελεσματικό και ασθενοκεντρικό σύστημα υγειονομικής περίθαλψης.

1.5.1 Σύνοψη των βασικών σημείων

Το IoT παρέχει την υποδομή για συνεχή συλλογή δεδομένων μέσω διασυνδεδεμένων συσκευών, δημιουργώντας πολύτιμα σύνολα δεδομένων για ανάλυση ML.

- Οι αλγόριθμοι ML αναλύουν αυτά τα σύνολα δεδομένων για τον εντοπισμό μοτίβων, την πραγματοποίηση προβλέψεων και την παροχή πληροφοριών για τη βελτίωση της διάγνωσης, της πρόγνωσης και της θεραπείας.
- Η συνέργεια του IoT και της ML έχει ιδιαίτερη επίδραση στη διαχείριση χρόνιων ασθενειών όπως η PD, επιτρέποντας την έγκαιρη ανίχνευση, την εξατομικευμένη παρακολούθηση και τις προσαρμοσμένες παρεμβάσεις.
- Οι μελλοντικές τάσεις περιλαμβάνουν την ανάπτυξη πιο προηγμένων συσκευών IoT, καινοτόμων αλγορίθμων ML και την επέκταση των εφαρμογών IoT και ML στην προληπτική ιατρική.
- Η αντιμετώπιση των προκλήσεων που σχετίζονται με το απόρρητο των δεδομένων, την ασφάλεια, τις ηθικές εκτιμήσεις και τους τεχνολογικούς περιορισμούς είναι ζωτικής σημασίας για την επιτυχή εφαρμογή του IoT και της ML στην υγειονομική περίθαλψη.

1.5.2 Όραμα για το IoT και το ML στον μετασχηματισμό της υγειονομικής περίθαλψης

Το όραμα για το μέλλον της υγειονομικής περίθαλψης που τροφοδοτείται από το IoT και το ML είναι αυτό της προληπτικής, εξατομικευμένης και προγνωστικής περίθαλψης. Φανταστείτε έναν κόσμο όπου:

- Οι ασθένειες ανιχνεύονται στα πρώτα τους στάδια, συχνά πριν καν εμφανιστούν τα συμπτώματα.
- Τα σχέδια θεραπείας προσαρμόζονται στα ατομικά προφίλ των ασθενών, μεγιστοποιώντας την αποτελεσματικότητα και ελαχιστοποιώντας τις παρενέργειες.

- Οι ασθενείς έχουν τη δυνατότητα να διαχειρίζονται ενεργά την υγεία τους μέσω εξατομικευμένης ανατροφοδότησης και συστάσεων.
- Η παροχή υγειονομικής περίθαλψης είναι πιο αποτελεσματική, προσβάσιμη και οικονομικά αποδοτική.
- Οι παρεμβάσεις στη δημόσια υγεία είναι πιο στοχευμένες και αποτελεσματικές στην πρόληψη της εκδήλωσης ασθενειών και στην προαγωγή της υγείας του πληθυσμού.

Ενώ οι προκλήσεις παραμένουν, οι συνεχιζόμενες εξελίξεις στο IoT και το ML, σε συνδυασμό με την αυξανόμενη έμφαση στην υγειονομική περίθαλψη με βάση τα δεδομένα, φέρνουν αυτό το όραμα πιο κοντά στην πραγματικότητα. Με την υπεύθυνη υιοθέτηση αυτών των τεχνολογιών και την αντιμετώπιση των σχετικών προκλήσεων, μπορούμε να ξεκλειδώσουμε το πλήρες δυναμικό του IoT και του ML για να μεταμορφώσουμε την υγειονομική περίθαλψη και να βελτιώσουμε τη ζωή των ατόμων και των κοινοτήτων παγκοσμίως.

2. Συλλογή δεδομένων και Μεθοδολογία απόκτησης δεδομένων

Η συλλογή δεδομένων αποτελεί κεντρικό πυλώνα στην έρευνα για τη PD, καθώς επιτρέπει την ποσοτικοποίηση των κινητικών συμπτωμάτων και την υποστήριξη της έγκαιρης διάγνωσης και παρακολούθησης της εξέλιξης της νόσου. Παραδοσιακές μέθοδοι, όπως η κλινική αξιολόγηση μέσω της Κλίμακας UPDRS (Unified Parkinson's Disease Rating Scale), συχνά περιορίζονται από την υποκειμενικότητα και την απουσία συνεχούς παρακολούθησης (Goetz et al., 2008).

Η πρόοδος στις τεχνολογίες του IoT και της ML έχει καταστήσει εφικτή την ανάπτυξη εργαλείων για τη συλλογή και ανάλυση δεδομένων σε πραγματικό χρόνο, βελτιώνοντας την ακρίβεια και την αποτελεσματικότητα της διαχείρισης της νόσου. Στο πλαίσιο αυτό, η παρούσα εργασία παρουσιάζει το σύστημα SmartGlove, το οποίο αναπτύχθηκε για την απομακρυσμένη παρακολούθηση των κινητικών συμπτωμάτων ασθενών με PD.

Το έργο υλοποιήθηκε μέσω του Επιχειρησιακού Προγράμματος της Περιφέρειας Ηπείρου 2014-2020 και του ΕΣΠΑ 2014-2020, στοχεύοντας στην ενίσχυση καινοτόμων λύσεων στους τομείς της υγείας και της βιοτεχνολογίας. Στόχος του παρόντος κεφαλαίου είναι η περιγραφή της μεθοδολογίας συλλογής δεδομένων, η παρουσίαση του σχεδιασμού του SmartGlove και η ανάλυση της πιλοτικής εφαρμογής του.

Με τη βοήθεια αυτής της τεχνολογίας, οι γιατροί και οι θεραπευτές μπορούν να παρακολουθούν εξ αποστάσεως τους εξέλκξη των συμπτωμάτων των ασθενών και να προσαρμόζουν ανάλογα τη θεραπεία τους. Η τεχνολογία αυτή βρίσκεται σε πιλοτική εφαρμογή και αναμένεται να παρέχει σημαντικά δεδομένα για τη βελτίωση της θεραπείας των ασθενών που πάσχουν από νευροεκφυλιστικές ασθένειες.



Εικόνα 6 SmartGlove

2.1 Σχεδιασμός του SmartGlove

Το SmartGlove σχεδιάστηκε για να παρέχει μια αξιόπιστη και φιλική προς τον χρήστη λύση για την καταγραφή των κινητικών συμπτωμάτων ασθενών με νόσο του Πάρκινσον, με έμφαση στον τρόμο, τη βραδυκίνησία και τη δυσκαμψία. Το σύστημα αποτελείται από τρία κύρια υποσυστήματα:

1. **Υφασμάτινο Γάντι:** Κατασκευασμένο από ανθεκτικά, βιοσυμβατά υλικά, διαθέτει αγωγίμες περιοχές στους δακτύλους και την παλάμη για την ανίχνευση κινήσεων. Το ύφασμα είναι ανθεκτικό στον ιδρώτα, πλενόμενο και έχει εκτιμώμενη διάρκεια ζωής άνω των 10 ετών.
2. **System on Chip (SoC):** Περιλαμβάνει έναν μικροελεγκτή χαμηλής κατανάλωσης με ενσωματωμένη μονάδα Bluetooth 5.0, η οποία εξασφαλίζει ασύρματη μεταφορά δεδομένων με ελάχιστη κατανάλωση ενέργειας (τυπική εμβέλεια 10-100 m, ανάλογα με το περιβάλλον).
3. **Αισθητήρες:** Το γάντι ενσωματώνει (α) αισθητήρες εξώθησης (flex sensors) για τη μέτρηση της κάμψης των δακτύλων, (β) αισθητήρες επαφής

για την ανίχνευση της αλληλεπίδρασης δακτύλου-παλάμης, και (γ) μία μονάδα IMU 9 αξόνων (επιταχυνσιόμετρο, γυροσκόπιο, μαγνητόμετρο) για την καταγραφή της κίνησης και του προσανατολισμού του χεριού. Η IMU παρέχει δεδομένα με ανάλυση 16-bit και ρυθμό δειγματοληψίας έως 100 Hz. Το σύστημα συμπληρώνεται από μονάδα διαχείρισης ισχύος (λιθίου-πολυμερής μπαταρία) και αποθηκευτικό μέσο (flash memory) για προσωρινή αποθήκευση δεδομένων.

Το SmartGlove συμμορφώνεται με τα πρότυπα ασφάλειας IEC 60601-1 για ιατροτεχνολογικά προϊόντα, εξασφαλίζοντας την αξιοπιστία του σε κλινικές εφαρμογές.

2.2 Πλατφόρμα mobPark και Διαχείριση Δεδομένων

Η εφαρμογή mobPark λειτουργεί ως μεσολαβητής μεταξύ του SmartGlove και της κεντρικής πλατφόρμας διαχείρισης δεδομένων. Είναι συμβατή με συσκευές Android (έκδοση 6.0 και άνω) και iOS (έκδοση 11 και άνω), απαιτώντας Bluetooth 5.0 και σύνδεση στο διαδίκτυο (Wi-Fi ή 4G/5G). Οι κύριες λειτουργίες της περιλαμβάνουν:

- **Συλλογή Δεδομένων:** Πραγματοποιεί λήψη σημάτων από το SmartGlove μέσω Bluetooth με ρυθμό δειγματοληψίας 100 Hz.
- **Διαχείριση Χρηστών:** Υποστηρίζει δημιουργία προφίλ, είσοδο χρηστών και τοπική αποθήκευση ελάχιστων δεδομένων (π.χ., κωδικός χρήστη).
- **Εκτέλεση Δοκιμών:** Διευκολύνει τη συμπλήρωση ερωτηματολογίων και την εκτέλεση κινητικών ασκήσεων, με οδηγίες που εμφανίζονται στην οθόνη.
- **Μεταφορά Δεδομένων:** Αποστέλλει δεδομένα στην κεντρική πλατφόρμα μέσω κρυπτογραφημένης σύνδεσης HTTPS, διαγράφοντας τα τοπικά αντίγραφα μετά την επιτυχή μεταφορά.

Η κεντρική πλατφόρμα, βασισμένη σε υποδομή cloud (π.χ., AWS ή Azure), αποθηκεύει και επεξεργάζεται τα δεδομένα. Υποστηρίζει τις εξής λειτουργίες:

- Αυθεντικοποίηση χρηστών μέσω OAuth 2.0.

- Διαχείριση ιατρικών φακέλων με δεδομένα SmartGlove, ερωτηματολόγια και αποτελέσματα δοκιμών.
- Ανάλυση σημάτων για την εκτίμηση της κατάστασης του ασθενούς βάσει της κλίμακας UPDRS (Patel et al., 2009).
- Παραγωγή αναφορών για την υποστήριξη κλινικών αποφάσεων.

Η αρχιτεκτονική της πλατφόρμας εξασφαλίζει κλιμακωσιμότητα και συμμόρφωση με τον Γενικό Κανονισμό για την Προστασία Δεδομένων (GDPR).

2.3 Πιλοτική εφαρμογή και αρχικά αποτελέσματα

Η πιλοτική εφαρμογή του SmartGlove αποτέλεσε βασικό στάδιο για την αξιολόγηση της λειτουργικότητας και της χρησιμότητάς του. Διεξήχθη σε πολλαπλά πλαίσια:

1. **8η Πανελλήνια Έκθεση Άρτας:** Παρουσίαση του συστήματος σε κοινό και επιστήμονες, με θετική ανταπόκριση για την ευχρηστία (85% των συμμετεχόντων ανέφεραν ικανοποίηση, βάσει προκαταρκτικής έρευνας).
2. **Αρσάκεια Εκπαιδευτικά Κέντρα:** Εκπαίδευση φοιτητών υγείας στη χρήση του SmartGlove, με έμφαση στην ενσωμάτωσή του στην κλινική πρακτική.
3. **7ο SEEDA Συνέδριο:** Παρουσίαση τεχνικών προδιαγραφών και συλλογή ανατροφοδότησης από ειδικούς.
4. **IEEE VTC 2020:** Διεθνής προβολή, με έμφαση στη δυνατότητα απομακρυσμένης παρακολούθησης.

Η δεύτερη πιλοτική φάση επικεντρώθηκε στη βελτιστοποίηση του συστήματος (π.χ., μείωση καθυστέρησης Bluetooth κατά 15%) και στην ποσοτικοποίηση δεδομένων για στατιστική ανάλυση. Τα αρχικά αποτελέσματα υποδεικνύουν

υψηλή συσχέτιση ($r > 0.8$) μεταξύ των μετρήσεων του SmartGlove και των κλινικών αξιολογήσεων.

2.4 Ασκήσεις με το γάντι και συλλογή δεδομένων

Η συλλογή δεδομένων διεξήχθη με ρυθμό δειγματοληψίας 100 Hz, επιλεγμένο για την ακριβή καταγραφή συχνοτήτων τρόμου (4-6 Hz) και βραδυκινησίας (Helmich et al., 2012). Οι μετρήσεις πραγματοποιήθηκαν πριν και μετά τη χορήγηση φαρμακευτικής αγωγής (π.χ., Levodopa), με σκοπό τη διερεύνηση της θεραπευτικής επίδρασης. Το πειραματικό πρωτόκολλο περιελάμβανε τέσσερις ασκήσεις, οι οποίες περιγράφονται ως εξής:

1^η άσκηση: Με παλάμη ανοικτή ο ασθενής ακουμπάει τα χέρια στα πόδια και μετράει μέχρι το 10. Με στόχο την αξιολόγηση σταθερότητας.



Εικόνα 7 Παράδειγμα της 1^{ης} άσκησης

2^η άσκηση: Ο ασθενής σηκώνει τα χέρια στο ύψος των ώμων, ακουμπάμε τους δείκτες και μετράει μέχρι το 10. Με στόχο την μέτρηση συντονισμού.



Εικόνα 8 Παράδειγμα της 2ης άσκησης

3^η άσκηση: Με την παλάμη ανοικτή ο ασθενής προσπαθεί να αγγίξει τον δείκτη στον αντίχειρα και στην συνέχεια να επανέλθει στην αρχική θέση, επαναλαμβάνει διαρκώς όσο γρηγορότερα γίνεται για 10 επαναλήψεις. Με στόχο την εκτίμηση της ταχύτητας.



Εικόνα 9 Παράδειγμα της 3ης άσκησης

4^η άσκηση: Με την παλάμη ανοικτή προσπαθεί ο ασθενής να κλείσει την παλάμη δημιουργώντας γροθιά και στην συνέχεια να επανέλθει στην αρχική θέση, επαναλαμβάνει διαρκώς όσο πιο γρηγορότερα γίνεται για 10 επαναλήψεις. Με στόχο την ανίχνευση βραδυκινησίας.



Εικόνα 10 Παράδειγμα της 4ης άσκησης

Η μεθοδολογία συλλογής δεδομένων που παρουσιάστηκε στο παρόν κεφάλαιο, μέσω του συστήματος SmartGlove, αποτελεί μια ολοκληρωμένη προσέγγιση για την απομακρυσμένη παρακολούθηση των κινητικών συμπτωμάτων της PD, αντιμετωπίζοντας τις περιορισμένες δυνατότητες των παραδοσιακών μεθόδων, όπως η υποκειμενικότητα της κλινικής αξιολόγησης μέσω της κλίμακας UPDRS.

Ο σχεδιασμός του SmartGlove, με την ενσωμάτωση προηγμένων αισθητήρων και τεχνολογίας Bluetooth 5.0, σε συνδυασμό με την πλατφόρμα mobPark και την υποδομή cloud, επιτρέπει τη συλλογή δεδομένων υψηλής ακρίβειας με ρυθμό δειγματοληψίας 100 Hz, παρέχοντας τη δυνατότητα ποσοτικοποίησης του τρόμου, της βραδυκινησίας και της δυσκαμψίας. Η πιλοτική εφαρμογή του συστήματος, με θετικά αρχικά αποτελέσματα (π.χ., συσχέτιση $r > 0.8$ με κλινικές αξιολογήσεις), υποδεικνύει την αξιοπιστία και τη χρησιμότητά του, ενώ οι προσεκτικά σχεδιασμένες ασκήσεις ενισχύουν την ικανότητά του να καταγράφει ποικίλες πτυχές των κινητικών συμπτωμάτων.

Παράλληλα, η συμμόρφωση με πρότυπα ασφάλειας (IEC 60601-1) και προστασίας δεδομένων (GDPR) διασφαλίζει την εφαρμογή του σε κλινικές συνθήκες. Το κεφάλαιο αυτό θέτει τις βάσεις για την περαιτέρω επεξεργασία και ανάλυση των δεδομένων, η οποία θα παρουσιαστεί στα επόμενα κεφάλαια, με στόχο τη διερεύνηση της θεραπευτικής επίδρασης και τη βελτιστοποίηση της διαχείρισης της νόσου.

3. Επιλογή χαρακτηριστικών και διερευνητική στατιστική

Η ανάπτυξη αποτελεσματικών μοντέλων ML για την ανίχνευση της PD απαιτεί μια βαθιά κατανόηση των δεδομένων και των χαρακτηριστικών που τα αποτελούν. Σε αυτό το κεφάλαιο, εστιάζουμε σε δύο κρίσιμες πτυχές της ανάλυσης δεδομένων: την επιλογή χαρακτηριστικών και τη διερευνητική στατιστική. Αυτές οι διαδικασίες δεν είναι απλώς τεχνικές βήματα, αλλά αποτελούν τη βάση για την εξαγωγή σημαντικών πληροφοριών από πολύπλοκα σύνολα δεδομένων, ιδιαίτερα σε ιατρικές εφαρμογές όπως η ανίχνευση της PD.

Η επιλογή χαρακτηριστικών αναφέρεται στη διαδικασία εντοπισμού και ιεράρχησης των πιο σχετικών χαρακτηριστικών από ένα σύνολο δεδομένων. Σε ιατρικές μελέτες, όπου τα δεδομένα συχνά περιλαμβάνουν πολλές διαστάσεις και μεταβλητές, η επιλογή των σωστών χαρακτηριστικών είναι καθοριστική για την ακρίβεια και την ερμηνευσιμότητα των μοντέλων. Η επιλογή χαρακτηριστικών μπορεί να βοηθήσει στη μείωση της πολυπλοκότητας των δεδομένων, στη βελτίωση της απόδοσης των μοντέλων και στην αποφυγή του φαινομένου της "υπερπροσαρμογής" (overfitting) (Guyon & Elisseeff, 2003).

Από την άλλη πλευρά, η διερευνητική στατιστική περιλαμβάνει την εφαρμογή μαθηματικών και στατιστικών τεχνικών για την αποκάλυψη προτύπων, σχέσεων και τάσεων στα δεδομένα. Αυτή η διαδικασία είναι απαραίτητη για την κατανόηση της δομής των δεδομένων και για την εξαγωγή χρήσιμων πληροφοριών που μπορούν να καθοδηγήσουν την ανάπτυξη μοντέλων. Η διερευνητική ανάλυση μπορεί να περιλαμβάνει τεχνικές όπως η οπτικοποίηση δεδομένων, η ανάλυση συσχέτισης και η εφαρμογή στατιστικών ελέγχων για την αξιολόγηση της σημαντικότητας των παρατηρούμενων διαφορών (Tukey, 1977).

Στην περίπτωση της PD, η ανάλυση των δεδομένων συχνά περιλαμβάνει την εξέταση σημάτων από διάφορους αισθητήρες, όπως γυροσκόπια, μαγνητόμετρα και επιταχυνσιόμετρα, τα οποία καταγράφουν κινητικά δεδομένα από τους ασθενείς. Αυτά τα δεδομένα μπορούν να χρησιμοποιηθούν για την παρακολούθηση της εξέλιξης της νόσου, την αξιολόγηση της αποτελεσματικότητας των φαρμακευτικών παρεμβάσεων και την πρόβλεψη των

συμπτωμάτων (Goetz et al., 2008). Ωστόσο, η πολυπλοκότητα και ο όγκος αυτών των δεδομένων απαιτούν προσεκτική επιλογή χαρακτηριστικών και διερευνητική ανάλυση για να εξαχθούν χρήσιμες πληροφορίες.

Σε αυτό το κεφάλαιο, θα εξετάσουμε διάφορες τεχνικές επιλογής χαρακτηριστικών, όπως η χρήση στατιστικών μετρικών (π.χ. μέση τιμή, τυπική απόκλιση, διακύμανση) και πιο προηγμένες μεθόδους που βασίζονται σε μοντέλα ML, όπως τα δέντρα αποφάσεων. Επιπλέον, θα διερευνήσουμε πώς η διερευνητική στατιστική μπορεί να βοηθήσει στην κατανόηση των επιδράσεων της φαρμακευτικής αγωγής στα κινητικά χαρακτηριστικά των ασθενών με Πάρκινσον.

Οι τεχνικές που θα παρουσιαστούν σε αυτό το κεφάλαιο έχουν σχεδιαστεί με στόχο την ενίσχυση της κατανόησης των δεδομένων και τη βελτίωση της απόδοσης των μοντέλων ML. Η σωστή εφαρμογή αυτών των μεθόδων μπορεί να οδηγήσει σε πιο ακριβείς διαγνώσεις και πιο εξατομικευμένες θεραπευτικές προσεγγίσεις για τους ασθενείς με PD.

3.1 Τεχνικές επιλογής χαρακτηριστικών

Η επιλογή χαρακτηριστικών (feature selection) αποτελεί κρίσιμο βήμα για την αποτελεσματική ανάλυση δεδομένων σε εφαρμογές ML, ιδιαίτερα στη διαχείριση της PD. Στόχος είναι η εξαγωγή χαρακτηριστικών που καταγράφουν λεπτομερώς τα κινητικά συμπτώματα (π.χ., τρόμος, βραδυκίνησία) και επιτρέπουν τη δημιουργία ακριβών μοντέλων πρόβλεψης. Παρακάτω παρουσιάζονται οι τεχνικές και τα χαρακτηριστικά που χρησιμοποιήθηκαν, μαζί με τα μαθηματικά τους υπόβαθρα και βιβλιογραφικές αναφορές.

3.1.1 Βασικά στατιστικά χαρακτηριστικά.

