



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

Μηχανική Μάθηση
για την Πρόγνωση Ψυχωσικής Υποτροπής,
σε ασθενείς με Σχιζοφρένεια, αξιοποιώντας Φορητές Συσσκευές

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Υπεύθυνος
Τασουλής Σωτήριος
Επίκουρος Καθηγητής

Παρασκευή Τσακμάκη

Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις⁽¹⁾, που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί **παράθεση χωρίς εισαγωγικά** χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.

2. Δέχομαι ότι η αυτολεξεί παράθεση χωρίς εισαγωγικά, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.

3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια

4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής

Ημερομηνία: 04/10/2022

Η Δηλούσα

Παρασκευή Τσακμάκη

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών

Μηχανική Μάθηση για την Πρόγνωση Ψυχωσικής Υποτροπής,
σε ασθενείς με σχιζοφρένεια, αξιοποιώντας Φορητές Συσκευές
Παρασκευή Τσακμάκη

Τριμελής Επιτροπή:

Τασουλής Σωτήριος, Επίκουρος Καθηγητής (Επιβλέπων)

Α. Κακαρούνας, Αναπληρωτής Καθηγητής

Πλαγιανάκος Βασίλειος, Καθηγητής

I

Περίληψη

Η ενσωμάτωση της τεχνολογίας στην κλινική περίθαλψη αυξάνεται ραγδαία και έχει γίνει ιδιαίτερα σημαντική. Η αναγνώριση των πρόδρομων συμπτωμάτων της σχιζοφρένειας προσφέρει τη δυνατότητα πρώιμης παρέμβασης για την αποφυγή υποτροπής και επανεισαγωγής στο νοσοκομείο.

Η παρούσα μελέτη αποτελείται από δύο πειράματα πρόγνωσης της υποτροπής σε πραγματικούς ασθενείς με σχιζοφρένεια, χρησιμοποιώντας φορητές συσκευές.

Η παρακολούθηση της μεταβλητότητας καρδιακού παλμού προσφέρεται ολοένα και περισσότερο μέσω φορητών συσκευών και η κατανόηση της αποκαλύπτει μερικά συναρπαστικά μονοπάτια μεταξύ του νου και σώματος αφού έχει αποδειχθεί ότι αποτελεί ισχυρό δείκτη ψυχολογικού στρες. Πολλά στοιχεία δείχνουν ότι τα υψηλά επίπεδα στρεσογόνων γεγονότων της ζωής μπορούν να αυξήσουν τον κίνδυνο υποτροπής σε άτομα με σχιζοφρένεια.

Το πρώτο πείραμα αφορά την εξαγωγή χαρακτηριστικών μέσω της μεταβλητότητας καρδιακού παλμού σε χρονικά διαστήματα κατά τη διάρκεια της ημέρας για την πρόγνωση της ψυχωσικής υποτροπής, με αλγόριθμους μηχανικής μάθησης. Το δεύτερο πείραμα αφορά την μεταβλητότητα καρδιακού παλμού κατά το νυχτερινό ύπνο για την πρόγνωση της ψυχωσικής υποτροπής, με αλγόριθμους μηχανικής μάθησης.

Η σύγκριση των πειραμάτων είχε ως αποτέλεσμα 77% ακρίβεια και 90% ανάκληση για το πρώτο πείραμα και 92% ακρίβεια και 51% ανάκληση για το δεύτερο.

II

Abstract

The integration of technology into clinical care is growing rapidly and has become particularly important. Recognizing the precursor symptoms of schizophrenia offers the possibility of early intervention to avoid relapse and re-hospitalization. The present study consists of two relapse prediction experiments in real patients, with schizophrenia, using wearable devices.

Heart rate variability monitoring is increasingly offered via wearable devices, and understanding it, reveals some fascinating pathways between mind and body as it has been shown to be a powerful indicator of psychological stress. Much evidence suggests that high levels of stressful life events can increase the risk of relapse in people with schizophrenia.

The first experiment is about the extraction of features through heart rate variability in day time intervals for the prediction of psychotic relapse via machine learning. The second experiment concerns the heart rate variability for night sleep for the prediction of psychotic relapse of each patient via machine learning.

The comparison of the two experiments resulted in 77% precision and 90% recall for the first experiment and 92% precision and 51% recall for the second

Πίνακας Περιεχομένων

I	Περίληψη	3
II	Abstract	5
	Εισαγωγή	11
1.1	Σχιζοφρένεια	11
1.1.1	Ιστορική Αναδρομή	12
1.1.2	Ψυχωσική Υποτροπή	12
1.1.3	Άγχος και Ψυχωσική Υποτροπή	13
1.2	Σχετικές έρευνες	13
1.3	Εξόρυξη δεδομένων μεγάλης κλίμακας	15
1.3.1	Μεγάλα Δεδομένα και Σχιζοφρένεια	15
	Χρόνος και Υλικό	17
2.1	Διακύμανση Καρδιακού Πάλμου	17
2.1.1	ΣΝΣ-ΠΣΝΣ	18
2.1.2	Ηλεκτροκαρδιογράφημα	20
2.2	Η Χρονοσειρά	23
2.3	Κατηγορίες Επεγερσίας HRV	23
2.4	Στατιστική Ανάλυση σε Χρονικό Πεδίο	24
2.4.1	SDNN	25
2.4.2	SDANN	25
2.4.3	RMSSD	25
2.4.4	HR Max- HR Min	26
2.4.5	NNx	26
2.4.6	SDD	26
2.5	Γεωμετρική απεικόνιση	27
2.5.1	Τριγωνοποίηση Ιστογράμματος	27
2.5.2	Poincare Plot	29

2.6	Εντροπίες	34
2.6.1	Approximate Entropy	34
2.6.2	Sample Entropy	36
2.6.3	Singular Value Decomposition (SVD) Entropy	36
2.6.4	Shannon Entropy	36
2.6.5	Permutation Entropy	36
2.7	Detrended Fluctuation Analysis – DFA	37
2.8	Fractal Dimension	38
2.8.1	Petrosian algorithm	38
2.8.2	Katz’s algorithm	39
2.8.3	Higuchi algorithm	40
2.9	Στατιστικός έλεγχος και συσχέτιση μεταβλητών	41
2.9.1	Στατιστικός Έλεγχος	41
2.9.2	Συσχέτιση Μεταβλητών	41
2.10	Μηχανική Μάθηση	42
2.10.1	Μηχανική Μάθηση με Επίβλεψη	43
2.10.2	Αλγόριθμοι/Μέθοδοι Μηχανικής Μάθησης χωρίς Επίβλεψη	44
2.10.3	Ημι-Εποπτευόμενοι Αλγόριθμοι Μηχανικής Μάθησης	45
2.10.4	Reinforcement Machine Learning Algorithms	45
2.10.5	Αλγόριθμοι Ταξινόμησης	45
Μεθοδολογία και Πειράματα		52
3.1	Δεδομένα από φορητές συσκευές	52
3.2	Μελέτη των Διαταραχών του Ύπνου	53
3.3	Περιγραφή του συνόλου δεδομένων	54
3.4	Αρχιτεκτονική Dataset	56
3.5	Τεχνικές Φιλτραρίσματος	59
3.5.1	Οπτικοποίηση φιλτραρίσματος	60
3.5.2	Ταχόγραμμα	61
3.6	Στατιστικός Έλεγχος	71
3.7	Συμπεράσματα Στατιστικού Ελέγχου	71
3.8	Συσχέτιση Μεταβλητών	73
3.9	Μηχανική Μάθηση	77
3.10	Αποτελέσματα	84
Συμπεράσματα		86

Κατάλογος Εικόνων

2.1	Δραστηριότητα ΚΝΣ	19
2.2	QRS	21
2.3	Χάρτης Απεικόνισης Διαστημάτων RR [60]	22
2.4	Πρότυπα ιστογράμματα διαστημάτων NN [51]	29
2.5	Poincare Plot	30
2.6	Κατηγορίες Αποτελεσμάτων Poincare Plot [19]	32
2.7	Αριθμός Δημοσιευμένων Εργασιών	33
3.1	Δεδομένα Ύπνου για τον 5019	56
3.2	Dataframe Χρονοσειράς	57
3.3	Οπτικοποίηση ενοποιημένης χρονοσειράς του 5006	58
3.4	Παρεμβαλλόμενο-Πρωτότυπο Σήμα	61
3.5	Ταχόγραμμα δεδομένων ύπνου	62
3.6	ECG Summary	63
3.7	Ημερήσια HRV ανάλυση για τον 5006	64
3.8	TriIndex νυχτερινού ύπνου του 5019	65
3.9	Poincare Plot νυχτερινού ύπνου του 5019	65
3.10	Poincare plot νύχτας και ημέρας υγιούς ατόμου	66
3.11	Pplot νυχτερινού ύπνου	67
3.12	Pplot ημερήσιας κατανομής	67
3.13	Κατανομή διαστημάτων RR του 5006	69
3.14	Σύγκριση μεταβλητών μεταξύ ασθενών	70
3.15	Κλάσεις ημέρησιων δεδομένων	77
3.16	Κλάσεις δεδομένων νυχτερινού ύπνου	77
3.17	Συνοπτικά στατιστικά στοιχεία των νοσούντων κατά τη διάρκεια της ημέρας	77
3.18	Συνοπτικά στατιστικά στοιχεία των μη νοσούντων κατά τη διάρκεια της ημέρας	78
3.19	KNN ημέρας	80
3.20	KNN νύχτας	80

3.21 ROC-AUC KNN ημέρας	80
3.22 ROC-AUC KNN νύχτας	80
3.23 Random Forest ημέρας	81
3.24 Random Forest νύχτας	81
3.25 ROC-AUC RF ημέρας	81
3.26 ROC-AUC RF νύχτας	81
3.27 SVM ημέρας	82
3.28 SVM νύχτας	82
3.29 ROC-AUC SVM ημέρας	82
3.30 ROC-AUC SVM νύχτας	82
3.31 GNB ημέρας	83
3.32 GNB νύχτας	83
3.33 ROC-AUC GNB ημέρας	83
3.34 ROC-AUC GNB νύχτας	83

Κατάλογος πινάκων

2.1 Πίνακας Στατιστικών Δεικτών	24
3.1 Πίνακας με τις ημερομηνίες καταγραφής των δεδομένων	54
3.2 Πίνακας Ψυχωσικών Επεισοδίων	55
3.3 Τιμές αναφοράς σε υγιή άτομα	68
3.4 Υσχυρή συσχέτιση μεταβλητών με τη μέθοδο Spearman's στο νυχτερινό ύπνο.	75
3.5 Υσχυρή συσχέτιση μεταβλητών με τη μέθοδο Spearman's σε δεδομένα ημέρας.	76
3.6 Μετρικές ακρίβειας, ανάκλησης, F1 για κάθε ταξινομητή	84

1.1 Σχιζοφρένεια

Οι διαταραχές στο φάσμα της σχιζοφρένειας SSD είναι οι πολύπλοκες και χρόνιες καταστάσεις που χαρακτηρίζονται από ένα ποικίλο σύνολο συμπτωμάτων που παρουσιάζονται ετερογενώς σε όλο τον πληγέντα πληθυσμό. Τα συμπτώματα συνήθως κατηγοριοποιούνται σε 2 ομάδες: Σε θετικά συμπτώματα, τα οποία είναι η υπερβολή της φυσιολογικής λειτουργίας. Μπορούν να περιλαμβάνουν παραληρηματικές ιδέες, αποδιοργανωμένη σκέψη και ομιλία, και ψευδαισθήσεις που ορίζονται σε οπτικό, οσφρητικό και γευστικό πεδίο που τυπικά θεωρούνται ως εκδηλώσεις της ψύχωσης. Οι ψευδαισθήσεις είναι επίσης τυπικά συνδεδεμένες με το περιεχόμενο του παραληρητικού θέματος. [26] Τα θετικά συμπτώματα γενικά ανταποκρίνονται καλά στη φαρμακευτική αγωγή.[3]

Τα αρνητικά συμπτώματα ορίζονται ως οι ελλείψεις των τυπικών συναισθηματικών αντιδράσεων ή άλλων διαδικασιών σκέψης και δεν ανταποκρίνονται τόσο καλά στη φαρμακευτική αγωγή. [8] Χαρακτηρίζονται συνήθως από επίπεδο ή αμβλύ συναίσθημα και συγκίνηση, φτωχό λόγο (αλογία), αδυναμία να νιώσουν ευχαρίστηση (ανηδονία), έλλειψη της επιθυμίας να δημιουργήσουν σχέσεις (ακοινωνία), και αβουλησία. Έρευνες δείχνουν ότι τα αρνητικά συμπτώματα συμβάλλουν στην λειτουργική ανικανότητα με αποτέλεσμα να ρίχνουν το βάρος της ευθύνης τους, στους άλλους από ότι τα θετικά συμπτώματα. [59] Η ανταπόκριση στην φαρμακευτική αγωγή των ασθενών με εμφανή αρνητικά συμπτώματα, είναι συχνά περιορισμένη. [8]

1.1.1 Ιστορική Αναδρομή

Έχει αναφερθεί το 2011 ότι η σχιζοφρένεια επηρεάζει τις ζωές περίπου 24 εκατομμυρίων ανθρώπων παγκοσμίως. [36]. Εμφανίζεται 1,4 φορές περισσότερο στους άντρες απ' ό,τι τις γυναίκες και συνήθως εμφανίζεται νωρίτερα στους πρώτους. Οι ηλικίες στις οποίες κατά μέσο όρο ξεκινάει η ασθένεια είναι 20-28 χρονών για τους άντρες και 26-32 χρονών για τις γυναίκες. Η εμφάνιση της ασθένειας στην παιδική ηλικία είναι πολύ σπανιότερη απ' ό,τι είναι στους μεσήλικες και στους ηλικιωμένους. Παρόλο που λέγεται ότι η σχιζοφρένεια εμφανίζεται ομοιογενώς σε όλο τον κόσμο, η αλήθεια είναι ότι η εμφάνισή της ποικίλλει από χώρα σε χώρα, αλλά και σε πιο τοπικό επίπεδο, από περιοχή σε περιοχή. Σ' αυτήν οφείλεται το 1% της αναπηρίας παγκοσμίως. Το ποσοστό της εμφάνισης της σχιζοφρένειας ποικίλλει και μπορεί να φτάσει μέχρι και στο τριπλάσιο αν καθοριστεί με πιο χαλαρά κριτήρια. Η σχιζοφρένεια είναι μια σημαντική αιτία αναπηρίας, που θεωρείται ως η τρίτη χειρότερη κατάσταση μετά από την τετραπληγία και την άνοια, ενώ βρίσκεται σε υψηλότερη θέση από την παραπληγία και την τύφλωση. Περίπου τα τρία τέταρτα των ανθρώπων με σχιζοφρένεια παρουσιάζουν ήδη ψύχωση που υποτροπιάζει. Οι περισσότεροι άνθρωποι με σχιζοφρένεια ζουν ανεξάρτητα, με μια μικρή βοήθεια από τους κοντινούς τους ανθρώπους. Στους πάσχοντες που βιώνουν ένα πρώτο επεισόδιο ψύχωσης, οι πιθανότητες για μια αισιόδοξη έκβαση της ασθένειας είναι γύρω στο 42%, για μια ενδιάμεση έκβαση στο 35% και για μια δυσάρεστη έκβαση στο 27%. Η έξαρση στα SSD οδηγεί σε ψυχωσική υποτροπή.(1)

1.1.2 Ψυχωσική Υποτροπή

Η υποτροπή έχει σοβαρές πιθανές συνέπειες, θέτοντας σε κίνδυνο πολλές φορές στη ζωή των ασθενών, με κίνδυνο να προκαλέσουν βλάβη στον εαυτό τους ή σε άλλους. Η ετερογένεια των συμπτωμάτων και ο χρόνος εμφάνισης της έξαρσης κάνουν την ανίχνευση πρώιμων προειδοποιητικών στοιχείων δύσκολη. Συμπτώματα υποτροπής, σε αντίθεση με το κοινό πρώτο επεισόδιο ψύχωσης, μπορεί να εμφανιστούν απότομα. Τα συμπτώματα δείχνουν ότι η εκδήλωση τους είναι εξαιρετικά ιδιοσυγκρασιακή. Καθε άτομο μπορεί να έχει τη δική του μοναδική υπογραφή υποτροπής και η αναγνώριση αυτής της υπογραφής θα μπορούσε να είναι ο πιο αποτελεσματικός τρόπος ανίχνευσης της. Ωστόσο, είναι συχνά απίθανο οι ασθενείς να παρουσιαστούν σε έναν γιατρό όταν τα συμπτώματά τους αρχίσουν να επανεμφανίζονται ή επιδεινώνονται, ιδιαίτερα σε μια ασθένεια που χαρακτηρίζεται από αποδιοργάνωση σε γνωσιακό επίπεδο, απώλεια ενόρασης και ασυνέπεια στη λήψη αγωγής. Όμως μπορούν να αναπτυχθούν εργαλεία για την πρόβλεψη των πρώιμων συμπτωμάτων ψυχωσικής υποτροπής. Η αποτελεσματική αναγνώριση μπορεί να προσφέρει τη δυνατότητα για έγκαιρη παρέμβαση για την πρόληψη της υποτροπής, όπως προσαρμογή στη φαρμακευτική αγωγή, ψυχοκοινωνικές θεραπείες,

κοινωνική υποστήριξη και μείωση του στρες. Η ικανότητα των ασθενών να προβλέπουν τις υποτροπές τους, μπορεί επίσης να επιδεινωθεί καθώς τα συμπτώματα προχωρούν και η διορατικότητα μειώνεται. Τα στοιχεία υποδηλώνουν ότι τα ποσοστά υποτροπής μετά από ένα πρώτο επεισόδιο ψύχωσης είναι περίπου 28% ένα χρόνο μετά τη θεραπεία και έως 54% σε 3 χρόνια μετά τη θεραπεία. Το ποσοστό υποτροπής μετά τη διακοπή των αντιψυχωσικών σε άτομα με χρόνια σχιζοφρένεια είναι περίπου 38%. Η υποτροπή είναι δεν είναι ένα σπάνιο συμβάν και η έλλειψη διαθέσιμων δεδομένων μπορεί να κάνει την πρόβλεψη προβληματική.

1.1.3 Άγχος και Ψυχωσική Υποτροπή

Υπάρχουν πολλές αιτίες που μπορεί να κάνουν κάποιον να υποτροπιάσει. Μεταξύ αυτών, ένας από τους πιο σημαντικούς λόγους πίσω από μια υποτροπή είναι το άγχος.[58] Σε μελέτη της αρχικής φάσης της σχιζοφρένειας, έχει βρεθεί ότι οι ψυχοκοινωνικοί στρεσογόνοι παράγοντες είναι σημαντικοί προγνωστικοί παράγοντες ψυχωσικής υποτροπής κατά τη διάρκεια μιας περιόδου κατά την οποία χρησιμοποιήθηκε φαρμακευτική αγωγή συντήρησης για τη μεγιστοποίηση της συμμόρφωσης των ασθενών.[39]

1.2 Σχετικές έρευνες

Σχετικές έρευνες[43] δείχνουν ότι η διακύμανση καρδιακού παλμού (Heart Rate Variability-HRV) μπορεί να παρέχει έναν πιθανό βιοδείκτη του κινδύνου καρδιαγγειακών νοσημάτων και της σοβαρότητας των συμπτωμάτων τους, σε σοβαρές ψυχικές ασθένειες. Υπολογίστηκε η HRV σε 47 ασθενείς με σχιζοφρένεια, 33 ασθενείς με διπολική διαταραχή και 212 υγιείς περιπτώσεις. Συλλέχθηκαν επίσης μετρήσεις της σοβαρότητας των συμπτωμάτων από τις ομάδες ασθενών. Τα αποτελέσματα έδειξαν ότι η HRV μειώθηκε σημαντικά τόσο στη διαταραχή του φάσματος της σχιζοφρένειας και της διπολικής διαταραχής, σε σύγκριση με τους υγιείς συμμετέχοντες. Ωστόσο, δεν υπήρχαν διαφορές HRV μεταξύ των διαταραχών. Είναι σημαντικό ότι αυτές οι μειώσεις ήταν ανεξάρτητες από τις επιδράσεις της φαρμακευτικής αγωγής, της ηλικίας ή του δείκτη μάζας σώματος. Υπήρχαν επίσης προκαταρκτικές ενδείξεις ότι οι ασθενείς με μειωμένη HRV είχαν αυξημένη συνολική και αρνητική σοβαρότητα συμπτωμάτων ψύχωσης ανεξάρτητα από τη διάγνωση τους. Παρά τα τρέχοντα διαγνωστικά συστήματα που διακρίνουν τη σχιζοφρένεια και τη διπολική διαταραχή ως ξεχωριστές ασθένειες, τα αναδυόμενα στοιχεία δείχνουν ότι μοιράζονται μια σειρά από κλινικά και επιδημιολογικά χαρακτηριστικά, όπως αυξημένο κίνδυνο καρδιαγγειακής νόσου.

Άλλη έρευνα[13] εντόπισε δείκτες που κάνουν διαφοροποίηση μεταξύ των υγιών ατόμων και των ατόμων με ψυχωσικές διαταραχές. Για το σκοπό αυτό συγκέντρωσε

συγκεκριμένα μεγάλο όγκο δεδομένων φυσικής δραστηριότητας και αυτόνομης λειτουργίας από φορητές συσκευές. Η στατιστική ανάλυση μεταξύ των δύο ομάδων, μέσω των περιγραφικών στατιστικών τους έδειξε σημαντικές διαφορές όσον αφορά τη συμπεριφορά της κίνησης, καθώς και σε ορισμένους δείκτες της καρδιακής λειτουργίας τόσο κατά τη διάρκεια της εγρήγορσης όσο και του ύπνου. Ενδεικτικά για την Sample Entropy που θα αναφέρουμε στη συνέχεια, στοιχεία που αφορούν το γκρουπ των υγείων ατόμων και των ασθενών κατά τη διάρκεια της φυσικής δραστηριότητας είναι 1.36(0.26) και 1.081(0.389) με p-value=0.02 και κατά τη διάρκεια του ύπνου 1.233(0.217) και 1.267(0.172) με p-value=0.26.

Σύμφωνα με σχετική έρευνα [1] η οποία προέβλεψε υψηλότερο ποσοστό ανωμαλιών σε ασθενείς που ανήκουν στην ομπρέλα της σχιζοφρένειας εντός των 30 ημερών υποτροπής, αποκαλύπτει πρότυπο συμπεριφορών σε ατομικό επίπεδο που αλλάζουν πριν από την υποτροπή. Τα ωριαία χαρακτηριστικά που προέρχονται από δεδομένα παθητικής ανίχνευσης smartphone εξήχθησαν από 60 ασθενείς με σχιζοφρένεια (42 μη υποτροπιάζουσες περιπτώσεις και 18 περιπτώσεις που έχουν υποτροπιάσει τουλάχιστον μια φορά σε όλη τη μελέτη) που χρησιμοποιήθηκαν για την εκπαίδευση μοντέλων και τη δοκιμή απόδοσης. Εκπαιδεύτηκαν 2 τύποι μοντέλων νευρωνικών δικτύων κωδικοποιητή-αποκωδικοποιητή και ένα μοντέλο cluster-based local outliers factor (CBLOF) για την πρόβλεψη ανωμαλιών συμπεριφοράς που εμφανίστηκαν εντός της περιόδου των 30 ημερών πριν από την ημερομηνία υποτροπής ενός συμμετέχοντα. Τα μοντέλα εκπαιδεύτηκαν να αναδημιουργούν τη συμπεριφορά των συμμετεχόντων σε ημέρες σχετικής υγείας (εκτός της κοντινής χρονικής περιόδου όπου εμφανίζουν την υποτροπή), μετά την οποία εφαρμόστηκε ένα κατώφλι στο σφάλμα αναπαράστασης για την πρόβλεψη ανωμαλιών. Η αρχιτεκτονική του μοντέλου νευρωνικού δικτύου και το ποσοστό των δεδομένων συμμετεχόντων στην υποτροπή που χρησιμοποιήθηκαν για την εκπαίδευση όλων των μοντέλων διέφεραν. Αναλύθηκαν συνολικά 20.137 ημέρες δεδομένων που συλλέχθηκαν, με δεδομένα 726 ημερών (0,037%) εντός οποιασδήποτε περιόδου σε κοντινή ημερομηνία υποτροπής 30 ημερών. Το μοντέλο με την καλύτερη απόδοση χρησιμοποίησε μια πλήρως συνδεδεμένη αρχιτεκτονική αυτοκωδικοποιητή νευρωνικού δικτύου και πέτυχε μέση ευαισθησία 0,25 (IQR 0,15-1,00) και ειδικότητα 0,88 (IQR 0,14-0,96, μέση αύξηση 108% στις ανωμαλίες συμπεριφοράς κοντά σε υποτροπή). Πραγματοποιήθηκε μια εκ των υστέρων ανάλυση χρησιμοποιώντας το μοντέλο με την καλύτερη απόδοση για να προσδιοριστούν χαρακτηριστικά συμπεριφοράς που είχαν μέτρια προς μεγάλη επίδραση (Cohen $d > 0,5$) στη διάκριση ανωμαλιών κοντά σε υποτροπή από DRH μεταξύ 4 συμμετεχόντων που υποτροπίασαν πολλές φορές σε όλη τη μελέτη. Η ποιοτική επικύρωση χρησιμοποιώντας κλινικές σημειώσεις που συλλέχθηκαν κατά τη διάρκεια της αρχικής μελέτης έδειξε ότι τα αναγνωρισμένα χαρακτηριστικά από την ανάλυσή μας παρουσιάστηκαν στους κλινικούς ιατρούς κατά

τη διάρκεια συμβάντων υποτροπής.

1.3 Εξόρυξη δεδομένων μεγάλης κλίμακας

Η ανίχνευση ανωμαλιών σε διάφορα πρότυπα (patterns) είναι ο κλάδος της εξόρυξης δεδομένων που ειδικεύεται στην πρόβλεψη ιδιομορφιών σε σπάνια περιστατικά. Προσεγγιστικά στην ανίχνευση ανωμαλιών στο τομέα των χρονοσειρών σχετίζονται με την πρόβλεψη και την αξιολόγηση στατιστικών μέτρων και μοντέλων που βασίζονται στην πρόβλεψη αλλαγών στη κατανομή της χρονοσειράς. Το να προβλέψεις την ανθρώπινη συμπεριφορά είναι ένα εξαιρετικά δύσκολο πρόβλημα, πόσο μάλλον σε ασθενείς με σχιζοφρένεια που δεν ακολουθούν τους κινητικούς ρυθμούς που παρατηρούνται σε ένα υγιές πληθυσμό. Αλγόριθμοι σχεδιασμένοι να εκπαιδεύονται σε πολύπλοκα χαρακτηριστικά εντός δεδομένων χρονοσειράς είναι πιθανό να έχουν μεγαλύτερη επιτυχία στην εύρεση ανωμαλιών και στον εντοπισμό συμπεριφορών που σχετίζονται με υποτροπή. Την τελευταία δεκαετία έχουν γίνει τρομακτικές βελτιώσεις στη τεχνολογία ανίχνευσης κίνησης με τη χρήση φορητών συσκευών ή κινητών τηλεφώνων νεάς τεχνολογίας που δημιουργούν νέες ευκαιρίες για εξ αποστάσεως μετρήσεις της ψυχικής υγείας και ασθένειας.