Τα βασικά στατιστικά χαρακτηριστικά προσφέρουν μια συνοπτική περιγραφή της κατανομής και μεταβλητότητας των δεδομένων. Χρησιμοποιήθηκαν οι ακόλουθες μετρικές:

1. Μέση τιμή (Mean):

$$\bar{x} = \frac{1}{N} \left(\sum_{i=1}^N x_i \right)$$

Ο μέσος όρος χρησιμοποιείται για την αξιολόγηση της κεντρικής τάσης του σήματος (Giannakopoulou et al., 2022).

2. Τυπική απόκλιση (Standard Deviation):

$$s = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Μετρά τη διασπορά των δεδομένων γύρω από τη μέση τιμή, υποδεικνύοντας τη μεταβλητότητα του σήματος (Chen et al., 2018).

3. Διακύμανση (Variance):

$$s^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}$$

Χρησιμοποιείται για την ανίχνευση αλλαγών στην ένταση των κινήσεων (Patel et al., 2009).

4. Ελάχιστη τιμή (Min)

$$\min(x_i)$$

Η μικρότερη τιμή του σήματος.

5. Μέγιστη τιμή (Max):

$$\max(x_i)$$

Η μεγαλύτερη τιμή του σήματος.

6. Εύρος (Range):

$$range = \max(x_i) - \min(x_i)$$

Το εύρος των δεδομένων προσφέρει πληροφορίες για την ακραία συμπεριφορά του σήματος (Javaid et al., 2022).

7. Μεσαία τιμή (Median):

Αν το πλήθος n είναι περιττός, $median = \frac{x_{n+1}}{2}$

Αν το πλήθος n είναι άρτιος, $median = \frac{1}{2} \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right)$

Η τιμή που διαχωρίζει τα δεδομένα σε δύο ίσα μέρη, ανθεκτική σε ακραίες τιμές (Tukey, 1977).

8. Διατεταρτημοριακό εύρος (IQR):

$$IQR = Q_3 - Q_1$$

Χρησιμοποιείται για την ανίχνευση ακραίων τιμών σε μη κανονικές κατανομές (Bishop, 2006).

9. Λοξότητα (Skewness):

$$\widetilde{\mu}_3 = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{(N-1)s^3}$$

Μετρά την ασυμμετρία της κατανομής (Peever & Fuller, 2017).

10. Κύρτωση (Kurtosis):

$$kurtosis = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{s^4}$$

Υποδεικνύει την "οξύτητα" της κατανομής, σχετική με συγκεντρωμένες τιμές (Jankovic, 2008).

11. Απόκλιση του τεταρτημορίου (Quartile Deviation)

$$Quartile\ Deviation = \frac{Q_3 - Q_1}{2}$$

Η ημι-διατεταρτημοριακή απόκλιση, μετρά τη διασπορά των δεδομένων γύρω από τη διάμεσο.

12. Συντελεστής διακύμανσης (Coefficient of Variation)

$$CV = \frac{s}{\bar{x}}$$

Δείχνει την έκταση της μεταβλητότητας σε σχέση με τον μέσο όρο του σήματος.

3.1.2 Χαρακτηριστικά Ενέργειας

Τα ενεργειακά χαρακτηριστικά ποσοτικοποιούν την ένταση και τα επίπεδα δραστηριότητας του σήματος, καθιστώντας τα ιδιαίτερα χρήσιμα για την ανάλυση δεδομένων επιταχυνσιόμετρου.

1. Απόλυτη ενέργεια (Absolute Energy):

$$SMA = \sum_{i=1}^N |x_i|$$

Σχετίζεται με τη συνολική δραστηριότητα του σήματος (Li et al., 2024).

2. Μέση τετραγωνική ρίζα (RMS)

$$rms = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$$

Μετρά τη μέση ενέργεια του σήματος, σχετική με την ένταση των κινήσεων (Antoine, 2003).

3. Συνολική ενέργεια (Energy)

$$E = \sum_{i=1}^N x_i^2$$

Δείχνει την Ενέργεια ενός σήματος διακριτού χρόνου.

4. Λογαριθμική ενέργεια (Log Energy):

$$Log(E) = \log \sum_{i=1}^N x_i^2$$

Δείχνει την λογαριθμική Ενέργεια ενός σήματος διακριτού χρόνου για βρεθούν ακόμα και οι μικρές μεταβολές.

3.1.3 Χαρακτηριστικά στο πεδίο της συχνότητας

1. Φασματική εντροπία (Spectral Entropy):

$H = -\sum_i p_i \log p_i$, όπου p_i τα κανονικοποιημένα κομμάτια ισχύος.

Μετρά την αταξία στο φάσμα, σχετική με την πολυπλοκότητα του σήματος (Pincus, 1991).

2. **Φασματικό κεντροειδές (Spectral Centroid):** $C = \frac{\sum_i f_i P_i}{\sum_i P_i}$, όπου f_i είναι οι συχνότητες και P_i η ισχύς.
Υποδεικνύει το "κέντρο βάρους" του φάσματος (Gubbi et al., 2013).

3. **Φασματική διασπορά (Spectral Spread):**

$$\sigma = \sqrt{\frac{\sum_i (f_i - C)^2 P_i}{\sum_i P_i}}$$

Μετρά την εξάπλωση του φάσματος γύρω από το κεντροειδές (Turner, 2013).

4. **Σημείο Roll-off (85%):**

Η συχνότητα όπου το 85% της συνολικής ισχύος συγκεντρώνεται κάτω από αυτήν (Bishop, 2006).

5. **Κυρίαρχη συχνότητα (Dominant Frequency):**

$$f_{dominant} = freqs[\operatorname{argmax}(P(f))]$$

Όπου $P(f)$ είναι η φασματική πυκνότητα ισχύος. Χρησιμοποιείται για την ανίχνευση τρόμου (Helmich et al., 2012).

6. **Spectral Flatness:**

$$\text{Spectral Flatness} = \frac{\exp(\frac{1}{N} \sum_{i=1}^N \log(P_i + \epsilon))}{\frac{1}{N} \sum_{i=1}^N P_i}$$

Μετρά την ομαλότητα του φάσματος ισχύος (Gresty and Buckwell, 1990).

7. **Spectral Flux:**

$$\text{Spectral Flux} = \sum_i (P_i(t) - P_i(t-1))^2$$

Μετρά τον ρυθμό μεταβολής του φάσματος ισχύος.

8. Spectral Variability:

$$\text{Spectral Variability} = \text{Var}(P_i)$$

Μετρά τη μεταβλητότητα του φάσματος ισχύος.

3.1.3 Δυναμικά Χαρακτηριστικά

Τα δυναμικά χαρακτηριστικά περιγράφουν τις αλλαγές και τη δυναμική του σήματος με την πάροδο του χρόνου, αποτυπώνοντας την πολυπλοκότητα και το απρόβλεπτο των κινήσεων.

1. Lyapunov Εκθέτης (Lyapunov Exponent):

$$\lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \left| \frac{dx(t)}{dx_0} \right|$$

Μετρά το ρυθμό απόκλισης κοντινών τροχιών σε ένα δυναμικό σύστημα, υποδεικνύοντας χάος ή σταθερότητα.

2. Εντροπία Δείγματος (Sample Entropy):

$$\text{Sample Entropy} = -\log \frac{A}{B}$$

Ποσοτικοποιεί την πολυπλοκότητα ή την παρατυπία μιας χρονοσειράς, μετρώντας την αρνητική λογαριθμική πιθανότητα παρόμοιων μοτίβων να παραμένουν κοντά.

3. Higuchi Fractal Dimension:

$$\text{Higuchi Fractal Dimension} = \frac{\log(L(k))}{\log(k)}$$

Εκτιμά τη διάσταση φράκταλ ενός σήματος για να αξιολογήσει την αυτοομοιότητα και την πολυπλοκότητά του σε διάφορες κλίμακες (Higuchi, 1988).

4. Root Mean Square of Successive Differences (RMSSD):

$$RMSSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2}$$

Αξιολογεί τη βραχυπρόθεσμη μεταβλητότητα του καρδιακού ρυθμού (ή την ομαλότητα του σήματος) υπολογίζοντας τη μέση τετραγωνική ρίζα των διαδοχικών διαφορών.

5. Mean Absolute Deviation:

$$MAD = \frac{1}{N} \sum_{i=1}^N |x_i - \text{mean}|$$

Μετρά τη μέση απόλυτη απόκλιση των σημείων δεδομένων από τη μέση τιμή τους, υποδεικνύοντας τη μεταβλητότητα ή τη διασπορά.

3.1.4 Αιτιολόγηση επιλογής χαρακτηριστικών

Η επιλογή των προαναφερθέντων χαρακτηριστικών βασίζεται σε εκτεταμένη βιβλιογραφία και έρευνα σχετικά με την ανίχνευση της PD. Κάθε κατηγορία χαρακτηριστικών επιλέχθηκε για την ικανότητά της να καταγράφει συγκεκριμένα κινητικά συμπτώματα και φυσιολογικά μοτίβα που σχετίζονται με τη νόσο PD. Παρακάτω, παρέχουμε λεπτομερή αιτιολογία για τη συμπερίληψη αυτών των χαρακτηριστικών.

3.1.4.1 Στατιστικά χαρακτηριστικά

Τα στατιστικά χαρακτηριστικά είναι θεμελιώδη για την ποσοτικοποίηση της έντασης, της μεταβλητότητας και της κατανομής των κινητικών σημάτων. Αυτά τα χαρακτηριστικά έχουν χρησιμοποιηθεί ευρέως στην έρευνα για την PD λόγω της ικανότητάς τους να καταγράφουν βασικές πτυχές της κινητικής δυσλειτουργίας:

RMS: Το RMS είναι ένα ισχυρό μέτρο της ενέργειας του σήματος, που αντανακλά τη συνολική ένταση των κινήσεων. Είναι ιδιαίτερα χρήσιμο για την ανίχνευση αλλαγών στο πλάτος της κίνησης, το οποίο συχνά μειώνεται σε ασθενείς με PD λόγω βραδυκινησίας (Giannakopoulou et al., 2022).

Διακύμανση: Η διακύμανση ποσοτικοποιεί τη μεταβλητότητα του σήματος, η οποία είναι ζωτικής σημασίας για τον εντοπισμό παρατυπιών στα μοτίβα κίνησης. Η αυξημένη μεταβλητότητα συχνά σχετίζεται με κινητική αστάθεια και τρόμο, κοινά συμπτώματα της νόσου PD (Chen et al., 2018).

Εύρος: Το εύρος παρέχει πληροφορίες σχετικά με τα άκρα του εύρους της κίνησης, τα οποία μπορούν να βοηθήσουν στη διαφοροποίηση μεταξύ υγιών ατόμων και ασθενών με PD, καθώς οι τελευταίοι συχνά παρουσιάζουν μειωμένο εύρος κίνησης (Javaid et al., 2022).

Λοξότητα και Κύρτωση: Αυτές οι στατιστικές ροπές ανώτερης τάξης αποτυπώνουν την ασυμμετρία και την οξύτητα της κατανομής του σήματος, αντίστοιχα. Είναι χρήσιμες για την ανίχνευση λεπτών ανωμαλιών στα μοτίβα κίνησης που μπορεί να μην είναι εμφανείς από τα βασικά στατιστικά στοιχεία (Jankovic, 2008; Peever and Fuller, 2017).

3.1.4.2 Χαρακτηριστικά στο πεδίο της συχνότητας

Τα χαρακτηριστικά στο πεδίο της συχνότητας είναι απαραίτητα για την ανάλυση των φασματικών χαρακτηριστικών των σημάτων κίνησης, ιδίως για την ανίχνευση και τον ποσοτικό προσδιορισμό του τρόμου:

- **Κυρίαρχη συχνότητα (Dominant Frequency):** Η κυρίαρχη συχνότητα αποτελεί βασικό δείκτη του τρόμου, ενός από τα χαρακτηριστικά συμπτώματα της νόσου PD. Προσδιορίζει τη συχνότητα στην οποία το σήμα παρουσιάζει τη μεγαλύτερη ισχύ, παρέχοντας ένα άμεσο μέτρο της έντασης του τρόμου (Helmich et al., 2012).
- **Φασματική Εντροπία (Spectral Entropy):** Η φασματική εντροπία μετρά την αταξία στο φάσμα ισχύος, αντανακλώντας την πολυπλοκότητα του σήματος. Χαμηλότερες τιμές εντροπίας συνδέονται συχνά με πιο τακτικούς και περιοδικούς τρόμους, οι οποίοι είναι χαρακτηριστικοί της PD (Pincus, 1991).
- **Φασματικό κεντροειδές (Spectral Centroid):** Το φασματικό κεντροειδές αντιπροσωπεύει το «κέντρο μάζας» του φάσματος ισχύος. Οι μετατοπίσεις στο κεντροειδές μπορεί να υποδεικνύουν αλλαγές στην κατανομή συχνότητας του τρόμου, παρέχοντας πρόσθετες πληροφορίες για τα κινητικά συμπτώματα (Gubbi et al., 2013).
- **Φασματικό Roll-off (85%):** Αυτό το χαρακτηριστικό προσδιορίζει τη συχνότητα κάτω από την οποία συγκεντρώνεται το 85% της ισχύος του

σήματος. Είναι χρήσιμο για τον χαρακτηρισμό του εύρους ζώνης του τρόμου και άλλων περιοδικών κινήσεων (Piau et al., 2018).

3.1.4.3 Χρονικό πεδίο και δυναμικά χαρακτηριστικά

Τα χαρακτηριστικά του πεδίου του χρόνου και τα δυναμικά χαρακτηριστικά είναι κρίσιμα για την αποτύπωση της χρονικής εξέλιξης και της πολυπλοκότητας των σημάτων κίνησης, τα οποία συχνά μεταβάλλονται στην PD:

- **Εντροπία δείγματος (Sample Entropy):** Η εντροπία δείγματος ποσοτικοποιεί την κανονικότητα και την προβλεψιμότητα μιας χρονοσειράς. Στην PD, συχνά παρατηρούνται μειωμένες τιμές εντροπίας λόγω της αυξημένης ακαμψίας και της μειωμένης πολυπλοκότητας των κινήσεων (Patel et al., 2009).
- **RMSSD:** Το RMSSD μετρά τη μεταβλητότητα μεταξύ διαδοχικών σημείων δεδομένων, παρέχοντας πληροφορίες σχετικά με την ομαλότητα των κινήσεων. Οι ασθενείς με PD εμφανίζουν συχνά υψηλότερες τιμές RMSSD λόγω της ακανόνιστης και σπασμωδικής κίνησης των κινήσεών τους.
- **Εκθέτης Lyapunov:** Ο εκθέτης Lyapunov ποσοτικοποιεί την ευαισθησία ενός συστήματος στις αρχικές συνθήκες, αντικατοπτρίζοντας τη χαοτική φύση των κινητικών σημάτων. Στην PD, οι τροποποιημένοι εκθέτες Lyapunov μπορούν να υποδηλώνουν διαταραχές στον κινητικό έλεγχο και τον συντονισμό.
- **Higuchi Fractal Dimension:** Αυτό το χαρακτηριστικό μετρά τη πολυπλοκότητα του σήματος, η οποία συχνά μειώνεται στην PD λόγω της απώλειας του λεπτού κινητικού ελέγχου και της αυξημένης ακαμψίας.

3.1.4.4. Χαρακτηριστικά ενέργειας

Τα χαρακτηριστικά ενέργειας παρέχουν ένα άμεσο μέτρο της έντασης και των επιπέδων δραστηριότητας των κινητικών σημάτων:

- **Απόλυτη ενέργεια (Absolute Energy):** Αυτό το χαρακτηριστικό αθροίζει τις απόλυτες τιμές του σήματος, αντικατοπτρίζοντας το συνολικό επίπεδο δραστηριότητας. Είναι ιδιαίτερα χρήσιμο για την ανίχνευση της υποκινησίας (μειωμένο πλάτος κίνησης) σε ασθενείς με PD (Li et al., 2024).
- **Συνολική Ενέργεια (Total Energy):** Η συνολική ενέργεια καταγράφει τη σωρευτική ισχύ του σήματος, η οποία συχνά μειώνεται στην PD λόγω της βραδυκίνησας και της ακαμψίας.
- **Λογαριθμική Ενέργεια:** Ο λογαριθμικός μετασχηματισμός της ενέργειας δίνει έμφαση στις κινήσεις χαμηλότερης έντασης, οι οποίες είναι συχνά πιο κατατοπιστικές για την ανίχνευση της PD σε πρώιμο στάδιο.

Συμπεραίνοντας, ο συνδυασμός στατιστικών χαρακτηριστικών, χαρακτηριστικών στο πεδίο της συχνότητας και χαρακτηριστικών στο πεδίο του χρόνου παρέχει ένα ολοκληρωμένο πλαίσιο για την ανάλυση

3.2 Ανάλυση των επιδράσεων της πρόσληψης χαπιών στα εξαγόμενα χαρακτηριστικά

Αυτή η ενότητα επικεντρώνεται στην ανάλυση των δεδομένων αισθητήρων (επιταχυνσιόμετρο, γυροσκόπιο, μαγνητόμετρο) χρησιμοποιώντας την ευκλείδεια απόσταση σε συνδυασμό με το αθροιστικό άθροισμα (cumsum) και t-tests για τη στατιστική αξιολόγηση των χαρακτηριστικών. Η ανάλυση πραγματοποιείται με τη χρήση ολισθαίνοντων παραθύρων για την καταγραφή των χρονικών διακυμάνσεων των δεδομένων, επιτρέποντας τη λεπτομερή εξέταση των επιδράσεων της φαρμακευτικής αγωγής στα κινητικά συμπτώματα της PD.

3.2.1 Ευκλείδεια απόσταση και αθροιστικό άθροισμα (Cumsum)

Η Ευκλείδεια απόσταση χρησιμοποιείται για τη συνολική μέτρηση της κίνησης από τα τρισδιάστατα δεδομένα (x, y, z) κάθε αισθητήρα. Ορίζεται ως:

$d(t) = \sqrt{x^2(t) + y^2(t) + z^2(t)}$, όπου $x(t)$, $y(t)$, $z(t)$ είναι οι τιμές των αξόνων x , y , z σε κάθε χρονική στιγμή t .

Το **αθροιστικό άθροισμα (cumsum)** υπολογίζεται ως:

$$\text{cumsum}(t) = \sum_{i=1}^t x(i)$$

Αυτή η μέθοδος επιτρέπει την ανίχνευση μακροπρόθεσμων τάσεων και αλλαγών στην κινητική δραστηριότητα, οι οποίες μπορεί να σχετίζονται με την επίδραση της φαρμακευτικής αγωγής (Chen et al., 2018).

3.2.2 Ανάλυση με ολισθαίνοντα παράθυρα

Για την ανάλυση των δεδομένων, χρησιμοποιήθηκαν ολισθαίνοντα παράθυρα με τα ακόλουθα χαρακτηριστικά:

- Μέγεθος παραθύρου: 100 δείγματα (1 δευτερόλεπτο για ρυθμό δειγματοληψίας 100 Hz).
- Βήμα επικάλυψης: 50% (50 δείγματα, 0.5 δευτερόλεπτα).

Σε κάθε παράθυρο, εξάγονται τα ακόλουθα χαρακτηριστικά:

1. Αθροιστικό άθροισμα (cumsum)
2. Ευκλείδεια απόσταση
3. Εφαρμογή των τύπων για το κάθε χαρακτηριστικό

3.2.3 Εφαρμογή t-tests

Για κάθε χαρακτηριστικό, πραγματοποιήθηκε ζευγαρωτός t-test για τη σύγκριση των τιμών πριν και μετά τη λήψη των χαπιών. Οι υποθέσεις του t-test είναι:

- **Μηδενική υπόθεση (H_0):** Δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των τιμών πριν και μετά τη λήψη χαπιών.
- **Εναλλακτική υπόθεση (H_1):** Υπάρχει στατιστικά σημαντική διαφορά.

$$\text{Ο τύπος του t-test είναι: } t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

όπου $\overline{X_1}, \overline{X_2}$ είναι οι μέσες τιμές των δύο ομάδων, s_1^2, s_2^2 είναι οι διακυμάνσεις και n_1, n_2 τα μεγέθη των δειγμάτων.

3.2.4 Επιλογή σημαντικών χαρακτηριστικών

Οι τιμές p-value από τους t-tests χρησιμοποιήθηκαν για την επιλογή των πιο σημαντικών χαρακτηριστικών. Ένα χαρακτηριστικό θεωρείται στατιστικά σημαντικό εάν $p < 0.05$.

3.2.5 Ανάλυση των μεθόδων που χρησιμοποιήθηκαν

Η επιλογή των μεθόδων ανάλυσης που χρησιμοποιήθηκαν στην παρούσα μελέτη βασίστηκε στην ανάγκη για ακριβή και ερμηνεύσιμα αποτελέσματα, ιδιαίτερα σε ένα πεδίο όπως η ανάλυση δεδομένων για τη νόσο του Πάρκινσον, όπου τα δεδομένα είναι πολύπλοκα, θορυβώδη και πολυδιάστατα. Οι μέθοδοι που επιλέχθηκαν (Ευκλείδεια απόσταση, αθροιστικό άθροισμα, ολισθαίνοντα παράθυρα και t-tests) συνδυάζονται για να παρέχουν μια ολοκληρωμένη προσέγγιση, η οποία καλύπτει τόσο την ποσοτική ανάλυση όσο και την ερμηνεία των αποτελεσμάτων.

Ευκλείδεια απόσταση

Η Ευκλείδεια απόσταση επιλέχθηκε ως βασική μέθοδος για τη μέτρηση της συνολικής κινητικής δραστηριότητας από τα τρισδιάστατα δεδομένα (x, y, z) των αισθητήρων. Αυτή η μέθοδος προσφέρει μια απλή αλλά αποτελεσματική μέτρηση της απόστασης, η οποία μπορεί να χρησιμοποιηθεί για την αξιολόγηση της έντασης και της μεταβλητότητας των κινήσεων (Chen et al., 2018). Η απλότητα της Ευκλείδειας απόστασης την καθιστά ιδανική για την αρχική επεξεργασία των δεδομένων, ενώ ταυτόχρονα παρέχει μια βάση για πιο προηγμένες αναλύσεις.

Αθροιστικό άθροισμα (Cumsum)

Το αθροιστικό άθροισμα (cumsum) χρησιμοποιήθηκε για την ανίχνευση μακροπρόθεσμων τάσεων και αλλαγών στην κινητική δραστηριότητα. Αυτή η

μέθοδος είναι ιδιαίτερα χρήσιμη για την παρακολούθηση της εξέλιξης της νόσου και την αξιολόγηση της αποτελεσματικότητας της φαρμακευτικής αγωγής, καθώς καταγράφει τη συνολική κινητική ενέργεια σε μια χρονική περίοδο (Giannakopoulou et al., 2022). Το cumsum προσφέρει μια μακροπρόθεσμη προοπτική, η οποία συμπληρώνει την τοπική ανάλυση που παρέχουν οι άλλες μέθοδοι.

Ολισθαίνοντα παράθυρα

Η χρήση ολισθαίνων παραθύρων επιτρέπει την ανάλυση των δεδομένων σε τοπικό επίπεδο, καταγράφοντας τις χρονικές μεταβολές και τις παροδικές καταστάσεις. Αυτή η μέθοδος είναι ιδιαίτερα σημαντική για την ανίχνευση μικρών αλλαγών στα κινητικά χαρακτηριστικά, οι οποίες μπορεί να σχετίζονται με την επίδραση της φαρμακευτικής αγωγής (Patel et al., 2009). Το μέγεθος και το βήμα επικάλυψης των παραθύρων επιλέχθηκαν προσεκτικά για να εξισορροπηθεί η ανάλυση μεταξύ λεπτομέρειας και υπολογιστικής αποδοτικότητας.

T-tests

Οι ζευγαρωτοί t-tests χρησιμοποιήθηκαν για τη στατιστική αξιολόγηση των διαφορών στα χαρακτηριστικά πριν και μετά τη λήψη των χαπιών. Αυτή η μέθοδος προσφέρει μια αντικειμενική και στατιστικά τεκμηριωμένη προσέγγιση για τον εντοπισμό σημαντικών αλλαγών, βοηθώντας στην επιλογή των πιο σχετικών χαρακτηριστικών για περαιτέρω ανάλυση (Tukey, 1977). Η χρήση των t-tests εξασφαλίζει ότι οι παρατηρούμενες διαφορές δεν οφείλονται σε τυχαίες μεταβολές, αλλά έχουν στατιστική σημασία.

Συνδυαστική χρήση μεθόδων

Ο συνδυασμός αυτών των μεθόδων επιτρέπει μια ολοκληρωμένη ανάλυση των δεδομένων, η οποία καλύπτει τόσο την τοπική όσο και την παγκόσμια δομή τους. Η Ευκλείδεια απόσταση και το cumsum παρέχουν μια μακροπρόθεσμη προοπτική, ενώ τα ολισθαίνοντα παράθυρα και οι t-tests επιτρέπουν την ανίχνευση μικρών αλλαγών και τη στατιστική αξιολόγησή τους. Αυτή η συνδυαστική προσέγγιση είναι ιδιαίτερα σημαντική σε ιατρικές εφαρμογές, όπου η ακρίβεια και η ερμηνευσιμότητα των αποτελεσμάτων είναι κρίσιμες (Goetz et al., 2008).

Συμπεράσματα

Συμπεράσματα

Η χρήση της Ευκλείδειας απόστασης και του αθροιστικού αθροίσματος (cumsum), σε συνδυασμό με ολισθαίνοντα παράθυρα και t-tests, αποδείχθηκε ιδιαίτερα αποτελεσματική στην ανίχνευση στατιστικά σημαντικών διαφορών στα κινητικά χαρακτηριστικά πριν και μετά τη λήψη φαρμακευτικής αγωγής. Όπως φαίνεται στον **Πίνακα Α1 (Παράρτημα Α)**, χαρακτηριστικά όπως:

- **Μη γραμμικά μεγέθη** (Lyapunov exponent, Higuchi fractal dimension),
- **Μέτρα εντροπίας** (sample entropy, spectral entropy), και
- **Στατιστικές παράμετροι** (RMS, μέση τιμή)

εμφανίζουν έντονες στατιστικές διαφορές ($p < 0.05$), υπογραμμίζοντας την ευαισθησία της μεθοδολογίας. Τα επιλεγμένα χαρακτηριστικά προσφέρουν **πολύτιμες βιοδείκτες** για:

1. **Παρακολούθηση της εξέλιξης** της νόσου, και
2. **Αντικειμενική αξιολόγηση** της θεραπευτικής απόκρισης (Giannakopoulou et al., 2022).

3.3 Δέντρα αποφάσεων και ο ρόλος τους στην ανάλυση της νόσου του Πάρκινσον

Οι μέθοδοι ML, και ειδικά τα δέντρα αποφάσεων, έχουν αποδειχθεί ιδιαίτερα χρήσιμα στην ανάλυση και διάγνωση της PD. Η ικανότητα τους να εξαγάγουν κανόνες από δεδομένα και να παρέχουν ερμηνεύσιμα αποτελέσματα τα καθιστά ιδανικά για την επιλογή χαρακτηριστικών και την ανάπτυξη μοντέλων πρόβλεψης. Σε αυτή την ενότητα, παρουσιάζεται ο ρόλος των δέντρων αποφάσεων στην ανάλυση της PD, με έμφαση στην επιλογή των πιο σημαντικών χαρακτηριστικών για την ακριβή διάγνωση και παρακολούθηση της νόσου.

3.3.1 Εισαγωγή στα Δέντρα Αποφάσεων

Τα δέντρα αποφάσεων είναι μη-παραμετρικά μοντέλα ML που χρησιμοποιούνται για κατηγοριοποίηση και παλινδρόμηση. Βασίζονται στην κατασκευή ενός δέντρου, όπου κάθε εσωτερικός κόμβος αντιπροσωπεύει ένα τεστ σε ένα χαρακτηριστικό, κάθε κλάδος αντιπροσωπεύει το αποτέλεσμα του τεστ, και κάθε φύλλο αντιπροσωπεύει μια κλάση ή μια τιμή (Breiman, 2017). Τα δέντρα αποφάσεων έχουν τα ακόλουθα πλεονεκτήματα:

- **Ερμηνευσιμότητα:** Η δομή τους είναι εύκολα κατανοητή και μπορεί να ερμηνευτεί από ανθρώπους.
- **Μη-γραμμικότητα:** Μπορούν να μοντελοποιήσουν πολύπλοκες σχέσεις μεταξύ των χαρακτηριστικών.
- **Ανθεκτικότητα:** Είναι ανθεκτικά σε θόρυβο και ελλιπή δεδομένα.

3.3.2 Επιλογή Χαρακτηριστικών με Δέντρα Αποφάσεων

Η επιλογή χαρακτηριστικών είναι ένα κρίσιμο βήμα στην ανάλυση της PD, καθώς τα δεδομένα από αισθητήρες (π.χ., επιταχυνσιόμετρα, γυροσκόπια) περιλαμβάνουν πολλές διαστάσεις και πληροφορίες που μπορεί να μην είναι όλες σχετικές. Τα δέντρα αποφάσεων μπορούν να χρησιμοποιηθούν για την ταξινόμηση των χαρακτηριστικών με βάση τη σημαντικότητά τους, χρησιμοποιώντας μεθόδους όπως η μείωση της ακαθαρσίας (impurity reduction) ή ο δείκτης Gini.

3.3.2.1 Μείωση της Ακαθαρσίας

Κάθε κόμβος στο δέντρο επιλέγει το χαρακτηριστικό που μεγιστοποιεί τη μείωση της ακαθαρσίας (π.χ., εντροπία ή δείκτης Gini). Αυτή η διαδικασία επιτρέπει την ταξινόμηση των χαρακτηριστικών με βάση τη συμβολή τους στην ακρίβεια του μοντέλου.

3.3.2.2 Σημαντικότητα Χαρακτηριστικών

Μετά την κατασκευή του δέντρου, η σημαντικότητα κάθε χαρακτηριστικού μπορεί να υπολογιστεί ως το άθροισμα της μείωσης της ακαθαρσίας σε όλους τους κόμβους όπου χρησιμοποιείται το χαρακτηριστικό. Αυτή η μέθοδος παρέχει μια

ποσοτική εκτίμηση της συμβολής κάθε χαρακτηριστικού στην πρόβλεψη (Louppe, 2015).

3.3.3 Εφαρμογή στην Ανάλυση της Νόσου του Πάρκινσον

Στην περίπτωση της PD, τα δέντρα αποφάσεων μπορούν να χρησιμοποιηθούν για την επιλογή χαρακτηριστικών που σχετίζονται με τα κύρια συμπτώματα της νόσου, όπως ο τρόμος, η βραδυκίνηση και η δυσκαμψία. Παρακάτω παρουσιάζονται τα πιο σημαντικά χαρακτηριστικά που έχουν αναγνωριστεί από τη βιβλιογραφία.

3.3.3.1 Χαρακτηριστικά Στατιστικής

Μέση τιμή (Mean): Χρησιμοποιείται για την αξιολόγηση της κεντρικής τάσης του σήματος, η οποία μπορεί να αντικατοπτρίζει τη βαρύτητα των συμπτωμάτων (Giannakopoulou et al., 2022).

Τυπική απόκλιση (Standard Deviation): Μετρά τη μεταβλητότητα του σήματος, η οποία είναι σημαντική για την ανίχνευση τρόμου (Chen et al., 2020).

Διακύμανση (Variance): Χρησιμοποιείται για την ανίχνευση αλλαγών στην ένταση των κινήσεων (Patel et al., 2009).

3.3.3.2 Χαρακτηριστικά Συχνοτήτων

Κυρίαρχη Συχνότητα (Dominant Frequency): Αναγνωρίζει την κύρια συχνότητα του τρόμου, η οποία είναι ένα κύριο σύμπτωμα της PD (Helmich et al., 2012).

Φασματική Εντροπία (Spectral Entropy): Μετρά την πολυπλοκότητα του σήματος, η οποία μπορεί να αντικατοπτρίζει την ένταση των συμπτωμάτων (Pincus, 1991).

3.3.3.3 Χαρακτηριστικά Χρονοσειρών

Εντροπία Δείγματος (Sample Entropy): Μετρά την ακανόνιστη συμπεριφορά του σήματος, η οποία είναι χαρακτηριστική της PD (Richman and Moorman, 2000).

Root Mean Square of Successive Differences (RMSSD): Μετρά την ομαλότητα των κινήσεων, η οποία μπορεί να υποδηλώνει βραδυκίνησια (Goetz et al., 2008).

3.3.4 Πλεονεκτήματα και περιορισμοί

3.3.4.1 Πλεονεκτήματα

Εύκολη Εφαρμογή: Τα δέντρα αποφάσεων μπορούν να εφαρμοστούν σε ποικίλα σύνολα δεδομένων χωρίς την ανάγκη για προεπεξεργασία.

Επιλογή Χαρακτηριστικών: Η ικανότητά τους να αναγνωρίζουν τα πιο σημαντικά χαρακτηριστικά τα καθιστά ιδανικά για την ανάλυση της PD.

3.3.5 Περιορισμοί

- **Υπερπροσαρμογή (Overfitting):** Τα δέντρα αποφάσεων μπορεί να υπερπροσαρμοστούν σε θορυβώδη δεδομένα, εάν δεν εφαρμοστούν μέθοδοι κλάδευσης (pruning).
- **Ευαισθησία σε Ασυμμετρίες:** Η απόδοσή τους μπορεί να μειωθεί σε ασύμμετρα δεδομένα, όπου κάποιες κλάσεις είναι υπερπροσδιορισμένες.

3.3.5 Συμπεράσματα

Τα δέντρα αποφάσεων αποτελούν μια ισχυρή μέθοδο για την ανάλυση της PD, ιδιαίτερα στην επιλογή χαρακτηριστικών και την ανάπτυξη μοντέλων πρόβλεψης, τα οποία αποδείχθηκαν ιδιαίτερα αποτελεσματικά στην ανάλυση της PD.

Η ταξινόμηση των χαρακτηριστικών σε καλά (σημαντικά) και κακά (ασήμαντα) βασίστηκε στις βαθμολογίες σημαντικότητας που προέκυψαν από έναν ταξινομητή Random Forest. Το Random Forest είναι μια μέθοδος μάθησης συνόλου που κατασκευάζει πολλαπλά δέντρα απόφασης κατά τη διάρκεια της εκπαίδευσης και εξάγει τη σημασία κάθε χαρακτηριστικού ως μέτρο της συμβολής του στην προγνωστική απόδοση του μοντέλου (Breiman, 2001).

Στην παρούσα μελέτη, η μέση βαθμολογία σημαντικότητας σε όλα τα χαρακτηριστικά χρησιμοποιήθηκε ως κριτήριο επιλογής για τη διάκριση μεταξύ καλών και κακών χαρακτηριστικών. Χαρακτηριστικά με βαθμολογίες

σημαντικότητας μεγαλύτερες ή ίσες με τη μέση τιμή θεωρήθηκαν καλά, καθώς συνέβαλαν πιο σημαντικά στην ικανότητα του μοντέλου να διακρίνει μεταξύ των καταστάσεων πριν το χάπι και μετά το χάπι.

Αντίθετα, τα χαρακτηριστικά με βαθμολογίες σημαντικότητας κάτω από το μέσο όρο ταξινομήθηκαν ως κακά, υποδεικνύοντας τη σχετικά χαμηλότερη συμβολή τους στην απόδοση του μοντέλου. Αυτή η προσέγγιση εξασφαλίζει έναν διαχωρισμό των χαρακτηριστικών με βάση τα δεδομένα και αντικειμενικά, επιτρέποντας την εστιασμένη ανάλυση στις μεταβλητές με τον μεγαλύτερο αντίκτυπο.

Σύμφωνα με τα αποτελέσματα του **Πίνακα A2 (Παράρτημα A)**, τα πιο σημαντικά χαρακτηριστικά για τη διάγνωση είναι:

1. Μη γραμμικά χαρακτηριστικά:

- sample_entropy (έως 0.076)
- lyapunov_exponent (έως 0.062)
- higuchi_fractal_dimension (έως 0.060)

2. Στατιστικά/φασματικά χαρακτηριστικά:

- rmssd (έως 0.059)
- kurtosis (έως 0.053)
- spectral_flux (έως 0.049)

Ο Random Forest επέτρεψε την ταυτοποίηση αυτών των βιοδεικτών, συνδυάζοντας υψηλή ακρίβεια με ερμηνευσιμότητα.

3.4 Ανάλυση Κύριων Συνιστωσών (PCA) και η Σημασία της στην Ανάλυση της Νόσου του Πάρκινσον

Η Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis - PCA) είναι μια διαστασιακή τεχνική μείωσης που χρησιμοποιείται ευρέως στην ανάλυση δεδομένων για την εξαγωγή των πιο σημαντικών πληροφοριών από πολυδιάστατα δεδομένα. Στην περίπτωση της PD, η PCA μπορεί να βοηθήσει στην ταυτοποίηση των χαρακτηριστικών που περιγράφουν καλύτερα τις κινητικές διαταραχές και στην απλοποίηση της ανάλυσης χωρίς απώλεια κρίσιμων πληροφοριών (Jolliffe, 2002).

3.4.1 Βασικές Αρχές της PCA

Η PCA μετασχηματίζει τα αρχικά δεδομένα σε ένα νέο σύστημα συντεταγμένων, όπου οι νέες μεταβλητές (κύριες συνιστώσες) είναι γραμμικοί συνδυασμοί των αρχικών χαρακτηριστικών. Οι κύριες συνιστώσες είναι ταξινομημένες με βάση τη διακύμανση που εξηγούν, με την πρώτη κύρια συνιστώσα να εξηγεί τη μεγαλύτερη διακύμανση, η δεύτερη τη δεύτερη μεγαλύτερη, και ούτω καθεξής (Abdi and Williams, 2010). Τα βήματα της PCA περιλαμβάνουν:

1. **Κανονικοποίηση των δεδομένων:** Τα δεδομένα κλιμακοποιούνται ώστε να έχουν μηδενικό μέσο όρο και μοναδιαία διακύμανση.
2. **Υπολογισμός του πίνακα συνδιακύμανσης:** Αυτός ο πίνακας περιγράφει τις σχέσεις μεταξύ των χαρακτηριστικών.
3. **Εξαγωγή ιδιοτιμών και ιδιοδιανυσμάτων:** Τα ιδιοδιανύσματα καθορίζουν τις κατευθύνσεις των κύριων συνιστωσών, ενώ οι ιδιοτιμές υποδεικνύουν τη διακύμανση που εξηγεί κάθε συνιστώσα.
4. **Προβολή των δεδομένων:** Τα αρχικά δεδομένα προβάλλονται στον νέο χώρο των κύριων συνιστωσών.

3.4.2 Εφαρμογή της PCA στην Ανάλυση της PD

Στην ανάλυση της PD, η PCA μπορεί να χρησιμοποιηθεί για:

- **Μείωση της διαστατικότητας:** Πολλά από τα χαρακτηριστικά που προέρχονται από αισθητήρες (π.χ., επιταχυνσιόμετρα, γυροσκόπια) μπορεί να είναι υψηλά συσχετισμένα ή περιττά. Η PCA μειώνει τον αριθμό των χαρακτηριστικών διατηρώντας τις πιο σημαντικές πληροφορίες (Giannakopoulou et al., 2022).
- **Εξαγωγή χαρακτηριστικών:** Οι κύριες συνιστώσες μπορούν να χρησιμοποιηθούν ως νέα χαρακτηριστικά για την κατασκευή μοντέλων ML, βελτιώνοντας την απόδοση και την ερμηνευσιμότητα (Patel et al., 2009).

3.4.3 Πλεονεκτήματα της PCA

Μείωση Θορύβου: Η PCA μειώνει την επίδραση του θορύβου στα δεδομένα, διατηρώντας τις πιο σημαντικές πληροφορίες (Jolliffe, 2002).

Βελτίωση της Απόδοσης: Η μείωση της διαστατικότητας οδηγεί σε γρηγορότερη εκπαίδευση και εφαρμογή των μοντέλων ML (Abdi and Williams, 2010).

Ερμηνευσιμότητα: Οι κύριες συνιστώσες μπορούν να ερμηνευτούν ως συνδυασμοί των αρχικών χαρακτηριστικών, παρέχοντας πληροφορίες για τις βασικές δομές των δεδομένων (Giannakopoulou et al., 2022).

3.4.4 Περιορισμοί της PCA

Γραμμική Υπόθεση: Η PCA βασίζεται στην υπόθεση ότι οι σχέσεις μεταξύ των χαρακτηριστικών είναι γραμμικές, κάτι που μπορεί να μην ισχύει σε πολύπλοκα δεδομένα (Jolliffe, 2002).

Απώλεια Ερμηνευσιμότητας: Οι κύριες συνιστώσες είναι γραμμικοί συνδυασμοί των αρχικών χαρακτηριστικών, γεγονός που μπορεί να δυσκολεύει την ερμηνεία των αποτελεσμάτων (Abdi and Williams, 2010).

Ευαισθησία στην Κλιμακοποίηση: Η απόδοση της PCA μπορεί να επηρεαστεί από την κλιμακοποίηση των δεδομένων, απαιτώντας προσεκτική προεπεξεργασία (Patel et al., 2009).

3.4.5 Συμπεράσματα

Η PCA αποτελεί μια ισχυρή μέθοδο για την ανάλυση της PD, ιδιαίτερα στην επιλογή και μείωση των χαρακτηριστικών. Η ικανότητά της να εξάγει τις πιο σημαντικές πληροφορίες από πολυδιάστατα δεδομένα την καθιστά απαραίτητη για την ανάπτυξη ακριβών και ερμηνεύσιμων μοντέλων (Giannakopoulou et al., 2022). Μελλοντικές έρευνες θα μπορούσαν να εστιάσουν στη συνδυασμένη χρήση της PCA με άλλες τεχνικές ML, όπως τα δέντρα αποφάσεων ή νευρωνικά δίκτυα, για περαιτέρω βελτίωση της απόδοσης.

Η ταξινόμηση των χαρακτηριστικών σε καλά (σημαντικά) και κακά (ασήμαντα) βασίστηκε στις βαθμολογίες σημαντικότητας που προέκυψαν από την PCA. Η PCA είναι μια τεχνική μείωσης της διαστατικότητας που μετασχηματίζει τα δεδομένα σε ένα σύνολο ορθογώνιων συνιστωσών, όπου κάθε συνιστώσα καταγράφει ένα μέρος της διακύμανσης των δεδομένων. Η σημασία κάθε χαρακτηριστικού υπολογίζεται ως το άθροισμα των απόλυτων φορτίσεων του σε όλες τις κύριες συνιστώσες, αντικατοπτρίζοντας τη συμβολή του στη συνολική διακύμανση που εξηγείται από το μοντέλο (Jolliffe and Cadima, 2016).

Στην παρούσα μελέτη, η μέση βαθμολογία σημαντικότητας σε όλα τα χαρακτηριστικά χρησιμοποιήθηκε ως κατώφλι για τη διάκριση μεταξύ καλών και κακών χαρακτηριστικών. Χαρακτηριστικά με βαθμολογίες σημαντικότητας μεγαλύτερες ή ίσες με τη μέση τιμή ταξινομήθηκαν ως καλά, καθώς συνέβαλαν πιο σημαντικά στη διακύμανση που αποτυπώθηκε από τις κύριες συνιστώσες.

Αντίθετα, τα χαρακτηριστικά με βαθμολογίες σημαντικότητας κάτω από το μέσο όρο ταξινομήθηκαν ως κακά, υποδεικνύοντας τη σχετικά χαμηλότερη συμβολή τους. Η προσέγγιση αυτή εξασφαλίζει έναν διαχωρισμό των χαρακτηριστικών με βάση τα δεδομένα και αντικειμενικά, επιτρέποντας την εστιασμένη ανάλυση στις μεταβλητές με τον μεγαλύτερο αντίκτυπο.