Προσεγγίσεις στην ανίχνευση ανωμαλιών γίνονται από διάφορα μοντέλα σε δεδομένα πολυμεταβλητών χρονοσειρών. Αυτοί οι αλγόριθμοι ταξινόμησης έχουν επιτυχία στην εκμάθηση πολύπλοκων χαρακτηριστικών, ειδικά σε υψηλά ακανόνιστα δεδομένα ανίχνευσης. Σε αντίθεση με τις στατιστικές προσεγγίσεις, οι αλγόριθμοι ταξινόμησης δεν απαιτούν θεωρήσεις σχετικά με την ομοιομορφία της κατανομής, με αποτέλεσμα να παρέχουν ακριβείς προβλέψεις χωρίς την ανάγκη περίπλοκων χαρακτηριστικών. Χαρακτηρίζονται ως μοντέλα μαύρου κουτιού. Το σύστημα ανίχνευσης ανωμαλιών στη συμπεριφορική υγεία, χρειάζεται όχι μόνο να παράγει ακριβή αποτελέσματα στην ανίχνευση αλλά και να βρει τα αίτια που προκαλούν την συγκεκριμένη συμπεριφορά ώστε να οδηγήσει στη κλινική μετάφρασή της. [1]

1.3.1 Μεγάλα Δεδομένα και Σχιζοφρένεια

Ερευνητές έχουν αρχίσει να χρησιμοποιούν την τεχνική ανίχνευσης ανωμαλιών για να προβλέψουν πρώιμα προειδοποιητικά σημάδια της ψύχωσης. Μια πιλοτική μελέτη που αξιολόγησε το συνδυασμό χαρακτηριστικών ανίχνευσης κινητών και self-report ερωτηματολογίου μπόρεσε να εντοπίσει μια αύξηση στις ανωμαλίες εντός 2 εβδομάδων υποτροπής σε μικρό πληθυσμό ασθενών [6]. Άλλη πρόσφατη μελέτη χρησιμοποίησε αναδρομικά χαρακτηριστικά που εξήχθησαν από το Facebook και δημιούργησε έναν ταξινομητή για τον εντοπισμό στην περίοδο 1 ενός μήνα πριν την υποτροπή και στη

συνέχεια ανέλυσε τις συμπεριφορές που ήταν σημαντικά διαφορετικές εντός αυτής της περιόδου. Ανάπτυξη ενός αλγόριθμου για την πρόβλεψη συγκεκριμένων ημερών έξαρσης των συμπτωμάτων πριν από την υποτροπή χρησιμοποιώντας αποκλειστικά δεδομένα παθητικής ανίχνευσης θα μπορούσε να παρέχει στους γιατρούς μια διακριτική μέθοδο μέτρησης των SSD συμπτωμάτων χωρίς την ανάγκη αυτοαναφοράς του ασθενούς. Τα άτομα με σχιζοφρένεια έχουν μειωμένο προσδόκιμο ζωής σε σύγκριση με τον γενικό πληθυσμό, κυρίως λόγω αυξημένου κινδύνου καρδιαγγειακών παθήσεων. [1]

2.1 Διακύμανση Καρδιακού Πάλμου

Η διακύμανση του καρδιακού ρυθμού (Heart Rate Variability-HRV) αναλύει τις μεταβολές στον καρδιακό ρυθμό. Η διαφορά μεταξύ καρδιακού παλμού και HRV είναι ότι ο καρδιακός ρυθμός, δείχνει πόσες φορές χτυπά η καρδιά σε ένα λεπτό, μετρούμενοι ως παλμοί ανά λεπτό (bpm), ενώ το HRV δείχνει το χρονικό διάστημα μεταξύ δύο καρδιακών παλμών, μετρημένο σε χιλιοστά του δευτερολέπτου (ms). Οι ταλαντώσεις μιας υγιούς καρδιάς είναι περίπλοκες και μεταβάλλονται έτσι ώστε να επιτρέπουν στο καρδιαγγειακό σύστημα να προσαρμοστεί γρήγορα, σε ξαφνικές σωματικές και ψυχολογικές προκλήσεις λόγω ομοιόστασης. Η HRV προσδιορίζεται από τη νευροκαρδιακή λειτουργία και γενικεύεται από τις αλληλεπιδράσεις καρδιάς-εγκεφάλου και τις δυναμικές μη γραμμικές διεργασίες του αυτόνομου νευρικού συστήματος (ΑΝΣ). Είναι μια αναδυόμενη ιδιότητα αλληλοεξαρτώμενων ρυθμιστικών συστημάτων που λειτουργούν σε διαφορετικές χρονικές κλίμακες για να μας βοηθήσει να προσαρμοστούμε στις περιβαλλοντικές και ψυχολογικές προκλήσεις. Αντανάκλα τη ρύθμιση της αυτόνομης ισορροπίας, της αρτηριακής πίεσης, της ανταλλαγής αερίων, του εντέρου, της καρδιάς και του αγγειακού τόνου, που αναφέρονται στην διάμετρο των αιμοφόρων αγγείων και ρυθμίζουν την αρτηριακή πίεση. Συνεχείς μεταβολές στο συμπαθητικό νευρικό σύστημα (ΣΝΣ) και στο παρασυμπαθητικό νευρικό σύστημα (ΠΣΝΣ) έχουν ως αποτέλεσμα περιοδικές μεταβολές στον καρδιακό ρυθμό γύρω από μία μέση τιμή. Κάνοντας μετρήσεις στην HRV , μπορούν να εξαχθούν συμπεράσματα

σχετικά με τη λειτουργία του ΣΝΣ και του ΠΣΝΣ. [54] Όταν η HRV είναι χαμηλή, σημαίνει ότι το σώμα αναρρώνει ή ανταποκρίνεται στο στρες. Το άγχος μπορεί να είναι είτε σωματικό (για παράδειγμα όταν αναρρώνουμε από μια καλή προπόνηση ή καταπολεμούμε έναν ιό) είτε φυσιολογικό. Όσο πιο ισορροπημένο είναι το ΑΝΣ, τόσο μεγαλύτερη μεταβλητότητα υπάρχει στον καρδιακό ρυθμό καθώς τα δύο συστήματα εναλλάσσονται ενεργά. Η HRV υψηλότερης συχνότητας συσχετίζεται με:

- Αυξημένο επίπεδο φυσικής κατάστασης
- Καλύτερη υγεία
- Καλύτερη ανθεκτικότητα
- Ήρεμα, θετικά συναισθήματα

Ενώ η HRV χαμηλότερης συχνότητας συνδέεται με:

- Μειωμένο επίπεδο φυσικής κατάστασης
- Κακή υγεία
- Αυξημένο άγχος και κατάθλιψη
- Αρνητικά συναισθήματα

2.1.1 ΣΝΣ-ΠΣΝΣ

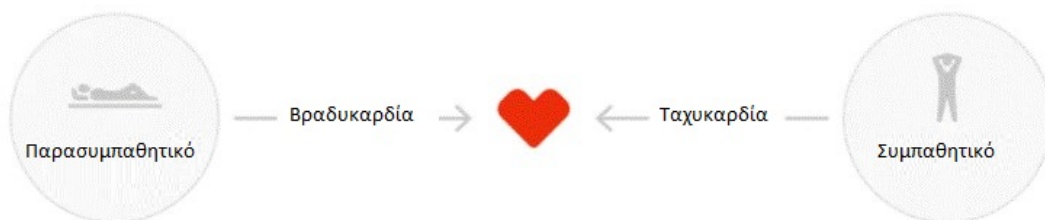
Το νευρικό σύστημα αποτελείται από το κεντρικό νευρικό σύστημα, που περιλαμβάνει τον εγκέφαλο και νωτιαίο μυελό και το περιφερικό νευρικό σύστημα, που περιλαμβάνει το εκούσιο και αυτόνομο νευρικό σύστημα. Το εκούσιο αποτελείται από το αισθητικό και κινητικό νευρικό σύστημα και το αυτόνομο από το συμπαθητικό και παρασυμπαθητικό νευρικό σύστημα. Η δράση του συμπαθητικού νευρικού συστήματος είναι διεγερτική, φέρνει το σώμα σε εγρήγορση και του δίνει την δυνατότητα για ανάληψη έντονης δραστηριότητας, προετοιμάζει τον οργανισμό να αντιδράσει στις διάφορες καταστάσεις που αντιμετωπίζει με ΠΑΛΗ ή ΦΥΓΗ. Το παρασυμπαθητικό νευρικό σύστημα πάλι καταστέλλει τις λειτουργίες για να επανέλθει ισορροπία μετά το τέλος της ΠΑΛΗΣ ή ΦΥΓΗΣ.

Όταν υπάρχει έντονο άγχος, τα επινεφρίδια εκκρίνουν διάφορες ορμόνες που επηρεάζουν άμεσα το συμπαθητικό νευρικό σύστημα και κατ' επέκταση διάφορες λειτουργίες όπως:

1) αύξηση του καρδιακού παλμού, 2) διάταση των βρογχιολίων, 3) αύξηση παροχής αίματος, 4) σύσπαση των αγγείων του δέρματος, 5) διαστολή των αγγείων των μυών, 6) διαστολή της κόρης του ματιού, 7) αναστολή των κινήσεων του εντέρου, 8) αναστολή στύσης και οργασμού, 9) ανόρθωση τριχών. Ελέγχει την τροφοδοσία των μυών με αίμα. Επίσης η πέψη δεν γίνεται σωστά και επόμενο είναι να μην απορροφούνται τα απαραίτητα θρεπτικά συστατικά που έχει ανάγκη ο οργανισμός μας.

Το παρασυμπαθητικό νευρικό σύστημα επηρεάζει τις ίδιες ακριβώς λειτουργίες με αντίθετο ακριβώς τρόπο, δηλαδή λειτουργεί κατασταλτικά στα διάφορα συστήματα που διεγείρονται για διάφορες αιτίες από το συμπαθητικό νευρικό σύστημα. Επίσης διεγείρει και ενεργοποιεί το αναπαραγωγικό, το ανοσοποιητικό και πεπτικό μας σύστημα, ελέγχει ακόμα και τη λειτουργία της ούρησης. Η δυσλειτουργία του αυτόνομου νευρικού συστήματος έχει τη δυνατότητα να επηρεάσει δυσμενώς την υγεία και είναι γνωστό ότι υπάρχει στις ψυχιατρικές διαταραχές. Αντανακλά αλλαγές στη λειτουργία πολλών περιοχών του κεντρικού νευρικού συστήματος. Η διακύμανση του καρδιακού ρυθμού παρέχει ένα εργαλείο για τη μελέτη της αυτόνομης λειτουργίας. Σε μια υγιή ανθρώπινη καρδιά, υπάρχει μια δυναμική σχέση μεταξύ του ΠΣΝΣ και του ΣΝΣ. Ο έλεγχος ΠΣΝΣ κυριαρχεί σε κατάσταση ηρεμίας, με αποτέλεσμα μέσο καρδιακό παλμό 75 bpm. Το ΠΣΝΣ μπορεί να επιβραδύνει την καρδιά στους 20 ή 30 bpm ή να τη σταματήσει για λίγο.

Τα παρασυμπαθητικά νεύρα λειτουργούν πιο γρήγορα (<1 δευτ.) από τα συμπαθητικά νεύρα (>5 δευτ.). Έτσι προκαλούνται αντιφατικές ενέργειες, όπως η επιτάχυνση και η επιβράδυνση της καρδιάς και η επίδρασή σε ένα όργανο εξαρτάται από την τρέχουσα ισορροπία δραστηριότητάς τους. Ενώ το ΣΝΣ μπορεί να καταστείλει τη δραστηριότητα του ΠΣΝΣ, μπορεί επίσης να αυξήσει την αντιδραστικότητα του ΠΣΝΣ. Παρασυμπαθητική ανάκαμψη μπορεί να συμβεί μετά από υψηλά επίπεδα στρες, με αποτέλεσμα αυξημένη γαστρική δραστηριότητα τη νύχτα και συμπτώματα άσθματος, όπως περιγράφεται από την εικόνα 2.1.



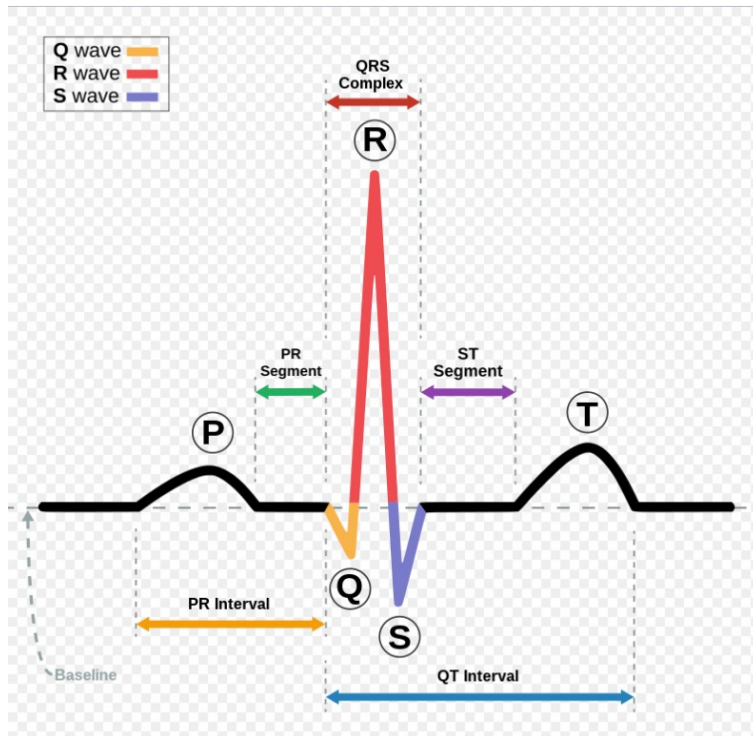
Σχήμα 2.1: Δραστηριότητα ΚΝΣ

Οι δυναμικές παράμετροι της HRV στο ψυχοφυσιολογικό στρες είναι ευαίσθητες και πρακτικές για τη διάγνωση της σχιζοφρένειας. [32] Πρόσφατο άρθρο έρευνας σχετικά με την ανάλυση HRV περιγράφει τεχνικές ανάλυσης που χρησιμοποιούνται στην ψυχοφυσιολογία ως δείκτης μέτρησης άγχους. [14] Καθιερωμένα κλινικά δεδομένα με βάση παγκόσμιων μελετών που δημοσιεύθηκαν κατά τη διάρκεια της τελευταίας δεκαετίας η μειωμένη μέτρηση HRV είναι ισχυρός προγνωστικός παράγοντας αυξημένης καρδιακής θνησιμότητας ή/και αρρυθμίας. [53] Υψηλές μετρήσεις δεν είναι πάντα καλύτερες, αφού προέρχονται από ανωμαλίες καρδιακής αγωγιμότητας καθώς οι παθολογικές συνθήκες μπορούν να παράγουν αυτές τις μετρήσεις. Η προσεκτική εξέταση της μορφολογίας του ηλεκτροκαρδιογραφήματος (ΗΚΓ) μπορεί να αποκαλύψει εάν οι αυξημένες τιμές οφείλονται σε προβλήματα όπως κολπική μαρμαρυγή. [52]

2.1.2 Ηλεκτροκαρδιογράφημα

Η καρδιακή δραστηριότητα σχετίζεται με τη ροή ιοντικού ρεύματος που προκαλείται από τη συστολή των καρδιακών ινών και στη συνέχεια τη χαλάρωση. Αυτή η ηλεκτρική δραστηριότητα μπορεί να καταγραφεί χρησιμοποιώντας ηλεκτρόδια κατάλληλα τοποθετημένα στην επιφάνεια του σώματος. Η διαφορά δυναμικού που μετράται μεταξύ αυτών των ηλεκτροδίων είναι γνωστή ως σήμα ηλεκτροκαρδιογραφήματος (ΗΚΓ). Η σύσπαση και η χαλάρωση της καρδιάς μπορούν να παρατηρηθούν ως διαδοχικές κυματομορφές στο σήμα ΗΚΓ που υποδηλώνουν κύματα P, QRS και T. Το φυσιολογικό ηλεκτροκαρδιογράφημα περιλαμβάνει ένα έπαρμα P, ένα σύμπλεγμα με όνομα "QRS", ένα έπαρμα T και ένα έπαρμα U. Στο ηλεκτροκαρδιογράφημα μπορούμε να χωρίσουμε αυτές τις καταγραφές ως επάρματα εκπόλωσης και επάρματα επαναπόλωσης. Επίσης υπάρχουν και κάποιοι χώροι μεταξύ αυτών των παρατηρήσεων, αυτοί οι χώροι ονομάζονται τμήματα, και είναι, το τμήμα P-Q και το τμήμα S-T. Τέλος, συναντάμε κάποιες αθροιστικές παρατηρήσεις όπου τις ονομάζουμε διαστήματα, αυτές οι παρατηρήσεις είναι το διάστημα P-Q το οποίο μπορούμε να το συναντήσουμε και ως διάστημα Q-R το διάστημα Q-T και το διάστημα R-R. [18]. Όπως φαίνεται και στην εικόνα 2.2.

- Έπαρμα P : Προκαλούνται από τα ηλεκτρικά δυναμικά που παράγονται όταν πραγματοποιείτε η εκπόλωση των κόλπων πριν γίνει η συστολή. Ουσιαστικά είναι αντιπροσωπευτικό ότι οι κόλποι θα συσπαστούν.
- Τμήμα P-Q: Είναι απλά μια ισοηλεκτρική γραμμή, δηλαδή έχει 0 mV
- Διάστημα P-Q : Η απόσταση της αρχής του επάρματος P ως την αρχή του Q είναι το διάστημα όπου οι κόλποι και οι κοιλίες διεγείρονται, υπό φυσιολογικές



Σχήμα 2.2: QRS

source by: <https://en.wikipedia.org/wiki/User:Agateller>

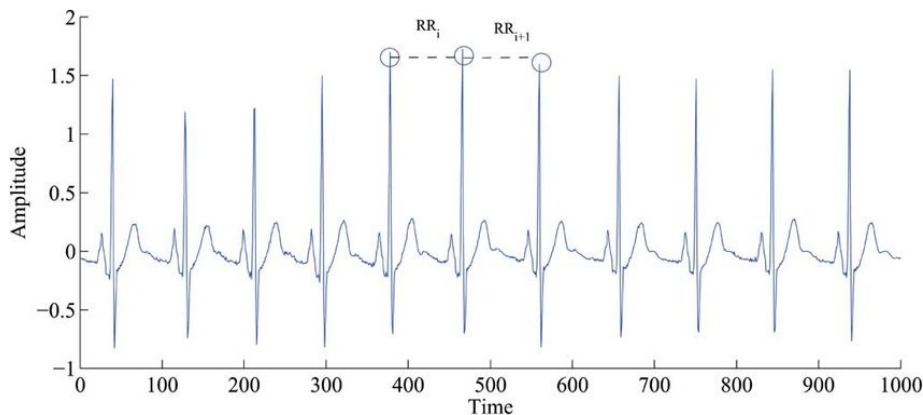
συνθήκες διαρκεί περίπου 0,16 sec.Επειδη είναι πιθανόν σε κάποιες περιπτώσεις να έχουμε απώλεια της κορυφής Q τότε το αναφέρουμε ως "Διάστημα P-R" και θεωρούμε σημείο τέλους την κορυφή R

- Σύμπλεγμα QRS: Ομοίως και αυτό το σύμπλεγμα προκαλείτε από τα ηλεκτρικά δυναμικά που παράγονται όταν πραγματοποιείτε η εκπόλωση των κόλπων πριν γίνει η συστολή, έτσι θεωρούνται και τα δύο επάρματα εκπόλωσης.
- Διάστημα P-T: Αντιπροσωπεύει την συστολή των κοιλιών, διαρκεί υπό φυσιολογικές συνθήκες 0,35 sec
- Τμήμα S-T :Είναι απλά μια ισοηλεκτρική γραμμή, δηλαδή έχει 0 mV
- Έπαρμα T : Προκύπτει από τα δυναμικά όταν οι κοιλίες επιστρέφουν σε κατάσταση πόλωσης, έτσι θεωρείτε έπαρμα επαναπόλωσης. Η επαναπόλωση των κόλπων γίνεται 0,15 με 0,20 sec μετά το έπαρμα P ,δηλαδή αρκετά συχνά συμπίπτει με

το σύμπλεγμα QRS και ως εκ τούτου θεωρείτε πολύ σπάνιο να είναι εμφανής στο ηλεκτροκαρδιογράφημα.

- Έπαρμα U :Είναι εξαιρετικά σπάνιο να παρατηρηθεί, προκαλείτε από την επαναπόλωση του μεσοκοιλιακού διαφράγματος.
- Διάστημα R-R : Είναι το διάστημα που μεσολαβεί από την μία κορυφή R στην ακριβώς επόμενη , ο μεταξύ τους χώρος μπορεί να μας δείξει τον καρδιακό ρυθμό, έτσι αν έχουν απόσταση 1 sec τότε ξέρουμε πως $60/1 = 60$ παλμοί το λεπτό , το σύνηθες διάστημα στον μέσο ενήλικα είναι 0,83 sec δηλαδή $60/0,83 = 72$ παλμοί το λεπτό [18]

Η χρήση του συμπλέγματος QRS και συγκεκριμένα του κύματος R γίνεται για να προσδιοριστεί η χρονική διάρκεια του καρδιακού παλμού επειδή είναι το πιο εμφανές στοιχείο στο ΗΚΓ και υπό κανονικές συνθήκες το μέγιστο σημείο κάθε παλμού. Έτσι η διαδικασία της μέτρησης αποτελείται από δύο στάδια: (α) μέτρηση του ηλεκτροκαρδιογραφήματος και (β) ανίχνευση των κυμάτων R [54] και εύρεση της απόστασης μεταξύ διαδοχικών RR(διάστημ RR). Το σήμα που προκύπτει από αυτή τη διαδικασία είναι γνωστό και ως Σήμα Διαστημάτων R (RR interval signal). Η δυναμική των καρδιακών παλμών μπορεί να αναπαρασταθεί με συνέπεια σε χάρτη απεικόνισης των διαστημάτων RR που αναπαρίσταται από την εικόνα 2.3 . Ο όρος NN



Σχήμα 2.3: Χάρτης Απεικόνισης Διαστημάτων RR [60]

χρησιμοποιείται στη θέση του RR για να δώσει έμφαση στο ότι τα σήματα των παλμών είναι επεξεργασμένα ως φυσιολογικά σήματα. Αυτό σημαίνει ότι μη φυσιολογικοί ρυθμοί, όπως οι έκτοποι ρυθμοί (καρδιακοί παλμοί που προέρχονται από τον σινοατριακό κόμβο του δεξιού αιθρίου) ή κάποια παρεμβολή που επικάλυψε πλήρως την καταγραφή, έχουν αφαιρεθεί.

2.2 Η Χρονοσειρά

Μια χρονοσειρά, ως σύνολο N διαδοχικών σημείων δεδομένων, ορίζεται ως

$$X_m = x(i), i = 1, \dots, N$$

Όπου, σε αυτήν την περίπτωση, κάθε σημείο δεδομένων θα αντιπροσωπεύει την τιμή ενός διαστήματος RR (ή NN). Από τον αρχικό χρόνο της σειράς X , ας ορίσουμε τα διανύσματα $m(i)$ ως εξής:

$$m(i) = (x(i), x(i + \tau), x(i + 2\tau), \dots, x(i + (m-1)\tau)), i = 1, \dots, K, K = [N - (m-1)\tau]$$

και m να είναι η διάσταση ενσωμάτωσης ενώ τ είναι η ενσωμάτωση καθυστέρησης. Η επιλογή των κατάλληλων παραμέτρων ενσωμάτωσης είναι κρίσιμη.

Διάσταση ενσωμάτωσης δ : Η διάσταση ενσωμάτωσης είναι η διάσταση του χώρου κατάστασης που χρησιμοποιείται για την ανακατασκευή. Σε αντίθεση με τη χρονική καθυστέρηση τ , η σημασία της διάστασης ενσωμάτωσης γίνεται αποδεκτή ομόφωνα. Μια πολύ μεγάλη διάσταση ενσωμάτωσης θα έχει ως αποτέλεσμα μεγάλους χρόνους υπολογισμού και υπερβολικό αριθμό σημείων δεδομένων.[19]

2.3 Κατηγορίες Επεξεργασίας HRV

Οι σύνηθες μετρήσεις που πραγματοποιούνται για την ανάλυση του HRV περιλαμβάνουν

- μετρήσεις στην περιοχή του χρόνου,
- στην περιοχή των συχνοτήτων,
- ανάλυση χρόνου-συχνότητας,
- μη γραμμική και χαοτική ανάλυση και
- ανάλυση με ευφυείς τεχνικές.

Τα φασματικά μέτρα δεν είναι κατάλληλα για χρονικές ακολουθίες για τιμές που λείπουν.[60] Οι μετρήσεις σε χρονικό πεδίο μπορούν να εφαρμοστούν σε διαστήματα RR ή ηλεκτροκαρδιογράφηματα, οποιασδήποτε χρονικής διάρκειας, τόσο βραχυπρόθεσμο, δηλαδή 5 ή 30 λεπτών, όσο και μακροπρόθεσμο, δηλαδή 24 ωρών. Ωστόσο, είναι συχνότερη η χρήση 24ωρων ηλεκτροκαρδιογραφημάτων είτε ενιαία είτε τμηματικά, όπου αυτό είναι απαραίτητο [34][Ιατρική_ Πληροφορική].

2.4 Στατιστική Ανάλυση σε Χρονικό Πεδίο

Η στατιστική ανάλυση στο πεδίο του χρόνου πραγματοποιείται για υπολογισμούς και ανεύρεση δεικτών.

Στατιστικός δείκτης	Μονάδα μέτρησης	Περιγραφή
SDNN	msec	Τυπική απόκλιση όλων των διαστημάτων RR
SDANN	msec	Τυπική απόκλιση όλων των διαστημάτων RR ανά 5 λεπτά σε εικοσιτετράωρη εγγραφή
RMSSD		Η τετραγωνική ρίζα του μέσου των τετραγώνων των διαφορών των κανονικών διαστημάτων RR
SDSD	msec	Τυπική απόκλιση των διαφορών όλων των κανονικών διαστημάτων RR σε ultra short διαστήματα που αφορούν έως 5 λεπτά
pNNx	%	Ποσοστό του αριθμού των περιπτώσεων όπου η διαφορά μεταξύ δυο διαδοχικών RR διαστημάτων είναι μεγαλύτερη από x msec σε σχέση με όλα τα κανονικά διαστήματα RR. Οι πιο συνηθισμένες χρήσεις είναι του δείκτη pnn50
HR Max- HR Min		Η διαφορά του μέγιστου καρδιακού παλμού από τον ελάχιστο.
Mean HR		Μέσος όρος καρδιακού παλμού
Max HR		Μέγιστος καρδιακός παλμός
Min HR		Ελάχιστος καρδιακός παλμός

Πίνακας 2.1: Πίνακας Στατιστικών Δεικτών

2.4.1 SDNN

Ο SDNN θεωρείται ένας από τους καλύτερους δείκτες για την ιατρική διαστρωμάτωση του καρδιακού κινδύνου. Με βάση την 24ωρη παρακολούθηση, οι ασθενείς με τιμές

- κάτω των 50 ms ταξινομούνται ως ασθενείς,
- με 50–100 ms διατρέχουν κίνδυνο στην υγεία και
- πάνω από 100 ms είναι υγιείς

Σε σχετικές έρευνες ο SDNN χρησιμοποιήθηκε ως βασικός δείκτης για τη θνησιμότητα σε ασθενείς μετά από έμφραγμα του μυοκαρδίου. Ο SDNN υπολογίζεται από τον παρακάτω μαθηματικό τύπο:

$$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (RR_i - \overline{RR})^2}$$

Όσο περισσότερο αυξάνεται αυτός ο δείκτης τόσο αυξάνεται η προσαρμοστικότητα του αυτόνομου νευρικού συστήματος. Όσο μειώνεται, τότε υπάρχει περιορισμένη αυτόνομη ρύθμιση. Άρα υψηλό SDNN, σημαίνει μεγαλύτερη παρασυμπαθητική δραστηριότητα.