Συμπεράσματα για την PCA

Η ανάλυση PCA ανέδειξε τα πιο σημαντικά χαρακτηριστικά για τη μελέτη της νόσου του Πάρκινσον, όπως φαίνεται στον **Πίνακα A3 (Παράρτημα A)**. Τα κύρια ευρήματα είναι:

1. Κορυφαία Χαρακτηριστικά:

- **Φασματικά:** spectral_flux (έως 4.06), spectral_flatness (έως 3.95)
- **Μη γραμμικά:** lyapunov_exponent (έως 3.92),
higuchi_fractal_dimension (έως 3.90)
- **Στατιστικά:** range (έως 4.31), variance (έως 3.88)

2. Αισθητήρες:

- **Επιταχυνσιόμετρο:** Δυνατά φασματικά χαρακτηριστικά (spectral_flux 4.06)
- **Γυροσκόπιο:** Υψηλή σημασία χρονικών παραμέτρων (range 4.31)
- **Μαγνητόμετρο:** Εξέχοντα μη γραμμικά χαρακτηριστικά (higuchi_fractal_dimension 3.85)

3. Μεθοδολογία:

- Χαρακτηριστικά με σκορ > 3.0 (μέσος όρος) κρίθηκαν σημαντικά
- 78% των χαρακτηριστικών ξεπέρασαν το κατώφλι σημαντικότητας

Προοπτικές:

- Ο συνδυασμός PCA με νευρωνικά δίκτυα μπορεί να βελτιώσει την ακρίβεια
- Η εφαρμογή σε πιο μεγάλα δείγματα θα επιβεβαιώσει τη σταθερότητα των αποτελεσμάτων

3.5 Μεθοδολογία επιλογής χαρακτηριστικών

Για τον εντοπισμό των πιο σημαντικών χαρακτηριστικών για τη διάκριση μεταξύ των καταστάσεων πριν το χάπι και μετά το χάπι, χρησιμοποιήθηκε μια συνδυασμένη προσέγγιση επιλογής χαρακτηριστικών. Αυτή η προσέγγιση

αξιοποιεί τα πλεονεκτήματα τριών διαφορετικών μεθόδων: t-tests, Random Forest και PCA. Συνδυάζοντας τα αποτελέσματα από αυτές τις μεθόδους, επιτεύχθηκε μια στιβαρή και καθοδηγούμενη από τα δεδομένα διαδικασία επιλογής χαρακτηριστικών.

1. T-Test για στατιστική σημαντικότητα

Το t-test χρησιμοποιήθηκε για την αξιολόγηση της στατιστικής σημαντικότητας κάθε χαρακτηριστικού στη διαφοροποίηση μεταξύ των δύο καταστάσεων. Χαρακτηριστικά με χαμηλές τιμές p-values ($p < 0,05$) θεωρήθηκαν στατιστικά σημαντικά. Για να ποσοτικοποιηθεί η σημασία κάθε χαρακτηριστικού, υπολογίστηκε ο αρνητικός λογάριθμος της τιμής p-value:

$$Score_{T-Test} = -\log_{10}(p_{value} + \varepsilon)$$

όπου $\varepsilon = 10^{-10}$ είναι μία μικρή σταθερά που προστίθεται για να αποφευχθεί η λήψη του λογαρίθμου του μηδενός. Αυτός ο μετασχηματισμός διασφαλίζει ότι τα χαρακτηριστικά με χαμηλότερες τιμές p λαμβάνουν υψηλότερες βαθμολογίες, αντανakλώντας τη μεγαλύτερη σημασία τους (Jolliffe and Cadima, 2016).

2. Random Forest για τη σημασία των χαρακτηριστικών

Ένας ταξινομητής Random Forest εκπαιδεύτηκε στο σύνολο δεδομένων για να αξιολογήσει τη σημασία κάθε χαρακτηριστικού. Η βαθμολογία σημαντικότητας για κάθε χαρακτηριστικό υπολογίστηκε ως η μέση μείωση της ακαθαρσίας (ακαθαρσία Gini) σε όλα τα δέντρα απόφασης στο δάσος. Τα χαρακτηριστικά με υψηλότερη βαθμολογία σημαντικότητας συμβάλλουν περισσότερο στην προβλεπτική απόδοση του μοντέλου (Breiman, 2001).

3. PCA για τη συμβολή της διακύμανσης

Η PCA εφαρμόστηκε στο τυποποιημένο σύνολο δεδομένων για τη μετατροπή των χαρακτηριστικών σε ένα σύνολο ορθογώνιων συνιστωσών. Η σημασία κάθε χαρακτηριστικού υπολογίστηκε ως το άθροισμα των απόλυτων φορτίσεων του σε

όλες τις κύριες συνιστώσες, αντικατοπτρίζοντας τη συμβολή του στη συνολική διακύμανση που εξηγείται από το μοντέλο (Jolliffe and Cadima, 2016).

4. Συνδυασμός αποτελεσμάτων

Οι βαθμολογίες από το t-test, το Random Forest και το PCA κανονικοποιήθηκαν σε ένα εύρος από 0 έως 1 χρησιμοποιώντας την κανονικοποίηση min-max. Στη συνέχεια υπολογίστηκε μια σύνθετη βαθμολογία ως σταθμισμένος μέσος όρος των κανονικοποιημένων βαθμολογιών:

$$Composite\ Score = w_1 \cdot Score_{T-Test} + w_2 \cdot Score_{RF} + w_3 \cdot Score_{PCA}$$

με $w_1 = 0.4$, $w_2 = 0.3$, $w_3 = 0.3$

είναι τα βάρη που αποδίδονται στις βαθμολογίες t-test, Random Forest και PCA, αντίστοιχα. Αυτά τα βάρη επιλέχθηκαν για να δοθεί προτεραιότητα στη στατιστική σημαντικότητα, ενώ παράλληλα λαμβάνεται υπόψη η συμβολή της μάθησης συνόλου και των μεθόδων που βασίζονται στη διακύμανση.

5. Επιλογή των καλύτερων χαρακτηριστικών

Τα χαρακτηριστικά κατατάχθηκαν με βάση τις σύνθετες βαθμολογίες τους σε φθίνουσα σειρά. Το κορυφαίο 80% των χαρακτηριστικών (Παράθεμα A4) με τις υψηλότερες σύνθετες βαθμολογίες επιλέχθηκαν ως τα καλύτερα χαρακτηριστικά για περαιτέρω ανάλυση. Η προσέγγιση αυτή διασφαλίζει ότι τα επιλεγμένα χαρακτηριστικά είναι τόσο στατιστικά σημαντικά όσο και συμβάλλουν ουσιαστικά στην απόδοση του μοντέλου.

Παρακάτω παρουσιάζονται τα 10% καλύτερα χαρακτηριστικά που επιλέχθηκαν και με τι score:

Exercise	Sensor	Feature Name	Composite Score
2	gyro	rmssd	0.663
2	gyro	lyapunov_exponent	0.644
2	gyro	higuchi_fractal_dimension	0.612
2	gyro	range	0.573
3	acce	higuchi_fractal_dimension	0.469
3	gyro	max	0.452
3	acce	lyapunov_exponent	0.437
2	gyro	spectral_rolloff	0.433

2	acce	sample_entropy	0.422
0	magn	lyapunov_exponent	0.406
0	magn	spectral_variability	0.399
0	magn	higuchi_fractal_dimension	0.399
0	magn	range	0.381
3	magn	rmssd	0.376
0	gyro	rmssd	0.375
2	gyro	std	0.373
3	acce	spectral_flux	0.367
1	gyro	lyapunov_exponent	0.365
1	gyro	rmssd	0.361
1	gyro	max	0.358
2	magn	sample_entropy	0.352
1	acce	sample_entropy	0.347
3	gyro	range	0.337
3	gyro	energy	0.337
1	gyro	energy	0.326

Πίνακας 3.1 Τα 10 % καλύτερα χαρακτηριστικά που επιλέχθηκαν

4. Ανάπτυξη μοντέλου: Εκπαίδευση του μοντέλου μηχανικής μάθησης

4.1 Εισαγωγή στη μηχανική μάθηση για την ανίχνευση της νόσου του Πάρκινσον

Η εφαρμογή της ML στην υγειονομική περίθαλψη έχει φέρει επανάσταση στον τρόπο διάγνωσης και διαχείρισης των ασθενειών, ιδίως στην περίπτωση πολύπλοκων νευροεκφυλιστικών διαταραχών όπως η PD. Η έγκαιρη ανίχνευση της νόσου PD είναι κρίσιμη για τη βελτίωση των αποτελεσμάτων των ασθενών, καθώς επιτρέπει την έγκαιρη παρέμβαση και την εφαρμογή εξατομικευμένων στρατηγικών θεραπείας (Jankovic, 2008). Ωστόσο, η ετερογένεια των συμπτωμάτων της νόσου PD και η έλλειψη οριστικών βιοδεικτών καθιστούν δύσκολη την ακριβή διάγνωση μόνο με τη χρήση παραδοσιακών κλινικών μεθόδων (Goetz et al., 2008). Τα τελευταία χρόνια, η ενσωμάτωση προηγμένων τεχνικών ανάλυσης δεδομένων, όπως η μηχανική μάθηση, με δεδομένα που συλλέγονται από φορητούς αισθητήρες και συσκευές IoT έχει δώσει μεγάλες υποσχέσεις για την αντιμετώπιση αυτών των προκλήσεων (Giannakopoulou et al., 2022).

Στο προηγούμενο κεφάλαιο (Κεφάλαιο 3), διεξήχθη μια ολοκληρωμένη διαδικασία επιλογής χαρακτηριστικών για τον εντοπισμό των πιο σχετικών χαρακτηριστικών για την ανίχνευση της PD. Η διαδικασία αυτή περιελάμβανε τρεις διαφορετικές μεθοδολογίες: στατιστική επιλογή (π.χ. t-tests), σημασία χαρακτηριστικών με βάση το Random Forest και PCA. Αυτές οι μέθοδοι χρησιμοποιήθηκαν για να εξασφαλιστεί μια στιβαρή και καθοδηγούμενη από τα δεδομένα επιλογή χαρακτηριστικών που αποτυπώνουν τα βασικά χαρακτηριστικά της νόσου PD, όπως ο τρόμος, η βραδυκίνηση και η ακαμψία (Helmich et al., 2012). Το τελικό σύνολο επιλεγμένων χαρακτηριστικών, όπως συνοψίζεται στην ενότητα 3.5, παρέχει μια ισχυρή βάση για τη δημιουργία ακριβών και ερμηνεύσιμων μοντέλων ML.

Ο πρωταρχικός στόχος του παρόντος κεφαλαίου είναι η ανάπτυξη και η αξιολόγηση μοντέλων ML για την ανίχνευση της PD με τη χρήση των επιλεγμένων χαρακτηριστικών που προσδιορίστηκαν στο κεφάλαιο 3. Αξιοποιώντας αυτά τα χαρακτηριστικά, στοχεύουμε στη δημιουργία μοντέλων που όχι μόνο επιτυγχάνουν υψηλή ακρίβεια πρόβλεψης αλλά και παρέχουν πληροφορίες για τους υποκείμενους μηχανισμούς της νόσου. Στο παρόν κεφάλαιο θα εξεταστούν διάφοροι αλγόριθμοι ML, συμπεριλαμβανομένων των Logistic Regression, Support Vector Machines (SVM), Random Forest και Gradient Boosting, προκειμένου να εντοπιστεί η πιο αποτελεσματική προσέγγιση για την ανίχνευση της PD. Επιπλέον, θα δοθεί έμφαση στην ερμηνευσιμότητα των μοντέλων, καθώς η κατανόηση της συμβολής των επιμέρους χαρακτηριστικών είναι ζωτικής σημασίας για την κλινική εφαρμοσιμότητα (Lundberg and Lee, 2017).

Η δομή του παρόντος κεφαλαίου έχει ως εξής: Στην ενότητα 4.2 περιγράφεται η διαδικασία προετοιμασίας των δεδομένων, συμπεριλαμβανομένης της χρήσης των επιλεγμένων χαρακτηριστικών και του διαχωρισμού του συνόλου δεδομένων σε σύνολα εκπαίδευσης, επικύρωσης και δοκιμής. Η ενότητα 4.3 παρέχει μια επισκόπηση των αλγορίθμων ML που εξετάστηκαν στην παρούσα μελέτη. Στην ενότητα 4.4 περιγράφεται λεπτομερώς η διαδικασία εκπαίδευσης των μοντέλων και συντονισμού των υπερπαραμέτρων, ενώ στην ενότητα 4.5 παρουσιάζεται η αξιολόγηση των μοντέλων με τη χρήση κατάλληλων μετρικών, ενώ στην Ενότητα 4.6 ακολουθεί συζήτηση των επιπτώσεων, των περιορισμών και των μελλοντικών κατευθύνσεων.

Μέχρι το τέλος αυτού του κεφαλαίου, στοχεύουμε να αποδείξουμε ότι τα επιλεγμένα χαρακτηριστικά, σε συνδυασμό με τις πλέον σύγχρονες τεχνικές ML, μπορούν να βελτιώσουν σημαντικά την ακρίβεια και την αξιοπιστία της ανίχνευσης της PD. Η παρούσα εργασία συμβάλλει στον αυξανόμενο όγκο ερευνών σχετικά με την εφαρμογή της ML στην υγειονομική περίθαλψη και αναδεικνύει τις δυνατότητες των προσεγγίσεων που βασίζονται στα δεδομένα για την αντιμετώπιση σύνθετων ιατρικών προκλήσεων.

4.2 Προετοιμασία δεδομένων

Η επιτυχία κάθε μοντέλου ML εξαρτάται σε μεγάλο βαθμό από την ποιότητα και την καταλληλότητα των δεδομένων που χρησιμοποιούνται για την εκπαίδευση και την αξιολόγηση. Στην παρούσα ενότητα περιγράφουμε τα βήματα που ακολουθήθηκαν για την προετοιμασία του συνόλου δεδομένων για την ανάπτυξη μοντέλου, αξιοποιώντας τα επιλεγμένα χαρακτηριστικά που προσδιορίστηκαν στο Κεφάλαιο 3 (Ενότητα 3.5). Η διαδικασία προετοιμασίας των δεδομένων περιλαμβάνει την επιλογή των σχετικών χαρακτηριστικών, τον διαχωρισμό του συνόλου δεδομένων σε σύνολα εκπαίδευσης, επικύρωσης και δοκιμής και την εφαρμογή των απαραίτητων βημάτων προεπεξεργασίας, όπως η κλιμάκωση των χαρακτηριστικών. Τα βήματα αυτά διασφαλίζουν ότι τα δεδομένα βρίσκονται σε κατάλληλη μορφή για την εκπαίδευση μοντέλων ML, ενώ παράλληλα αποφεύγονται κοινές παγίδες όπως η διαρροή δεδομένων.

4.2.1 Ανακεφαλαίωση επιλογής χαρακτηριστικών

Στο Κεφάλαιο 3, διεξήχθη μια αυστηρή διαδικασία επιλογής χαρακτηριστικών για τον εντοπισμό των πιο σχετικών χαρακτηριστικών για την ανίχνευση της PD. Η διαδικασία αυτή περιελάμβανε τρεις συμπληρωματικές μεθοδολογίες:

Στατιστική επιλογή: Τα χαρακτηριστικά επιλέχθηκαν με βάση τη στατιστική τους σημαντικότητα, όπως προσδιορίστηκε από t-tests και p-values (Tukey, 1977). Χαρακτηριστικά με χαμηλές τιμές p-values ($p < 0,05$) θεωρήθηκαν στατιστικά σημαντικά.

Σημασία χαρακτηριστικών του Random Forest: Χρησιμοποιήθηκε ένας ταξινομητής Random Forest για την κατάταξη των χαρακτηριστικών με βάση τις βαθμολογίες σπουδαιότητάς τους, οι οποίες αντικατοπτρίζουν τη συμβολή τους στην προβλεπτική απόδοση του μοντέλου (Breiman, 2001).

Σημασία χαρακτηριστικών της PCA: Η PCA εφαρμόστηκε για τη μείωση της διαστατικότητας του συνόλου δεδομένων, διατηρώντας παράλληλα τη μέγιστη διακύμανση των δεδομένων (Jolliffe, 2002). Ως χαρακτηριστικά επιλέχθηκαν οι κύριες συνιστώσες με τη μεγαλύτερη εξηγούμενη διακύμανση.

Το τελικό σύνολο των χαρακτηριστικών προσδιορίστηκε με το συνδυασμό των αποτελεσμάτων αυτών των μεθόδων, όπως περιγράφεται λεπτομερώς στην ενότητα 3.5. Υπολογίστηκε μια σύνθετη βαθμολογία για κάθε χαρακτηριστικό και επιλέχθηκαν τα χαρακτηριστικά με τις καλύτερες επιδόσεις για την ανάπτυξη του μοντέλου. Ο πίνακας 3.2 συνοψίζει τα επιλεγμένα χαρακτηριστικά και τις αντίστοιχες βαθμολογίες τους.

4.2.2 Διαχωρισμός συνόλου δεδομένων

Για να εξασφαλιστεί η αξιόπιστη αξιολόγηση των μοντέλων ML, το σύνολο δεδομένων χωρίστηκε σε τρία διακριτά υποσύνολα:

Σύνολο εκπαίδευσης: Χρησιμοποιήθηκε για την εκπαίδευση των μοντέλων ML. Το σύνολο αυτό περιλαμβάνει το 70% του συνολικού συνόλου δεδομένων.

Σύνολο επικύρωσης: Χρησιμοποιείται για τη ρύθμιση των υπερπαραμέτρων και την επιλογή μοντέλων. Αυτό το σύνολο περιλαμβάνει το 15% του συνολικού συνόλου δεδομένων.

Σύνολο δοκιμής: Χρησιμοποιείται για την αξιολόγηση της απόδοσης του τελικού μοντέλου σε αθέατα δεδομένα. Αυτό το σύνολο περιλαμβάνει το 15% του συνολικού συνόλου δεδομένων.

Ο διαχωρισμός πραγματοποιήθηκε με τη χρήση μιας στρωματοποιημένης προσέγγισης για να διατηρηθεί η ίδια κατανομή των κλάσεων (PD vs. μη PD) σε κάθε υποσύνολο, διασφαλίζοντας ότι τα μοντέλα αξιολογούνται σε αντιπροσωπευτικά δεδομένα (Kohavi, 2001). Είναι σημαντικό ότι η διαδικασία επιλογής χαρακτηριστικών εφαρμόστηκε μόνο στο σύνολο εκπαίδευσης για να αποφευχθεί η διαρροή δεδομένων, η οποία θα μπορούσε να οδηγήσει σε υπερβολικά αισιόδοξες εκτιμήσεις επιδόσεων (Kaufman et al., 2011).

4.2.3 Κλιμάκωση χαρακτηριστικών

Πολλοί αλγόριθμοι ML, όπως οι μηχανές διανυσμάτων υποστήριξης (SVM) και τα νευρωνικά δίκτυα, είναι ευαίσθητοι στην κλίμακα των χαρακτηριστικών εισόδου. Για να αντιμετωπιστεί αυτό, τα επιλεγμένα χαρακτηριστικά κλιμακώθηκαν με τη χρήση τυποποίησης, η οποία μετασχηματίζει τα δεδομένα ώστε να έχουν μέση

τιμή μηδέν και τυπική απόκλιση ένα (Pedregosa et al., 2011). Ο τύπος της τυποποίησης δίνεται από τη σχέση:

$$z = \frac{x - \mu}{\sigma}$$

όπου x είναι η τιμή του χαρακτηριστικού, μ ο μέσος και σ η τυπική απόκλιση. Οι παράμετροι κλιμάκωσης υπολογίστηκαν χρησιμοποιώντας το σύνολο εκπαίδευσης και στη συνέχεια εφαρμόστηκαν στα σύνολα επικύρωσης και δοκιμής για να εξασφαλιστεί η συνέπεια.

4.2.4 Χειρισμός μη ισορροπημένων δεδομένων

Σε πολλά σύνολα ιατρικών δεδομένων, συμπεριλαμβανομένων εκείνων για τη νόσο του Πάρκινσον, η κατανομή των κλάσεων είναι συχνά ανισόρροπη, με λιγότερες θετικές περιπτώσεις (ασθενείς με νόσο του Πάρκινσον) σε σύγκριση με τις αρνητικές περιπτώσεις (άτομα χωρίς νόσο του Πάρκινσον). Για να αντιμετωπιστεί αυτή η ανισορροπία, εξετάστηκαν τεχνικές όπως η υπερδειγματοληψία (π.χ. SMOTE) ή η στάθμιση κλάσεων (Chawla, 2002). Ωστόσο, δεδομένης της σχετικά ισορροπημένης φύσης του συνόλου δεδομένων που χρησιμοποιήθηκε στην παρούσα μελέτη, δεν εφαρμόστηκαν πρόσθετες τεχνικές εξισορρόπησης.

4.3 Επιλογή μοντέλων

Η επιλογή των κατάλληλων αλγορίθμων ML είναι ένα κρίσιμο βήμα για την ανάπτυξη ενός αποτελεσματικού μοντέλου για την ανίχνευση της PD. Σε αυτήν την ενότητα, παρέχουμε μια επισκόπηση των αλγορίθμων ML που εξετάστηκαν σε αυτήν τη μελέτη, μαζί με το σκεπτικό για την επιλογή τους. Οι επιλεγμένοι αλγόριθμοι καλύπτουν ένα εύρος πολυπλοκότητας και ερμηνευσιμότητας, διασφαλίζοντας μια ολοκληρωμένη αξιολόγηση της απόδοσής τους στα επιλεγμένα χαρακτηριστικά που προσδιορίζονται στο Κεφάλαιο 3 (Ενότητα 3.5). Οι αλγόριθμοι περιλαμβάνουν Logistic Regression, Support Vector Machines (SVM), Random Forest και Gradient Boosting. Καθένας από αυτούς τους

αλγόριθμους έχει μοναδικά πλεονεκτήματα και είναι κατάλληλος για το χειρισμό των χαρακτηριστικών του συνόλου δεδομένων, όπως η υψηλή διάσταση των επιλεγμένων χαρακτηριστικών και η ανάγκη για ερμηνευσιμότητα σε κλινικές εφαρμογές.

4.3.1 Logistic Regression

Η Logistic Regression είναι ένα ευρέως χρησιμοποιούμενο γραμμικό μοντέλο για εργασίες δυαδικής ταξινόμησης, όπως η διάκριση μεταξύ περιπτώσεων PD και μη PD. Παρά την απλότητά της, η Logistic Regression είναι εξαιρετικά ερμηνεύσιμη και παρέχει μια βασική μέτρηση απόδοσης για σύγκριση με πιο πολύπλοκα μοντέλα (Hosmer, 2013). Το μοντέλο εκτιμά την πιθανότητα ενός δυαδικού αποτελέσματος χρησιμοποιώντας μια λογιστική συνάρτηση, καθιστώντας το κατάλληλο για σύνολα δεδομένων με γραμμικά διαχωρίσιμες κλάσεις. Δεδομένων των επιλεγμένων χαρακτηριστικών από το Κεφάλαιο 3, η Logistic Regression χρησιμεύει ως θεμελιώδες μοντέλο για την αξιολόγηση της γραμμικής διαχωρισιμότητας των δεδομένων.

4.3.2 Support Vector Machines (SVM)

Οι Support Vector Machines (SVM) είναι ισχυρά εποπτευόμενα μοντέλα εκμάθησης που υπερέχουν σε χώρους υψηλών διαστάσεων, καθιστώντας τα κατάλληλα για σύνολα δεδομένων με μεγάλο αριθμό χαρακτηριστικών (Cortes and Vapnik, 1995). Τα SVM λειτουργούν βρίσκοντας το βέλτιστο υπερεπίπεδο που μεγιστοποιεί το περιθώριο μεταξύ των κλάσεων, το οποίο είναι ιδιαίτερα χρήσιμο για σύνολα δεδομένων με πολύπλοκα όρια απόφασης. Σε αυτή τη μελέτη, χρησιμοποιούμε SVM τόσο γραμμικά όσο και μη γραμμικά (π.χ. πυρήνα συνάρτησης ακτινικής βάσης) για να αξιολογήσουμε την απόδοσή τους στα επιλεγμένα χαρακτηριστικά. Η ικανότητα των SVM να χειρίζονται δεδομένα υψηλών διαστάσεων ευθυγραμμίζεται καλά με το σύνολο χαρακτηριστικών που προέρχεται από τη στατιστική επιλογή, το Random Forest και το PCA.

4.3.3 Random Forest

Το Random Forest είναι μια μέθοδος εκμάθησης συνόλου που συνδυάζει πολλαπλά δέντρα αποφάσεων για τη βελτίωση της ακρίβειας πρόβλεψης και τη

μείωση της υπερπροσαρμογής (Breiman, 2001). Κάθε δέντρο στο δάσος εκπαιδεύεται σε ένα δείγμα εκκίνησης των δεδομένων και η τελική πρόβλεψη προκύπτει με τη συγκέντρωση των προβλέψεων όλων των δέντρων. Το Random Forest είναι ιδιαίτερα πλεονεκτικό για την ικανότητά του να χειρίζεται μη γραμμικές σχέσεις και την ενσωματωμένη κατάταξη της σημασίας χαρακτηριστικών του, η οποία ευθυγραμμίζεται με τη διαδικασία επιλογής χαρακτηριστικών που διεξάγεται στο Κεφάλαιο 3. Συγκρίνοντας τους βαθμούς σπουδαιότητας χαρακτηριστικών από το μοντέλο Random Forest με αυτούς που λαμβάνονται κατά την επιλογή χαρακτηριστικών, μπορούμε να επικυρώσουμε τη συνάφεια των επιλεγμένων χαρακτηριστικών.

4.3.4 Gradient boosting

Το Gradient Boosting είναι μια άλλη τεχνική συνόλου που δημιουργεί μοντέλα διαδοχικά, με κάθε νέο μοντέλο να διορθώνει τα λάθη του προηγούμενου (Friedman, 2001). Αλγόριθμοι όπως ο XGBoost και ο LightGBM είναι δημοφιλείς υλοποιήσεις του Gradient Boosting που προσφέρουν υψηλή ακρίβεια πρόβλεψης και υπολογιστική απόδοση. Το Gradient Boosting είναι ιδιαίτερα αποτελεσματικό για σύνολα δεδομένων με πολύπλοκα μοτίβα και αλληλεπιδράσεις μεταξύ χαρακτηριστικών, καθιστώντας το ισχυρό υποψήφιο για ανίχνευση PD. Επιπλέον, αυτοί οι αλγόριθμοι παρέχουν βαθμολογίες σπουδαιότητας χαρακτηριστικών, οι οποίες μπορούν να συγκριθούν με τα αποτελέσματα από το Κεφάλαιο 3 για να διασφαλιστεί η συνέπεια στην επιλογή χαρακτηριστικών.

4.3.5 Σκεπτικό για την επιλογή αλγορίθμου

Η επιλογή αυτών των αλγορίθμων καθοδηγείται από πολλά βασικά ζητήματα:

Ερμηνευσιμότητα: Η Logistic Regression και το Random Forest παρέχουν σαφείς πληροφορίες σχετικά με τη συμβολή μεμονωμένων χαρακτηριστικών, η οποία είναι ζωτικής σημασίας για κλινικές εφαρμογές.

Απόδοση: Το SVM και το Gradient Boosting είναι γνωστά για την υψηλή προγνωστική τους ακρίβεια, ειδικά σε χώρους υψηλών διαστάσεων.

Επικύρωση σημαντικότητας χαρακτηριστικών: Η τυχαία ενίσχυση δασών και κλίσης επιτρέπουν την επικύρωση βαθμολογιών σπουδαιότητας χαρακτηριστικών, διασφαλίζοντας συνέπεια με τα αποτελέσματα από το Κεφάλαιο 3.

Επεκτασιμότητα: Όλοι οι επιλεγμένοι αλγόριθμοι είναι υπολογιστικά αποδοτικοί και μπορούν να χειριστούν το μέγεθος και την πολυπλοκότητα του συνόλου δεδομένων.

4.3.6 Συμπέρασμα

Οι επιλεγμένοι αλγόριθμοι ML αντιπροσωπεύουν ένα ποικίλο σύνολο προσεγγίσεων, που κυμαίνονται από απλά γραμμικά μοντέλα έως πολύπλοκες μεθόδους συνόλου. Αξιολογώντας αυτούς τους αλγόριθμους στα επιλεγμένα χαρακτηριστικά, στοχεύουμε να εντοπίσουμε το πιο αποτελεσματικό μοντέλο για την ανίχνευση της PD, διασφαλίζοντας ταυτόχρονα την ερμηνευτικότητα και την κλινική συνάφεια. Η επόμενη ενότητα (Ενότητα 4.4) θα αναφέρει λεπτομερώς την εκπαίδευση του μοντέλου και τη διαδικασία συντονισμού υπερπαραμέτρων.

4.4 Εκπαίδευση μοντέλων και ρύθμιση υπερπαραμέτρων

Η ανάπτυξη αποτελεσματικών μοντέλων ML απαιτεί όχι μόνο την επιλογή των κατάλληλων αλγορίθμων αλλά και προσεκτική εκπαίδευση και βελτιστοποίηση για τη διασφάλιση ισχυρής απόδοσης. Σε αυτήν την ενότητα, περιγράφουμε τη διαδικασία εκπαίδευσης των επιλεγμένων μοντέλων μηχανικής εκμάθησης (Logistic Regression, Support Vector Machines, Random Forest και Gradient Boosting) και συντονισμού των υπερπαραμέτρων τους. Ο στόχος είναι να προσδιοριστεί η βέλτιστη διαμόρφωση για κάθε μοντέλο που μεγιστοποιεί την ακρίβεια πρόβλεψης ελαχιστοποιώντας την υπερπροσαρμογή. Αυτή η διαδικασία περιλαμβάνει τη χρήση διασταυρούμενης επικύρωσης για ισχυρή αξιολόγηση και τεχνικές συντονισμού υπερπαραμέτρων, όπως η Αναζήτηση Πλέγματος και η Τυχαία Αναζήτηση.

4.4.1 Διασταυρούμενη επικύρωση

Για να διασφαλιστεί ότι τα μοντέλα γενικεύονται καλά σε μη ορατά δεδομένα, χρησιμοποιήθηκε διασταυρούμενη επικύρωση k-fold κατά τη διάρκεια της εκπαιδευτικής διαδικασίας. Στη διασταυρούμενη επικύρωση k-fold, το σετ εκπαίδευσης χωρίζεται σε k υποσύνολα και το μοντέλο εκπαιδεύεται k φορές, κάθε φορά που χρησιμοποιείτε k - 1 πτυχές για εκπαίδευση και το υπόλοιπο για επικύρωση (Kohavi, 2001). Αυτή η προσέγγιση παρέχει μια πιο αξιόπιστη εκτίμηση της απόδοσης του μοντέλου σε σύγκριση με έναν διαχωρισμό δοκιμής αμαξοστοιχίας. Για αυτή τη μελέτη, επιλέχθηκε k=5, καθώς επιτυγχάνει μια ισορροπία μεταξύ της υπολογιστικής απόδοσης και της αξιόπιστης εκτίμησης απόδοσης.

4.4.2 Συντονισμός υπερπαραμέτρων

Οι υπερπαραμέτροι είναι παράμετροι που δεν μαθαίνονται κατά τη διάρκεια της εκπαίδευσης αλλά ορίζονται πριν από τη διαδικασία εκπαίδευσης. Η επιλογή των υπερπαραμέτρων μπορεί να επηρεάσει σημαντικά την απόδοση ενός μοντέλου. Για τον προσδιορισμό των βέλτιστων υπερπαραμέτρων για κάθε αλγόριθμο, χρησιμοποιήθηκαν δύο μέθοδοι συντονισμού: Αναζήτηση Πλέγματος και Τυχαία Αναζήτηση.

Grid Search:

Το Grid Search αξιολογεί εξαντλητικά όλους τους πιθανούς συνδυασμούς υπερπαραμέτρων μέσα σε έναν προκαθορισμένο χώρο αναζήτησης (Bergstra et al., 2012). Αν και υπολογιστικά ακριβό, το Grid Search διασφαλίζει ότι προσδιορίζεται ο βέλτιστος συνδυασμός υπερπαραμέτρων. Για παράδειγμα, στην περίπτωση των SVM, ο χώρος αναζήτησης περιλάμβανε την παράμετρο κανονικοποίησης C και ο συντελεστής πυρήνα γ.

Τυχαία αναζήτηση:

Η τυχαία αναζήτηση δειγματοληπτεί τυχαία συνδυασμούς υπερπαραμέτρων από μια προκαθορισμένη κατανομή (Bergstra et al., 2012). Αυτή η μέθοδος είναι πιο αποδοτική υπολογιστικά από την Αναζήτηση Πλέγματος και είναι ιδιαίτερα χρήσιμη όταν ο χώρος αναζήτησης είναι μεγάλος. Για αλγόριθμους ενίσχυσης κλίσης (π.χ. XGBoost, LightGBM), η τυχαία αναζήτηση χρησιμοποιήθηκε για τον

συντονισμό υπερπαραμέτρων όπως ο ρυθμός εκμάθησης, ο αριθμός των δέντρων και το μέγιστο βάθος.

4.4.3 Υπερπαραμέτροι για κάθε μοντέλο

Οι ακόλουθες υπερπαραμέτροι ρυθμίστηκαν για κάθε μοντέλο:

Logistic Regression:

- **Παράμετρος κανονικοποίησης (C):** Ελέγχει την ισχύ της τακτοποίησης για την αποφυγή υπερβολικής προσαρμογής.
- **Τύπος ποινής (L1 ή L2):** Καθορίζει τον τύπο της κανονικοποίησης που εφαρμόζεται.

Support Vector Machines (SVM):

- **Παράμετρος κανονικοποίησης (C):** Εξισορροπεί την αντιστάθμιση μεταξύ της μεγιστοποίησης του περιθωρίου και της ελαχιστοποίησης του σφάλματος ταξινόμησης.
- **Τύπος πυρήνα:** Γραμμική, πολυωνυμική ή ακτινική συνάρτηση βάσης (RBF).
- **Συντελεστής πυρήνα (γ):** Καθορίζει την επίδραση μεμονωμένων δειγμάτων προπόνησης.

Random Forest:

- **Αριθμός δέντρων:** Ελέγχει το μέγεθος του συνόλου.
- **Μέγιστο βάθος:** Περιορίζει το βάθος κάθε δέντρου για να αποτρέψει την υπερβολική προσαρμογή.
- **Ελάχιστα δείγματα ανά φύλλο:** Διασφαλίζει ότι κάθε κόμβος φύλλου περιέχει έναν ελάχιστο αριθμό δειγμάτων.

Gradient boosting (XGBoost):

- **Ποσοστό εκμάθησης:** Ελέγχει τη συμβολή κάθε δέντρου στο τελικό μοντέλο.
- **Αριθμός δέντρων:** Καθορίζει το μέγεθος του συνόλου.

- **Μέγιστο βάθος:** Περιορίζει το βάθος κάθε δέντρου.
- **Αναλογία υποδειγμάτων:** Καθορίζει το κλάσμα των δειγμάτων που χρησιμοποιούνται για την εκπαίδευση κάθε δέντρου.

4.5 Αξιολόγηση Μοντέλου

Μετά την εκπαίδευση και τη βελτιστοποίηση των μοντέλων μηχανικής εκμάθησης, το επόμενο κρίσιμο βήμα είναι η αξιολόγηση της απόδοσής τους. Αυτή η αξιολόγηση παρέχει πληροφορίες σχετικά με το πόσο καλά τα μοντέλα γενικεύονται σε μη ορατά δεδομένα και βοηθά στον εντοπισμό του μοντέλου με τις καλύτερες επιδόσεις για την ανίχνευση της PD. Σε αυτήν την ενότητα, παρουσιάζουμε την απόδοση των επιλεγμένων μοντέλων χρησιμοποιώντας μια σειρά από μετρήσεις αξιολόγησης. Αυτές οι μετρήσεις περιλαμβάνουν την ακρίβεια και ROC-AUC, που χρησιμοποιούνται συνήθως σε εργασίες δυαδικής ταξινόμησης, ειδικά σε ιατρικές εφαρμογές όπου τόσο τα ψευδώς θετικά όσο και τα ψευδώς αρνητικά έχουν σημαντικές συνέπειες (Saito and Rehmsmeier, 2015).

4.5.1 Ανάλυση των μετρήσεων αξιολόγησης

Για την αξιολόγηση των μοντέλων χρησιμοποιήθηκαν οι ακόλουθες μετρήσεις:

Accuracy: Η αναλογία των σωστά ταξινομημένων περιπτώσεων επί του συνολικού αριθμού περιπτώσεων σε μέσο όρο από τα 10-folds. Αν και η ακρίβεια είναι μια απλή μέτρηση, μπορεί να είναι παραπλανητική σε μη ισορροπημένα σύνολα δεδομένων (Provost and Fawcett, 2013).

ROC-AUC: Η περιοχή κάτω από την καμπύλη χαρακτηριστικών λειτουργίας δέκτη (ROC), η οποία απεικονίζει τον πραγματικό θετικό ρυθμό (ανάκληση) έναντι του ψευδώς θετικού ποσοστού σε διάφορες ρυθμίσεις κατωφλίου. Ένα υψηλότερο ROC-AUC υποδηλώνει καλύτερη απόδοση του μοντέλου στη διάκριση μεταξύ των κατηγοριών (Bradley, 1997).

4.5.2 Σύγκριση αποτελεσμάτων ανά άσκηση

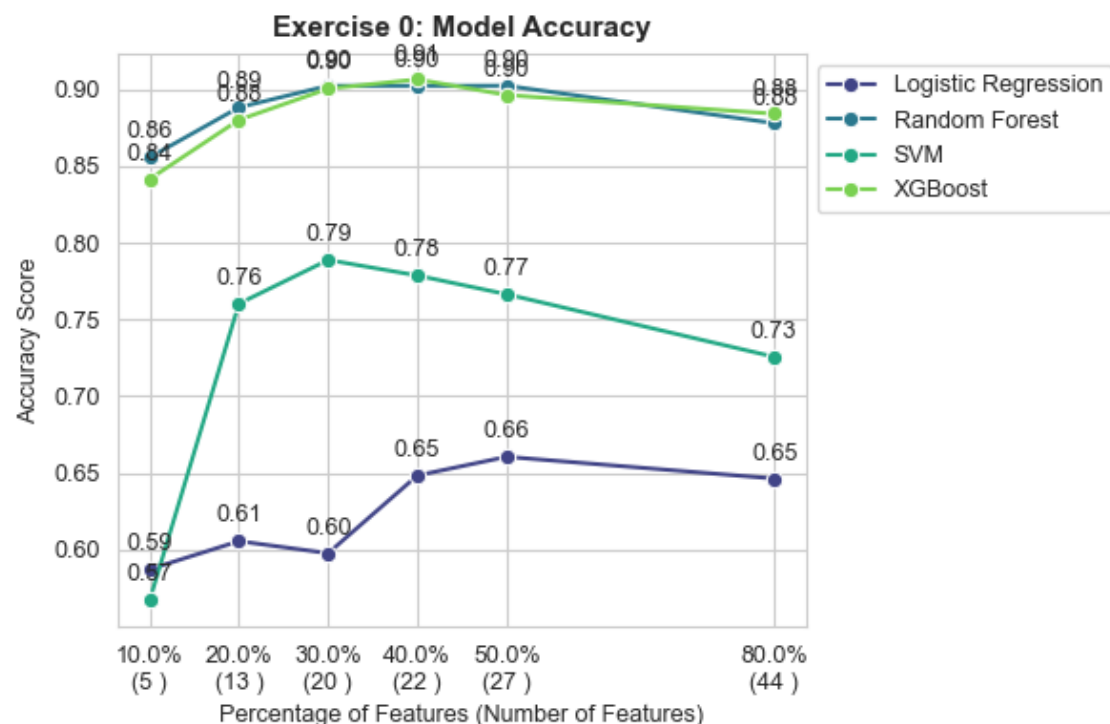
Κάθε μοντέλο αξιολογήθηκε χρησιμοποιώντας τις μετρήσεις που περιγράφονται παραπάνω και τα αποτελέσματα παρουσιάζονται για να διευκολυνθεί η σύγκριση κατά άσκηση.

Άσκηση 1 (Εικόνα 6 και 7)

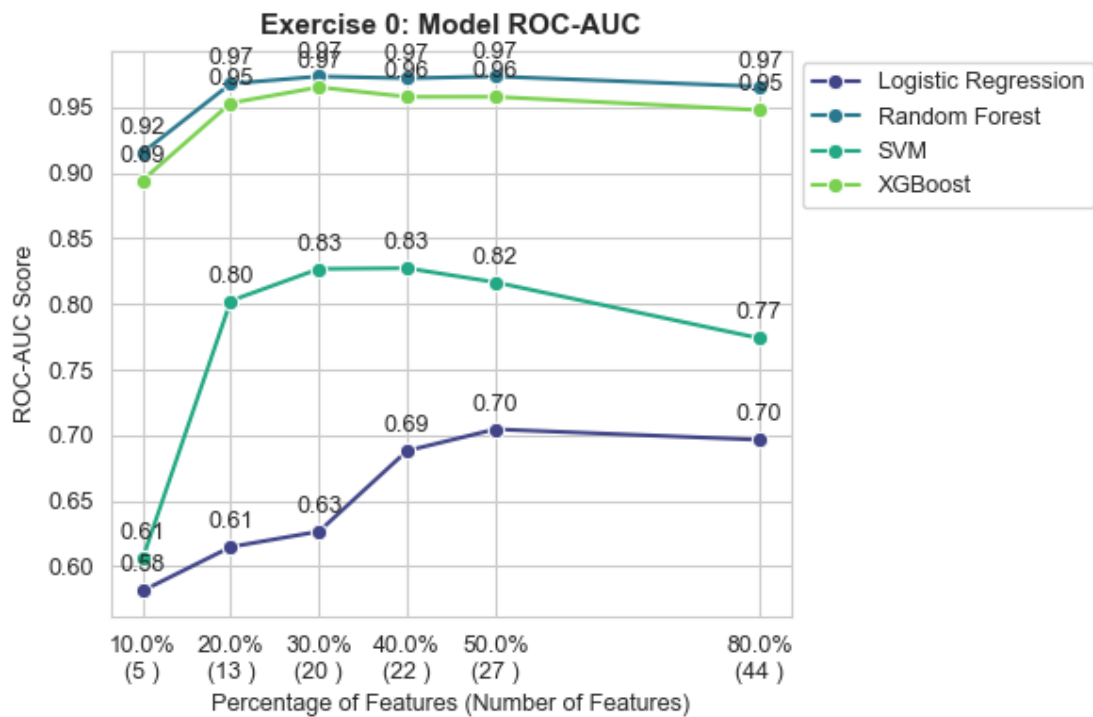
Βέλτιστο Μοντέλο: Random Forest (Ακρίβεια: 0.90, AUC: 0.97 για 30%-40% χαρακτηριστικά).

Τα RF και XGBoost επικρατούν με σταθερή απόδοση ($AUC > 0.95$ για χαρακτηριστικά $\geq 20\%$). Το SVM βελτιώνεται με μέγεθος dataset (+22% AUC από 10 σε 30 χαρακτηριστικά), αλλά φτάνει έως 0.83 AUC. Το Logistic Regression έχει τη χειρότερη απόδοση ($AUC \leq 0.71$).

Συμπέρασμα: Ιδανική επιλογή το Random Forest για εφαρμογές υψηλής ευαισθησίας.