2.4.2 SDANN

Ο SDANN αφορά την τμηματοποίηση σε διαστήματα των 5 λεπτών. Η χρονοσειρά τμηματοποιείται σε περιόδους διάρκειας έως 5 λεπτών και μετά υπολογίζεται η παράμετρος του κάθε τμήματος. Τέλος, υπολογίζεται η μέση τιμή όλων των υπολογισμένων τιμών. Μετράται σε ms και υπολογίζεται όπως ο SDNN. Παρέχει μακροπρόθεσμες πληροφορίες. Είναι ένας ευαίσθητος δείκτης χαμηλών συχνοτήτων όπως οι αλλαγές στη θέση και ο κικράδιος ρυθμός.

2.4.3 RMSSD

Ο RMSSD αφορά την μέση τιμή του τετραγώνου της ρίζας των διαδοχικών διαφορών μεταξύ των κανονικών καρδιακών παλμών που προκύπτει υπολογίζοντας πρώτα κάθε διαδοχική χρονική διαφορά μεταξύ των καρδιακών παλμών σε ms. Στη συνέχεια, καθεμία από τις τιμές τετραγωνίζεται και υπολογίζεται ο μέσος όρος του αποτελέσματος πριν ληφθεί η τετραγωνική ρίζα του συνόλου. Είναι πανομοιότυπος με το μη γραμμικό μετρικό SD1, που θα οριστεί στη συνέχεια, το οποίο ανταναχλά το βραχυπρόθεσμο HRV. Επηρεάζεται λιγότερο από την αναπνοή σε διάφορες διεργασίες [29]. Ο δείκτης

RMSSD επηρεάζεται περισσότερο από το ΠΣΝΣ παρά το SDNN. Υπολογίζεται ως εξής:

$$\sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} (RR_{i+1} - RR_i)^2}$$

Αυτό το μέτρο αντανακλά την ακεραιότητα του αυτόνομου ελέγχου της καρδιάς με τη μεσολάβηση του πνευμονογαστρικού νεύρου. Χαμηλές τιμές σχετίζονται με ασθένεια.

2.4.4 HR Max- HR Min

Η μέση διαφορά μεταξύ του υψηλότερου και του χαμηλότερου HR κατά τη διάρκεια κάθε αναπνευστικού κύκλου είναι ιδιαίτερα ευαίσθητη στις επιπτώσεις του ρυθμού αναπνοής, ανεξάρτητα από την κυκλοφορία του πνευμονογαστρικού νεύρου. Η μέση διαφορά μεταξύ του υψηλότερου και του χαμηλότερου HR κατά τη διάρκεια κάθε αναπνευστικού κύκλου (HR Max – HR Min) επηρεάζεται εύκολα από τις επιδράσεις του ρυθμού αναπνοής. . Απαιτείται τουλάχιστον ένα δείγμα 2 λεπτών για τον υπολογισμό HR Max – HR Min. Τιμές χαμηλότερες από 3% υποδηλώνουν υψηλό κίνδυνο ασθένειας.

2.4.5 NNx

Εκφράζει τον αριθμό των έγκυρων παρακείμενων τιμών NN που δεν διαχωρίζονται με διαλείμματα δεδομένων ή εσφαλμένα σημάδια δεδομένων που διαφέρουν κατά περισσότερο από x χιλιοστά του δευτερολέπτου.

2.4.6 SDDSD

Είναι η τυπική απόκλιση των διαδοχικών διαφορών μεταξύ γειτονικών διαστημάτων NN.

$$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (DNNJ_i - \overline{DNNI})^2}$$

Οι pNN50 και RMSSD είναι οι πιο κοινές παράμετροι που βασίζονται στις διαφορές των μεσοδιαστημάτων NN.

Οι μετρήσεις στο πεδίο των συχνοτήτων γίνεται με ανάλυση Fourier και υπολογίζεται η απόλυτη ή σχετική ποσότητα ενέργειας σήματος. Η έλλειψη όμως λεπτομέρειας στη στατιστική ανάλυση και η μη στατικότητα του σήματος σε σχέση με την φασματική ανάλυση, οδήγησε γρήγορα στην αναζήτηση νέων, πιο σύνθετων, μεθόδων που χρησιμοποιούνται επιλεκτικά ή συμπληρωματικά. Γενικά οι στατιστικές μετρήσεις

στην περιοχή του χρόνου παρέχουν περιορισμένες πληροφορίες σε σχέση με την κατάσταση υγείας του ατόμου, γι' αυτό και συνήθως λαμβάνονται υπόψιν συμπληρωματικά, γεωμετρικές μετρήσεις, μη γραμμικές μέθοδοι και η χαοτική ανάλυση. Χαρακτηριστικό των μετρήσεων είναι ότι έχουν μεγάλη ανοχή σε λάθη στην εξαγωγή χαρακτηριστικών, όπως απώλειες παλμών και λάθη κατά την αναγνώριση του QRS, με αποτέλεσμα να μπορούν να εφαρμοστούν και σε "θορυβώδες" σήμα, όπως το σήμα του dataset που έχουμε χρησιμοποιήσει, κάτι που δεν ισχύει για άλλα είδη ανάλυσης. Έτσι η γεωμετρική αποτίμηση αποτελεί μέθοδος για την οπτικοποίηση των δεδομένων. Μη γραμμικές μέθοδοι, ποσοτικοποιούν την πολυπλοκότητα ανάμεσα στα διαστήματα RR, με αποτέλεσμα να αναλύουν και να εξάγουν τις μη γραμμικές δομές του σήματος. Οι ευφυείς τεχνικές ανάλυσης είναι από τις πιο σύγχρονες τάσεις ανάλυσης και επεξεργασίας του βιοϊατρικού σήματος. Η ανάλυση στην περιοχή των συχνοτήτων δεν είναι κατάλληλη μέθοδος όταν υπάρχουν εκλειπόμενες τιμές, όπως στην παρούσα μελέτη [60]. Οι τιμές στις στατιστικές μετρήσεις εξαρτώνται από την ηλικία, το φύλο, το εύρος των δεδομένων, τη σωματική κατάσταση, τη δραστηριότητα, το κύκλο του ύπνου, την επίρεια των φαρμάκων ή κάποια ασθένεια.

2.5 Γεωμετρική απεικόνιση

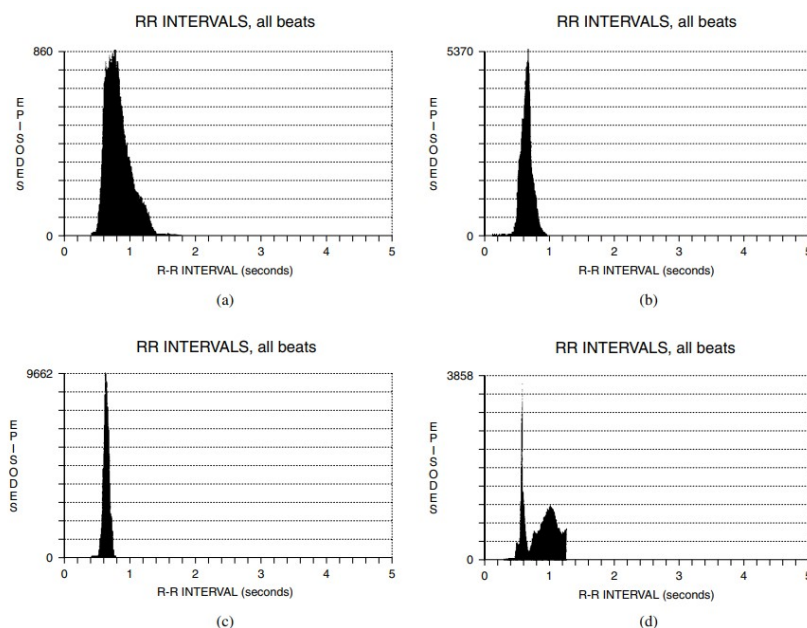
2.5.1 Τριγωνοποίηση Ιστογράμματος

Στη γεωμετρική απεικόνιση μπορεί να εφαρμοστεί μία μέθοδος που ονομάζεται «τριγωνοποίηση». Προσεγγίζει την κατανομή των διαστημάτων RR με μία γραμμική συνάρτηση παρεμβολής χρησιμοποιώντας ως κορυφή του τριγώνου τη μεγαλύτερη κορυφή του ιστογράμματος και πλευρές τις ευθείες που ελαχιστοποιούν το εμβαδόν μεταξύ αυτών και του ιστογράμματος, υπολογίζοντας τη βάση του τριγώνου, που χρησιμοποιείται ως δείκτης για τη HRV [55]. Όπως αποδεικνύεται παρακάτω, αυτός ο δείκτης έχει υψηλή συσχέτιση με τον SDNN στην περιοχή του χρόνου με τη βασική διαφορά ότι δεν επηρεάζεται τόσο από λάθη, όπως μετατόπιση ή απώλεια παλμών διότι αυτά μένουν έξω από το τρίγωνο υπολογισμού. Η θέση των άκρων του τριγώνου καθορίζεται με την ακόλουθη διαδικασία: Η πρώτη ακμή τοποθετείται στο σημείο $(D(X), X)$ (κορυφή του τριγώνου) με το $D(X)$ να είναι το μέγιστο του ιστογράμματος και το X να περιέχει το μέγιστο. Οι άλλες δύο ακμές είναι τοποθετημένες στα σημεία με συντεταγμένες $(N, 0)$ και $(M, 0)$, όπως φαίνεται στο σχήμα. Τέλος, τα N και M προσδιορίζονται με την εύρεση του παρεμβαλλόμενου τριγώνου με την καλύτερη προσαρμογή στο ιστογράμμο NNI χρησιμοποιώντας τη μέθοδο των ελαχίστων τετραγώνων, όπως παρουσιάζεται από τον ακόλουθο τύπο:

$$(n, N, M) = \min \sum_{N_{min}}^{M_{max}} (D(X) - q(n, N, M))^2$$

Όπου:

- $E(n)$: Σφάλμα της τριγωνικής παρεμβολής με την καλύτερη προσαρμογή στην κατανομή
- $D(X)$: NNI κατανομή
- $q(n, N, m)$: Συνάρτηση τριγωνικής παρεμβολής
- n : Κατακόρυφη στήλη
- N : Τιμή N που καθορίζει την αριστερή γωνία του παρεμβαλλόμενου τριγώνου
- N_{min} : Το ύψος της χαμηλότερης στήλης όπου $D(x) \neq 0$
- M : Τιμή M που καθορίζει τη δεξιά γωνία του παρεμβαλλόμενου τριγώνου
- M_{min} : Το ύψος της υψηλότερης κατακόρυφης στήλης, όπου $D(x) \neq 0$



Σχήμα 2.4: Πρότυπα ιστογράμματα διαστημάτων NN [51]

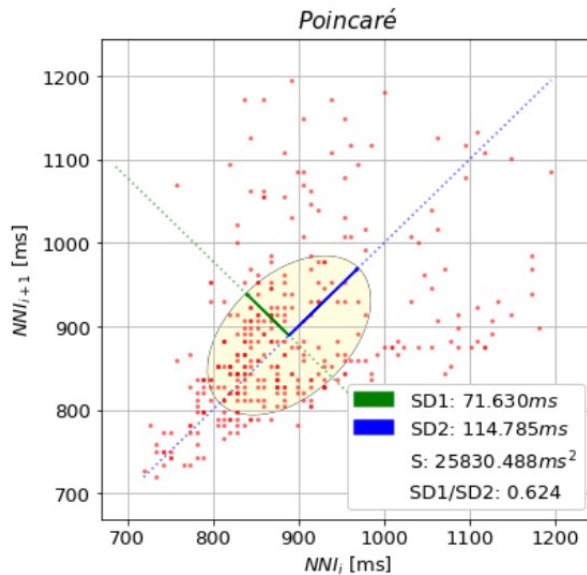
Η εικόνα 2.4 [51] αποτελεί παραδείγματα (a) για ένα φυσιολογικό περιστατικό, (b) για έναν καρδιακό ασθενή με μειωμένο HRV, (c) έναν καρδιακό ασθενή με πολύ χαμηλό HRV και (d) έναν ασθενή με εξαιρετικά μη φυσιολογική κατανομή διαστήματος RR με μεγάλο αριθμός κοιλιακών έκτοπων παλμών.

Οι δείκτες HRV Triangular Index και RMSSD μπορούν να διακρίνουν από κοινού μεταξύ φυσιολογικών καρδιακών ρυθμών και αρρυθμιών.

2.5.2 Poincare Plot

Μια γραφική παράσταση Πουανκαρέ, γίνεται σχεδιάζοντας κάθε διάστημα σε συνάρτηση με το προηγούμενο διάστημα, δημιουργώντας ένα διάγραμμα διασποράς. Η ανάλυση της γραφικής παράστασης, επιτρέπει στους ερευνητές να αναζητήσουν οπτικά μοτίβα μέσα σε μια χρονοσειρά. Στατιστικά, η γραφική παράσταση εμφανίζει τη συσχέτιση μεταξύ διαδοχικών διαστημάτων με γραφικό τρόπο. Η μη γραμμική δυναμική θεωρεί το γράφημα ως το διδιάστατο ανακατασκευασμένο διάστημα φάση-χώρος, που περιγράφει τη δυναμική του καρδιακού συστήματος. Απεικονίζεται από μια έλλειψη, στην οποία ο μεγάλος άξονας SD2 αντανακλά το επίπεδο της βραχύπροθεσμης μεταβλητότητας και η διασπορά των σημείων στο μικρο άξονα SD1 δείχνει το επίπεδο της μακροπρόθεσμης μεταβλητότητας. Ακολούθως απεικονίζεται παράδειγμα γραφικής παράστασης Πουανκαρέ στην εικόνα 2.5, περιστατικού που περιγράφει λεπτομερώς τη

διαδικασία προσαρμογής της έλλειψης. Το σύστημα συντεταγμένων x_1 και x_2 είναι στις “45°” ως προς τον κανονικό άξονα. Η τυπική απόκλιση της απόστασης των σημείων από κάθε άξονα καθορίζει το πλάτος SD1 και το μήκος SD2



Σχήμα 2.5: Poincare Plot

Προσαρμόζοντας στην έλλειψη στα σημεία που σχεδιάζονται, εξαγάγουμε τρεις μη γραμμικές μετρήσεις, τις S, SD1 και SD2. Η περιοχή της έλλειψης που αντιπροσωπεύει το συνολικό HRV (S) συσχετίζεται με την ευαισθησία baroreflex (BRS), και το RMSSD. Το BRS (baroreflex sensitivity) ορίζεται ως η μεταβολή του διαστήματος interbeat (RR) σε χιλιοστά του δευτερολέπτου ανά μονάδα μεταβολής. Για παράδειγμα, όταν η αρτηριακή πίεση, αυξάνεται κατά 10 mmHg και η RR αυξάνεται κατά 100 ms, το BRS θα είναι $100/10=10$ ms/mmHg. Χρησιμοποιείται ως μέτρο αυτόνομου ελέγχου του καρδιαγγειακού συστήματος. Γενικά, είναι ένα μέτρο της απόκρισης του αυτόνομου τελεστή σε μια δεδομένη αλλαγή στην αρτηριακή πίεση. Το baroreflex είναι ο ταχύτερος μηχανισμός για τη ρύθμιση των οξειών αλλαγών της αρτηριακής πίεσης μέσω του ελέγχου του καρδιακού ρυθμού, της συσταλτικότητας και της περιφερικής αντίστασης.

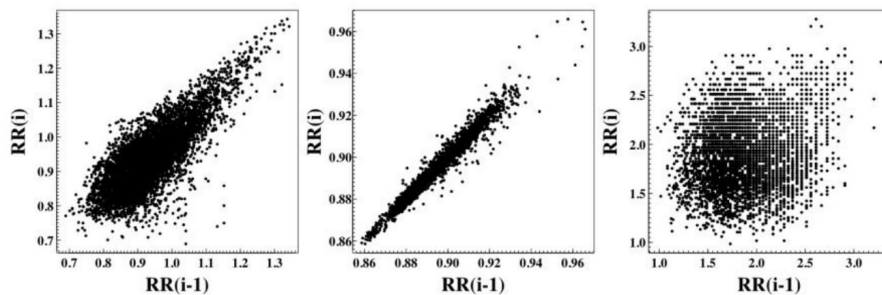
Η τυπική απόκλιση (άρα SD) της απόστασης κάθε σημείου από τον άξονα $y=x$ (SD1), καθορίζει το πλάτος της έλλειψης. Το SD1 μετρά το βραχυπρόθεσμο HRV σε ms. Το RMSSD είναι πανομοιότυπο με το μη γραμμικό μετρικό SD1, το οποίο αντανακλά το βραχυπρόθεσμο διάστημα ανάλυσης της HRV. Το SD1 προβλέπει διαστο-

λική αρτηριακή πίεση, κατά τη διάρκεια εγγραφών 5λεπτών. Η αναλογία SD1/SD2, η οποία μετρά την απρόβλεπτη χρονοσειρά RR, χρησιμοποιείται για τη μέτρηση της αυτόνομης ισορροπίας όταν η περίοδος παρακολούθησης είναι αρκετά μεγάλη και υπάρχει συμπαθητική ενεργοποίηση [48].

Υπολογίζεται ως εξής:

$$SD1/SD2 = \frac{\sqrt{1/2}\sigma(RR_{i+1} - RR_i)}{\sqrt{1/2}\sigma(RR_{i+1} + RR_i)}$$

Για μια υγιή καρδιά, το σύννεφο των σημείων παρουσιάζει ένα σχήμα κομήτη προσανατολισμένο κατά μήκος της γραμμής της ταυτότητας. η δυναμική της καρδιακής καρδιακής ανεπάρκειας χαρακτηρίζεται από ένα τεντωμένο ελλειπτικού σχήματος σύννεφο σημείων επίσης κατά μήκος της γραμμής ταυτότητας. Στην περίπτωση της κολπικής μαρμαρυγής, το νέφος των σημείων παρουσιάζει πιο κυκλικό σχήμα, παρόμοιο με αυτό που συμβαίνει με τη χρονοσειρά λευκού θορύβου [19]. Πιο ξεκάθαρα τα αποτελέσματα μπορούν να διακριθούν σε τρεις κατηγορίες που απεικονίζονται στην εικόνα 2.6.



Σχήμα 2.6: Κατηγορίες Αποτελεσμάτων Poincare Plot [19]

Συμπερασματικά αυτό το γράφημα μπορεί να συνοψίσει ένα πλήρες ρεκόρ Holter (ακόμα και αρκετών ημερών) σε μια εικόνα. Οι βασικές του ιδιότητες είναι:

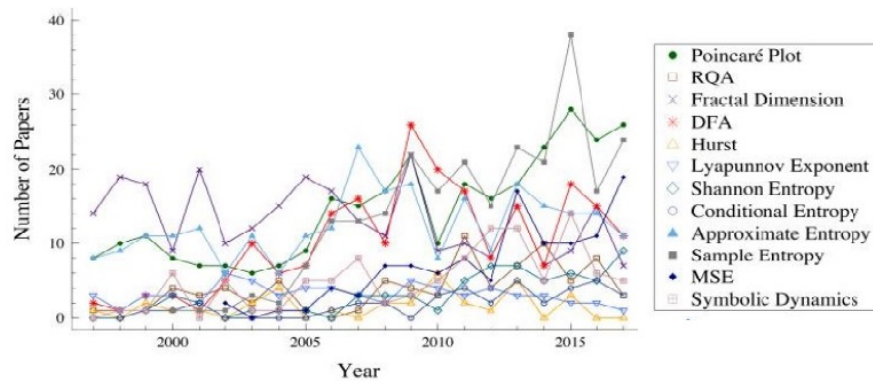
1. Ανίχνευση αλλαγών στη μεταβλητότητα του καρδιακού ρυθμού (HRV)
2. Ανίχνευση κολπικής μαρμαρυγής ή πτερυγισμού
3. Αναγνώριση έκτοπων παλμών
4. Ανίχνευση παύσεων
5. Επαλήθευση των διαγνωστικών
6. Ανάλυση ποιότητας καταγραφής ΗΚΓ

Μη γραμμική ανάλυση

Ο όρος χάος στην επιστήμη, αναφέρεται σε μια μαθηματική προσέγγιση που ασχολείται με συστήματα που μπορούν να περιγραφούν αλλά παράγουν τυχαία εμφανιζόμενα αποτελέσματα. Η θεωρία του χάους ασχολείται με μοτίβα στη χρονική εξέλιξη ενός μη γραμμικού συστήματος που είναι ευαίσθητο στις αρχικές συνθήκες. Μικρές διαφορές στις αρχικές συνθήκες που μπορεί να οφείλονται σε σφάλματα στρογγυλοποίησης σε αριθμητικούς υπολογισμούς, αποδίδουν πολύ διαφορετικά αποτελέσματα, κάνοντας τη μακροπρόθεσμη πρόβλεψη αδύνατη σε γενικές γραμμές [25]. Από την άλλη πλευρά ένα φράκταλ είναι ένα αντικείμενο που αποτελείται από υπομονάδες που μοιάζουν με τη δομή μεγαλύτερης κλίμακας. Αυτή η ιδιότητα της αυτο-ομοιότητάς της αμετάβλητης κλίμακας μπορεί να οδηγήσει σε ένα πιο άκαμπτο και λιγότερο προσαρμόσιμο σύστημα είτε με τυχαία είτε με υψηλή συσχέτιση δομής. Ωστόσο, ο ορισμός του φράκταλ ξεπερνά την ίδια την ομοιότητα για να συμπεριλάβει την ιδέα ενός λεπτομερούς μοτίβου που επαναλαμβάνεται σε όλο και πιο μικρές κλίμακες. Αν και τα φράκταλ είναι

ακανόνιστα, δεν είναι όλες οι ακανόνιστες χρονοσειρές φράκταλ. Η αυτο-ομοιότητα των διακυμάνσεων του συστήματος μπορεί να παρατηρηθεί όταν ένα σήμα αναλύεται σε διαφορετικές χρονικές κλίμακες. Μπορεί να είναι δύσκολο να ειπωθεί από τα δεδομένα εάν μία φυσική ή άλλη διαδικασία που παρατηρείται είναι τυχαία ή χαοτική, διότι στην πράξη οι χρονοσειρές δεν αποτελούνται από καθαρό «σήμα». Πάντα θα υπάρχει κάποια μορφή διαβρωτικού θορύβου, ακόμη και αν αυτό είναι αποτέλεσμα στρογγυλοποίησης ή σφάλματος αποκοπής. Έτσι, οποιαδήποτε πραγματική χρονοσειρά, ακόμη και αν είναι ως επί το πλείστον ντετερμινιστική, θα περιέχει κάποια τυχαιότητα [42]. Πολλές μη γραμμικές μέθοδοι που βασίζονται σε έννοιες του χάους, της κλασματικότητας και της πολυπλοκότητας έχουν χρησιμοποιηθεί για την αξιολόγηση της μεταβλητότητας του καρδιακού ρυθμού για την κατανόηση της καρδιαγγειακής δυναμικής σε υγιείς καθώς και σε παθολογικές καταστάσεις.

Ο αριθμός των εργασιών που δημοσιεύθηκαν τα τελευταία 20 χρόνια για κάθε μη γραμμικό μέτρο εμφανίζεται στην εικόνα 2.7 [19].



Σχήμα 2.7: Αριθμός Δημοσιευμένων Εργασιών

Από την οπτική του Schrodinger, η ζωή είναι απεριοδική (ταλαντώσεις συμβαίνουν χωρίς καθορισμένη περίοδο) και λειτουργεί μεταξύ τυχαιότητας και περιοδικότητας. Η εικοσιτετράωρη παρακολούθηση δεδομένων αποδίδει κάθε φορά σειρά διαστημάτων R-R. Μη γραμμικότητα σημαίνει ότι μια σχέση μεταξύ μεταβλητών δεν μπορεί να απεικονιστεί ως ευθεία γραμμή. Μη γραμμικές μετρήσεις αναδεικνύουν την μη πρόβλεψιμότητα μιας χρονοσειράς, η οποία προκύπτει από την πολυπλοκότητα των μηχανισμών που ρυθμίζουν την HRV. Στρεσογόνοι παράγοντες και άλλες διαταραχές μπορεί να

εμφανίσουν χαμηλές τιμές σε μη γραμμικές μετρήσεις και οι αυξημένες τιμές δεν σηματοδοτούν πάντα υγεία.

Η ανάπτυξη τεχνικών μη-γραμμικής ανάλυσης σημάτων επιτρέπει την εξαγωγή μη-γραμμικών χαρακτηριστικών από το σήμα [40].

Οι γραμμικές μέθοδοι παρέχουν μόνο περιορισμένες πληροφορίες για κάποιο πολύπλοκο σύστημα. Ως εκ τούτου, πολλές μη γραμμικές μέθοδοι και αλγόριθμοι έχουν αναπτυχθεί για την ανάλυση του σήματος HRV [31].

2.6 Εντροπίες

Η μέτρηση εντροπίας είναι ένας καλός τρόπος ώστε να αξιολογήσουμε ένα μέρος του συστήματος για την τυχαιότητα που θα συνεισφέρει στο σύστημα. Η εντροπία μπορεί να ερμηνευθεί ως το προϊόν της σημασίας ενός γεγονότος και της αβεβαιότητας που έχει το γεγονός. Όσο πιο σημαντικό είναι ένα γεγονός στο σύστημα, τόσο περισσότερη αβεβαιότητα, δηλαδή τυχαιότητα, μπορεί να εισαγάγει στο σύστημα. Επομένως, εάν εξάγουμε αυτήν την τυχαιότητα για κάθε γεγονός και τη συνοψίσουμε, τότε μπορούμε να κατανοήσουμε την τυχαιότητα σε ολόκληρο το σύστημα, με τη μεγαλύτερη εντροπία να σχετίζεται με περισσότερη τυχαιότητα και λιγότερη σειρά συστήματος.

2.6.1 Approximate Entropy

Η προσεγγιστική εντροπία (ApEn) μετρά την κανονικότητα και την πολυπλοκότητα σε μια χρονοσειρά. Ποσοτικοποιεί την αταξία του συστήματος και την συσχετίζει με παθολογικές καταστάσεις, αφού ανιχνεύει υποκείμενες μεταβολές που δεν απεικονίζονται στο πλάτος των κορυφών και δεν γίνονται αντιληπτές από κάποιου άλλου είδους ανάλυση. Εφαρμόζεται σε δεδομένα HRV, με τις αυξημένες τιμές να υποδεικνύουν χαμηλή προβλεψιμότητα των διακυμάνσεων στα διαδοχικά διαστήματα. Χαμηλότερες τιμές ApEn σημαίνουν ότι το σήμα είναι τακτικό και μπορεί να προβλεφθεί. Είναι μια παράμετρος που μετρά τη συσχέτιση, ή την κανονικότητα με την έννοια ότι οι χαμηλές τιμές δηλώνουν ότι το σύστημα είναι επαναλαμβανόμενο χωρίς σημάδια ανωμαλίας στο μοτίβο του, σε όλη τη σειρά, ενώ οι υψηλές τιμές σημαίνουν ανεξαρτησία μεταξύ των δεδομένων, χαμηλό αριθμό επαναλαμβανόμενων μοτίβων και τυχαιότητα. Αυτός ο ορισμός δημιουργεί την πεποίθηση ότι τα συστήματα που εμφανίζουν τυχαιότητα έχουν πιθανότητες να έχουν υψηλότερη εντροπία και είναι εκείνα που έχουν υψηλότερο ρυθμό εντροπίας. Ποσοτικοποιεί την έννοια της μεταβλητής πολυπλοκότητας χωρίς τις δυσκολίες των ακριβών στατιστικών της κανονικότητας. Δεν είναι απαραίτητο να ανακατασκευαστεί πλήρως η δυναμική του συστήματος για να το ταξινομήσει

[41]. Έστω μία χρονοσειρά S_N αποτελούμενη από N στιγμιαίες μετρήσεις: $HR(1), HR(2), \dots, HR(N)$. Επιλέγω μη αρνητικό ακέραιο m που ορίζει το μήκος ενός μοτίβου και r θετικό, πραγματικό αριθμό που ορίζει το κριτήριο ομοιότητας που απαιτούμε για τα δύο μοτίβα. Δηλώνω ένα υποσύνολο μοτίβο χρονοσειράς $Pm(i)$ που ανήκει στο σύνολο μετρήσεων της χρονοσειράς S_N με i την αρχή της μέτρησης. Δύο πρότυπα $Pm(i)$ και $Pm(j)$ χαρακτηρίζονται όμοια όταν η διαφορά μεταξύ οποιοδήποτε ζεύγων αντίστοιχων μετρήσεων στα μοτίβα είναι μικρότερη από την παράμετρο r που έχουμε ορίσει.

$$|HR(i+k) - HR(j+k)| < r$$

$$0 \leq k < m$$

Για κάθε υποσύνολο σετ μοτίβων

$$[Pm(1), Pm(2), \dots, P(N-m+1)]$$

που ανήκει στην S_N μπορούμε να ορίσουμε:

$$C_{im}(r) = \frac{n_{im}(r)}{N-m+1}$$

Η ποσότητα $C_m(r)$ εκφράζει την επικράτηση των επαναλαμβανόμενων μοτίβων. Υπολογίζουμε την προσεγγιστική εντροπία βάσει του τύπου:

$$ApEn(S_N, m, r) = \ln \frac{C_m(r)}{C_{m+1}(r)}$$

που είναι ο φυσικός λογάριθμος της σχετικής επικράτησης επαναλαμβανόμενων μοτίβων μήκους m σε σύγκριση με αυτά του μήκους $m+1$. Έτσι, αν βρούμε όμοια μοτίβα σε μια χρονοσειρά καρδιακών παλμών, η $ApEn$ υπολογίζει τη λογαριθμική πιθανότητα να διαφέρουν τα επόμενα διαστήματα μετά από κάθε ένα από τα μοτίβα (δηλαδή ότι η ομοιότητα των μοτίβων είναι απλή σύμπτωση και στερείται προγνωστικής αξίας). Μικρότερες τιμές του υποδηλώνουν μεγαλύτερη πιθανότητα όμοια μοτίβα μετρήσεων να ακολουθούνται από πρόσθετες παρόμοιες μετρήσεις. Εάν η χρονική σειρά είναι πολύ ακανόνιστη, η εμφάνιση όμοιων μοτίβων δεν θα είναι προγνωστική για τις ακόλουθες μετρήσεις και η $ApEn$ θα είναι σχετικά μεγάλη.

Θα πρέπει να σημειωθεί ότι η $ApEn$ έχει σημαντικές αδυναμίες, ιδίως την ισχυρή εξάρτησή της από το μήκος ακολουθίας και την κακή αυτοσυνέπειά της (δηλαδή, η παρατήρηση ότι η $ApEn$ για ένα σύνολο δεδομένων είναι μεγαλύτερη από την $ApEn$ για ένα άλλο για μια δεδομένη επιλογή των m και r . Εξαρτάται σε μεγάλο βαθμό από το μήκος του δείγματος εισόδου. Για αυτό τον λόγο εισάγεται και το μέγεθος $SampEn$.

2.6.2 Sample Entropy

Προκειμένου να ξεπεραστούν οι περιορισμοί που θέτει η προσεγγιστική εντροπία, οι συγγραφείς πρότειναν μια νέα οικογένεια στατιστικών, τη SampEn(m,r), η οποία, μεταξύ άλλων διαφορών, εξαλείφει τις αυτο-ταιριάσεις, μειώνοντας έτσι τον υπολογιστικό χρόνο κατά το ήμισυ σε σύγκριση με την ApEn. Επίσης, λόγω της μη ανεξαρτησίας των προτύπων, έχει υπολειπόμενη προκατάληψη για πολύ σύντομες χρονικές σειρές. Χρησιμοποιεί τις ίδιες μεταβλητές με την ApEn. Η εντροπία δείγματος (SampEn) είναι η εντροπία ενσωμάτωσης που ποσοτικοποιεί την κανονικότητα της χρονοσειράς ως μέτρο της πολυπλοκότητας του συστήματος. Ταχτικές ακολουθίες θα οδηγήσουν σε χαμηλότερες τιμές της SampEn ενώ υψηλότερες τιμές της SampEn σχετίζονται με τυχαίες συμπεριφορές στη χρονοσειρά.

2.6.3 Singular Value Decomposition (SVD) Entropy

Είναι ένα θεμελιώδες εργαλείο στη γραμμική άλγεβρα. Χρησιμοποιείται για την ανάλυση βραχυπρόθεσμων, μη-στάσιμων δεδομένων συνόλου HRV. Για να υποδηλωθούν τα χαρακτηριστικά διασποράς των μοναδικών τιμών, η SvdEn ορίζεται με τον τύπο της εντροπίας Shannon.

2.6.4 Shannon Entropy

Ο Shannon πρότεινε την έννοια της εντροπίας (εντροπία Shannon (SE)) για να μετρήσει πώς οι πληροφορίες, μέσα σε ένα σήμα μπορούν να ποσοτικοποιηθούν με απόλυτη ακρίβεια ως προς την ποσότητα των απροσδόκητων δεδομένων που περιέχονται στο σήμα [49]. Η εντροπία Shannon προκύπτει:

$$SE = - \sum_i p(x(i)) \log(p(x(i)))$$

Η εντροπία Shannon μπορεί να μετρήσει την αβεβαιότητα μιας τυχαίας διαδικασίας. Η πιθανότητα δίνει μια "τοπική εικόνα" ολόκληρου του συστήματος. Το μέρος p του τύπου της εντροπίας μας δίνει πληροφορίες για τη σημασία του γεγονότος σε ολόκληρο το σύστημα. Σε ένα περιοδικό σήμα, ορισμένες ακολουθίες που εμφανίζονται συχνά, έχουν χαμηλότερη εντροπία Shannon [62].

2.6.5 Permutation Entropy

Η εντροπία μετάθεσης (PE) είναι ένα ισχυρό εργαλείο ανάλυσης χρονοσειρών που παρέχει ένα ποσοτικό μέτρο της πολυπλοκότητας ενός δυναμικού συστήματος

καταγράφοντας τις σχέσεις τάξης μεταξύ τιμών μιας χρονοσειράς και εξάγοντας μια κατανομή πιθανότητας των τακτικών προτύπων. Μεταξύ των κύριων χαρακτηριστικών της: Είναι μη παραμετρική και δεν περιέχει περιοριστικές παραμετρικές υποθέσεις. Είναι ισχυρό εργαλείο, όσον αφορά το θόρυβο, υπολογιστικά αποδοτικό, εύελιχτο και αμετάβλητο σε σχέση με μη γραμμικούς μονοτονικούς μετασχηματισμούς των δεδομένων. Βασίζεται στις έννοιες της εντροπίας και της symbol dynamics. Υπολογίζει τη δομή της χρονικής διάταξης (χρονική αιτιότητα) μιας δεδομένης χρονικής σειράς πραγματικών τιμών. Επιτρέπει στο χρήστη να ξεκλειδώσει το σύνθετο δυναμικό περιεχόμενο μη γραμμικών χρονοσειρών. Όσο μικρότερη τιμή έχει, τόσο πιο κανονική και πιο ντετερμινιστική είναι η χρονοσειρά. Αντίθετα, όσο πιο κοντά στο 1 είναι, τόσο πιο θορυβώδης και τυχαία είναι η χρονοσειρά [20].

2.7 Detrended Fluctuation Analysis – DFA

Η ανάλυση διακύμανσης μετά την αφαίρεση της τάσης (Detrended Fluctuation Analysis – DFA) έχει αποδειχθεί χρήσιμη, για την εύρεση της έκτασης των συσχετίσεων μακράς εμβέλειας στις χρονοσειρές. Εν συντομία, πρώτα ενσωματώνεται η χρονοσειρά που θα αναλυθεί (με N δείγματα). Στη συνέχεια, η ολοκληρωμένη χρονοσειρά χωρίζεται σε χρονικά παράθυρα ίσου μήκους, n . Σε κάθε πλαίσιο μήκους n , υπολογίζεται το ευθύγραμμο τμήμα με τη μέθοδο των ελάχιστων τετραγώνων που μοντελοποιεί την τάση των τιμών στα δεδομένα (που αντιπροσωπεύει την τάση σε αυτό το πλαίσιο). Η συντεταγμένη y των ευθύγραμμων τμημάτων συμβολίζεται με $y(k)$.

Στη συνέχεια, αφαιρούμε την ολοκληρωμένη χρονοσειρά, αφαιρώντας την τοπική τάση σε κάθε πλαίσιο. Ο αλγόριθμος πραγματοποιείται σε δύο βήματα. Αρχικά υπολογίζεται η ακολουθία:

$$y(k) = \sum_{i=1}^k (x(i) - \bar{x})$$

με $k=1,2,\dots,N$ και υπολογίζεται η διακύμανση:

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N (y(k) - y_n(k))^2}$$

Εκφράζει τη συμπεριφορά της HRV ως προς την κλασματική συμπίεση: Σήματα που τείνουν να έχουν συμπεριφορά κλασματικής συμπίεσης έχουν τιμές που πλησιάζουν τη μονάδα και αφορούν υγιή άτομα ενώ μειώνεται σε σήματα που παρουσιάζουν καρδιακές

παθήσεις. Είναι ένα δείκτης "τραχύτητας" του σήματος. Όσο μεγαλύτερη τιμή έχει τόσο πιο ομαλό είναι το σήμα. Συμπερασματικά:

Αν η τιμή του α τείνει στο 0,5 τότε η χρονοσειρά αντιπροσωπεύει ασύνδετη τυχαιότητα (white noise) η τιμή του α προσεγγίζει τη μονάδα τότε η χρονοσειρά έχει συσχετίσεις μεγάλης εμβέλειας και παρουσιάζει αμετάβλητη κλίμακα.

Αν το α προσεγγίζει τη τιμή 1,5 τότε αντιπροσωπεύει ένα τυχαίο βήμα (κίνηση Brown).

2.8 Fractal Dimension

Στα μαθηματικά, το φράκταλ είναι ένας όρος που χρησιμοποιείται για να περιγράψει γεωμετρικά σχήματα που περιέχουν λεπτομερή δομή σε αυθαίρετα μικρές κλίμακες, που συνήθως έχουν μια φράκταλ διάσταση που υπερβαίνει αυστηρά την τοπολογική διάσταση. Πολλά φράκταλ εμφανίζονται παρόμοια σε διάφορες κλίμακες. Στη γεωμετρία φράκταλ, μια φράκταλ διάσταση είναι μια αναλογία που παρέχει έναν στατιστικό δείκτη πολυπλοκότητας που συγκρίνει πώς αλλάζει η λεπτομέρεια σε ένα μοτίβο με την κλίμακα στην οποία μετράται. Σε αντίθεση με τις τοπολογικές διαστάσεις, ο δείκτης φράκταλ μπορεί να λάβει μη ακέραιες τιμές, υποδεικνύοντας ότι ένα σύνολο γεμίζει το χώρο του ποιοτικά και ποσοτικά διαφορετικά από αυτό που κάνει ένα συνηθισμένο γεωμετρικό σύνολο. Έχει επίσης χαρακτηριστεί ως μέτρο της ικανότητας πλήρωσης χώρου ενός σχεδίου που λέει πώς ένα φράκταλ κλιμακώνεται διαφορετικά από τον χώρο στον οποίο είναι ενσωματωμένο. Πολλά φράκταλ εμφανίζονται παρόμοια σε διάφορες κλίμακες, όπως φαίνεται στις διαδοχικές μεγεθύνσεις του συνόλου Mandelbrot [47][50]. Αυτή η έκθεση παρόμοιων μοτίβων σε όλο και μικρότερες κλίμακες ονομάζεται αυτο-ομοιότητα. Η διάσταση φράκταλ (FD) είναι ένας δείκτης για τον χαρακτηρισμό μοτίβων ή συνόλων φράκταλ ποσοτικοποιώντας την πολυπλοκότητά τους ως αναλογία της αλλαγής στη λεπτομέρεια προς την αλλαγή στην κλίμακα. Αρκετοί τύποι φράκταλ διάστασης μπορούν να μετρηθούν θεωρητικά και εμπειρικά με διάφορους αλγόριθμους [45][17]. Τα βιολογικά σήματα είναι τα αυτο-επαναλαμβανόμενα, μη γραμμικά μοτίβα φράκταλ. Η χρήση διακριτού δείκτη Fractal και η ενσωμάτωση αυτού με τη θεωρία πιθανοτήτων αποτελεί πλέον ένα πολύ μεγάλο πλεονέκτημα στην προγνωστική καρδιολογία. Μια πρόσφατη έρευνα απέδειξε τη δύναμη της προσέγγισης που βασίζεται στη διάσταση φράκταλ σε προβλήματα αναγνώρισης προτύπων [44].

2.8.1 Petrosian algorithm

Ο αλγόριθμος του Petrosian μπορεί να χρησιμοποιηθεί για να παρέχει έναν γρήγορο υπολογισμό της FD ενός σήματος μεταφράζοντας τη χρονοσειρά σε μια δυαδική ακολουθία. Συνεχόμενα δείγματα στη χρονοσειρά αποσπώνται και η δυαδική ακολουθία

δημιουργείται με βάση το αποτέλεσμα της αφαίρεσης. Ένα «+1» ή «-1» εκχωρείται για κάθε θετικό ή αρνητικό αποτέλεσμα αντίστοιχα. Η δυαδική ακολουθία σχηματίζεται εκχωρώντας ένα «1» για κάθε διαφορά μεταξύ διαδοχικών δειγμάτων στις χρονοσειρές που υπερβαίνουν ένα τυπικό μέγεθος απόκλισης ενώ σε διαφορετική περίπτωση εκχωρείται το «0». Η FD υπολογίζεται ως:

$$P = \frac{\log(N)}{\log(N) + \log\left(\frac{N}{N+0.4N_d}\right)}$$

Όπου N είναι το μήκος της χρονοσειράς και N_d είναι ο αριθμός των αλλαγών του προσήμου στην παράγωγο του σήματος. (Η πρώτη παράγωγος ενός σήματος είναι ο ρυθμός μεταβολής του y με το x , δηλαδή το dy/dx , το οποίο ερμηνεύεται ως η κλίση της εφαπτομένης στο σήμα σε κάθε σημείο) ή αλλιώς ο αριθμός των αλλαγών του προσήμου στη δυαδική ακολουθία.

2.8.2 Katz's algorithm

Σε αντίθεση με τη μέθοδο Petrosian, ο υπολογισμός της κλασματικής διάστασης Katz είναι ελαφρώς πιο αργός, αλλά προέρχεται απευθείας από τη κυματομορφή, εξαλείφοντας το βήμα προεπεξεργασίας της δημιουργίας δυαδικής ακολουθίας. Η κλασματική διάσταση Katz [24] της κυματομορφής που αντιπροσωπεύει τη χρονοσειρά υπολογίζεται χρησιμοποιώντας τη σχέση:

$$FD_K = \frac{\log L}{\log d}$$

όπου L είναι το συνολικό μήκος της καμπύλης που υπολογίζεται ως το άθροισμα της απόστασης μεταξύ των διαδοχικών σημείων δεδομένων και d είναι η διάμετρος ή η επίπεδη έκταση της καμπύλης, που υπολογίζεται ως η απόσταση μεταξύ του πρώτου σημείου και του σημείου της ακολουθίας που δίνει την πιο απομακρυσμένη απόσταση. Για τα σήματα που δεν διασταυρώνονται, μπορεί να εκφραστεί ως:

$$d = \max(\text{distance}(1, i))$$

Λαμβάνοντας υπόψη την απόσταση μεταξύ κάθε σημείου της ακολουθίας με i το πρώτο, σημείο να είναι αυτό που μεγιστοποιεί την απόσταση σε σχέση με το αρχικό σημείο. Η κλασματική διάσταση συγκρίνει τον πραγματικό αριθμό μονάδων που συνθέτουν την καμπύλη με τον ελάχιστο αριθμό μονάδων που απαιτούνται για την αναπαραγωγή μοτίβου της ίδιας χωρικής έκτασης. Εξαρτώνται από τις μονάδες μέτρησης που χρησιμοποιούνται. Εάν οι μονάδες είναι διαφορετικές, τότε το ίδιο και η προσέγγιση της κλασματικής διάστασης.

2.8.3 Higuchi algorithm

Η μέθοδος του Higuchi [22] [21] είναι ένας πολύ αποτελεσματικός αλγόριθμος για τον υπολογισμό της κλασματικής διάστασης μιας καμπύλης και χρησιμοποιείται όλο και περισσότερο για την ανάλυση χρονοσειρών. Έστω μια χρονική σειρά που εκφράζεται από:

$$X : (1, 2, \dots, N)$$

που ανήκει στους παραγματικούς αριθμούς με πλήθος δεδομένων N και παράμετρος

$$k_{max} \geq 2$$

η κλασματική διάσταση Higuchi της X υπολογίζεται:

$$\forall k \in (1, \dots, k_{max}) \text{ και } m \in (1, \dots, k)$$

ορίζουμε το μήκος

$$L_m(k) = \frac{N-1}{\frac{N-m}{k}k^2} \sum_{i=1}^{\frac{N-m}{k}} |X_N(m+ik) - X_N(m+(i-1)k)|$$

Το μήκος της καμπύλης $L(k)$ ορίζεται από το μέσο όρο των μηκών ($L_1(k), \dots, L_k(k)$)

$$\text{και υπολογίζεται ως } L(k) = \frac{1}{k} \sum_{m=1}^k L_m(k)$$

Η κλίση της γραμμικής συνάρτησης που ταιριάζει καλύτερα στα σημεία είναι

$$\left(\log \frac{1}{k}, \log L(k) \right)$$

και ορίζεται ως η φραγματική διάσταση Higuchi της δεδομένης χρονοσειράς X που αναζητούμε. Η ακριβής προσομοίωση διαδικασίας και επιλογής από λογισμικό υπολογιστή εξαρτάται άμεσα από την κατάλληλη επιλογή των τιμών k που χρησιμοποιούνται. Οι τιμές k είναι το βασικό συστατικό για το σχεδιασμό και την προσομοίωση ενός τέτοιου συστήματος. Στην παρούσα μελέτη ορίζεται αυτόματα από τον αλγόριθμο που καθορίζει η `ryentrp Library` της Python.

2.9 Στατιστικός έλεγχος και συσχέτιση μεταβλητών

2.9.1 Στατιστικός Έλεγχος

Το τεστ Kruskal-Wallis είναι μια μη παραμετρική έκδοση του one-way ANOVA test. Ονομάστηκε από τους προγραμματιστές της μεθόδου, William Kruskal και Wilson Wallis. Αυτό το τεστ μπορεί να χρησιμοποιηθεί για να προσδιοριστεί εάν περισσότερα από δύο ανεξάρτητα δείγματα έχουν διαφορετική κατανομή. Μπορεί να θεωρηθεί ως η γενίκευση του τεστ Mann-Whitney U.

Η προεπιλεγμένη υπόθεση ή η μηδενική υπόθεση είναι ότι όλα τα δείγματα δεδομένων αντλήθηκαν από την ίδια κατανομή. Συγκεκριμένα, ότι οι διάμεσοι πληθυσμού όλων των ομάδων είναι ίσοι. Η απόρριψη της μηδενικής υπόθεσης υποδηλώνει ότι υπάρχουν αρκετά στοιχεία που να υποδηλώνουν ότι ένα ή περισσότερα δείγματα κυριαρχούν σε άλλο δείγμα, αλλά το τεστ δεν υποδεικνύει ποια δείγματα ή κατά πόσο.

Όταν το Kruskal-Wallis H-test οδηγεί σε σημαντικά αποτελέσματα, τότε τουλάχιστον ένα από τα δείγματα είναι διαφορετικό από τα άλλα δείγματα. Ωστόσο, η δοκιμή δεν προσδιορίζει πού εμφανίζονται οι διαφορές. Επιπλέον, δεν προσδιορίζει πόσες διαφορές υπάρχουν. Για να εντοπίσει τις ιδιαίτερες διαφορές μεταξύ των ζευγών δειγμάτων, ένας ερευνητής μπορεί να χρησιμοποιήσει αντιθέσεις δειγμάτων ή δοκιμές post hoc, για να αναλύσει τα συγκεκριμένα ζεύγη δειγμάτων για σημαντικές διαφορές. Είναι μια χρήσιμη μέθοδος για την εκτέλεση τεστ με διαφορές των δειγμάτων μεταξύ μεμονωμένων συνόλων δειγμάτων. Άρα διερευνώνται οι παρακάτω συνθήκες:

-Αποτυχία της απόρριψης της H_0 : Όλες οι κατανομές δειγμάτων είναι ίσες.
-Απόρριψη H_0 : Μία ή περισσότερες κατανομές δειγμάτων δεν είναι ίσες. Κάθε δείγμα δεδομένων πρέπει να είναι ανεξάρτητο, να έχει 5 ή περισσότερες παρατηρήσεις και τα δείγματα δεδομένων μπορεί να διαφέρουν σε μέγεθος. ¹

2.9.2 Συσχέτιση Μεταβλητών

Ο συντελεστής συσχέτισης Pearson είναι ένα μέτρο της γραμμικής συσχέτισης μεταξύ δύο μεταβλητών X και Y. Παίρνει τιμές στο διαστήματος (-1,1) όπου:

- Το -1 υποδηλώνει μια απόλυτα αρνητική γραμμική συσχέτιση μεταξύ δύο μεταβλητών

¹Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach του Gregory W. Corder

- Το 0 υποδηλώνει ότι δεν υπάρχει γραμμική συσχέτιση μεταξύ δύο μεταβλητών
- 1 δείχνει μια απόλυτα θετική γραμμική συσχέτιση μεταξύ δύο μεταβλητών

Έστω r η συσχέτιση δύο μεταβλητών X και Y τότε ισχύει η παρακάτω μαθηματική φόρμουλα:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Ωστόσο, υπάρχουν κάποιοι περιορισμοί:

- Η σχέση μεταξύ των μεταβλητών είναι μη γραμμική.
- Επηρεάζεται από την παρουσία ακραίων τιμών
- Η κατανομή πρέπει να είναι κανονική.

Σε αυτή τη περίπτωση ελέγχουμε αν η σχέση είναι γραμμική ή όχι. Ένας τρόπος είναι να εξετάσουμε τα διαγράμματα διασποράς. Εάν οι τιμές των μεταβλητών παραβιάζουν οποιαδήποτε από τις 3 συνθήκες που αναφέρονται παραπάνω, τότε μπορούμε να χρησιμοποιήσουμε τη "μέθοδο του Spearman" για την εύρεση της συσχέτισης. Ο λόγος που παραμένει ανεπηρέαστη σε παρουσία ακραίων τιμών είναι επειδή δεν υπολογίζει τις τιμές αλλά την κατάταξη των αντίστοιχων τιμών. Είναι ένα μη-παραμετρικό μέτρο της στατιστικής εξάρτησης μεταξύ δύο μεταβλητών, αξιολογώντας τη μεταξύ τους σχέση με μια μονότονη συνάρτηση. Ο συντελεστής Spearman, όπως κάθε συντελεστής συσχέτισης, είναι κατάλληλος και για συνεχείς και για διακριτές μεταβλητές, συμπεριλαμβανομένων των τακτικών διακριτών μεταβλητών και ορίζεται όπως και ο συντελεστής Pearson [2].

2.10 Μηχανική Μάθηση

Ενώ οι περισσότερες στατιστικές αναλύσεις βασίζονται στη λήψη αποφάσεων που βασίζονται σε κανόνες, η μηχανική μάθηση υπερέχει σε εργασίες που είναι δύσκολο να καθοριστούν με ακριβείς κανόνες βήμα προς βήμα. Η μηχανική μάθηση μπορεί να εφαρμοστεί σε πολλά σενάρια στα οποία ένα αποτέλεσμα εξαρτάται από εκατοντάδες παράγοντες – παράγοντες που είναι δύσκολο ή αδύνατο να παρακολουθήσει ένας άνθρωπος. Ως αποτέλεσμα, η μηχανική μάθηση χρησιμοποιείται για την πρόβλεψη μελλοντικών τάσεων. Η μηχανική μάθηση είναι ένα υποπεδίο της τεχνητής νοημοσύνης που παρέχει στα συστήματα τη δυνατότητα να μαθαίνουν αυτόματα και να βελτιώνονται από την εμπειρία χωρίς να προγραμματίζονται ρητά. Για τη διαδικασία μάθησης

(προσαρμογή μοντέλου) πρέπει να έχουμε διαθέσιμες κάποιες παρατηρήσεις ή δείγματα προκειμένου να εξερευνήσουμε πιθανά υποκείμενα μοτίβα, κρυμμένα στα δεδομένα μας. Αυτά τα εκπαιδευόμενα μοτίβα δεν είναι τίποτα περισσότερο από ορισμένες λειτουργίες και εκπαιδεύονται από συστήματα υπολογιστών, αυτόματα χωρίς ανθρώπινη παρέμβαση ή εισροή. Ο επιβλέπων, είναι εκείνος που παρέχει τη σωστή τιμή εξόδου της συνάρτησης, για τη βάση γνώσης που αναλύουμε. Σε αυτήν την κατηγορία μηχανικής μάθησης το σύστημα πρέπει να "μάθει" με επαγωγική μέθοδο μια συνάρτηση που ονομάζεται συνάρτηση στόχος (target function) και που αποτελεί έκφραση του μοντέλου που περιγράφει τα δεδομένα. Η συνάρτηση στόχος χρησιμοποιείται για την πρόβλεψη της τιμής μιας μεταβλητής, που ονομάζεται εξαρτημένη μεταβλητή ή μεταβλητή εξόδου. Η επιβλεπόμενη μάθηση είναι μία κατηγορία μηχανικής μάθησης, στόχος της οποίας είναι ο χαρακτηρισμός δεδομένων με βάση κάποια δεδομένα εκπαίδευσης. Οι αλγόριθμοι μηχανικής μάθησης συνήθως κατηγοριοποιούνται ως *εποπτευόμενοι* ή *μη εποπτευόμενοι*.