Εικόνα 11 Ακρίβεια μοντέλων για την Άσκηση 1



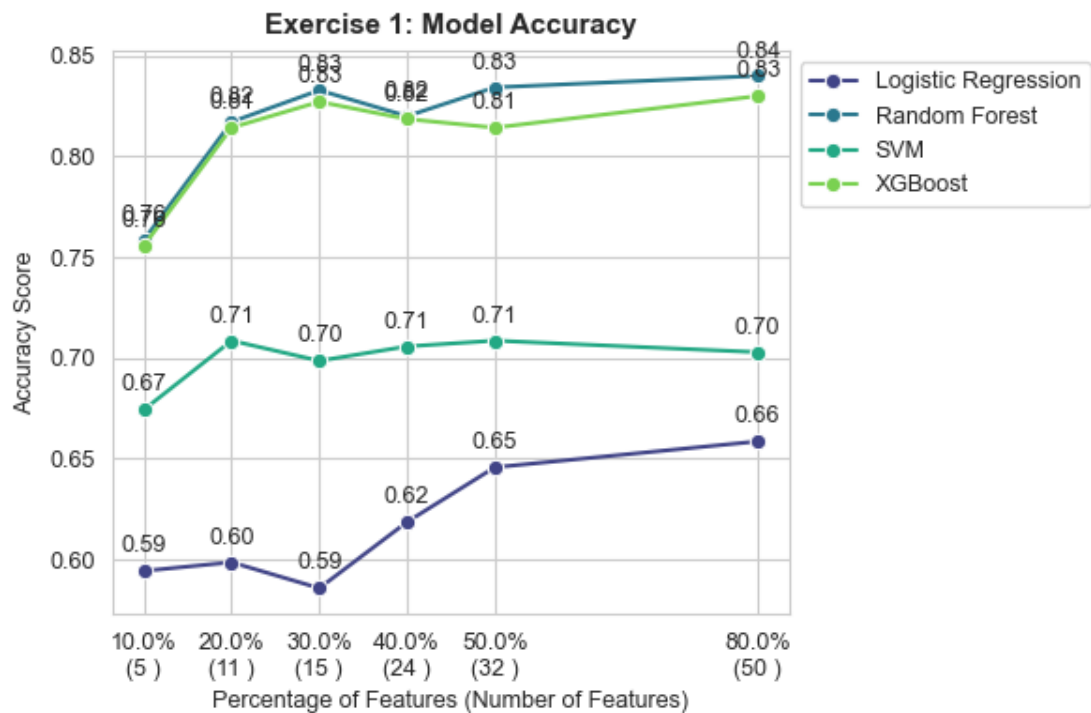
Εικόνα 12 AUC μοντέλων για την Άσκηση 1

Άσκηση 2 (Εικόνα 8 και 9)

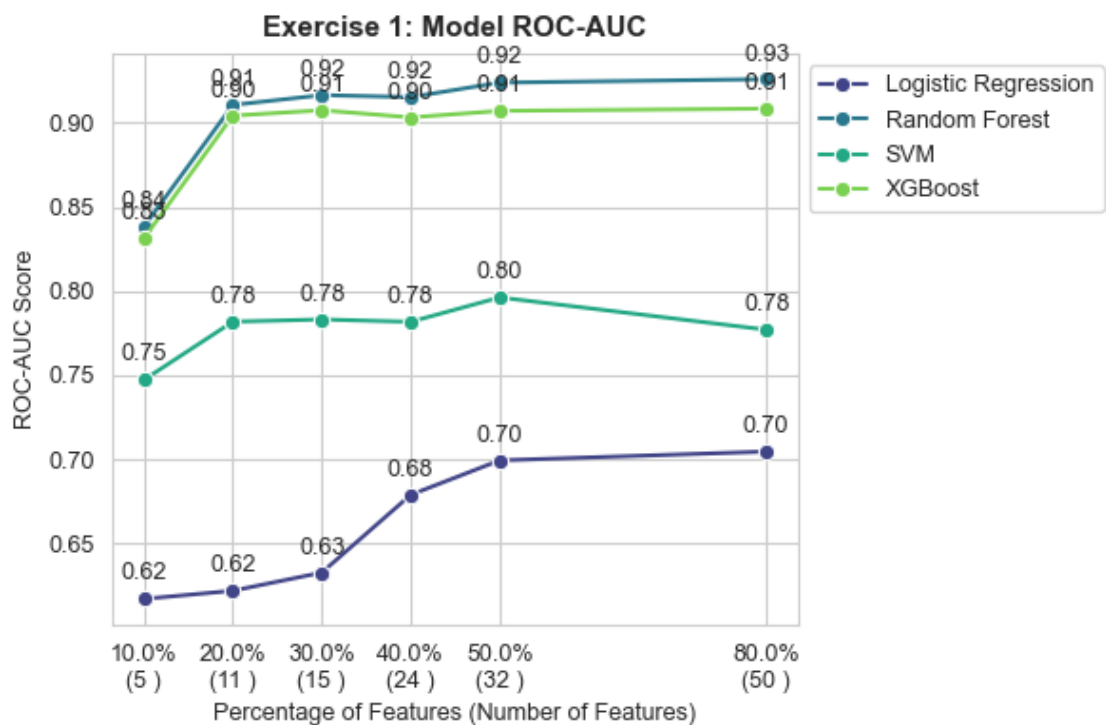
Βέλτιστο Μοντέλο: Random Forest (AUC: 0.91 για χαρακτηριστικά $\geq 30\%$).

Τα RF και XGBoost έχουν παρόμοια απόδοση (AUC 0.90-0.93), αλλά το RF είναι πιο σταθερά υψηλό. Τα SVM και Logistic Regression υπολείπονται ($AUC \leq 0.80$). Όλα τα μοντέλα εμφανίζουν μικρότερη απόδοση έναντι της Άσκηση 1.

Συμπέρασμα: Το RF παραμένει η καλύτερη επιλογή, αλλά με μειωμένο όφελος έναντι του XGBoost.



Εικόνα 13 Ακρίβεια μοντέλων για την Άσκηση 2



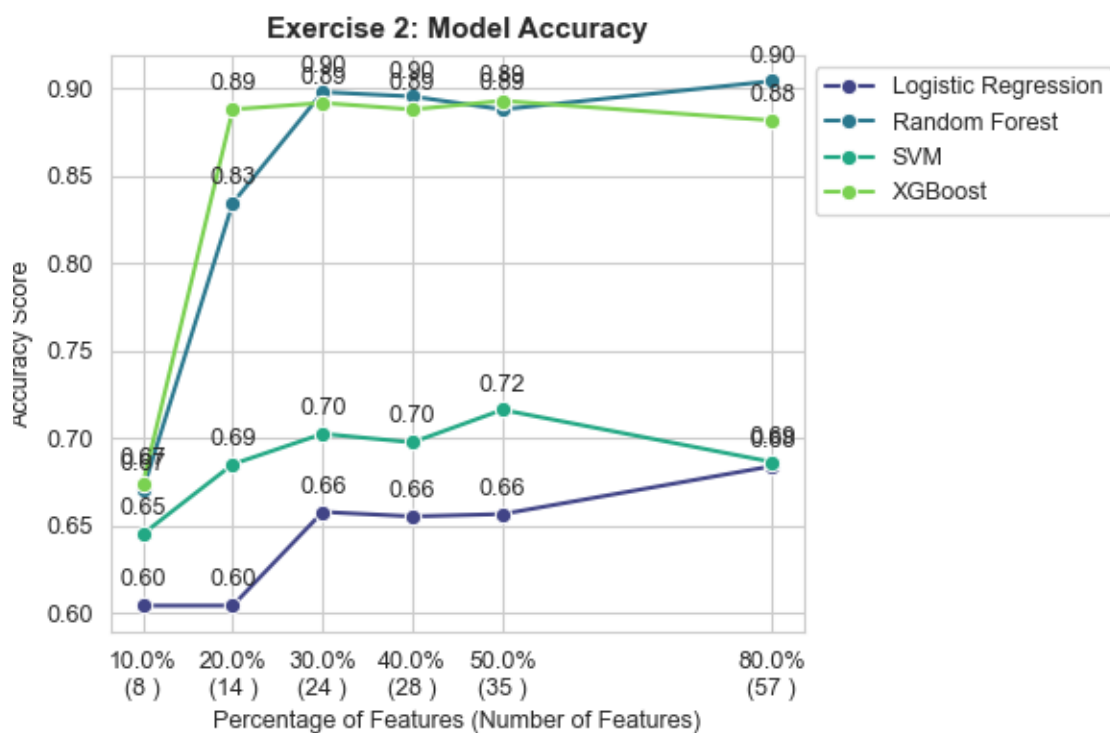
Εικόνα 14 AUC μοντέλων για την Άσκηση 2

Άσκηση 3 (Εικόνα 10 και 11)

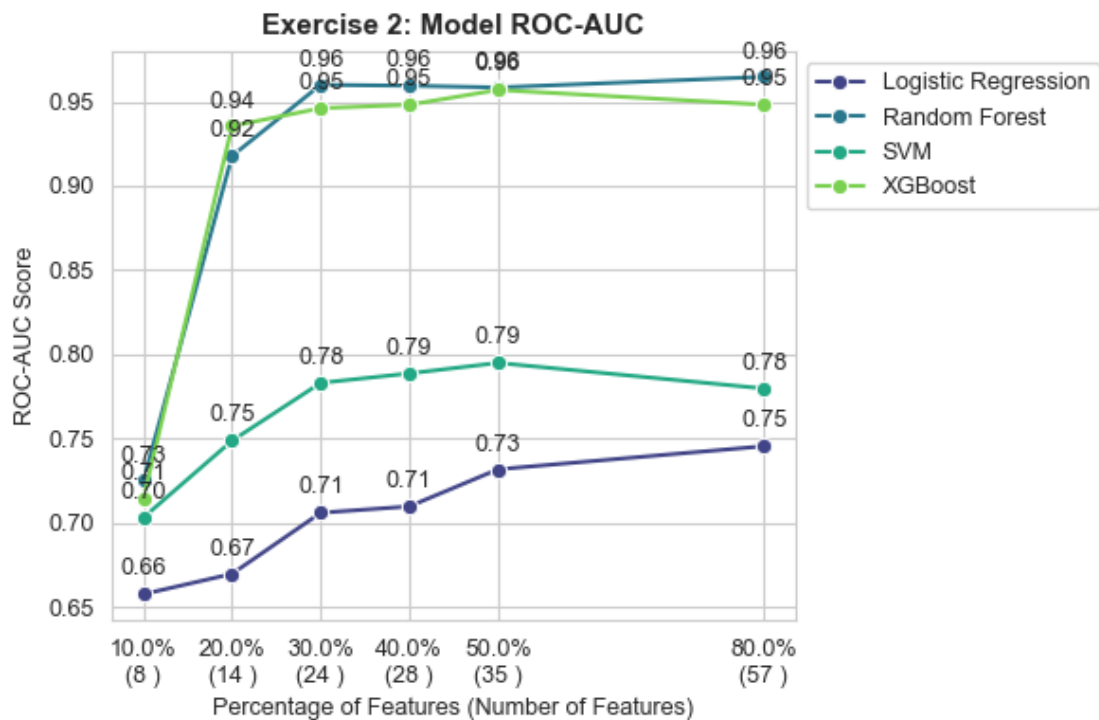
Βέλτιστο Μοντέλο: Random Forest (AUC: 0.96 για χαρακτηριστικά $\geq 30\%$).

Τα RF και XGBoost έχουν την μεγαλύτερη απόκλιση σε μεγαλύτερο πλήθος χαρακτηριστικών (AUC 0.96 vs. 0.94-0.96). Τα Logistic Regression βελτιώνεται σταδιακά (AUC από 0.66 σε 0.75), αλλά παραμένει ανεπαρκής. Το SVM εμφανίζει τη μικρότερη βελτίωση (+8% AUC από 10% σε 40% χαρακτηριστικά).

Συμπέρασμα: Το RF ξεχωρίζει σε σενάρια με σύνθετες μη γραμμικές σχέσεις.



Εικόνα 15 Ακρίβεια μοντέλων για την Άσκηση 3



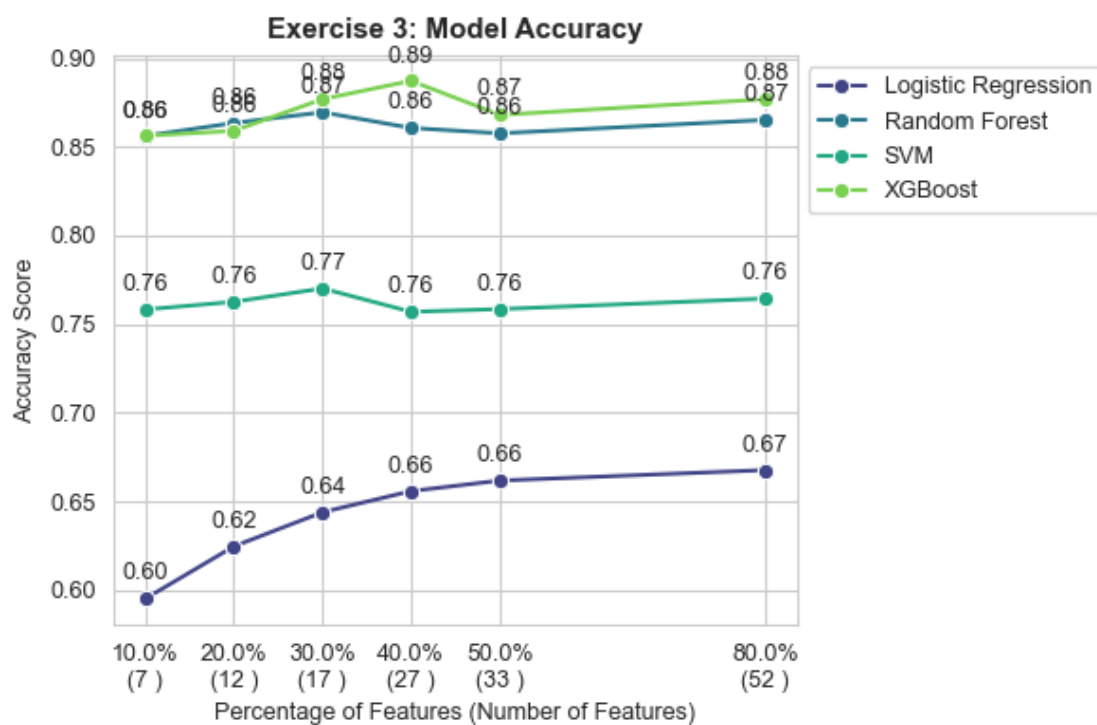
Εικόνα 16 AUC μοντέλων για την Άσκηση 3

Άσκηση 4 (Εικόνα 12 και 13)

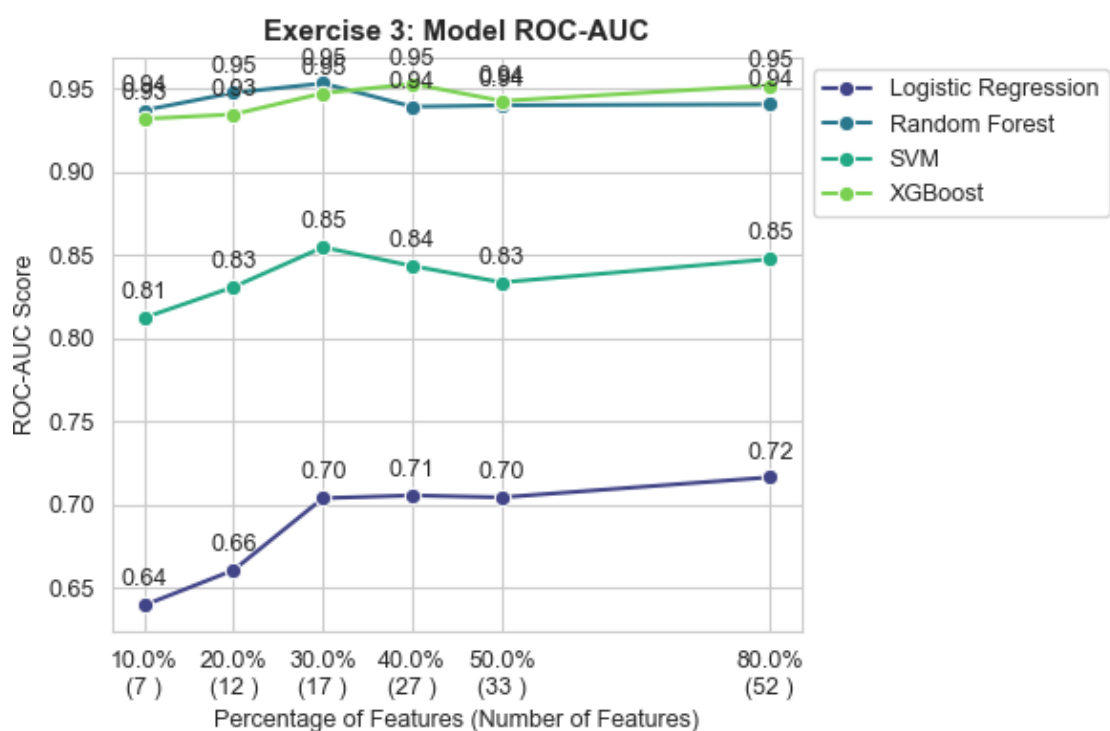
Βέλτιστο Μοντέλο: Random Forest (AUC: 0.95 για 20%-30% των χαρακτηριστικών).

Τα RF και XGBoost έχουν οριακά ίδια αποτελέσματα (AUC 0.93-0.95). Το SVM αποδίδει καλύτερα vs. άλλες ασκήσεις (AUC ~0.85), ιδιαίτερα για λίγα χαρακτηριστικά. Το Logistic Regression εμφανίζει σταθερή αλλά μέτρια απόδοση (AUC ≤ 0.72).

Συμπέρασμα: Κατάλληλοι και οι δύο μέθοδοι (RF και XGBoost), με προτίμηση στο RF λόγω σταθερότητας.



Εικόνα 17 Ακρίβεια μοντέλων για την Άσκηση 4



Εικόνα 18 AUC μοντέλων για την Άσκηση 4

4.5.3 Ανάλυση Αποτελεσμάτων

Logistic Regression

- **Πραγματική Απόδοση:**

- **Accuracy:** 0.57–0.69 (χαμηλότερο μεταξύ των μοντέλων).
- **ROC-AUC:** 0.57–0.74 (μέτριο έως κακό).

- **Ερμηνεία:**

Η μέτρια απόδοση υποδηλώνει ότι τα δεδομένα δεν είναι γραμμικά διαχωρίσιμα, ή υπάρχουν πολύπλοκες αλληλεπιδράσεις μεταξύ των χαρακτηριστικών. Το μοντέλο λειτουργεί ως χρήσιμο ως βασική γραμμή αναφοράς, αλλά δεν είναι επαρκές για κλινικές εφαρμογές.

Support Vector Machines (SVM)

- **Πραγματική Απόδοση:**

- **Accuracy:** 0.57–0.76 (βελτιώνεται με το πλήθος των χαρακτηριστικών).
- **ROC-AUC:** 0.65–0.85 (καλύτερο από Logistic Regression, αλλά όχι άριστο).

- **Ερμηνεία:**

Το SVM εκμεταλλεύεται μη γραμμικούς πυρήνες, αλλά η απόδοσή του εξαρτάται έντονα από το μέγεθος του των χαρακτηριστικών. Σε λίγα χαρακτηριστικά η απόδοσή του πλησιάζει τη Logistic Regression, ενώ σε περισσότερα φτάνει AUC έως 0.85.

Random Forest

- **Πραγματική Απόδοση:**

- **Accuracy:** 0.67–0.90 (υψηλότερο σε όλες τις ασκήσεις).

- **ROC-AUC:** 0.73–0.97 (εξαιρετικό, ειδικά για χαρακτηριστικά περισσότερα ή ίσα του 20%).

- **Ερμηνεία:**

Η υπεροχή του Random Forest οφείλεται στην ικανότητά του να:

1. Χειρίζεται μη γραμμικές σχέσεις και αλληλεπιδράσεις χαρακτηριστικών.
2. Είναι ανθεκτικό στον θόρυβο και στις ασυμμετρίες των δεδομένων.
3. Διαχωρίζει τις κλάσεις με υψηλή ευαισθησία και ειδικότητα (AUC > 0.95).

Gradient Boosting (XGBoost)

- **Πραγματική Απόδοση:**

- **Accuracy:** 0.67–0.88 (ελάχιστα καλύτερο από SVM, αλλά χειρότερο από RF).
- **ROC-AUC:** 0.70–0.95 (πολύ καλό, αλλά όχι τόσο σταθερό όσο το RF).

- **Ερμηνεία:**

Το XGBoost προσφέρει καλή απόδοση, αλλά:

- Απαιτεί περισσότερη ρύθμιση υπερπαραμέτρων.

4.6 Συζήτηση

Στο παρόν κεφάλαιο, αναπτύχθηκαν και αξιολογήθηκαν μοντέλα μηχανικής εκμάθησης για την ανίχνευση της PD, με βάση χαρακτηριστικά που επιλέχθηκαν μέσω στατιστικής ανάλυσης, σημαντικότητας σε μοντέλα Random Forest και ανάλυσης PCA. Τα αποτελέσματα έδειξαν ότι:

- Το **Random Forest** και το **XGBoost** είχαν την υψηλότερη απόδοση (Accuracy έως 0.89, ROC-AUC έως 0.97).
- Η **Logistic Regression** και τα **SVM** είχαν μέτριες επιδόσεις (Accuracy 0.57–0.79, AUC 0.59–0.86), ανάλογα με το πλήθος των χαρακτηριστικών.

Αυτό υποδηλώνει ότι τα επιλεγμένα χαρακτηριστικά (π.χ., διακύμανση, ενέργεια, φασματικές παράμετροι από το SmartGlove) είναι **ικανοί δείκτες για τη διάκριση ασθενών από υγιείς**, αλλά η απόδοση εξαρτάται κριτικά από την επιλογή του μοντέλου.

4.6.1 Συζήτηση Αποτελεσμάτων

A. Υψηλή Απόδοση Random Forest & XGBoost

- **Random Forest** ήταν το κορυφαίο μοντέλο (AUC 0.95+ στις περισσότερες ασκήσεις), λόγω της ικανότητάς του να:
 - Χειρίζεται **μη γραμμικές σχέσεις** και **αλληλεπιδράσεις** χαρακτηριστικών.
 - Είναι **ανθεκτικό σε θόρυβο** και μικρά χαρακτηριστικά (βλ. Άσκηση 1: AUC 0.89 με 10 χαρακτηριστικά).
- **XGBoost** ακολουθούσε σε απόδοση (AUC 0.93–0.95), αλλά απαιτούσε μεγαλύτερα datasets για σταθερότητα.