2.10.1 Μηχανική Μάθηση με Επίβλεψη

Για αυτήν την οικογένεια μοντέλων, η έρευνα πρέπει να έχει στη διάθεσή της ένα σύνολο δεδομένων με ορισμένες παρατηρήσεις και τις ετικέτες/τάξεις των παρατηρήσεων. Για παράδειγμα, οι παρατηρήσεις θα μπορούσαν να είναι εικόνες ζώων και οι ετικέτες το όνομα του ζώου (π.χ. γάτα, σκύλος κ.λπ.). Αυτά τα μοντέλα μαθαίνουν από το επισημασμένο σύνολο δεδομένων και στη συνέχεια χρησιμοποιούνται για την πρόβλεψη μελλοντικών γεγονότων. Για τη διαδικασία εκπαίδευσης, η είσοδος είναι ένα γνωστό σύνολο δεδομένων εκπαίδευσης (training data) με τις αντίστοιχες ετικέτες του και ο αλγόριθμος εκμάθησης παράγει μια συνάρτηση για να κάνει τελικά προβλέψεις σχετικά με κάποιες νέες αόρατες παρατηρήσεις που μπορεί κανείς να δώσει στο μοντέλο. Το μοντέλο είναι σε θέση να παρέχει στόχους (targets) για οποιαδήποτε νέα εισαγωγή μετά από επαρκή εκπαίδευση. Ο αλγόριθμος εκμάθησης μπορεί επίσης να συγκρίνει την έξοδό του με τη σωστή προβλεπόμενη έξοδο (ground truth level) και να βρει λάθη προκειμένου να τροποποιηθεί ανάλογα (via back-propagation). Τα εποπτευόμενα μοντέλα μπορούν να ομαδοποιηθούν περαιτέρω σε περιπτώσεις παλινδρόμησης και ταξινόμησης:

- **Regression** (Παλινδρόμηση): Ένα πρόβλημα παλινδρόμησης είναι όταν η μεταβλητή εξόδου είναι μια πραγματική συνεχής τιμή π.χ. πρόβλεψη τιμής μετοχής
- **Classification** (Ταξινόμηση): Πρόβλημα ταξινόμησης είναι όταν η μεταβλητή εξόδου είναι κατηγορία π.χ. «ασθένεια» / «καμία ασθένεια».

Μερικά παραδείγματα μοντέλων που ανήκουν στην ταξινόμηση είναι τα ακόλουθα:

- Μάθηση Εννοιών (Concept Learning)
- Δέντρα Απόφασης (Decision Trees)
- Μάθηση Κανόνων (Rule Learning)
- Μάθηση κατά Περίπτωση (Instance Based Learning)
- Μάθηση κατά Bayes
- Γραμμική Παρεμβολή (Linear Regression)
- Νευρωνικά Δίκτυα (Neural Networks)
- Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)

2.10.2 Αλγόριθμοι/Μέθοδοι Μηχανικής Μάθησης χωρίς Επίβλεψη

Για αυτήν την οικογένεια μοντέλων, η έρευνα πρέπει να έχει στη διάθεσή της ένα σύνολο δεδομένων με ορισμένες παρατηρήσεις χωρίς να χρειάζεται να υπάρχουν και οι ετικέτες/τάξεις των παρατηρήσεων. Η μάθηση χωρίς επίβλεψη μελετά πώς τα συστήματα μπορούν να εισάγουν μια συνάρτηση για την περιγραφή μιας κρυφής δομής από δεδομένα χωρίς να ανήκουν σε κάποια κατηγορία. Το σύστημα δεν προβλέπει τη σωστή έξοδο, αλλά αντ' αυτού, ερευνά τα δεδομένα και μπορεί να αντλήσει συμπεράσματα από σύνολα δεδομένων για να περιγράψει κρυφές δομές από δεδομένα χωρίς ετικέτα. Τα μοντέλα χωρίς επίβλεψη μπορούν να ομαδοποιηθούν περαιτέρω σε περιπτώσεις ομαδοποίησης και συσχέτισης.

- **Clustering**(Ομαδοποίηση): Ένα πρόβλημα ομαδοποίησης είναι όταν θέλουμε να βρούμε τις εγγενείς ομαδοποιήσεις στα δεδομένα, όπως η ομαδοποίηση ζώων με βάση ορισμένα χαρακτηριστικά, όπως ο αριθμός ποδιών.
- **Association**(Συσχέτιση): Η εκμάθηση κανόνων συσχέτισης είναι όταν θέλουμε να βρούμε κανόνες συσχέτισης, όπως αν οι άνθρωποι που αγοράζουν X τείνουν επίσης να αγοράζουν Y. Μερικά παραδείγματα μοντέλων που ανήκουν σε αυτήν την οικογένεια είναι τα ακόλουθα: PCA, K-means, DBSCAN κ.λπ.

Συμπερασματικά, οι κύριες διαφορές μεταξύ εποπτευόμενης και μη επιτηρούμενης μάθησης είναι οι παρακάτω: Οι αλγόριθμοι εποπτευόμενης μάθησης εκπαιδεύονται χρησιμοποιώντας δεδομένα με ετικέτα. Οι αλγόριθμοι μάθησης χωρίς επίβλεψη εκπαιδεύονται χρησιμοποιώντας δεδομένα χωρίς επιτήρηση. Το εποπτευόμενο μοντέλο μάθησης

λαμβάνει άμεση ανατροφοδότηση για να ελέγξει αν προβλέπει σωστή έξοδο ή όχι. Ο στόχος της εποπτευόμενης μάθησης, είναι να γίνει προσέγγιση στη λειτουργία αντιστοίχισης τόσο καλά ώστε όταν έχετε νέα δεδομένα εισόδου (x) ώστε να μπορείτε να προβλέψετε τις μεταβλητές εξόδου (Y) για αυτά τα δεδομένα.

2.10.3 Ημι-Εποπτευόμενοι Αλγόριθμοι Μηχανικής Μάθησης

Αυτή η οικογένεια βρίσκεται μεταξύ των εποπτευόμενων και των μη εποπτευόμενων οικογενειών μάθησης. Τα ημι-εποπτευόμενα μοντέλα χρησιμοποιούν δεδομένα τόσο με τάξη όσο και χωρίς τάξη για εκπαίδευση. Η ημι-εποπτευόμενη μάθηση είναι μια προσέγγιση στη μηχανική μάθηση που συνδυάζει μια μικρή ποσότητα δεδομένων με ετικέτα με μια μεγάλη ποσότητα δεδομένων χωρίς ετικέτα κατά τη διάρκεια της εκπαίδευσης. Η ημι-εποπτευόμενη μάθηση εμπίπτει μεταξύ της μάθησης χωρίς επίβλεψη (χωρίς επισημασμένα δεδομένα κατάρτισης) και της εποπτευόμενης μάθησης (με μόνο επισημασμένα δεδομένα εκπαίδευσης).

2.10.4 Reinforcement Machine Learning Algorithms

Αυτή η οικογένεια μοντέλων αποτελείται από αλγόριθμους που χρησιμοποιούν τα εκτιμώμενα σφάλματα ως ανταμοιβές ή ποινές. Εάν το λάθος είναι μεγάλο, τότε η ποινή είναι υψηλή και η ανταμοιβή χαμηλή. Εάν το λάθος είναι μικρό, τότε η ποινή είναι χαμηλή και η ανταμοιβή υψηλή. Η αναζήτηση δοκιμαστικού σφάλματος και "η καθυστερημένη" ανταμοιβή είναι τα πιο σχετικά χαρακτηριστικά της ενισχυτικής μάθησης. Αυτή η οικογένεια μοντέλων επιτρέπει τον αυτόματο προσδιορισμό της ιδανικής συμπεριφοράς μέσα σε ένα συγκεκριμένο πλαίσιο προκειμένου να μεγιστοποιηθεί η επιθυμητή απόδοση. Απαιτείται ανάδραση ανταμοιβής (reward feedback) για να μάθει το μοντέλο ποια ενέργεια είναι καλύτερη και αυτό είναι γνωστό ως «σήμα ενίσχυσης».

2.10.5 Αλγόριθμοι Ταξινόμησης

Χρησιμοποιώντας μοντέλα μηχανικής μάθησης, είμαστε σε θέση να εκτελούμε αναλύσεις τεράστιων ποσοτήτων δεδομένων. Μοτίβα δεδομένων που θα ήταν αδύνατο να αναγνωριστούν από έναν άνθρωπο, μπορούν να εξαχθούν με ακρίβεια χρησιμοποιώντας αυτά τα μοντέλα μέσα σε δευτερόλεπτα. Στην παρούσα μελέτη χρησιμοποιήθηκαν:

1. K-nearest neighbors (KNN)

Ο KNN είναι ένας τύπος εποπτευόμενου αλγόριθμου μάθησης που χρησιμοποιείται τόσο για παλινδρόμηση όσο και για ταξινόμηση. Ο KNN προσπαθεί να προβλέψει τη σωστή κατηγορία (class) για τα δεδομένα δοκιμής (test data) υπολογίζοντας την απόσταση μεταξύ των δεδομένων δοκιμής και όλων των σημείων εκπαίδευσης. Στη συνέχεια, επιλέγουμε μέσω κάποιων τεχνικών υπολογισμού, τον αριθμό K των σημείων που βρίσκεται κοντά στα δεδομένα της δοκιμής. Ο αλγόριθμος KNN υπολογίζει την πιθανότητα τα δεδομένα δοκιμής να ανήκουν στις κλάσεις των δεδομένων εκπαίδευσης « K » και επιλέγει την κατηγορία που έχει την υψηλότερη πιθανότητα. Στην περίπτωση παλινδρόμησης, η τιμή είναι ο μέσος όρος των « K » επιλεγμένων σημείων εκπαίδευσης. Η λειτουργία του K-NN μπορεί να εξηγηθεί με βάση τον παρακάτω αλγόριθμο:

Βήμα-1: Γίνεται επιλογή του αριθμού K των γειτόνων

Βήμα-2: Υπολογίζεται η ευκλείδεια απόσταση του K αριθμού γειτόνων

Βήμα-3: Οι K πλησιέστεροι γείτονες υπολογίζονται σύμφωνα με την Ευκλείδεια απόσταση.

Βήμα-4: Μεταξύ αυτών των k γειτόνων, μετρείται ο αριθμός των σημείων δεδομένων σε κάθε κατηγορία.

Βήμα-5: Αντιστοιχίζονται τα νέα σημεία δεδομένων σε εκείνη την κατηγορία για την οποία ο αριθμός του γείτονα είναι μέγιστος.

Βήμα-6: Το μοντέλο μας είναι έτοιμο.

Δεν υπάρχουν προκαθορισμένες στατιστικές μέθοδοι για την εύρεση της πιο ευνοϊκής τιμής του K . Αρχικά παίρνουμε μια τυχαία τιμή K και ξεκινάμε τον υπολογισμό. Η επιλογή μιας σχετικά μικρής τιμής του K οδηγεί σε ασταθή όρια απόφασης. Ωστόσο αν εξάγουμε μια γραφική παράσταση μεταξύ του ποσοστού σφάλματος και του K που δηλώνει τιμές σε ένα καθορισμένο αριθμητικό εύρος είναι σαφώς πιο βοηθητικό. Επιλέγουμε την τιμή με το μικρότερο ποσοστό λάθους. Ωστόσο για να βελτιώσουμε το ποσοστό επιτυχίας του αλγορίθμου, εφαρμόζουμε τεχνικές όπως η παρακάτω:

2. Naive Bayes

Ο Naive Bayes είναι ένας από τους δημοφιλείς αλγόριθμους μηχανικής μάθησης που κάνει ταξινόμηση των δεδομένων με βάση τον υπολογισμό των τιμών πιθανότητας υπό όρους. Εφαρμόζει το θεώρημα Bayes για τον υπολογισμό και χρησιμοποιεί τα επίπεδα κλάσεων που αντιπροσωπεύονται ως τιμές χαρακτηριστικών ή διανύσματα προγνωστικών παραγόντων για ταξινόμηση. Ο αλγόριθμός του, ταιριάζει πολύ σε περιπτώσεις χρήσης πρόβλεψης σε πραγματικό χρόνο, πρόβλεψης πολλών κλάσεων, συστήματος συστάσεων, ταξινόμησης κειμένου και ανάλυ-

σης συναισθήματος. Ο αλγόριθμος Naive Bayes μπορεί να κατασκευαστεί χρησιμοποιώντας κατανομή Gaussian, Multinomial και Bernoulli. Αυτός ο αλγόριθμος είναι κλιμακωτός και εύκολος στην εφαρμογή για ένα μεγάλο σύνολο δεδομένων.

Ο τύπος ή η εξίσωση για τον υπολογισμό της μεταγενέστερης πιθανότητας είναι:

$$P(c|x) = \frac{(P(x|c)P(c))}{P(x)}$$

Χρησιμοποιείται στις παρακάτω περιπτώσεις:

- Πρόβλεψη σε πραγματικό χρόνο: Ο αλγόριθμος Naive Bayes είναι γρήγορος και πάντα έτοιμος για μάθηση, επομένως είναι ο καταλληλότερος για προβλέψεις σε πραγματικό χρόνο.
- Πρόβλεψη πολλαπλών κλάσεων: Η πιθανότητα πολλαπλών κλάσεων οποιασδήποτε μεταβλητής στόχου μπορεί να προβλεφθεί χρησιμοποιώντας έναν αλγόριθμο Naive Bayes.
- Σύστημα συστάσεων: Ο ταξινομητής Naive Bayes με τη βοήθεια του συνεργατικού φιλτραρίσματος δημιουργεί ένα Σύστημα Προτάσεων. Αυτό το σύστημα χρησιμοποιεί τεχνικές εξόρυξης δεδομένων και μηχανικής εκμάθησης για να φιλτράρει τις πληροφορίες που δεν έχουν δει προηγουμένως και στη συνέχεια να προβλέψει εάν ένας χρήστης θα εκτιμούσε έναν δεδομένο πόρο ή όχι.
- Ταξινόμηση κειμένου/ Ανάλυση συναισθήματος/ Φιλτράρισμα ανεπιθύμητης αλληλογραφίας: Λόγω της καλύτερης απόδοσής του με προβλήματα πολλαπλών κλάσεων και του κανόνα ανεξαρτησίας του, ο αλγόριθμος Naive Bayes αποδίδει καλύτερα ή έχει υψηλότερο ποσοστό επιτυχίας στην ταξινόμηση κειμένου. Ως εκ τούτου, χρησιμοποιείται στην ανάλυση συναισθήματος και στο φιλτράρισμα ανεπιθύμητων μηνυμάτων.

Τα πλεονεκτήματά του είναι:

- Εύκολος και γρήγορος στην εφαρμογή.
- Εάν ισχύει η υπόθεση της ανεξαρτησίας, τότε λειτουργεί πιο αποτελεσματικά από άλλους αλγόριθμους.
- Απαιτεί λιγότερα δεδομένα εκπαίδευσης.
- Είναι εξαιρετικά επεκτάσιμος.
- Μπορεί να κάνει πιθανολογικές προβλέψεις.

- Μπορεί να χειριστεί τόσο συνεχή όσο και διακριτά δεδομένα.
- Ανεπηρέαστος σε άσχετα χαρακτηριστικά.
- Μπορεί να λειτουργήσει εύκολα με τιμές που λείπουν.
- Εύκολη ενημέρωση κατά την άφιξη νέων δεδομένων.
- Ταιριάζει καλύτερα σε προβλήματα ταξινόμησης κειμένου.
Τα μειονεκτήματα είναι:
- Η ισχυρή υπόθεση ότι τα χαρακτηριστικά είναι ανεξάρτητα δεν είναι αληθής σε πραγματικές εφαρμογές.
- Έλλειψη δεδομένων.
- Πιθανότητες απώλειας ακρίβειας.
- Μηδενική συχνότητα, δηλαδή εάν η κατηγορία οποιασδήποτε κατηγορικής μεταβλητής δεν εμφανίζεται στο σύνολο δεδομένων εκπαίδευσης, τότε το μοντέλο εκχωρεί μηδενική πιθανότητα σε αυτήν την κατηγορία και, στη συνέχεια, δεν μπορεί να γίνει πρόβλεψη.

3. Support Vector Machine

ο SVM είναι ένας από τους πιο δημοφιλείς αλγόριθμους εποπτευόμενης μάθησης, ο οποίος χρησιμοποιείται για προβλήματα ταξινόμησης και παλινδρόμησης. Ωστόσο, κατά κύριο λόγο, χρησιμοποιείται για προβλήματα ταξινόμησης στη μηχανική μάθηση.

Ο στόχος του αλγόριθμου SVM είναι να δημιουργήσει τις καλύτερες προϋποθέσεις απόφασης που μπορεί να διαχωρίσει το χώρο n -διαστάσεων σε κλάσεις, ώστε να μπορούμε εύκολα να βάλουμε το νέο σημείο δεδομένων στη σωστή κατηγορία στο μέλλον. Αυτό το όριο της καλύτερης απόφασης ονομάζεται υπερεπίπεδο.

Ο SVM επιλέγει τα ακραία σημεία/διανύσματα που βοηθούν στη δημιουργία του υπερεπίπεδου. Αυτές οι ακραίες περιπτώσεις ονομάζονται διανύσματα υποστήριξης και ως εκ τούτου ο αλγόριθμος ονομάζεται Μηχανή Διανυσμάτων Υποστήριξης (ΜΔΥ).

Ο SVM μπορεί να είναι δύο τύπων:

Γραμμικός SVM: Ο γραμμικός SVM χρησιμοποιείται για γραμμικά δεδομένα που θέλουμε να διαχωρίσουμε, πράγμα που σημαίνει ότι εάν ένα σύνολο δεδομένων μπορεί να ταξινομηθεί σε δύο κατηγορίες χρησιμοποιώντας μία ευθεία γραμμή, τότε αυτά τα δεδομένα ονομάζονται γραμμικά δεδομένα και ο ταξινομητής χρησιμοποιείται ως γραμμικός ταξινομητής SVM.

Μη γραμμικός SVM: Ο μη γραμμικός SVM χρησιμοποιείται για μη γραμμικά δεδομένα που θέλουμε να διαχωρίσουμε, πράγμα που σημαίνει ότι εάν ένα σύνολο δεδομένων δεν μπορεί να ταξινομηθεί χρησιμοποιώντας μια ευθεία γραμμή, τότε αυτά τα δεδομένα ορίζονται ως μη γραμμικά δεδομένα και ο ταξινομητής που χρησιμοποιείται ονομάζεται μη- γραμμικός ταξινομητής SVM.

Πλεονεκτήματα:

- Λειτουργεί σχετικά καλά όταν υπάρχει ένα σαφές περιθώριο διαχωρισμού μεταξύ των κλάσεων.
- Είναι πιο αποτελεσματικό σε χώρους υψηλών διαστάσεων.
- Είναι αποτελεσματικό σε περιπτώσεις όπου ο αριθμός των διαστάσεων είναι μεγαλύτερος από τον αριθμό των δειγμάτων.
- Είναι σχετικά αποδοτικό στη μνήμη.

Μειονεκτήματα:

- Δεν είναι κατάλληλος για μεγάλα σύνολα δεδομένων.
- Δεν αποδίδει πολύ καλά όταν το σύνολο δεδομένων έχει περισσότερο θόρυβο, δηλαδή οι κλάσεις στόχου αλληλεπικαλύπτονται.
- Σε περιπτώσεις όπου ο αριθμός των δυνατοτήτων για κάθε σημείο δεδομένων υπερβαίνει τον αριθμό των δειγμάτων δεδομένων εκπαίδευσης, ο SVM θα έχει χαμηλή απόδοση.
- Καθώς ο ταξινομητής διανύσματος υποστήριξης λειτουργεί βάζοντας σημεία δεδομένων, πάνω και κάτω από το υπερεπίπεδο ταξινόμησης δεν υπάρχει πιθανολογική εξήγηση για την ταξινόμηση.

4. Random Forest

Ο αλγόριθμος Random Forest χρησιμοποιείται πολύ συχνά σε προβλήματα ταξινόμησης μιας και η δενδρική δόμη είναι πιο κατανοητή. Βασίζεται στην έννοια της εκμάθησης συνόλου, η οποία είναι μια διαδικασία συνδυασμού πολλαπλών ταξινομητών για την επίλυση ενός σύνθετου προβλήματος και τη βελτίωση της απόδοσης του μοντέλου.

Όπως υποδηλώνει το όνομα, "Random Forest" είναι ένας ταξινομητής που περιέχει έναν αριθμό δέντρων αποφάσεων σε διάφορα υποσύνολα του δεδομένου συνόλου δεδομένων και παίρνει τον μέσο όρο για να βελτιώσει την προγνωστική ακρίβεια αυτού του συνόλου δεδομένων. Αντί να βασίζεται σε ένα δέντρο απόφασης,

το τυχαίο δάσος παίρνει την πρόβλεψη από κάθε δέντρο και βασίζεται στην πλειοψηφία των ψήφων των προβλέψεων και προβλέπει την τελική έξοδο.

Ο μεγαλύτερος αριθμός δέντρων στο δάσος οδηγεί σε μεγαλύτερη ακρίβεια και αποτρέπει το πρόβλημα της υπερπροσαρμογής. Η διαδικασία εργασίας του αλγορίθμου μπορεί να εξηγηθεί στα παρακάτω βήματα:

Βήμα-1: Επιλέγουμε K τυχαία σημεία δεδομένων από το σετ εκπαίδευσης.

Βήμα-2: Δημιουργούμε τα δέντρα αποφάσεων που σχετίζονται με τα επιλεγμένα σημεία δεδομένων (Υποσύνολα).

Βήμα-3: Επιλέγουμε τον αριθμό N για τα δέντρα αποφάσεων που θέλετε να δημιουργήσουμε.

Βήμα-4: Επαναλαμβάνουμε τα βήματα 1 και 2

Βήμα-5: Για νέα σημεία δεδομένων, βρίσκουμε τις προβλέψεις κάθε δέντρου αποφάσεων και αντιστοιχίζουμε τα νέα σημεία δεδομένων στην κατηγορία που κερδίζει την πλειοψηφία.

Με λίγα λόγια, ξεκινάμε από τη ρίζα του δέντρου και ακολουθούμε την ακμή που αντιστοιχεί στην τιμή της μεταβλητής του άγνωστου στιγμιοτύπου. Έπειτα, πηγαίνουμε στον επόμενο κόμβο και επαναλαμβάνουμε την ίδια διαδικασία μέχρι να φτάσουμε στον τελικό κόμβο που οδηγεί και στην κατηγορία της κλάσης του στιγμιοτύπου. Έχοντας ένα σύνολο απόφασης μπορούμε να δημιουργήσουμε διαφορετικά δέντρα. Το κριτήριο που τα διαφοροποιεί είναι σε ποιους κόμβους θα βάλουμε τα γνωρίσματα. Κάθε φορά επιλέγουμε ένα γνώρισμα ανάλογα με το πόσο καλά διαχωρίζει τα δείγματα του συνόλου με βάση τις κατηγορίες.

Πλεονεκτήματα του Random Forest

- Είναι σε θέση να χειρίζεται μεγάλα σύνολα δεδομένων με υψηλή διάσταση.
- Βελτιώνει την ακρίβεια του μοντέλου και αποτρέπει το πρόβλημα overfitting. (Η υπερπροσαρμογή συμβαίνει όταν ένα μοντέλο μαθαίνει τις λεπτομέρειες και τον θόρυβο στα δεδομένα εκπαίδευσης, στο βαθμό που επηρεάζει αρνητικά την απόδοση του μοντέλου σε νέα δεδομένα. Αυτό σημαίνει ότι ο θόρυβος ή οι τυχαίες διακυμάνσεις στα δεδομένα εκπαίδευσης συλλέγονται και μαθαίνονται ως έννοιες από το μοντέλο)

Μειονεκτήματα του Random Forest

- Αν και το τυχαίο δάσος μπορεί να χρησιμοποιηθεί τόσο για εργασίες ταξινόμησης όσο και για εργασίες παλινδρόμησης, δεν είναι τόσο κατάλληλο για εργασίες παλινδρόμησης.

Κατά τη δημιουργία ενός μοντέλου μηχανικής μάθησης, στοχεύουμε στην υψηλή

προγνωστική απόδοση. Για να το κάνουμε αυτό χρειαζόμαστε κάποιο είδος τεχνικής αξιολόγησης, δηλαδή κάποιον τρόπο να ελέγξουμε αν το μοντέλο μας λειτουργεί καλά ή όχι. Το K-fold είναι μια από τις τεχνικές που μας βοηθούν να αξιολογήσουμε το μοντέλο μας.

k-Fold Cross-Validation

Η διασταυρούμενη επικύρωση είναι όταν το σύνολο δεδομένων χωρίζεται τυχαία σε ομάδες «k». Μία από τις ομάδες χρησιμοποιείται ως σετ δοκιμών και οι υπόλοιπες ως σετ εκπαίδευσης. Το μοντέλο εκπαιδεύεται στο σετ εκπαίδευσης και βαθμολογείται στο σετ δοκιμής. Στη συνέχεια, η διαδικασία επαναλαμβάνεται έως ότου κάθε μοναδική ομάδα χρησιμοποιηθεί ως σύνολο δοκιμής. Για παράδειγμα, για πενταπλάσια διασταυρούμενη επικύρωση, το σύνολο δεδομένων θα χωριστεί σε 5 ομάδες και το μοντέλο θα εκπαιδευτεί και θα δοκιμαστεί 5 ξεχωριστές φορές, ώστε κάθε ομάδα να έχει την ευκαιρία να αποτελέσει το σετ δοκιμής. Η μέθοδος **train-test-split** που χρησιμοποιείται κάποιες φορές ονομάζεται 'holdout'. Η διασταυρούμενη επικύρωση είναι καλύτερη από τη χρήση της μεθόδου 'holdout', επειδή η απόδοση της μεθόδου 'holdout' εξαρτάται από τον τρόπο με τον οποίο τα δεδομένα χωρίζονται σε σύνολα εκπαίδευσης και δοκιμής. Η διασταυρούμενη επικύρωση δίνει στο μοντέλο την ευκαιρία να δοκιμάσει σε πολλαπλούς διαχωρισμούς, ώστε να έχουμε μια καλύτερη ιδέα για την απόδοση του μοντέλου σε άρατα δεδομένα.

Μεθοδολογία και Πειράματα

3.1 Δεδομένα από φορητές συσκευές

Τα τελευταία χρόνια, οι φορητοί αισθητήρες έχουν προχωρήσει με εξειδικευμένη χρήση στη καθημερινή ζωή. Αυτοί οι σύγχρονοι αισθητήρες είναι μικροί σε μέγεθος, διακριτικοί και φοριούνται ακόμη και στους καρπούς του χρήστη. Παρά το μικρό τους μέγεθος, μπορούν να καταγράψουν μεγάλο αριθμό ποικίλων δεδομένων από καρδιακούς παλμούς έως και τη γεωγραφική θέση του χρήστη. Μπορούν να αποθηκεύσουν και να μεταδώσουν πληθώρα πληροφοριών. Αυτός ο πλούτος πληροφοριών οδήγησε σε δεδομένα πολύ μεγάλης κλίμακας. Τα big data είναι ένας όρος που περιγράφει μεγάλο όγκο δεδομένων που μπορούν να υποστούν επεξεργασία, να αναλυθούν και να αντληθούν γνώσεις, μοτίβα και άλλες τάσεις. Μπορούν να εκτελέσουν αναλύσεις που θα μπορούσαν να εντοπίσουν και να προβλέψουν την υγεία και τη συνολική φυσική κατάσταση με την πάροδο του χρόνου.

Οι μετρήσεις πραγματοποιούνται γενικά με βάση τις 24ωρες εγγραφές (μακροπρόθεσμες εγγραφές) ή σε μικρότερες περιόδους που κυμαίνονται από 0,5 έως 5 λεπτά (βραχυπρόθεσμες ηχογραφήσεις). Η χρήση μακροπρόθεσμων ή βραχυπρόθεσμων καταγραφών εξαρτάται από το είδος της μελέτης που πρέπει να πραγματοποιηθεί [53]. Στη παρούσα μελέτη, χρησιμοποιούνται μακροπρόθεσμες εγγραφές χωρισμένες σε τμήματα των έξι ωρών. Οι μεγαλύτερες περιόδους καταγραφής δεδομένων παρέχουν καλύτερα αποτελέσματα σε συνθήκες με απότομες καρδιακές αντιδράσεις σε έντονη περιβαλλοντική διέγερση. Οι παρατεταμένες περιόδους μέτρησης μπορούν να αναδείξουν

την απόκριση της καρδιάς σε μεταβαλλόμενο περιβάλλον , την κεντρική νευρική δραστηριότητα που περιλαμβάνει τα αντανακλαστικά, τον κινκάρδιο ρυθμό, συμπεριλαμβανομένου του κύκλου του ύπνου. Ωστόσο εικοσιτετράωρες ώρες εγγραφές αποκαλύπτουν τη συμβολή του ΚΝΣ στη διακύμανση καρδιακού παλμού και μαζί με τις παραπάνω διεργασίες, τις μετρήσεις της θερμοκρασίας σώματος και το σύστημα ρενίνης - αγγειοτενσίνης - αλδοστερόνης (που είναι υπεύθυνο για την αντιμετώπιση δυσμενών καταστάσεων υπέρτασης, καρδιακής ανεπάρκειας και ασθενειών που σχετίζονται με τα νεφρά) είναι το golden-standard της κλινικής αξιολόγησης HRV [48].

3.2 Μελέτη των Διαταραχών του Ύπνου

Η πεποίθηση ότι ο ύπνος, τα όνειρα και η ψύχωση συνδέονται μεταξύ τους σηματοδοτείται από τον Hughlings Jackson, ο οποίος, πριν από πάνω από έναν αιώνα, είπε: «Μάθετε για τα όνειρα και θα μάθετε για την παραφροσύνη» [23]. Η ανακάλυψη το 1953 ότι ο ύπνος ταχείας κίνησης των ματιών (REM) και η συσχέτισή του με το όνειρο οδήγησε σε ένα κύμα ενδιαφέροντος για τη συσχέτιση ύπνου και ψυχώσεων. Έχουν καταγραφεί, σε ηλεκτροεγκεφαλογραφικές εγγραφές (EEG), ανωμαλίες στον ύπνο για τη σχιζοφρένεια. Αυτές περιλαμβάνουν μειωμένο συνολικό ύπνο, μειωμένη συνέχεια ύπνου, καθυστέρηση (REM) και μειωμένη ή αργή δραστηριότητα δέλτα που αφορά στα στάδια βαθύτερου ύπνου [26]. Επίσης οι διαταραχές ύπνου συχνά εμφανίζονται πριν από τα κύρια κλινικά διαγνωστικά συμπτώματα και μπορεί να χρησιμεύσουν ως ισχυροί προγνωστικοί παράγοντες για την εκδήλωση της ψύχωσης, καθώς και των γνωστικών διαταραχών και της διάθεσης. Αυτό το ειδικό χαρακτηριστικό εστιάζει στη σχέση μεταξύ των προβλημάτων ύπνου και των διαταραχών στη μάθηση και τη μνήμη, ένα βασικό γνωστικό σύμπτωμα της σχιζοφρένειας. [10] Φυσιολογικά κατά τη διάρκεια του βαθύ ύπνου, η HRV δεν επηρεάζεται από εσωτερικούς ή εξωτερικούς παράγοντες όπως η άσκηση ή το ψυχικό στρες. Παρέχει μια καθαρή εικόνα για τη μέτρηση και την παρακολούθηση της HRV για μια χρονική περίοδο.

Ως επί το πλείστον, η μεταβλητότητα του καρδιακού παλμού θα πρέπει να είναι χαμηλότερη κατά τη διάρκεια του ύπνου από ό,τι όταν κάποιος δεν κοιμάται. Η καρδιά επιβραδύνεται καθώς το μεγαλύτερο μέρος του σώματος γίνεται ανενεργό. Η εξαίρεση σε αυτό είναι ο ύπνος (REM) , όταν ο εγκέφαλος γίνεται πιο ενεργός και το σώμα αναλόγως «ξυπνάει» περισσότερο. Ωστόσο, ο καρδιακός ρυθμός κατά τη διάρκεια του ύπνου (REM) δεν πρέπει να υπερβαίνει αυτόν που είναι κατά τη διάρκεια της ημέρας. Στη παρούσα εργασία, δημιουργείται μια βάση δεδομένων με τις εγγραφές του νυχτερινού ύπνου και άλλη μια βάση δεδομένων που έχει παρθεί από εξάωρα εγγραφών κατά τη διάρκεια της ημέρας με εύρος, από τις έξι το πρωί έως τις δώδεκα το βράδυ,

για κάθε περιστατικό ξεχωριστά. Οι δύο διαφορετικές βάσεις δεδομένων συγκρίνονται μεταξύ τους όσον αφορά τα γραμμικά και τα μη γραμμικά στοιχεία HRV.

3.3 Περιγραφή του συνόλου δεδομένων

Ανακτήθηκαν raw data από επτά φακέλους περιστατικών με id 5001, 5002, 5003, 5004, 5005, 5006 και 5019b. Τα δεδομένα είναι σε μορφή χρονοσειρών που περιγράφεται από την πλήρη ημερομηνία και ώρα με ακρίβεια millisecond. Το σύνολο δεδομένων αποτελείται κυρίως από πρωτογενή δεδομένα σε τριαξονική επιτάχυνση που μετράται στις τρεις χωρικές διαστάσεις X, Y, Z. Έτσι αναγνωρίζουν τη κίνηση που μπορεί να είναι διαφορετική από ασθενή σε ασθενή ή από τη κάθε δραστηριότητα τους. Ακόμη περιέχει φάκελο με τις καρδιολογικές μετρήσεις τους, που αφορούν το καρδιακό παλμό, τα διαστήματα RR σε καταγεγραμμένη χρονοσειρά με επικρατούσα τιμή συχνότητας 0,2 s, ένα φάκελο καταμέτρησης βημάτων και ένα φάκελο που έχει καταγράψει τον ύπνο τους, δηλαδή την ακριβή ώρα και ημερομηνία του πότε κοιμούνται και πότε ξυπνούν. Τα δεδομένα αφορούν τις ημερομηνίες συνολικής καταγραφής δεδομένων, που είτε ο ασθενής εμφανίζει υποτροπή καθόλη τη διάρκεια της καταγραφής, είτε δεν παρουσιάζει υποτροπή καθόλη τη διάρκεια υποτροπής και παρουσιάζονται στον πίνακα 3.1:

ID	Διάρκεια καταγραφής
5001	05/01/2020 - 20/01/2020
5002	23/05/2020 - 29/05/2020
5003	16/07/2020 - 22/07/2020
5004	11/08/2020 - 19/08/2020
5005	30/06/2020 - 07/07/2020
5006	07/05/2020 - 13/05/2020
5019	15/08/2020-22/08/2020

Πίνακας 3.1: Πίνακας με τις ημερομηνίες καταγραφής των δεδομένων

Τέσσερις ασθενείς δεν εμφάνισαν ψυχωσικά επεισόδια κατά τη διάρκεια καταγραφής των δεδομένων, ενώ οι υπόλοιποι τρεις υποτροπίασαν, με βάση την κλινική τους εκτίμηση. Λεπτομέρειες παρουσιάζονται στο πίνακα 3.2.

ID	Ημερομηνία Έναρξης	Ημερομηνία Τέλους
5003	2020-06-30	2020-08-30
5004	2020-08-11	2020-08-18
5019	2020-04-19 2020-07-31	2020-05-30 2020-09-09

Πίνακας 3.2: Πίνακας Ψυχωσικών Επεισοδίων

Γενικότερα κατά την ανάλυση των δεδομένων των ασθενών, παρατηρήθηκε ότι η μέγιστη διάρκεια ύπνου δεν ξεπέρασε τις τέσσερις ώρες, ενώ τις περισσότερες φορές διήρκησε από μισή ώρα έως δύο ώρες. Αυτό έδωσε το έναυσμα να γίνει σύγκριση του ύπνου των δύο ομάδων. Κατά τη διάρκεια του νυχτερινού ύπνου των ασθενών με υποτροπή, φάνηκε ότι οι δύο από τους τρεις ασθενείς είχαν κατά μεγάλη πλειοψηφία, διάρκεια ύπνου κάθε φορά λιγότερη από τέσσερις ώρες.

Για το περιστατικό 5019 παρατίθεται και ο αναλυτικός πίνακας καταγραφής ύπνου.

start_time	end_time	duration
2020-08-15 19:04:53.582400	2020-08-15 19:43:05.214953	0 days 00:38:11.632553
2020-08-16 01:34:17.098476	2020-08-16 03:32:21.838882	0 days 01:58:04.740406
2020-08-16 08:15:33.179219	2020-08-16 08:54:34.747184	0 days 00:39:01.567965
2020-08-16 23:35:53.688959	2020-08-17 01:16:57.736693	0 days 01:41:04.047734
2020-08-17 01:52:59.177213	2020-08-17 03:47:03.747182	0 days 01:54:04.569969
2020-08-17 10:55:09.822690	2020-08-17 12:00:12.429641	0 days 01:05:02.606951
2020-08-17 15:08:19.962879	2020-08-17 16:06:22.294879	0 days 00:58:02.332000
2020-08-18 09:01:40.378003	2020-08-18 09:43:42.061818	0 days 00:42:01.683815
2020-08-18 11:57:47.428322	2020-08-18 12:30:48.750924	0 days 00:33:01.322602
2020-08-19 01:17:11.403740	2020-08-19 04:01:17.978758	0 days 02:44:06.575018
2020-08-19 04:20:18.730436	2020-08-19 05:52:22.423290	0 days 01:32:03.692854
2020-08-19 14:12:35.529581	2020-08-19 15:18:38.178002	0 days 01:06:02.648421
2020-08-20 10:00:54.274015	2020-08-20 10:34:55.638199	0 days 00:34:01.364184
2020-08-20 15:11:06.690873	2020-08-20 16:11:09.099413	0 days 01:00:02.408540
2020-08-21 00:34:19.089027	2020-08-21 02:31:23.772922	0 days 01:57:04.683895
2020-08-21 06:05:32.363752	2020-08-21 07:13:35.096513	0 days 01:08:02.732761
2020-08-21 07:33:35.895367	2020-08-21 07:43:36.300328	0 days 00:10:00.404961
2020-08-22 01:57:55.309640	2020-08-22 06:10:05.426006	0 days 04:12:10.116366

Σχήμα 3.1: Δεδομένα Ύπνου για τον 5019

3.4 Αρχιτεκτονική Dataset

Η διαδικασία εξόρυξης δεδομένων χωρίζεται σε δύο μέρη την προεπεξεργασία δεδομένων και την εξόρυξη δεδομένων. Η προεπεξεργασία δεδομένων περιλαμβάνει καθαρισμό, ενωποίηση, μείωση και μετασχηματισμό δεδομένων. Το τμήμα εξόρυξης δεδομένων εκτελεί την, αξιολόγηση προτύπων και την οπτικοποίηση της γνώσης των δεδομένων.

Κάθε υποφάκελος αποτελείται από 10 έως 20 υποφακέλους που έχουν πάρει το όνομά τους από την αρχή και το τέλος της ημερομηνίας καταγραφής. Η παρούσα μελέτη αναλύει τις καρδιολογικές μετρήσεις των διαστημάτων RR σε εξάωρα διαστήματα, την καρδιολογική ανάλυση κατά τη διάρκεια του ύπνου τους, και συγκρίνει τα αποτελέσματα των αλγορίθμων μηχανικής μάθησης για την πρόγνωση των ψυχωσικών υποτροπών των περιστατικών.

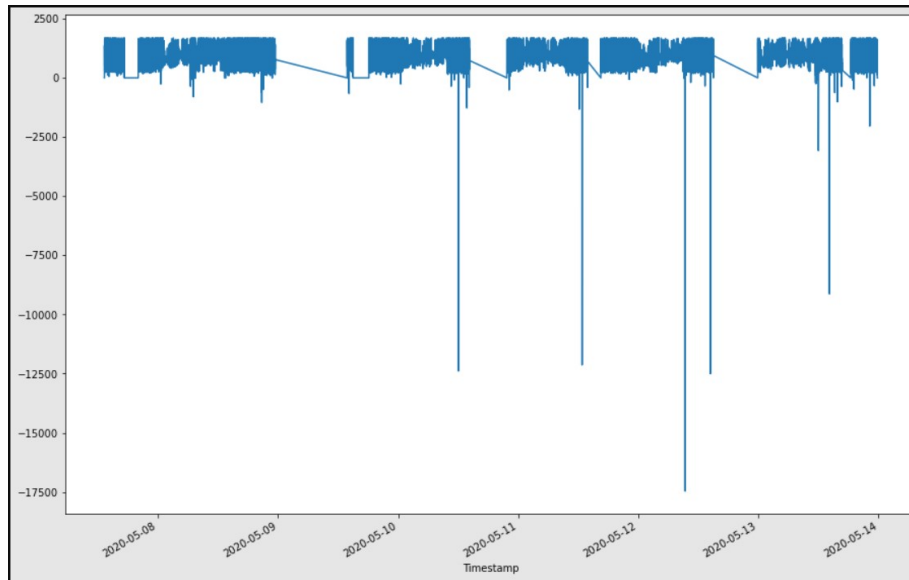
Αφού φορτώθηκαν οι κατάλληλες βιβλιοθήκες σε περιβάλλον προγραμματισμού Python, ακολουθήθηκε η παρακάτω διαδικασία για τις καρδιολογικές μετρήσεις:

- Φόρτωση και ανάλυση αρχείου parquet και μετασχηματισμού του σε csv.
- Μετασχηματίστηκε σε dataframe χρονοσειράς όπου και ανακτήθηκαν πληροφορίες που αφορούν το περιεχόμενο, τη δομή, το εύρος των ημερομηνιών και τους τύπους των δεδομένων για περαιτέρω ανάλυση. Ενδεικτικά, η εικόνα 3.2 παρουσιάζει το πίνακα με raw data από φορητή συσκευή, για το περιστατικό 5006 και η οπτικοποίηση:

	Timestamp	heartRate	rRInterval
0	2020-05-07 13:18:01.015483	0	0
1	2020-05-07 13:18:01.214990	0	0
2	2020-05-07 13:18:01.418369	0	0
3	2020-05-07 13:18:01.615287	0	0
4	2020-05-07 13:18:01.815439	0	0
...
95950	2020-05-13 23:48:48.165870	-3	0
95951	2020-05-13 23:48:48.366127	-3	0
95952	2020-05-13 23:48:48.565965	-3	0
95953	2020-05-13 23:48:48.766303	-3	0
95954	2020-05-13 23:48:48.966620	-3	0

[2095527 rows x 3 columns]

Σχήμα 3.2: Dataframe Χρονοσειράς



Σχήμα 3.3: Οπτικοποίηση ενοποιημένης χρονοσειράς του 5006

Από την εικόνα 3.3 βγαίνει το συμπέρασμα ότι υπάρχουν πολλές τιμές εκτός φυσιολογικού εύρους τιμών το οποίο ορίζεται στο διάστημα από 0,6-1,2 s που όμως τις κανονικοποιεί ο αλγόριθμος με συγκεκριμένη τεχνική, που η περιγραφή της, ακολουθεί στη συνέχεια. Οι ασυνέχειες οφείλονται στο γεγονός ότι δεν υπάρχουν εγγραφές ανάμεσα σε αυτές τις ημερομηνίες. Η επιλογή των διαστημάτων χρονοσειράς που επιλέγονται, αξιολογούνται από συγκεκριμένα κριτήρια που ορίζουμε για να βγουν ασφαλή συμπεράσματα.

- Στην αρχή είχε γίνει η επιλογή των εικοσιτετράωρων εγγραφών, μιας και έχει αποδειχθεί πειραματικά από δημοσιεύσεις ότι αποτελεί πηγή πιο έγκυρης καρδιολογικής πληροφόρησης, ωστόσο μεγαλύτερο ενδιαφέρον προκάλεσε η αξιολόγηση της καρδιολογικής μελέτης σε μη-γραμμικά αλλά και σε γραμμικά μεγέθη στους ασθενείς με σχιζοφρένεια, μεταξύ των ωρών της νύχτας και της ημέρας. Για να γίνει αυτό θα έπρεπε η εξόρυξη καρδιολογικών χαρακτηριστικών να πηγάζει από ισόποσης χρονικά επεξεργασίας RR διαστημάτων.
- Για τη βάση δεδομένων του ύπνου για τη καρδιολογική ανάλυση, ορίστηκε η τμηματοποίηση της χρονοσειράς από το εξής κριτήριο: επιλέχθηκαν συνεχόμενες τιμές με ελάχιστο εύρος χρόνου τις τέσσερις ώρες με ημερομηνία έναρξης και λήξης του νυχτερινού ύπνου από το φάκελο καταγραφής της φορητής συσκευής.
- Δε λήφθηκε υπόψιν ο ύπνος της ημέρας στα δεδομένα ύπνου αλλά ούτε συμπεριλή-

φθηξε και στις χρονοσειρές ημέρας.

- Ο ύπνος που ξεκινάει αργά το πρωί και τελειώνει έως τις 12 το μεσημέρι, θεωρείται νυχτερινός ύπνος και συμπεριλαμβάνεται στα δεδομένα νυχτερινού ύπνου.
- Οι χρονοσειρές ημέρας και νύχτας δεν ταυτίζονται.
- Όσο οι χρονοσειρές ημέρας αλλά και νυχτερινού ύπνου είναι σχετικά της ίδιας διάρκειας.

Οι λόγοι που δεν χρησιμοποιήσουμε ένα μόνο σημείο δεδομένων (για παράδειγμα 5 λεπτά) που συλλέγονται κατά τη διάρκεια της νύχτας είναι επειδή ο κικαάδιος ρυθμός και τα στάδια του ύπνου, επηρεάζουν τη φυσιολογία του ανθρώπινου οργανισμού στην αυτόνομη δραστηριότητα. Αυτό σημαίνει ότι υπάρχουν αλλαγές σε φυσιολογικές μεταβλητές όπως ο καρδιακός ρυθμός και η HRV, που εξαρτώνται από την ώρα της ημέρας (ή της νύχτας). Εάν χρησιμοποιήσουμε ένα σημείο δεδομένων που συλλέγεται στη 1 π.μ. μια μέρα και στις 4 το πρωί μια άλλη μέρα, τότε μπορεί να έχουμε μεγάλη μεταβλητότητα μεταξύ των βαθμολογιών απλώς και μόνο επειδή απέχουν πολύ μεταξύ τους. Με βάση τα παραπάνω κριτήρια, εξήχθηκαν από ένα έως δύο συνεχή εξάωρα τμήματα χρονοσειρών από κάθε ημέρα καταγραφής των ημερησίων δεδομένων, ενώ στο νυχτερινό ύπνο υπήρχαν νύχτες που δεν παρήγαγαν συνεχή τμήματα χρονοσειρών και σε όσες νύχτες είχαμε δεδομένα από αυτά, εξήχθηκαν ένα τετράωρο η εξάωρο συνεχών δεδομένων.

3.5 Τεχνικές Φιλτραρίσματος

Οι θόρυβοι που μπορεί να επηρεάσουν το σήμα χωρίζονται σε δύο κατηγορίες. Η πρώτη κατηγορία αναφέρεται σε τεχνικά χαρακτηριστικά που μπορεί να προκληθούν από την κίνηση του ασθενούς ή τη κακή επαφή δέρματος-φορητής συσκευής. Μπορούν επίσης να προκληθούν από το λογισμικό ανίχνευσης QRS όταν αποτυγχάνει να ανιχνεύσει σωστά τις κορυφές R. Η δεύτερη κατηγορία θορύβου περιλαμβάνει βιολογικά χαρακτηριστικά που είναι οι έκτοποι ρυθμοί. Στην πραγματικότητα, οι συσπάσεις και οι χαλαρώσεις της καρδιάς οφείλονται σε περιοδικές κινήσεις που προέρχονται από τον φλεβοκομβικό κόμβο. Όταν αυτές οι παρορμήσεις αποτυπώνονται σε ασυνήθιστα σημεία, οι παλμοί εμφανίζονται ακανόνιστοι και μη φυσιολογικοί. Για τις μεθόδους φιλτραρίσματος ακραίων και έκτοπων τιμών χρησιμοποιήθηκε το πακέτο HRV-analysis. Η παρουσία των έκτοπων τιμών μπορεί να επηρεάσει εξαιρετικά την HRV καθώς προκαλούν ασάφειες. Επομένως, οι έκτοποι παλμοί πρέπει πρώτα να ανιχνευθούν

και στη συνέχεια να διορθωθούν με αφαίρεση, αντικατάσταση ή με χρήση μεθόδων παρεμβολής. [38] Αυτές οι μέθοδοι είναι χρήσιμες για να μεταφερθεί το κανονικό διάστημα (NN-διαστήματα) από τα διαστήματα RR.

- Εισάγεται η τμηματοποιημένη χρονοσειρά,
- Αφαιρούνται οι ακραίες τιμές του σήματος που ορίζεται από ένα σύνολο τιμών από 300 έως 2000.
- Γίνεται αντικατάσταση των ακραίων τιμών αυτών καθώς και των κενών τιμών μέσω γραμμικής παρεμβολής. Είναι μια μέθοδος "filling the gaps" που περιγράφεται από την φόρμουλα:

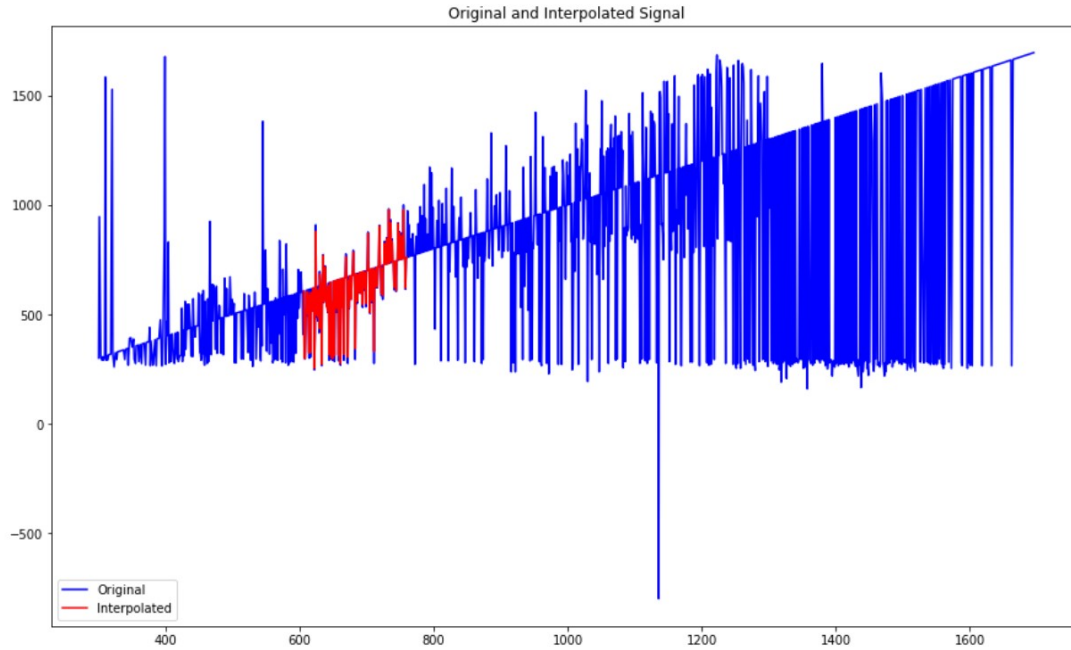
$$y = y1 + \frac{(x - x1)(y2 - y1)}{x2 - x1}$$

Με $(x1, y1)$ να είναι οι συντεταγμένες των "προβληματικών" σημείων, $(x2, y2)$ τα τελικά σημεία όπου x ορίζεται η θέση της παρεμβολής και y είναι η τελική τιμή που παίρνει. Έτσι γίνεται η προσαρμογή της καμπύλης χρησιμοποιώντας γραμμικά πολυώνυμα για την κατασκευή νέων σημείων δεδομένων, εντός του εύρους, ενός διακριτού συνόλου γνωστών σημείων δεδομένων.

- Στη συνέχεια αφαιρούνται οι έκτοπες τιμές με τη μέθοδο malik.
- Και αντικαθίστανται με τη μέθοδο της γραμμικής παρεμβολής

3.5.1 Οπτικοποίηση φιλτραρίσματος

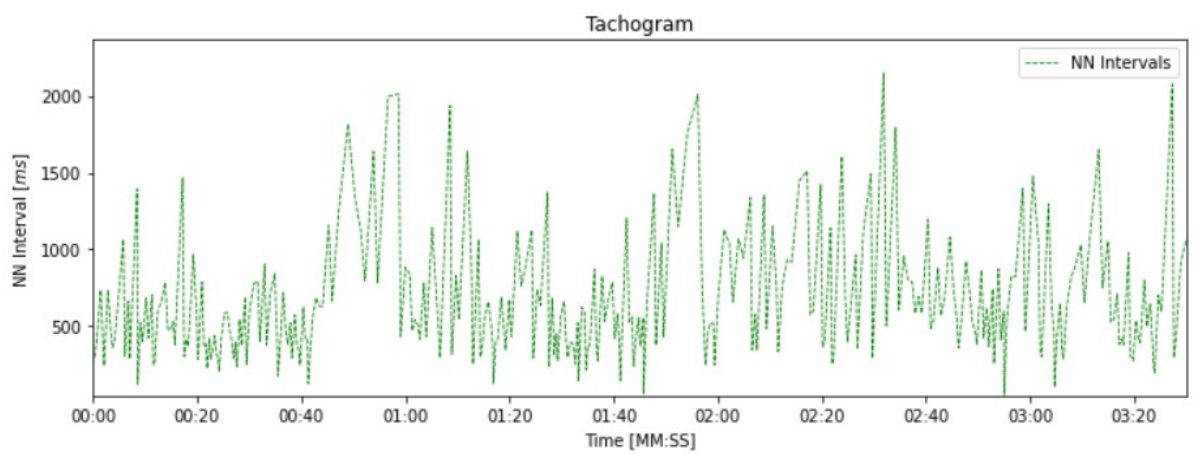
Πρώτα δημιουργούμε μια ομοιόμορφη χρονική γραμμή για τα διαστήματα R-R. Για να γίνει αυτό παίρνουμε τη θέση δείγματος όλων των κορυφών R, η οποία βρίσκεται στη λίστα που υπολογίστηκε με βάση τα βήματα της τεχνικής που περιγράφηκε προηγουμένως. Στη συνέχεια εκχωρούμε τις τιμές y σε αυτές τις θέσεις δείγματος από τη λίστα με τα δεδομένα από τη φορητή συσκευή, που περιέχει τη διάρκεια όλων των διαστημάτων R-R. Τέλος παρεμβάλλουμε το σήμα. Σημειώστε ότι η χρονική σειρά δεν ξεκινά από την πρώτη κορυφή, αλλά από τη θέση δείγματος της δεύτερης κορυφής R. Επειδή εργαζόμαστε με τα διαστήματα, το πρώτο διάστημα είναι διαθέσιμο στη δεύτερη κορυφή. Μπορούμε τώρα να χρησιμοποιήσουμε τις κατάλληλες συναρτήσεις που δημιουργήθηκαν για να βρούμε την τιμή y οποιασδήποτε θέσης x εντός της περιοχής του σήματος μας. Το συγκεκριμένο παράδειγμα του σχήματος 3.4 αφορά στον ύπνο του περιστατικού 5006.