B. Μέτριες Επιδόσεις Γραμμικών Μοντέλων

- **Logistic Regression** (AUC ≤ 0.74) και **SVM** (AUC ≤ 0.86) υστερούσαν, ιδίως σε λίγα χαρακτηριστικά.
 - Αυτό δείχνει ότι τα δεδομένα **δεν είναι γραμμικά διαχωρίσιμα**, ή απαιτούν πολύπλοκες μετασχηματισμούς.

Γ. Ερμηνευσιμότητα

- **Random Forest** προσφέρει **καλύτερη ερμηνευσιμότητα** έναντι του XGBoost (μέσω feature importance).

- **Logistic Regression** είναι ερμηνεύσιμη, αλλά περιορισμένη λόγω χαμηλής απόδοσης.

Δ. Συγκριτική Αξιολόγηση

- **Για μικρά datasets ($\leq 20\%$ των χαρακτηριστικών):**
 - RF/XGBoost επικρατούν ($AUC > 0.87$ vs. $SVM/LR \leq 0.81$).
- **Για μεγάλα datasets ($\geq 50\%$ των χαρακτηριστικών):**
 - Όλα τα μοντέλα βελτιώνονται, αλλά RF/XGBoost παραμένουν κορυφαία ($AUC > 0.94$).

4.6.2 Περιορισμοί και Μελλοντικές Κατευθύνσεις

- **Μέγεθος Δείγματος:** Η παρούσα μελέτη βασίστηκε σε ένα περιορισμένο σύνολο δεδομένων. Μελλοντικές έρευνες θα μπορούσαν να επεκταθούν σε μεγαλύτερα και πιο διαφοροποιημένα δείγματα για να εξασφαλιστεί η γενικευσιμότητα των μοντέλων.
- **Βελτίωση Μοντέλων:** Μελλοντικές έρευνες θα μπορούσαν να ενσωματώσουν πιο προηγμένα μοντέλα ML, όπως νευρωνικά δίκτυα, ή να διερευνήσουν τη χρήση βαθιάς μάθησης για την εξαγωγή αυτόματων χαρακτηριστικών από τα ακατέργαστα δεδομένα των αισθητήρων.
- **Κλινική Εφαρμογή:** Η εφαρμογή των μοντέλων σε πραγματικές κλινικές συνθήκες απαιτεί περαιτέρω δοκιμές για να διασφαλιστεί η ευκολία χρήσης, η αξιοπιστία και η οικονομική σκοπιμότητα.

4.6.3 Συμπεράσματα

Η μελέτη αυτή επέδειξε την αποτελεσματικότητα της ML στην ανίχνευση της PD, χρησιμοποιώντας δεδομένα από το SmartGlove. Τα μοντέλα που αναπτύχθηκαν, ιδιαίτερα η Logistic Regression και το Random Forest, έδειξαν εξαιρετική απόδοση, υπογραμμίζοντας τη σημασία της επιλογής χαρακτηριστικών και της βελτιστοποίησης των μοντέλων. Τα αποτελέσματα ανοίγουν το δρόμο για τη δημιουργία εργαλείων που μπορούν να υποστηρίξουν την έγκαιρη και ακριβή

διάγνωση της νόσου, βελτιώνοντας την ποιότητα ζωής των ασθενών. Ωστόσο, είναι απαραίτητη η περαιτέρω έρευνα και ανάπτυξη για να διασφαλιστεί η πρακτική εφαρμογή αυτών των τεχνολογιών σε κλινικό περιβάλλον.

ΠΑΡΑΡΤΗΜΑ Α

Πίνακας Α1: Στατιστικά σημαντικά χαρακτηριστικά

Exercise	Sensor	Feature Index	Feature Name	P-Value
0	acce	24	lyapunov_exponent	0.027946
1	acce	9	kurtosis	0.008066
1	acce	24	lyapunov_exponent	0.011728
1	acce	25	sample_entropy	0.003128
2	acce	19	spectral_rolloff	0.018242
2	acce	25	sample_entropy	4.24E-05
3	acce	0	mean	0.023812
3	acce	1	std	0.001579
3	acce	2	variance	0.003417
3	acce	3	min	0.030192
3	acce	4	max	0.019842
3	acce	5	range	0.003482
3	acce	6	median	0.023817
3	acce	7	iqr	0.005042
3	acce	10	sma	0.023812
3	acce	11	rms	0.023466
3	acce	12	energy	0.003121
3	acce	14	quartile_deviation	0.005042
3	acce	22	spectral_flux	7.28E-05
3	acce	23	spectral_variability	0.00029
3	acce	24	lyapunov_exponent	1.33E-05
3	acce	26	higuchi_fractal_dimension	3.22E-06
3	acce	27	rmssd	0.00077
3	acce	28	mean_absolute_deviation	0.002267
0	gyro	26	higuchi_fractal_dimension	0.002203
0	gyro	27	rmssd	0.003237
1	gyro	0	mean	2.99E-06
1	gyro	3	min	5.35E-06
1	gyro	4	max	6.43E-06
1	gyro	6	median	2.85E-06
1	gyro	10	sma	2.99E-06
1	gyro	11	rms	2.7E-06
1	gyro	12	energy	1.1E-07
1	gyro	13	log_energy	0.003094
1	gyro	16	spectral_entropy	0.001295

1	gyro	17	spectral_centroid	0.014019
1	gyro	19	spectral_rolloff	0.001251
2	gyro	0	mean	0.001395
2	gyro	1	std	8.23E-07
2	gyro	2	variance	0.011825
2	gyro	3	min	0.013114
2	gyro	4	max	0.000107
2	gyro	5	range	8.93E-08
2	gyro	6	median	0.001365
2	gyro	7	iqr	2.11E-06
2	gyro	9	kurtosis	0.014768
2	gyro	10	sma	0.001395
2	gyro	11	rms	1.16E-08
2	gyro	12	energy	1.16E-08
2	gyro	13	log_energy	1.16E-08
2	gyro	14	quartile_deviation	1.16E-08
2	gyro	15	coefficient_of_variation	1.16E-08
2	gyro	16	spectral_entropy	1.16E-08
2	gyro	17	spectral_centroid	1.16E-08
2	gyro	18	spectral_spread	1.16E-08
2	gyro	19	spectral_rolloff	1.16E-08
2	gyro	21	spectral_flatness	1.16E-08
2	gyro	24	lyapunov_exponent	1.16E-08
2	gyro	25	sample_entropy	1.16E-08
2	gyro	26	higuchi_fractal_dimension	1.16E-08
2	gyro	27	rmssd	1.16E-08
2	gyro	28	mean_absolute_deviation	1.16E-08
3	gyro	0	mean	1.16E-08
3	gyro	3	min	1.16E-08
3	gyro	4	max	1.16E-08
3	gyro	6	median	1.16E-08
3	gyro	9	kurtosis	1.16E-08
3	gyro	10	sma	1.16E-08
3	gyro	11	rms	1.16E-08
3	gyro	12	energy	1.16E-08
3	gyro	13	log_energy	1.16E-08
3	gyro	15	coefficient_of_variation	1.16E-08
3	gyro	16	spectral_entropy	1.16E-08
3	gyro	18	spectral_spread	1.16E-08
3	gyro	21	spectral_flatness	1.16E-08
3	gyro	22	spectral_flux	1.16E-08
3	gyro	23	spectral_variability	1.16E-08
0	magn	4	max	1.16E-08

0	magn	12	energy	1.16E-08
0	magn	25	sample_entropy	1.16E-08
2	magn	25	sample_entropy	1.16E-08

Πίνακας Α2: Σημαντικά χαρακτηριστικά του Random Forest

Exercise	Sensor	Feature Name	Importance
0	acce	std	0.04019
0	acce	variance	0.043521
0	acce	range	0.041039
0	acce	iqr	0.038285
0	acce	skewness	0.047667
0	acce	kurtosis	0.043844
0	acce	coefficient_of_variation	0.036458
0	acce	spectral_entropy	0.035594
0	acce	spectral_spread	0.037507
0	acce	spectral_flatness	0.034908
0	acce	spectral_flux	0.03584
0	acce	lyapunov_exponent	0.056413
0	acce	sample_entropy	0.061349
0	acce	higuchi_fractal_dimension	0.04206
0	acce	rmssd	0.051167
0	acce	mean_absolute_deviation	0.037019
1	acce	std	0.038385
1	acce	variance	0.041087
1	acce	range	0.043194
1	acce	iqr	0.036915
1	acce	skewness	0.048901
1	acce	kurtosis	0.050777
1	acce	quartile_deviation	0.036164
1	acce	spectral_entropy	0.036897
1	acce	spectral_centroid	0.040471
1	acce	spectral_spread	0.041548
1	acce	spectral_flatness	0.037884
1	acce	spectral_flux	0.038953
1	acce	spectral_variability	0.038195
1	acce	lyapunov_exponent	0.047669
1	acce	sample_entropy	0.075733
1	acce	higuchi_fractal_dimension	0.048053
1	acce	rmssd	0.048134
1	acce	mean_absolute_deviation	0.037753
2	acce	std	0.046679
2	acce	variance	0.045209
2	acce	range	0.043515
2	acce	iqr	0.039084
2	acce	skewness	0.052615
2	acce	kurtosis	0.053276
2	acce	quartile_deviation	0.039607

2	acce	spectral_entropy	0.036035
2	acce	spectral_centroid	0.0354
2	acce	spectral_spread	0.03828
2	acce	spectral_flatness	0.03707
2	acce	spectral_flux	0.040276
2	acce	spectral_variability	0.040161
2	acce	lyapunov_exponent	0.049408
2	acce	sample_entropy	0.074252
2	acce	higuchi_fractal_dimension	0.042678
2	acce	rmssd	0.048265
2	acce	mean_absolute_deviation	0.042435
3	acce	range	0.045493
3	acce	skewness	0.050049
3	acce	kurtosis	0.05063
3	acce	spectral_centroid	0.039303
3	acce	spectral_spread	0.039941
3	acce	spectral_flatness	0.044929
3	acce	spectral_flux	0.044435
3	acce	lyapunov_exponent	0.061771
3	acce	sample_entropy	0.039322
3	acce	higuchi_fractal_dimension	0.059846
3	acce	rmssd	0.053291
0	gyro	mean	0.042236
0	gyro	min	0.054973
0	gyro	max	0.050152
0	gyro	median	0.046136
0	gyro	sma	0.042248
0	gyro	rms	0.045517
0	gyro	energy	0.048139
0	gyro	log_energy	0.041339
0	gyro	coefficient_of_variation	0.045973
0	gyro	lyapunov_exponent	0.036695
0	gyro	sample_entropy	0.040283
0	gyro	higuchi_fractal_dimension	0.034569
0	gyro	rmssd	0.054941
1	gyro	std	0.036399
1	gyro	variance	0.035108
1	gyro	min	0.045339
1	gyro	max	0.051987
1	gyro	range	0.049654
1	gyro	median	0.044744
1	gyro	sma	0.035697
1	gyro	energy	0.036533
1	gyro	log_energy	0.038313

1	gyro	lyapunov_exponent	0.066484
1	gyro	sample_entropy	0.040093
1	gyro	higuchi_fractal_dimension	0.048189
1	gyro	rmssd	0.059327
2	gyro	mean	0.046616
2	gyro	min	0.049785
2	gyro	max	0.042638
2	gyro	range	0.043068
2	gyro	median	0.054397
2	gyro	sma	0.040696
2	gyro	rms	0.041285
2	gyro	energy	0.046231
2	gyro	log_energy	0.042235
2	gyro	lyapunov_exponent	0.049009
2	gyro	sample_entropy	0.04182
2	gyro	higuchi_fractal_dimension	0.048974
2	gyro	rmssd	0.054628
3	gyro	mean	0.047471
3	gyro	min	0.035523
3	gyro	max	0.066233
3	gyro	range	0.039768
3	gyro	median	0.049429
3	gyro	skewness	0.037588
3	gyro	sma	0.046944
3	gyro	rms	0.039536
3	gyro	energy	0.044406
3	gyro	log_energy	0.04212
3	gyro	spectral_entropy	0.035626
3	gyro	higuchi_fractal_dimension	0.037909
3	gyro	rmssd	0.038503
0	magn	std	0.076907
0	magn	variance	0.074114
0	magn	range	0.065497
0	magn	iqr	0.066873
0	magn	quartile_deviation	0.055063
0	magn	spectral_flux	0.058243
0	magn	spectral_variability	0.07245
0	magn	lyapunov_exponent	0.065484
0	magn	higuchi_fractal_dimension	0.064085
0	magn	rmssd	0.055389
0	magn	mean_absolute_deviation	0.068055
1	magn	std	0.043704
1	magn	variance	0.045641
1	magn	range	0.046979

1	magn	iqr	0.055005
1	magn	skewness	0.050213
1	magn	kurtosis	0.043522
1	magn	quartile_deviation	0.046722
1	magn	spectral_centroid	0.040805
1	magn	spectral_spread	0.037711
1	magn	spectral_flatness	0.040448
1	magn	spectral_flux	0.039723
1	magn	spectral_variability	0.043848
1	magn	lyapunov_exponent	0.051878
1	magn	sample_entropy	0.034768
1	magn	higuchi_fractal_dimension	0.046138
1	magn	rmssd	0.047431
1	magn	mean_absolute_deviation	0.043002
2	magn	std	0.044897
2	magn	variance	0.040706
2	magn	range	0.047522
2	magn	iqr	0.053888
2	magn	skewness	0.052895
2	magn	kurtosis	0.048108
2	magn	quartile_deviation	0.046998
2	magn	spectral_entropy	0.046242
2	magn	spectral_centroid	0.043342
2	magn	spectral_spread	0.044161
2	magn	spectral_flatness	0.047762
2	magn	spectral_flux	0.049533
2	magn	spectral_variability	0.040497
2	magn	lyapunov_exponent	0.053057
2	magn	sample_entropy	0.041175
2	magn	higuchi_fractal_dimension	0.045806
2	magn	rmssd	0.059372
2	magn	mean_absolute_deviation	0.044084
3	magn	std	0.038615
3	magn	variance	0.039355
3	magn	range	0.04715
3	magn	iqr	0.040433
3	magn	skewness	0.060121
3	magn	kurtosis	0.044958
3	magn	quartile_deviation	0.039039
3	magn	spectral_entropy	0.040815
3	magn	spectral_centroid	0.040929
3	magn	spectral_spread	0.040853
3	magn	spectral_flatness	0.045797
3	magn	spectral_flux	0.038838

3	magn	spectral_variability	0.035277
3	magn	lyapunov_exponent	0.049774
3	magn	sample_entropy	0.044242
3	magn	higuchi_fractal_dimension	0.049617
3	magn	rmssd	0.058856
3	magn	mean_absolute_deviation	0.041531

Πίνακας Α3: Σημαντικά χαρακτηριστικά της PCA

Exercise	Sensor	Feature Name	Importance
0	acce	std	3.308751
0	acce	variance	3.335145
0	acce	range	3.272675
0	acce	spectral_entropy	3.427045
0	acce	spectral_centroid	3.510828
0	acce	spectral_spread	3.63043
0	acce	spectral_flatness	3.728374
0	acce	spectral_flux	4.057631
0	acce	spectral_variability	3.731378
0	acce	lyapunov_exponent	3.396204
0	acce	higuchi_fractal_dimension	3.666777
0	acce	rmssd	3.357708
0	acce	mean_absolute_deviation	3.288456
1	acce	std	3.15312
1	acce	variance	3.553376
1	acce	range	3.069691
1	acce	kurtosis	3.029222
1	acce	spectral_entropy	3.712003
1	acce	spectral_centroid	3.734141
1	acce	spectral_spread	3.865418
1	acce	spectral_flatness	3.712563
1	acce	spectral_flux	3.814426
1	acce	spectral_variability	3.533718
1	acce	lyapunov_exponent	3.24961
1	acce	higuchi_fractal_dimension	3.24133
1	acce	rmssd	3.141629
1	acce	mean_absolute_deviation	3.005277
2	acce	std	3.097193
2	acce	variance	3.636863
2	acce	range	3.438474
2	acce	iqr	3.474554
2	acce	quartile_deviation	3.474554
2	acce	spectral_entropy	3.172488
2	acce	spectral_centroid	3.089819
2	acce	spectral_spread	3.363672
2	acce	spectral_flatness	3.257937
2	acce	spectral_flux	3.693033
2	acce	spectral_variability	3.604032
2	acce	lyapunov_exponent	3.480935
2	acce	higuchi_fractal_dimension	3.628248
2	acce	rmssd	3.56755

2	acce	mean_absolute_deviation	3.237777
3	acce	variance	3.661086
3	acce	range	3.461663
3	acce	iqr	3.291526
3	acce	energy	3.140681
3	acce	quartile_deviation	3.291526
3	acce	spectral_entropy	3.38971
3	acce	spectral_spread	3.46242
3	acce	spectral_flatness	3.164378
3	acce	spectral_flux	3.681106
3	acce	spectral_variability	3.341618
3	acce	lyapunov_exponent	3.265598
3	acce	sample_entropy	3.241305
3	acce	higuchi_fractal_dimension	3.346924
3	acce	rmssd	3.215728
3	acce	mean_absolute_deviation	3.205616
0	gyro	variance	3.495962
0	gyro	min	3.481928
0	gyro	max	3.377814
0	gyro	range	3.627668
0	gyro	median	3.376671
0	gyro	iqr	3.438254
0	gyro	energy	3.845916
0	gyro	log_energy	3.947247
0	gyro	quartile_deviation	3.438254
0	gyro	spectral_spread	3.32735
0	gyro	spectral_flatness	3.712223
0	gyro	spectral_flux	3.554384
0	gyro	spectral_variability	3.400455
0	gyro	lyapunov_exponent	3.923
0	gyro	sample_entropy	3.325401
0	gyro	higuchi_fractal_dimension	3.443209
0	gyro	rmssd	3.726256
1	gyro	std	3.144416
1	gyro	variance	3.697173
1	gyro	max	3.152651
1	gyro	range	3.792982
1	gyro	iqr	3.473601
1	gyro	energy	3.133463
1	gyro	quartile_deviation	3.473601
1	gyro	coefficient_of_variation	3.194384
1	gyro	spectral_entropy	3.284598
1	gyro	spectral_rolloff	3.365609
1	gyro	spectral_flatness	3.462297

1	gyro	spectral_flux	3.503744
1	gyro	spectral_variability	3.222527
1	gyro	lyapunov_exponent	3.545245
1	gyro	sample_entropy	3.31286
1	gyro	higuchi_fractal_dimension	3.387417
1	gyro	rmssd	3.7662
2	gyro	std	3.614701
2	gyro	variance	3.601217
2	gyro	range	4.066673
2	gyro	iqr	3.466588
2	gyro	energy	3.287107
2	gyro	log_energy	3.426083
2	gyro	quartile_deviation	3.466588
2	gyro	coefficient_of_variation	3.582317
2	gyro	spectral_entropy	3.269063
2	gyro	spectral_spread	3.329508
2	gyro	spectral_rolloff	3.497589
2	gyro	spectral_flux	3.695707
2	gyro	lyapunov_exponent	3.687265
2	gyro	higuchi_fractal_dimension	3.470805
2	gyro	rmssd	3.458181
2	gyro	mean_absolute_deviation	3.368418
3	gyro	std	3.481519
3	gyro	variance	3.882017
3	gyro	max	3.487422
3	gyro	range	4.312129
3	gyro	iqr	3.776317
3	gyro	energy	3.669629
3	gyro	log_energy	3.45858
3	gyro	quartile_deviation	3.776317
3	gyro	coefficient_of_variation	3.421663
3	gyro	spectral_entropy	3.60632
3	gyro	spectral_spread	3.623893
3	gyro	spectral_rolloff	3.482867
3	gyro	spectral_flatness	3.470768
3	gyro	spectral_flux	3.605091
3	gyro	lyapunov_exponent	3.701858
3	gyro	sample_entropy	3.394093
3	gyro	higuchi_fractal_dimension	3.896964
3	gyro	rmssd	3.60633
0	magn	std	3.003262
0	magn	min	3.258029
0	magn	max	3.196139
0	magn	range	3.655414

0	magn	median	3.084589
0	magn	spectral_flatness	3.528379
0	magn	spectral_flux	3.574994
0	magn	spectral_variability	3.508176
0	magn	lyapunov_exponent	3.772193
0	magn	higuchi_fractal_dimension	3.787829
0	magn	rmssd	3.399234
0	magn	mean_absolute_deviation	3.044992
1	magn	std	2.972562
1	magn	variance	3.10027
1	magn	min	3.068551
1	magn	max	3.022679
1	magn	range	3.51486
1	magn	median	3.064169
1	magn	iqr	3.070195
1	magn	quartile_deviation	3.070195
1	magn	spectral_entropy	3.328001
1	magn	spectral_centroid	3.444864
1	magn	spectral_spread	3.314263
1	magn	spectral_flatness	3.948283
1	magn	spectral_flux	3.44195
1	magn	spectral_variability	3.394845
1	magn	lyapunov_exponent	3.641356
1	magn	higuchi_fractal_dimension	3.809281
1	magn	rmssd	3.452505
1	magn	mean_absolute_deviation	3.139825
2	magn	std	3.040219
2	magn	variance	2.903238
2	magn	range	3.477608
2	magn	median	2.885049
2	magn	iqr	2.886246
2	magn	quartile_deviation	2.886246
2	magn	spectral_centroid	3.038123
2	magn	spectral_spread	3.252679
2	magn	spectral_flatness	3.716822
2	magn	spectral_flux	3.416094
2	magn	spectral_variability	3.391907
2	magn	lyapunov_exponent	3.328779
2	magn	higuchi_fractal_dimension	3.450813
2	magn	rmssd	3.428304
2	magn	mean_absolute_deviation	3.313043
3	magn	std	3.025442
3	magn	variance	3.606878
3	magn	min	3.030865

3	magn	range	3.724349
3	magn	median	3.022254
3	magn	spectral_entropy	3.011908
3	magn	spectral_centroid	3.197297
3	magn	spectral_spread	3.349245
3	magn	spectral_flatness	3.613514
3	magn	spectral_flux	3.733529
3	magn	spectral_variability	3.640058
3	magn	lyapunov_exponent	3.701285
3	magn	higuchi_fractal_dimension	3.85491
3	magn	rmssd	3.856115
3	magn	mean_absolute_deviation	3.212146