Σχήμα 3.4: Παρεμβαλλόμενο-Πρωτότυπο Σήμα

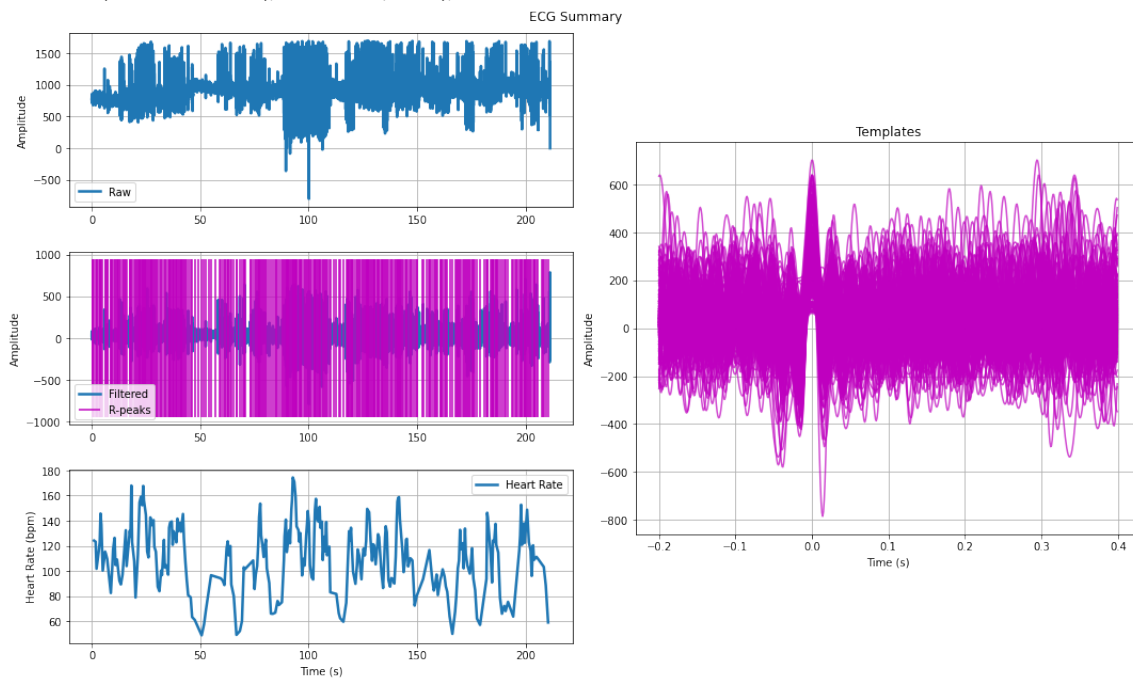
3.5.2 Ταχύγραμμα

Ο ταχογράφος αντιπροσωπεύει την ταχύτητα κάθε διαστήματος RR. Σχεδιάζοντας τον δείκτη κάθε υπολογιζόμενου διαστήματος RR στον άξονα x και την τιμή στον άξονα y. Για το περιστατικό 5006, το ταχύγραμμα των 3.5 λεπτών, είναι της 3.5 εικόνας.



Σχήμα 3.5: Ταχύγραμμα δεδομένων ύπνου
Ο ταχογράφος παρέχει πληροφορίες χρησιμοποιώντας μεταβλητότητα διαστήματος παλμό σε παλμό και μπορεί να βοηθήσει στον εντοπισμό διαφορετικών αρρυθμιών.

Στην καρδιολογική απεικονιστική σύνοψη 3.6 έχει γίνει επεξεργασία του ακατέργαστου σήματος και έχουν εξαχθεί τα σχετικά χαρακτηριστικά χρησιμοποιώντας προεπιλεγμένες παραμέτρους. Παρατηρούμε, τον καρδιακό παλμό από [0-200] sec κατά τη διάρκεια του ύπνου του ασθενούς 5006 κατά το χρονικό διάστημα της 2020-05-08 και τις ώρες 01:32:51.362040- 01:32:58.166712. Με τα πακέτα Biosppy, neurokit2 και pyhrn [15] της python, μπορούν να απαντηθούν οποιαδήποτε επιστημονικά ερωτήματα.



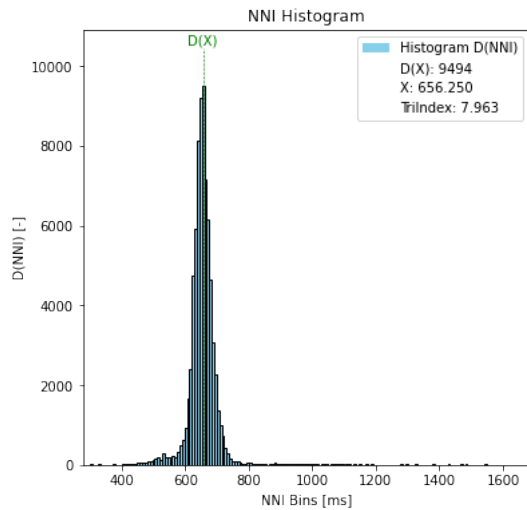
Σχήμα 3.6: ECG Summary

Από την τεχνική κανονικοποίησης των διαστημάτων RR γίνεται ο υπολογισμός των στατιστικών μετρικών της διακύμανσης καρδιακού παλμού και των μη γραμμικών στοιχείων που περιγράφηκαν στην προηγούμενη ενότητα. Ενδεικτικά για το περιστατικό 5006 αποτυπώνονται στον πίνακα της εικόνας 3.7.

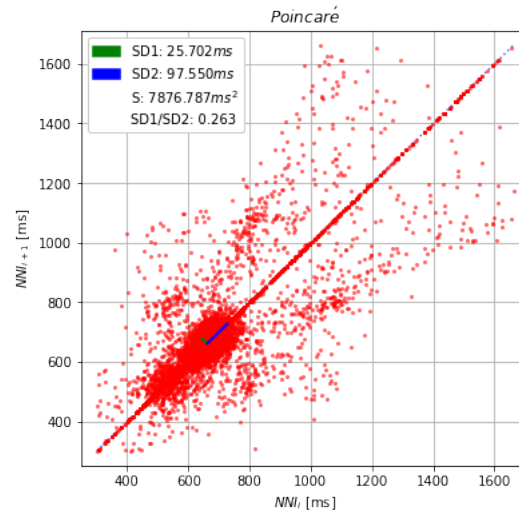
	5006_d2a	5006_d2b	5006_d2c	5006_d3	5006_d4	5006_d5a	5006_d5b	5006_d6	5006_d7a
PermEnt	0.439870	0.531549	0.587458	0.588921	0.479764	0.496135	0.556534	0.598096	0.48321
SVDEnt	0.275166	0.381058	0.44159	0.431925	0.289686	0.320311	0.39988	0.447109	0.326604
AppEnt	0.554771	0.744586	0.93413	0.893802	0.553329	0.622929	0.793115	0.924825	0.593743
SampleEnt	0.229561	0.338624	0.528319	0.439061	0.194093	0.218642	0.335507	0.547832	0.248015
ShannEn	7.975629	9.721945	9.695538	10.087308	7.793873	8.272391	9.65032	8.101402	8.153912
PetrosianFD	1.006431	1.00789	1.008262	1.008511	1.007407	1.007486	1.008487	1.008328	1.007493
KatzFD	3.372728	4.304581	3.837452	4.270904	3.591689	3.716012	3.825185	4.227923	3.700395
HiguchiFD	1.306493	1.430732	1.450208	1.46376	1.408193	1.428066	1.442174	1.463074	1.392545
DFA	0.870315	0.796358	0.729452	0.790793	0.887123	0.830307	0.798311	0.807322	0.827909
mean_nni	911.233220	765.265243	717.408888	700.99404	851.677686	822.995056	717.297898	696.089035	841.523083
snnn	189.033056	236.263529	254.896996	241.965081	216.67025	217.28068	220.182994	260.133279	210.659865
sdsd	77.207501	108.057983	130.339392	122.353	78.375835	87.971315	109.298659	129.830426	92.382507
nni_50	16176.000000	25716.0	34267.0	32602.0	17869.0	20591.0	28523.0	38474.0	19254.0
pnni_50	14.946363	24.832461	35.249038	32.953615	16.51082	19.025751	26.428049	35.549355	17.790385
nni_20	23254.000000	30931.0	38348.0	37264.0	25174.0	28122.0	33746.0	43893.0	26749.0
pnni_20	21.486320	29.868286	39.446993	37.665895	23.260584	25.984274	31.267431	40.556423	24.715644
rmssd	77.207501	108.057983	130.339392	122.353	78.375835	87.971315	109.298659	129.830427	92.382507
median_nni	915.000000	708.0	673.0	657.0	909.0	881.0	679.0	637.0	890.0
range_nni	1396.000000	1395.0	1398.0	1399.0	1393.0	1396.0	1398.0	1393.0	1396.0
cvsd	0.084729	0.141203	0.181681	0.174542	0.092025	0.106892	0.152376	0.186514	0.10978
cvnni	0.207448	0.308734	0.355302	0.345174	0.254404	0.264012	0.306962	0.373707	0.250332

Σχήμα 3.7: Ημερήσια HRV ανάλυση για τον 5006

Αφορά την επταήμερη καταγραφή δεδομένων. Έχουν συλλεχθεί 3 αρχεία εξάωρων εγγραφών από την δεύτερη μέρα, ένα από την τρίτη και την τέταρτη μέρα, δύο αρχεία από την πέμπτη μέρα, ένα αρχείο από την έκτη μέρα και ένα τελευταίο από την έβδομη μέρα.



Σχήμα 3.8: TriIndex
νυχτερινού ύπνου του
5019

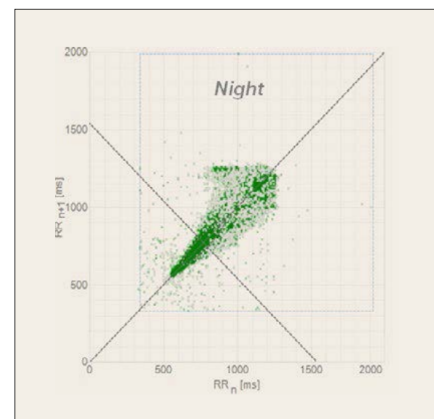
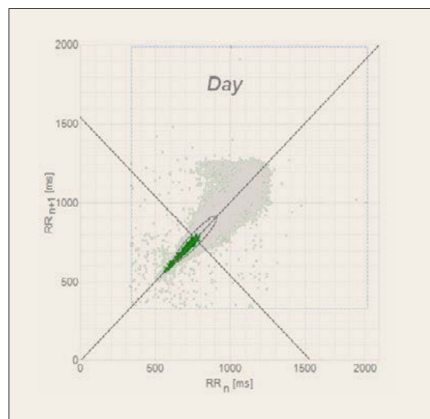


Σχήμα 3.9: Poincaré
Plot νυχτερινού ύπνου
του 5019

Το ιστόγραμμα αξιολογεί τη σχέση μεταξύ του συνολικού αριθμού των διαστημάτων RR που ανιχνεύθηκαν και της διακύμανσης του διαστήματος RR, με τον άξονα των ψ να απεικονίζει τον αριθμό των παλμών και τον άξονα των χ να απεικονίζει τη διάρκεια του παλμού. Η γραφική παράσταση των περιοχών των καρδιακών παλμών θα πρέπει να είναι σχετικά ομοιόμορφη με την πάροδο του χρόνου, αν και είναι σύνηθες να φαρδύνει τη νύχτα. Όταν το εύρος HR είναι συνήθως διασκορπισμένο κατά τη διάρκεια της ημέρας είναι πιθανό να αντιπροσωπεύει έναν μη φυσιολογικό ρυθμό. Στενότερα ιστογράμματα όπως το ιστόγραμμα του 5019 συσχετίζεται με μειωμένο HRV. Ο τριγωνικός δείκτης HRV <15 μπορεί να χαρακτηρίσει κάποιο άτομο σοβαρά καταθλιπτικό και HRV τριγωνικός δείκτης <20 μπορεί να το χαρακτηρίσει με μέτρια κατάθλιψη. [51] Στη συγκεκριμένη περίπτωση έχουμε δείκτη ίσο με 7,953. Αναφορικά με το poincare plot, κάθε ζευγάρι διαστημάτων RR εμφανίζεται ως μία κουκκίδα στο γράφημα. Όσο μεγαλύτερη είναι η εγγραφή, τόσο περισσότερες κουκκίδες εμφανίζονται στο γράφημα. Η διάρκεια του τρέχοντος καρδιακού παλμού (RRn) εμφανίζεται στον άξονα x και η διάρκεια του επόμενου παλμού (RRn+1) στον άξονα y. Με αυτόν τον τρόπο κάθε σημείο της γραφικής παράστασης Poincaré περιγράφεται στον χώρο (RRn , RRn+1). Όλα τα σημεία που περιγράφονται από διαδοχικούς καρδιακούς παλμούς ίσης διάρκειας (RRn = RRn+1) βρίσκονται στη γραμμή ταυτότητας. Τα σημεία πάνω από τη γραμμή ταυτότητας αντιστοιχούν σε όλες τις παρατάσεις (RRn < RRn+1) και τα σημεία κάτω από αυτή τη γραμμή αντιπροσωπεύουν όλες τις συντομεύσεις του διαστήματος μεταξύ 2 διαδοχικών παλμών (RRn > RRn+1)

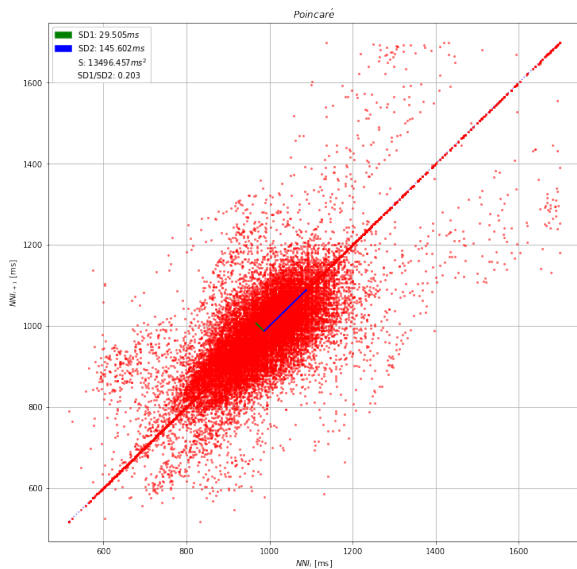
Αξίζει να σημειωθεί ένα πρότυπο HRV σε poincare plot υγιούς ατόμου που συγκρίνει καρδιολογικά δεδομένα ημέρας και νύχτας.¹

LOW HRV DURING THE DAY, NORMAL HRV AT NIGHT:

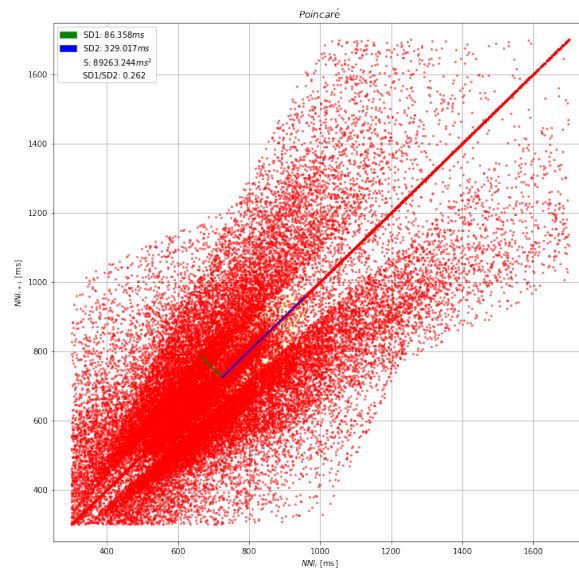


Σχήμα 3.10: Poincare plot νύχτας και ημέρας υγιούς ατόμου

Για να υπάρχει μια αίσθηση σύγκρισης παρατίθεται το γράφημα poincare , ασθενούς που υποτροπιάζει και συγκεκριμένα του 5003. Για το χρονικό διάστημα στις 18/07/21 από τις 00:20 έως τις 06:29 κατά τη διάρκεια του νυχτερινού του και δεξιά, για το χρονικό διάστημα στις 18/7/21 από τις 18:00 έως τις 23:59.



Σχήμα 3.11: Pplot νυχτερινού ύπνου



Σχήμα 3.12: Pplot ημερήσιας κατανομής

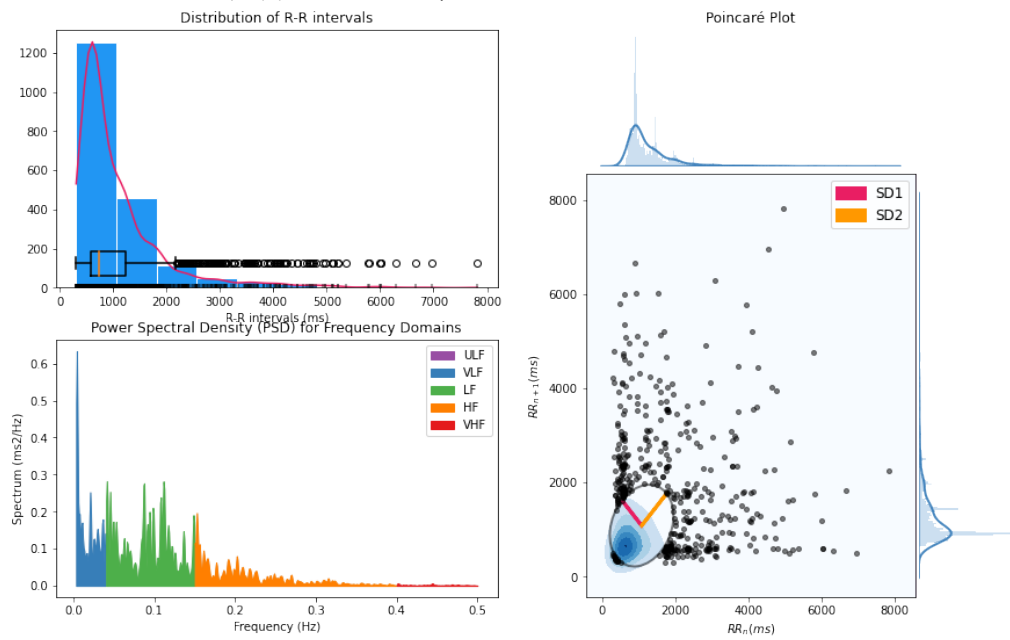
Κατά τις ώρες της ημέρας που περιγράφονται από το 3.12 γράφημα παρατηρούμε ότι το κύριο σύμπλεγμα των κουκκίδων εμφανίζεται στην επάνω δεξιά γωνία. Μια τέτοια εμφάνιση είναι η απόδειξη ότι το παρασυμπαθητικό νευρικό σύστημα έχει υψηλή επίδραση στον καρδιακό ρυθμό του ασθενούς. Τότε υπάρχει βραδυκαρδία. Συχνά, το κύριο σύμπλεγμα που βρίσκεται στην επάνω δεξιά θέση που τείνει να είναι φαρδιά ανταναχλά το γεγονός ότι το HRV γίνεται υψηλότερο με χαμηλότερους καρδιακούς παλμούς/υψηλότερο RR. Η ύπαρξη άλλων συμπλεγμάτων (ή ακόμα και μεμονωμένες κουκκίδες) όπως φαίνεται και από το γράφημα, που διαχωρίζονται σημαντικά από την κύρια συστάδα, μπορεί να αναμένει την πιθανότητα εμφάνισης αρρυθμίας. Συνήθως αυτές οι ομάδες (μεμονωμένες κουκκίδες) αντιπροσωπεύουν πρόωρους υπερκοιλιακούς ή κοιλιακούς παλμούς ή παύσεις. Ο μεγάλος αριθμός κουκκίδων που δεν έχουν την τάση να δημιουργούν συστάδες και μάλλον φαίνεται να κατανέμονται τυχαία (φαίνεται θρυμματισμένο) στο γράφημα, μπορεί να είναι μόνο ένα σημάδι θορύβου, αλλά μερικές φορές αποτελούν απόδειξη της ύπαρξης κολπικής μαρμαρυγής ή κολπικού πτερυγισμού.

Σχετικά σημαντική είναι η αναφορά τιμών σε κάποιες παραμέτρους όπως τις αναφέρει το pyhrn από την Python [15] του πίνακα 3.3. Να σημειωθεί ότι πρόκειται για μετρήσεις υγιών ατόμων που δεν παρουσιάζουν παθολογική αρρυθμία, ενώ στην παρούσα μελέτη συγκρίνουμε ασθενείς με σχιζοφρένεια που κάποιοι εμφανίζουν ψυχωσικές υποτροπές και κάποιοι όχι.

Παράμετρος	Μονάδα Μέτρησης	pyHRV
Mean NNI	ms	836.6884 ±143.429
Mean HR		74.505 ±13.545
SDNN	ms	65.937 ±26.291
SDANN	ms	41.397 ±18.281
rmssd	ms	42.485 ±22.611
NN50	-	58.120 ±22.611
pNN50	%	17.187 ±15.440
TINN	ms	110.313 ±59.1
SD1	ms	30.041 ±15.988
SD2	ms	87.171 ±35.907
SD2 SD1	-	3.204 ±1.306
Sample En	-	1.332 ±0.365
DFA a1	-	1.204 ±0.264
DFA a2	-	0.886 ±0.215

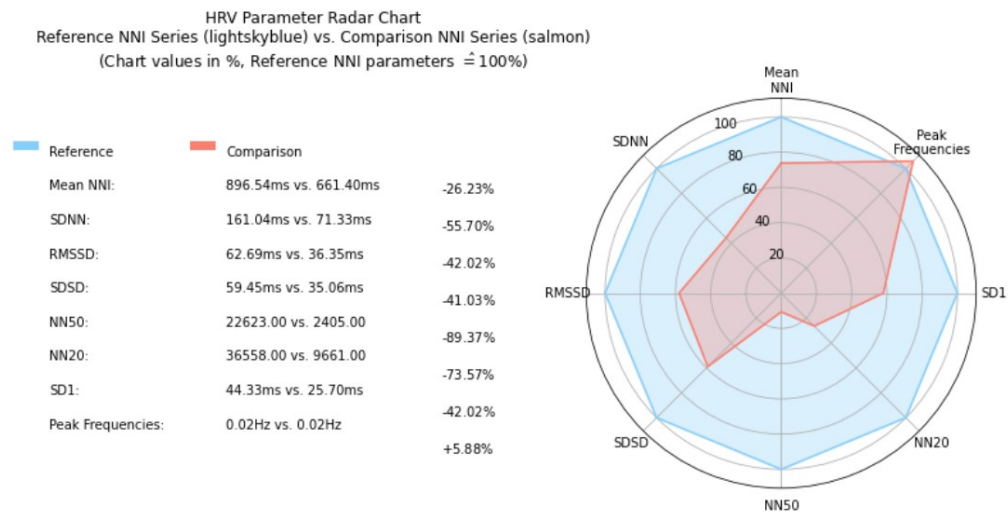
Πίνακας 3.3: Τιμές αναφοράς σε υγιή άτομα

Αξίζει να σημειωθεί το αποτύπωμα της κατανομής των RR διαστημάτων, ο χάρτης Poincare και το διάγραμμα Welch της εικόνας 3.13.



Σχήμα 3.13: Κατανομή διαστημάτων RR του 5006

Ακολουθεί η σύγκριση των παραμέτρων SDNN, MeanNNi, rmssd, SDSD, NN50 μεταξύ του ασθενούς 5006 κατά τη διάρκεια του ύπνου του που είναι ασθενής χωρίς υποτροπή, ως σημείο αναφοράς και του 5019 κατά τη διάρκεια του δικού του ύπνου που είναι ασθενής με υποτροπή. Η σύγκριση γίνεται μέσω της εικόνας 3.14 που ακολουθεί.



Σχήμα 3.14: Σύγκριση μεταβλητών μεταξύ ασθενών

3.6 Στατιστικός Έλεγχος

Τα ερωτήματα που τίθενται είναι τα εξής:

- Έχουν διαφορά οι μετρήσεις ενός νοσούντα μεταξύ της ημέρας και της νύχτας;
- Έχουν διαφορά οι μετρήσεις ενός μη νοσούντα μεταξύ της ημέρας και της νύχτας;
- Έχουν διαφορά οι μετρήσεις της νύχτας στους υποτροπιάζοντες και στους μη-υποτροπιάζοντες;
- Έχουν διαφορά οι μετρήσεις της ημέρας στους υποτροπιάζοντες και στους μη-υποτροπιάζοντες;
- Ποιες μετρήσεις έχουν μεγαλύτερη βαρύτητα οι γραμμικές ή οι μη γραμμικές τόσο στην ημέρα αλλά και στη νύχτα μεταξύ των δύο γκρουπ;

Πραγματοποιήθηκε στατιστικός έλεγχος Kruskal-Wallis για τις άνισες διακυμάνσεις ανά δύο μεταβλητών στα δύο γκρουπ την ημέρα και τη νύχτα. Αυτή η μη παραμετρική στατιστική δοκιμή, έγινε ώστε να ελεγχθεί η υπόθεση εάν οι δύο πληθυσμοί έχουν ίσους μέσους όρους. Η απόρριψη της μηδενικής υπόθεσης υποδηλώνει ότι υπάρχουν αρκετά στοιχεία που αποδεικνύουν ότι ένα ή περισσότερα δείγματα κυριαρχούν σε άλλο δείγμα, αλλά το τεστ δεν υποδεικνύει ποια δείγματα. Όταν το Kruskal-Wallis H-test οδηγεί σε σημαντικά αποτελέσματα, τότε τουλάχιστον ένα από τα δείγματα είναι διαφορετικό από τα άλλα δείγματα.

3.7 Συμπεράσματα Στατιστικού Ελέγχου

Για τον έλεγχο υποθέσεων ίσων μέσων όρων μεταξύ των γκρουπ που έχουν άνισες διακυμάνσεις:

- των ασθενών χωρίς υποτροπές τη νύχτα και των ασθενών χωρίς υποτροπές την ημέρα, που έχουν άνισες διακυμάνσεις, διαπιστώθηκε ότι παρουσιάζουν πολύ μεγάλη στατιστική σημαντικότητα, σχεδόν σε όλες τις μεταβλητές με $0.01 > p\text{-value} > 0.000000005$. Η υπόθεση H_0 , $\alpha=0.05$, ότι τα δείγματα ανήκουν στην ίδια κατανομή, απορρίπτεται. Με εξαίρεση την μεταβλητή SDANN με $p\text{-value}=0.571$.
- των ασθενών με ψύχωση την ημέρα και των ασθενών με ψύχωση τη νύχτα, προκύπτει στατιστικά σημαντική διαφορά στις περισσότερες μεταβλητές από $0.03 > p\text{-value} > 0.0000003$ όπου απορρίπτεται η αρχική υπόθεση των δειγμάτων της ίδιας κατανομής, με εξαίρεση στις μεταβλητές SDANN με $p\text{-value}=0.451$, την Tri-Index με $p\text{-value}=0.203$ και την Petrosian με $p\text{-value}=0.053$.
- των ασθενών χωρίς υποτροπές την ημέρα και των ασθενών με υποτροπές την ημέρα, που έχουν άνισες διακυμάνσεις, διαπιστώθηκε ότι δεν παρουσιάζουν στατιστική σημαντικότητα. Άρα η αρχική υπόθεση H_0 , $\alpha=0.05$ των ίσων μέσων όρων των γκρουπ, είναι αποδεκτή.
- των ασθενών χωρίς υποτροπές και ασθενών με υποτροπές κατά τη διάρκεια της νύχτας, διαπιστώθηκε μικρότερη σημαντικότητα από την παραπάνω περίπτωση.

Άρα η αρχική υπόθεση H_0 , $\alpha=0.05$ των ίσων μέσων όρων των γκρουπ, είναι αποδεκτή.

Έτσι και στην τελευταία περίπτωση η υπόθεση ότι είναι ίσοι οι μέσοι όροι των μεταβλητών καρδιακής διακύμανσης απορρίπτεται.

Αυτό σημαίνει ότι διαχωρίζεται έντονα η HRV της ημέρας από την HRV κατά τη διάρκεια του νυχτερινού ύπνου μεταξύ των ασθενών που δεν παρουσιάζουν ψυχωσικές υποτροπές και λιγότερο μεταξύ των ασθενών με ψύχωση, κατά τη διάρκεια της ημέρας από τη διάρκεια του νυχτερινού τους ύπνου.