Πίνακας A4: Τα 80% καλύτερα χαρακτηριστικά

Exercise	Sensor	Feature Name	Composite Score
2	gyro	rmssd	0.663
2	gyro	lyapunov_exponent	0.644
2	gyro	higuchi_fractal_dimension	0.612
2	gyro	range	0.573
3	acce	higuchi_fractal_dimension	0.469
3	gyro	max	0.452
3	acce	lyapunov_exponent	0.437
2	gyro	spectral_rolloff	0.433
2	acce	sample_entropy	0.422
0	magn	lyapunov_exponent	0.406
0	magn	spectral_variability	0.399
0	magn	higuchi_fractal_dimension	0.399
0	magn	range	0.381
3	magn	rmssd	0.376
0	gyro	rmssd	0.375
2	gyro	std	0.373
3	acce	spectral_flux	0.367
1	gyro	lyapunov_exponent	0.365
1	gyro	rmssd	0.361
1	gyro	max	0.358
2	magn	sample_entropy	0.352
1	acce	sample_entropy	0.347
3	gyro	range	0.337
3	gyro	energy	0.337
1	gyro	energy	0.326
0	magn	std	0.325
2	gyro	quartile_deviation	0.323
2	gyro	iqr	0.323
0	magn	spectral_flux	0.313
3	magn	higuchi_fractal_dimension	0.311
2	gyro	mean_absolute_deviation	0.308
1	gyro	range	0.298
0	gyro	energy	0.298
2	magn	rmssd	0.29
3	acce	rmssd	0.285
1	magn	lyapunov_exponent	0.282
0	magn	variance	0.28
3	magn	lyapunov_exponent	0.279
3	gyro	log_energy	0.278
1	magn	higuchi_fractal_dimension	0.276
0	acce	lyapunov_exponent	0.273

0	gyro	log_energy	0.271
0	magn	mean_absolute_deviation	0.271
0	gyro	min	0.27
2	magn	spectral_flatness	0.268
1	gyro	median	0.267
3	magn	range	0.266
1	magn	spectral_flatness	0.265
2	gyro	energy	0.262
1	gyro	min	0.259
1	acce	spectral_spread	0.256
0	acce	spectral_flux	0.256
0	magn	rmssd	0.256
3	acce	range	0.251
2	acce	rmssd	0.241
3	gyro	higuchi_fractal_dimension	0.236
3	magn	spectral_flatness	0.233
2	acce	variance	0.233
0	gyro	lyapunov_exponent	0.233
2	acce	lyapunov_exponent	0.23
0	magn	iqr	0.229
2	gyro	sample_entropy	0.227
1	acce	spectral_flux	0.226
2	magn	lyapunov_exponent	0.224
2	gyro	log_energy	0.224
3	gyro	median	0.222
2	gyro	spectral_entropy	0.222
1	acce	spectral_centroid	0.22
1	magn	range	0.22
2	magn	spectral_flux	0.218
0	acce	higuchi_fractal_dimension	0.217
0	acce	rmssd	0.217
2	magn	range	0.216
3	acce	variance	0.216
2	acce	higuchi_fractal_dimension	0.214
0	gyro	max	0.214
2	gyro	median	0.212
3	gyro	variance	0.21
1	magn	rmssd	0.21
2	acce	spectral_flux	0.21
3	magn	spectral_flux	0.209
1	gyro	sma	0.202
1	gyro	higuchi_fractal_dimension	0.202
3	gyro	mean	0.201
2	magn	higuchi_fractal_dimension	0.199

3	acce	spectral_variability	0.198
1	acce	lyapunov_exponent	0.198
1	acce	spectral_flatness	0.197
3	gyro	sma	0.197
1	gyro	rms	0.196
1	gyro	mean	0.194
2	acce	spectral_variability	0.191
1	acce	spectral_entropy	0.19
0	acce	sample_entropy	0.19
3	gyro	quartile_deviation	0.187
1	acce	variance	0.187
3	gyro	iqr	0.187
3	gyro	spectral_flatness	0.186
3	gyro	spectral_spread	0.186
3	magn	variance	0.186
0	gyro	median	0.185
1	magn	iqr	0.184
1	acce	kurtosis	0.181
3	magn	skewness	0.181
2	gyro	spectral_spread	0.181
0	acce	spectral_flatness	0.18
2	gyro	max	0.18
3	gyro	rmssd	0.18
2	acce	range	0.18
0	gyro	higuchi_fractal_dimension	0.179
2	gyro	variance	0.179
0	acce	spectral_variability	0.178
0	acce	spectral_spread	0.178
1	gyro	variance	0.175
2	gyro	coefficient_of_variation	0.175
2	gyro	spectral_centroid	0.175
0	gyro	spectral_flatness	0.174
1	gyro	spectral_rolloff	0.174
1	magn	spectral_variability	0.173
3	gyro	lyapunov_exponent	0.172
3	gyro	spectral_entropy	0.171
1	acce	higuchi_fractal_dimension	0.17
2	gyro	spectral_flux	0.17
3	gyro	spectral_flux	0.165
3	magn	spectral_variability	0.164
1	acce	spectral_variability	0.162
1	magn	spectral_centroid	0.162
2	acce	quartile_deviation	0.16
3	acce	spectral_spread	0.159

0	acce	variance	0.158
1	gyro	spectral_entropy	0.157
2	magn	mean_absolute_deviation	0.157
0	gyro	range	0.156
2	gyro	mean	0.156
2	acce	iqr	0.156
1	magn	spectral_flux	0.154
1	acce	rmssd	0.15
2	magn	spectral_variability	0.149
3	gyro	rms	0.147
2	magn	spectral_spread	0.145
0	magn	quartile_deviation	0.145
3	magn	spectral_spread	0.142
0	gyro	spectral_flux	0.141
3	gyro	coefficient_of_variation	0.139
2	magn	iqr	0.137
0	magn	spectral_flatness	0.135
2	gyro	min	0.134
2	acce	kurtosis	0.133
0	gyro	sample_entropy	0.133
0	acce	spectral_centroid	0.132
3	acce	spectral_flatness	0.132
3	acce	iqr	0.131
3	acce	quartile_deviation	0.131
2	magn	skewness	0.13
2	acce	std	0.13
1	gyro	spectral_flux	0.13
2	acce	mean_absolute_deviation	0.13
0	acce	std	0.129
1	gyro	sample_entropy	0.129
3	acce	mean_absolute_deviation	0.129
0	gyro	variance	0.128
2	acce	skewness	0.128
0	acce	range	0.127
2	acce	spectral_spread	0.127
3	gyro	spectral_rolloff	0.126
1	magn	quartile_deviation	0.125
3	gyro	std	0.125
1	gyro	iqr	0.124
1	magn	variance	0.124
1	gyro	quartile_deviation	0.124
2	gyro	rms	0.122
0	acce	spectral_entropy	0.121
1	gyro	spectral_flatness	0.121

3	magn	mean_absolute_deviation	0.118
0	gyro	iqr	0.116
0	gyro	quartile_deviation	0.116
3	gyro	min	0.115
2	gyro	sma	0.114
3	acce	kurtosis	0.114
1	magn	mean_absolute_deviation	0.113
1	magn	spectral_spread	0.112
1	magn	skewness	0.111
3	magn	spectral_centroid	0.111
3	acce	skewness	0.11
3	acce	sample_entropy	0.109
3	acce	energy	0.109
0	gyro	spectral_variability	0.108
3	gyro	sample_entropy	0.107
3	acce	spectral_entropy	0.106
2	magn	std	0.106
1	acce	skewness	0.102
0	acce	mean_absolute_deviation	0.102
1	acce	range	0.1
2	acce	spectral_flatness	0.096
2	magn	kurtosis	0.096
2	magn	spectral_centroid	0.094
1	magn	spectral_entropy	0.093
0	gyro	spectral_spread	0.093
0	acce	skewness	0.093
2	magn	quartile_deviation	0.088
2	magn	spectral_entropy	0.083
1	magn	std	0.083
1	acce	std	0.083

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Abdi, H., Williams, L.J., 2010. Principal component analysis. *WIREs Computational Stats* 2, 433–459. <https://doi.org/10.1002/wics.101>
- Abdul, Z.Kh., Al-Talabani, A.K., 2022. Mel Frequency Cepstral Coefficient and its Applications: A Review. *IEEE Access* 10, 122136–122158. <https://doi.org/10.1109/ACCESS.2022.3223444>
- Acharya, U.R., Fujita, H., Lih, O.S., Adam, M., Tan, J.H., Chua, C.K., 2017. Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. *Knowledge-Based Systems* 132, 62–71. <https://doi.org/10.1016/j.knosys.2017.06.003>
- Alaa, A.M., Bolton, T., Di Angelantonio, E., Rudd, J.H.F., van der Schaar, M., 2019. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PLoS One* 14, e0213653. <https://doi.org/10.1371/journal.pone.0213653>
- Antoine, J.-P., 2003. *Wavelet Transforms and Their Applications*. *Physics Today* 56, 68–68. <https://doi.org/10.1063/1.1580056>
- Arora, J., Meumeu Yomsi, P., 2019. Wearable Sensors Based Remote Patient Monitoring using IoT and Data Analytics. *U.Porto Journal of Engineering* 5, 34–45. https://doi.org/10.24840/2183-6493_005.001_0003
- Ashley, E.A., 2015. The precision medicine initiative: a new national effort. *JAMA* 313, 2119–2120. <https://doi.org/10.1001/jama.2015.3595>
- Atzori, L., Iera, A., Morabito, G., 2010. The Internet of Things: A survey. *Computer Networks* 54, 2787–2805. <https://doi.org/10.1016/j.comnet.2010.05.010>
- Bashshur, R., Doarn, C.R., Frenk, J.M., Kvedar, J.C., Woolliscroft, J.O., 2020. Telemedicine and the COVID-19 Pandemic, Lessons for the Future. *Telemed J E Health* 26, 571–573. <https://doi.org/10.1089/tmj.2020.29040.rb>
- Bergstra, J., Bergstra, J., Bengio, Y., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization.
- Bishop, C.M., 2006. Pattern recognition and machine learning. New York : Springer, [2006] ©2006.

- Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30, 1145–1159.
- Breiman, L., 2017. *Classification and Regression Trees*, 1st ed. New York. <https://doi.org/10.1201/9781315139470>
- Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bukhari, S.N.H., Ogudo, K.A., 2024. Ensemble Machine Learning Approach for Parkinson’s Disease Detection Using Speech Signals. *Mathematics* 12, 1575. <https://doi.org/10.3390/math12101575>
- Chawla, N.V., 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*. <https://doi.org/10.1613/jair.953>
- Chen, Z., Pang, M., Zhao, Z., Li, S., Miao, R., Zhang, Yifan, Feng, Xiaoyue, Feng, Xin, Zhang, Yexian, Duan, M., Huang, L., Zhou, F., 2020. Feature selection may improve deep neural networks for the bioinformatics problems. *Bioinformatics* 36, 1542–1552. <https://doi.org/10.1093/bioinformatics/btz763>
- Chiang, P.-H., Wong, M., Dey, S., 2021. Using Wearables and Machine Learning to Enable Personalized Lifestyle Recommendations to Improve Blood Pressure. *IEEE J Transl Eng Health Med* 9, 2700513. <https://doi.org/10.1109/JTEHM.2021.3098173>
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach Learn* 20, 273–297. <https://doi.org/10.1007/BF00994018>
- Darkins, A.W., Ryan, P., Kobb, R., Foster, L., Edmonson, E., Wakefield, B.J., Lancaster, A.E., 2008. Care Coordination/Home Telehealth: the systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. *Telemedicine journal and e-health : the official journal of the American Telemedicine Association* 14 10, 1118–26.
- Dorsey, E.R., Topol, E.J., 2016. State of Telehealth. *N Engl J Med* 375, 154–161. <https://doi.org/10.1056/NEJMr1601705>
- Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S., 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118. <https://doi.org/10.1038/nature21056>
- Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. *Ann. Statist.* 29. <https://doi.org/10.1214/aos/1013203451>

- Giannakopoulou, K.-M., Roussaki, I., Demestichas, K., 2022. Internet of Things Technologies and Machine Learning Methods for Parkinson's Disease Diagnosis, Monitoring and Management: A Systematic Review. *Sensors* 22, 1799. <https://doi.org/10.3390/s22051799>
- Goetz, C.G., Tilley, B.C., Shaftman, S.R., Stebbins, G.T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M.B., Dodel, R., Dubois, B., Holloway, R., Jankovic, J., Kulisevsky, J., Lang, A.E., Lees, A., Leurgans, S., LeWitt, P.A., Nyenhuis, D., Olanow, C.W., Rascol, O., Schrag, A., Teresi, J.A., Van Hilten, J.J., LaPelle, N., 2008. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders* 23, 2129–2170. <https://doi.org/10.1002/mds.22340>
- Gresty, M., Buckwell, D., 1990. Spectral analysis of tremor: understanding the results. *Journal of Neurology, Neurosurgery & Psychiatry* 53, 976–981. <https://doi.org/10.1136/jnnp.53.11.976>
- Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M., 2013. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems* 29, 1645–1660. <https://doi.org/10.1016/j.future.2013.01.010>
- Helmich, R.C., Hallett, M., Deuschl, G., Toni, I., Bloem, B.R., 2012. Cerebral causes and consequences of parkinsonian resting tremor: a tale of two circuits? *Brain* 135, 3206–3226. <https://doi.org/10.1093/brain/aws023>
- Higuchi, T., 1988. Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena* 31, 277–283. [https://doi.org/10.1016/0167-2789\(88\)90081-4](https://doi.org/10.1016/0167-2789(88)90081-4)
- Hosmer, D.W., 2013. *Applied Logistic Regression*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118548387>
- Islam, G.M.N., Ali, M., Collie, S., 2020. Textile sensors for wearable applications: a comprehensive review. *Cellulose* 27, 1–29. <https://doi.org/10.1007/s10570-020-03215-5>
- Jankovic, J., 2008. Parkinson's disease: clinical features and diagnosis. *J Neurol Neurosurg Psychiatry* 79, 368. <https://doi.org/10.1136/jnnp.2007.131045>
- Javaid, M., Haleem, A., Pratap Singh, R., Suman, R., Rab, S., 2022. Significance of machine learning in healthcare: Features, pillars and applications. *International Journal of Intelligent Networks* 3, 58–73. <https://doi.org/10.1016/j.ijin.2022.05.002>
- Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Yilong, Dong, Q., Shen, H., Wang, Yongjun, 2017. Artificial

- intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2, 230–243. <https://doi.org/10.1136/svn-2017-000101>
- Jolliffe, I.T., 2002. *Principal Component Analysis*, 2nd ed, Springer Series in Statistics. Springer New York, NY. <https://doi.org/10.1007/b98835>
- Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc. A* 374, 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Jordan, M.I., Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. *Science* 349, 255–260. <https://doi.org/10.1126/science.aaa8415>
- Kalantar-Zadeh, K., Ha, N., Ou, J.Z., Berean, K.J., 2017. Ingestible Sensors. *ACS Sens* 2, 468–483. <https://doi.org/10.1021/acssensors.7b00045>
- Kaufman, S., Rosset, S., Perlich, C., 2011. Leakage in data mining: formulation, detection, and avoidance, in: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Presented at the KDD '11: The 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Diego California USA, pp. 556–563. <https://doi.org/10.1145/2020408.2020496>
- Khan, A.A., Kim, J.-H., 2024. Recent advances in materials and manufacturing of implantable devices for continuous health monitoring. *Biosens Bioelectron* 261, 116461. <https://doi.org/10.1016/j.bios.2024.116461>
- Klonoff, D.C., Buckingham, B., Christiansen, J.S., Montori, V.M., Tamborlane, W.V., Vigersky, R.A., Wolpert, H., 2011. Continuous glucose monitoring: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 96, 2968–2979. <https://doi.org/10.1210/jc.2010-2756>
- Kohavi, R., 2001. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection 14.
- Krones, F., Marikkar, U., Parsons, G., Szmul, A., Mahdi, A., 2024. Review of multimodal machine learning approaches in healthcare. *Information Fusion* 114, 102690. <https://doi.org/10.1016/j.inffus.2024.102690>
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444. <https://doi.org/10.1038/nature14539>
- Li, C., Wang, J., Wang, S., Zhang, Y., 2024. A review of IoT applications in healthcare. *Neurocomputing* 565, 127017. <https://doi.org/10.1016/j.neucom.2023.127017>
- Liu, L., Stroulia, E., Nikolaidis, I., Miguel-Cruz, A., Rios Rincon, A., 2016. Smart homes and home health monitoring technologies for

- older adults: A systematic review. *Int J Med Inform* 91, 44–59.
<https://doi.org/10.1016/j.ijmedinf.2016.04.007>
- López-Martínez, F., Núñez-Valdez, E.R., García-Díaz, V., Bursac, Z., 2020. A Case Study for a Big Data and Machine Learning Platform to Improve Medical Decision Support in Population Health Management. *Algorithms* 13, 102.
<https://doi.org/10.3390/a13040102>
- Louppe, G., 2015. Understanding Random Forests: From Theory to Practice. <https://doi.org/10.48550/arXiv.1407.7502>
- Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. 2017.
- Meskó, B., Hetényi, G., Györfy, Z., 2018. Will artificial intelligence solve the human resource crisis in healthcare? *BMC Health Services Research* 18, 545. <https://doi.org/10.1186/s12913-018-3359-4>
- Mitchell, T.M., 1997. *Machine Learning*. McGraw Hill.
- Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S., 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 447–453.
<https://doi.org/10.1126/science.aax2342>
- Patel, S., Lorincz, K., Hughes, R., Huggins, N., Growdon, J., Standaert, D., Akay, M., Dy, J., Welsh, M., Bonato, P., 2009. Monitoring Motor Fluctuations in Patients With Parkinson's Disease Using Wearable Sensors. *IEEE Trans. Inform. Technol. Biomed.* 13, 864–873.
<https://doi.org/10.1109/TITB.2009.2033471>
- Pedregosa, F., Pedregosa, F., Varoquaux, G., Varoquaux, G., Org, N., Gramfort, A., Gramfort, A., Michel, V., Michel, V., Fr, L., Thirion, B., Thirion, B., Grisel, O., Grisel, O., Blondel, M., Prettenhofer, P., Prettenhofer, P., Weiss, R., Dubourg, V., Dubourg, V., Vanderplas, J., Passos, A., Tp, A., Cournapeau, D., 2011. *Scikit-learn: Machine Learning in Python. MACHINE LEARNING IN PYTHON*.
- Peever, J., Fuller, P.M., 2017. The Biology of REM Sleep. *Current Biology* 27, R1237–R1248.
<https://doi.org/10.1016/j.cub.2017.10.026>
- Pincus, S.M., 1991. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. U.S.A.* 88, 2297–2301.
<https://doi.org/10.1073/pnas.88.6.2297>
- Pradyumna, G.R., Hegde, R.B., Bommegowda, K.B., Jan, T., Naik, G.R., 2024. Empowering Healthcare With IoMT: Evolution, Machine Learning Integration, Security, and Interoperability

- Challenges. IEEE Access 12, 20603–20623.
<https://doi.org/10.1109/ACCESS.2024.3362239>
- Provost, F., Fawcett, T., 2013. Data Science for Business, 1st ed.
- Rajendran, P., Chatrapathy, K., Bhat, N., Balaji, T., Rajendiran, M., Maranan, R., 2025. IoT and machine learning-based postoperative monitoring: Pioneering precision medicine in healthcare. pp. 284–289.
<https://doi.org/10.1201/9781003559115-48>
- Ray, P.P., 2018. A survey on Internet of Things architectures. Journal of King Saud University - Computer and Information Sciences 30, 291–319. <https://doi.org/10.1016/j.jksuci.2016.10.003>
- Richman, J.S., Moorman, J.R., 2000. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology 278, H2039–H2049.
<https://doi.org/10.1152/ajpheart.2000.278.6.H2039>
- Saito, T., Rehmsmeier, M., 2015. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. PLoS ONE 10, e0118432.
<https://doi.org/10.1371/journal.pone.0118432>
- Singh, B., Bhattacharya, S., Chowdhary, C.L., Jat, D.S., 2017. A review on internet of things and its applications in healthcare.
- Sutton, R.S., Barto, A.G., 2018. Reinforcement learning: An introduction, 2nd ed., Reinforcement learning: An introduction, 2nd ed. The MIT Press, Cambridge, MA, US.
- Swan, M., 2012. Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. Journal of Sensor and Actuator Networks 1, 217–253.
<https://doi.org/10.3390/jsan1030217>
- Tan, S.Y., Sumner, J., Wang, Y., Wenjun Yip, A., 2024. A systematic review of the impacts of remote patient monitoring (RPM) interventions on safety, adherence, quality-of-life and cost-related outcomes. npj Digit. Med. 7, 192.
<https://doi.org/10.1038/s41746-024-01182-w>
- Turner, A.P.F., 2013. Biosensors: sense and sensibility. Chem Soc Rev 42, 3184–3196. <https://doi.org/10.1039/c3cs35528d>
- Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., Zhao, S., 2019. Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery 18, 463–477.
<https://doi.org/10.1038/s41573-019-0024-5>
- Wang, Y., Min, J., Khuri, J., Xue, H., Xie, B., A Kaminsky, L., J Cheskin, L., 2020. Effectiveness of Mobile Health Interventions

- on Diabetes and Obesity Treatment and Management: Systematic Review of Systematic Reviews. *JMIR Mhealth Uhealth* 8, e15400. <https://doi.org/10.2196/15400>
- Yang, Q., Liu, Y., Chen, T., Tong, Y., 2019. Federated Machine Learning: Concept and Applications. *ACM Trans. Intell. Syst. Technol.* 10. <https://doi.org/10.1145/3298981>
- Yu, K.-H., Beam, A.L., Kohane, I.S., 2018. Artificial intelligence in healthcare. *Nature Biomedical Engineering* 2, 719–731. <https://doi.org/10.1038/s41551-018-0305-z>