Όσο οι τιμές των μεταβλητών της καρδιακής διακύμανσης μειώνονται, τότε υπάρχει στρες. Αυτό γίνεται εμφανές σε όλες τις μεταβλητές σχεδόν. Αξίζει να σημειωθεί ότι, κατά την διάρκεια του ύπνου των τυπικών ανθρώπων, ο καρδιακός ρυθμός επιβραδύνεται. Όταν εισέρχονται στα βαθύτερα στάδια του ύπνου, ο καρδιακός παλμός ηρεμίας μειώνεται κατά περίπου 20 έως 30 τοις εκατό σε σχέση με τον ημερήσιο καρδιακό παλμό ηρεμίας². Μελέτες δείχνουν ότι τα HRV χαμηλών τιμών ύπνου μπορεί να υποδεικνύουν βαριές διαταραχές του ύπνου, αδυναμία βύθισης σε πιο βαθειά στάδια REM ή λήψη αντιψυχωσικών φαρμάκων [35].

²Όπως αναφέρεται από την Alicia Roth, PhD, DBSM στο άρθρο της στην ιστοσελίδα: <https://amerisleep.com/blog/heart-rate-variability-and-sleep/>

3.8 Συσχέτιση Μεταβλητών

Έγινε test που ελέγχει τη μηδενική υπόθεση ότι ένα δείγμα προέρχεται από μια κανονική κατανομή. Στη βάση δεδομένων που αφορά το νυχτερινό ύπνου οι 7 από τις 32 μεταβλητές δεν έχουν κανονική κατανομή με $p\text{-value} < 0.0001$ και στη βάση δεδομένων που αφορά τα ημερήσια καρδιολογικά δεδομένα οι 20 από τις 32 μεταβλητές δεν έχουν κανονική κατανομή με $p\text{-value} < 0.0001$. Συνεπώς έγινε έλεγχος συσχέτισης Spearman ανά δύο μεταβλητών στα δύο γκρουπ. Για το γκρουπ των δεδομένων νυχτερινού ύπνου παρατηρούνται οι υψηλές συσχετίσεις στον πίνακα 3.4 που αφορά το νυχτερινό ύπνο και στον πίνακα 3.5 για τα δεδομένα ημέρας.

Πιο αναλυτικά για το γκρουπ του νυχτερινού ύπνου υπάρχει απόλυτη συσχέτιση των μεταβλητών:

- rmssd-sdsd ($p\text{-value}=1.33\text{e-}174$)
- sdsd-sd1 ($p\text{-value}=0.000000\text{e+}00$)
- rmssd-sd1 ($p\text{-value}=0.000000\text{e+}00$) που είναι σύμφωνο με τη βιβλιογραφία [11]

Ακολουθεί η συσχέτιση με συντελεστή 0,99 για τις μεταβλητές,

- sdnn-sd2 ($p\text{-value}=3.115400\text{e-}43$)
- cvsd-SVDEnt ($p\text{-value}=2.887850\text{e-}28$)
- cvnni-std hr ($p\text{-value}=3.806249\text{e-}12$)
- PermEnt-HiguchiFD ($p\text{-value}=2.197646\text{e-}08$)

και με συντελεστή συσχέτισης 0,98 οι μεταβλητές:

- sdsd-ellipse area ($p\text{-value}=1.939018\text{e-}13$)
- sdsd-KatzFD ($p\text{-value}=9.199603\text{e-}10$)
- rmssd-KatzFD ($p\text{-value}=9.199603\text{e-}10$)
- sd1-KatzFD ($p\text{-value}=9.199603\text{e-}10$)

Για το γκρουπ των καρδιολογικών δεδομένων ημέρας υπάρχει απόλυτη ταύτιση μεταβλητών στις:

- rmssd-sdsd ($p\text{-value}=0.000000\text{e+}00$)

Ακολουθεί η συσχέτιση με τιμή συντελεστή 0,99 για τις μεταβλητές,

- sdnn-sd2 ($p\text{-value}=5.685470\text{e-}68$)
- sdsd-pnni 50 ($p\text{-value}=2.103276\text{e-}36$)
- sdsd pnni 20 ($p\text{-value}= 1.237852\text{e-}30$)

- sdsd- App Ent (p-value=7.439965e-34)
- sdsd- Sample Ent (p-value=9.501643e-19)
- sdsd-sd1 (p-value=1.104518e-86)
- nni 50-nni 20 (p-value=3.145773e-59)
- pnni 50-pnni 20 (p-value=1.232123e-52)
- pnni 50-rmssd (p-value=2.103275e-36)
- pnni 50- Sample Ent (p-value=3.346908e-29)
- pnni50 - sd1 (p-value=6.533227e-36)
- pnni 20- rmssd (p-value=1.237851e-30)
- pnni 20-App Ent (p-value=2.853965e-42)
- rmssd-App Ent (p-value=7.439962e-34)
- rmssd-Sample Ent (p-value= 9.501643e-19)
- rmssd-sd1 (p-value=1.104518e-86)
- cvsd-Perm Ent (p-value=1.637576e-30)
- cvsd-SVD Ent (p-value=3.298435e-50)
- cvnni-std hr (p-value=1.444513e-36)

και με συντελεστή συσχέτισης 0,98 οι μεταβλητές:

- pnni 50-App Ent (p-value=1.471338e-49)
- pnni 20-cvsd (p-value=1.816588e-30)
- pnni20-sd1 (p-value=4.620787e-30)
- pnni20-SVD Ent (p-value=7.023808e-24)

sdnn	cvnni	0.95
sdnn	sd2	0.99
sdnn	ellipse area	0.96
sdsd	nni 50	0.96
sdsd	pnni 50	0.95
sdsd	nni 20	0.93
sdsd	pnni 20	0.96
sdsd	rmssd	1.0
sdsd	KatzFD	0.98
sdsd	sd1	1.0
sdsd	ellipse area	0.98
nni 50	pnni 50	0.93
nni 50	nni 20	0.96
nni 50	rmssd	0.967
nni 50	KatzFD	0.97
nni 50	sd1	0.97
nni 50	ellipse area	0.97
pnni 50	rmssd	0.96
pnni 50	KatzFD	0.97
pnni 50	sd1	0.96
pnni 50	ellipse area	0.95
nni 20	rmssd	0.93
nni 20	sd1	0.93
pnni 20	rmssd	0.96
pnni 20	cvsd	0.96
pnni 20	SVDEnt	0.96
pnni 20	KatzFD	0.93
pnni 20	sd1	0.96
rmssd	KatzFD	0.98
rmssd	sd1	1.0
rmssd	ellipse area	0.98
range nni	std hr	0.95
cvsd	SVDEnt	0.99
cvnni	std hr	0.99
cvnni	sd2	0.95
cvnni	ellipse area	0.93
mean hr	PetrosianFD	0.96
std hr	sd2	0.93
PermEnt	HiguchiFD	0.99
KatzFD	sd1	0.98
KatzFD	ellipse area	0.95
sd1	ellipse area	0.98
sd2	ellipse area	0.96

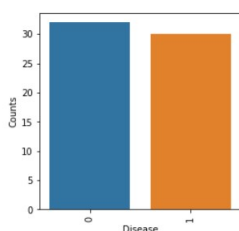
Πίνακας 3.4: Υσχυρή συσχέτιση μεταβλητών με τη μέθοδο Spearman's στο νυχτερινό ύπνο.

sdsn	sd2	0.997
sdsd	pnni 50	0.996
sdsd	pnni 20	0.99
sdsd	rmssd	1.0
sdsd	cvnni	0.96
sdsd	AppEnt	0.99
sdsd	SampleEnt	0.99
sdsd	sd1	0.99
sdsd	ellipse area	0.96
nni 50	nni 20	0.99
pnni 50	pnni 20	0.99
pnni 50	rmssd	0.99
pnni 50	cvsd	0.95
pnni 50	cvnni	0.97
pnni 50	SVDEnt	0.95
pnni 50	AppEnt	0.98
pnni 50	SampleEnt	0.99
pnni 50	sd1	0.99
pnni 50	ellipse area	0.95
pnni 20	rmssd	0.99
pnni 20	cvsd	0.98
pnni 20	cvnni	0.96
pnni 20	PermEnt	0.95
pnni 20	SVDEnt	0.98
pnni 20	AppEnt	0.99
pnni 20	SampleEn	0.97
pnni 20	sd1	0.98
rmssd	cvnni	0.95
rmssd	AppEnt	0.99
rmssd	SampleEnt	0.99
rmssd	sd1	0.999
rmssd	ellipse area	0.96
cvsd	PermEnt	0.99
cvsd	SVDEnt	0.99
cvsd	AppEnt	0.96
cvnni	std hr	0.99
cvnni	AppEnt	0.96
cvnni	SampleEnt	0.96
cvnni	sd1	0.95
mean hr	HiguchiFD	0.97
std hr	SampleEnt	0.95
std hr	ellipse area	0.96

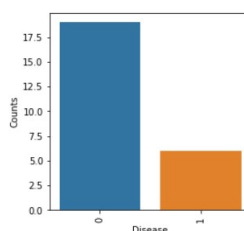
Πίνακας 3.5: Υσχυρή συσχέτιση μεταβλητών με τη μέθοδο Spearman's σε δεδομένα ημέρας.

3.9 Μηχανική Μάθηση

Κατά την προεπεξεργασία δεδομένων ελέγχουμε την ισορροπία του δείγματος των κλάσεων της ασθένειας που ορίζουμε ως ψυχωσική υποτροπή. Παρατηρούμε από τις εικόνες 3.15 και 3.16 ότι η βάση των δεδομένων της ημερήσιας ανάλυσης έχει ισορροπημένο δείγμα με 30 εγγραφές της κλάσης 1 που είναι οι νοσούντες από 30 εγγραφές της κλάσης 0, που είναι οι μη νοσούντες. Ενώ στη δεύτερη βάση δεδομένων που αφορούν το νυχτερινό ύπνο, έχουμε ένα δείγμα που αποτελείται από 6 εγγραφές της κλάσης 1 των νοσούντων και 19 εγγραφές της κλάσης 0, των μη νοσούντων.



Σχήμα 3.15: Κλάσεις ημερήσιων δεδομένων



Σχήμα 3.16: Κλάσεις δεδομένων νυχτερινού ύπνου

Ακολουθεί μια πρώτη ματιά των δεδομένων στις πιο αντιπροσωπευτικές μεταβλητές μετά από τη συσχέτισή τους στο γκρουπ των ημερησίων δεδομένων όπου υπάρχει μια σχετική ομοιομορφία σε σχέση με τους νοσούντες και μη νοσούντες. Ίσως μπορεί να ειπωθεί, ότι μειωμένο HRV σχετίζεται με τη σοβαρότητα των ψυχωτικών συμπτωμάτων και υποδηλώνοντας πιθανή εμπλοκή της νευροαυτόνομης δυσλειτουργίας στην παθοφυσιολογία συγκεκριμένων συμπτωμάτων της σχιζοφρένειας σύμφωνα με σχετική έρευνα [32]. Λεπτομέρειες παρατίθενται στις εικόνες 3.17 και 3.18.

Cardio	rmssd	cvsd	snn	PermEnt	mean_hr	SDANN	SampleEnt	ShannEn	AppEnt	DFA	TriIndex
count	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000
mean	108.167700	0.142511	230.456961	0.523746	86.311351	83.290325	0.374154	9.699768	0.766826	0.803893	41.807492
std	10.901572	0.019399	27.025844	0.038595	7.756537	37.098055	0.083026	0.592589	0.086906	0.050788	14.787351
min	90.231222	0.107713	170.061410	0.442626	71.754627	35.685069	0.247116	8.609929	0.618696	0.727187	15.932684
25%	98.287939	0.131314	211.287087	0.492883	81.385374	56.999368	0.309291	9.174555	0.694676	0.762744	29.040434
50%	107.465866	0.142657	233.315748	0.528776	86.969295	75.266011	0.367560	9.994412	0.762098	0.803898	44.708359
75%	117.050827	0.154070	245.465469	0.546522	91.843337	99.991086	0.417788	10.117887	0.843010	0.832300	55.286069
max	129.990501	0.177792	281.565473	0.583220	103.729927	174.531353	0.582782	10.388658	0.941982	0.896657	66.113012

Σχήμα 3.17: Συνοπτικά στατιστικά στοιχεία των νοσούντων κατά τη διάρκεια της ημέρας

Cardio	rmsdd	cvstd	sdnn	PermEnt	mean_hr	SDANN	SampleEnt	ShannEn	AppEnt	DFA	TriIndex
count	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000
mean	113.515644	0.154487	240.155383	0.556426	89.904752	86.770015	0.461626	9.906068	0.828947	0.813500	46.521396
std	23.713686	0.037300	40.133082	0.060536	9.617053	22.331228	0.139539	1.195969	0.146972	0.048353	17.247585
min	38.949246	0.040445	125.215440	0.353529	57.171059	56.526710	0.230247	7.556053	0.451939	0.729452	12.094242
25%	101.244782	0.139521	217.802212	0.539831	86.631699	72.549779	0.337845	9.253848	0.742165	0.785526	34.273998
50%	119.693752	0.164813	256.312931	0.570461	90.762852	81.864063	0.463314	10.119545	0.870374	0.797335	44.827110
75%	129.957668	0.182588	266.376907	0.594494	96.024920	94.125476	0.604037	10.365690	0.947442	0.835652	58.668773
max	137.480506	0.195855	287.504823	0.626048	102.610306	145.963413	0.691342	12.486767	1.013121	0.932524	73.179821

Σχήμα 3.18: Συνοπτικά στατιστικά στοιχεία των μη νοσούντων κατά τη διάρκεια της ημέρας

Το επόμενο βήμα προκύπτει από τεχνικές αξιολόγησης της πρόγνωσης του μοντέλου. Χρησιμοποιούμε την διασταυρούμενη επικύρωση για να εντοπίσουμε την υπερπροσαρμογή (overfitting), δηλαδή την αποτυχία της γενίκευσης ενός μοτίβου που δίνει λανθασμένα αποτελέσματα στον αλγόριθμο. Η διασταυρούμενη επικύρωση κατά την οποία αφήνουμε μία εγγραφή εκτός, είναι μια ειδική περίπτωση διασταυρούμενης επικύρωσης όπου ο αριθμός των αναδιπλώσεων ισούται με τον αριθμό των παρουσιών στο σύνολο δεδομένων. Έτσι, ο αλγόριθμος εκμάθησης εφαρμόζεται μία φορά για κάθε περίπτωση, χρησιμοποιώντας όλα τα άλλα στιγμιότυπα ως σύνολο εκπαίδευσης και χρησιμοποιώντας το επιλεγμένο παράδειγμα ως σύνολο δοκιμής ενός στοιχείου. Εκπαιδεύοντας πολλά μοντέλα σε υποσύνολα των διαθέσιμων δεδομένων εισόδου και αξιολογώντας τα στο συμπληρωματικό υποσύνολο των δεδομένων. Στη δική μας περίπτωση παίρνουμε δύο εγγραφές ταυτόχρονα, μια εγγραφή της κλάσης 1 και μια εγγραφή της κλάσης 0 για μια πιο αντικειμενική πρόβλεψη. Έτσι διαχωρίζεται το σύνολο δεδομένων σε ένα σύνολο εκπαίδευσης και ένα σύνολο δοκιμών, χρησιμοποιώντας όλες τις παρατηρήσεις εκτός από δύο ως μέρος του σετ εκπαίδευσης. Τελικά η μέθοδος παίρνει το όνομα διασταυρούμενη επικύρωση "leave-one-out". Επαναλαμβάνουμε αυτή τη διαδικασία αφήνοντας έξω δυο διαφορετικές παρατηρήσεις από το σετ εκπαίδευσης κάθε φορά.

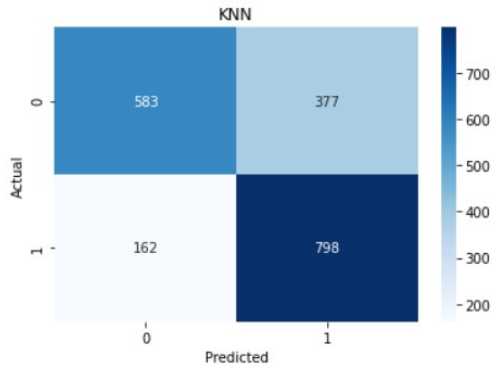
Η GridSearchCV είναι μια τεχνική αναζήτησης, που οπτικοποιείται διαγραμματικά, των καλύτερων τιμών παραμέτρων από το δεδομένο σύνολο του πλέγματος παραμέτρων που ορίζουμε, ώστε το μοντέλο να αποκτήσει τη μέγιστη απόδοση που μπορεί. Δημιουργεί ένα μοντέλο για κάθε συνδυασμό υπερπαραμέτρων που καθορίζονται και μετά το αξιολογεί. Μια πιο αποτελεσματική τεχνική για τον συντονισμό υπερπαραμέτρων είναι η τυχαίοποιημένη αναζήτηση — όπου χρησιμοποιούνται συνδυασμοί των υπερπαραμέτρων για την εύρεση της καλύτερης λύσης.

Μετά από εφαρμογή της GridSearchCV η καλύτερη παράμετρος για τους κοντινούς γείτονες για τον αλγόριθμο KNN είναι ο αριθμός 1 και στα δύο πειράματα. Η παράμετρος C για την βελτιστοποίηση SVM επιλέχθηκε στην τιμή $c=1000$ για τα περισσότερα folds σε γραμμικό kernel, που υποδεικνύει πόσο επιδιώκουμε να αποφύγουμε την εσφαλμένη ταξινόμηση κάθε παραδείγματος εκπαίδευσης. Έτσι η βελτιστοποίηση επιλέ-

γει ένα υπερεπίπεδο μικρότερου περιθωρίου, για να ταξινομήσει σωστά όλα τα σημεία εκπαίδευσης. Αντίθετα, μια πολύ μικρή τιμή του C θα ανάγκαζε τον βελτιστοποιητή να αναζητήσει ένα διαχωριστικό υπερεπίπεδο μεγαλύτερου περιθωρίου, ακόμα κι αν αυτό το υπερεπίπεδο ταξινομούσε εσφαλμένα περισσότερα σημεία. Αναφορικά με τον αλγόριθμο του Random Forest και την παράμετρο που ορίζει τον αριθμό επανεκτίμησης των υποδειμάτων τέθηκε ως $\text{max depth}=2$. Για την αποφυγή της προκατάληψης στην εκπαίδευση του μοντέλου λόγω διαφορετικών κλιμάκων των μεταβλητών έγινε κανονικοποίηση σε κλίμακα MinMaxScaler, η οποία επιλέχθηκε ως η καλύτερη μέθοδος κανονικοποίησης των δεδομένων μετά από πειραματική σύγκριση με τη μέθοδο Standar Scaler.

Στο τέλος υπολογίζονται οι μέσοι όροι των μετρικών αξιολόγησης κατηγοριοποίησης για τους αλγορίθμους μας που παρουσιάζουν και την συνολική εικόνα σε επίπεδο ακρίβειας.

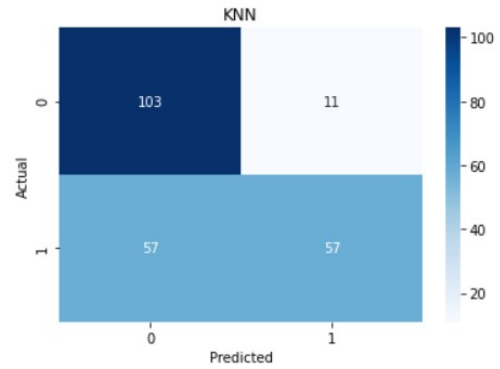
Predicted	0	1
Actual 0	583	377
Actual 1	162	798



F1 score: 0.75

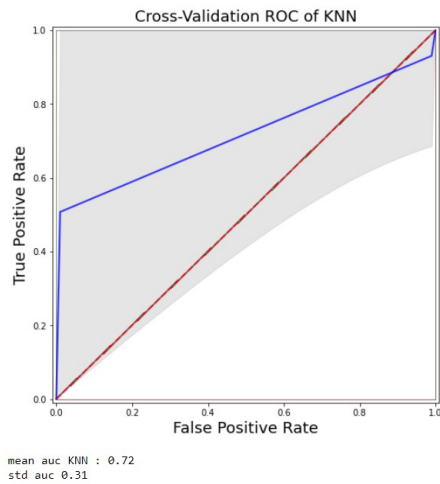
Σχήμα 3.19: KNN ημέρας

Predicted	0	1
Actual 0	103	11
Actual 1	57	57

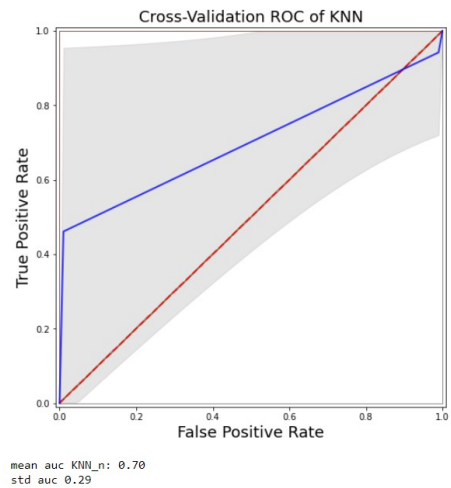


F1 score: 0.63

Σχήμα 3.20: KNN νύχτας

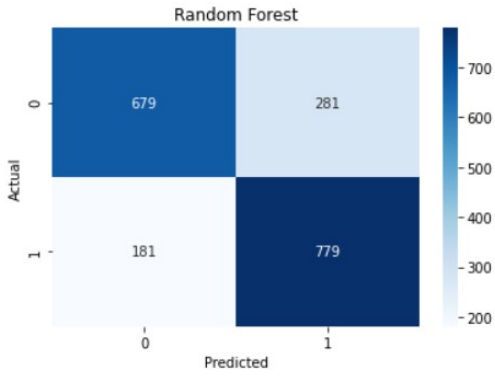


Σχήμα 3.21: ROC-AUC KNN ημέρας



Σχήμα 3.22: ROC-AUC KNN νύχτας

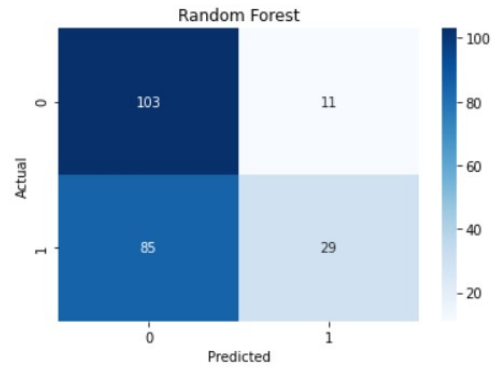
Predicted \ Actual	0	1
0	679	281
1	181	779



F1 score: 0.77

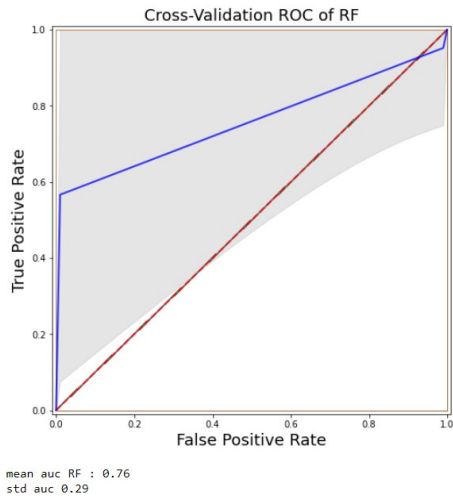
Σχήμα 3.23: Random Forest ημέρας

Predicted \ Actual	0	1
0	103	11
1	85	29

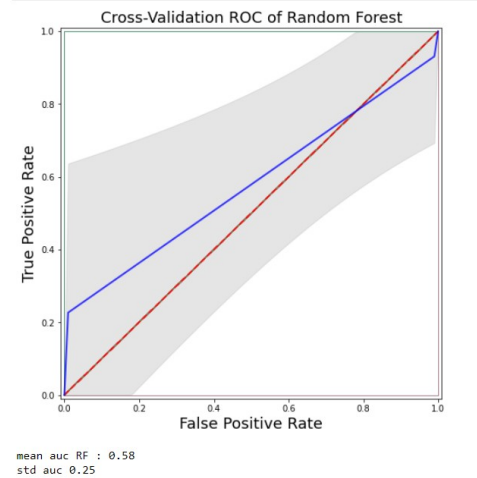


F1 score: 0.38

Σχήμα 3.24: Random Forest νύχτας

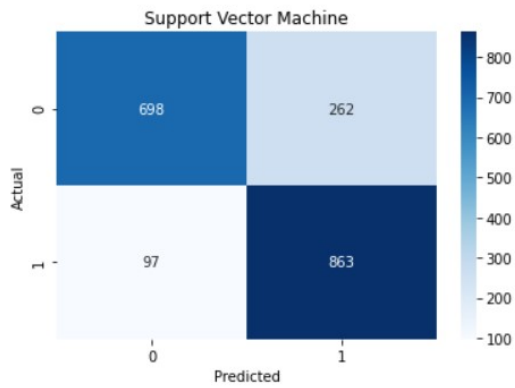


Σχήμα 3.25: ROC-AUC RF ημέρας



Σχήμα 3.26: ROC-AUC RF νύχτας

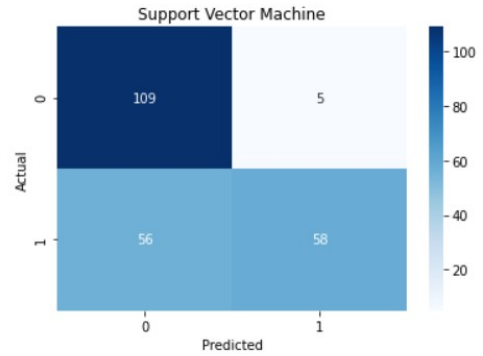
Predicted	0	1
Actual		
0	698	262
1	97	863



F1 score: 0.83

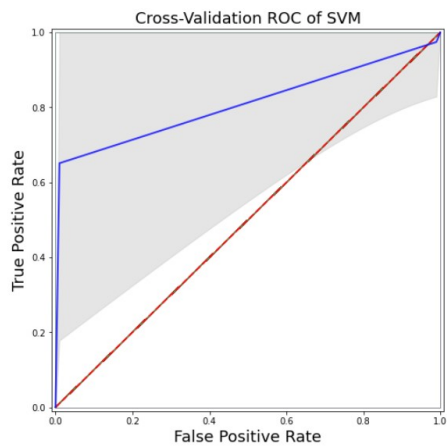
Σχήμα 3.27: SVM ημέρας

Predicted	0	1
Actual		
0	109	5
1	56	58



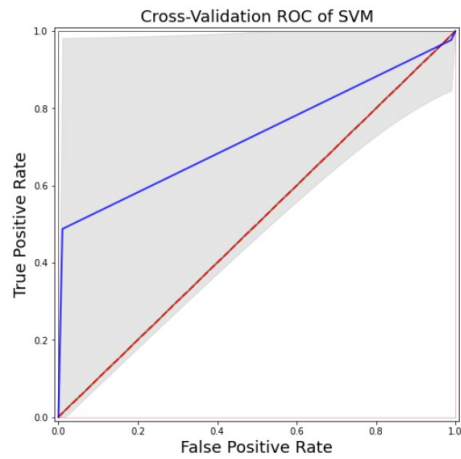
F1 score: 0.66

Σχήμα 3.28: SVM νύχτας



mean auc SVM : 0.81
std auc 0.26

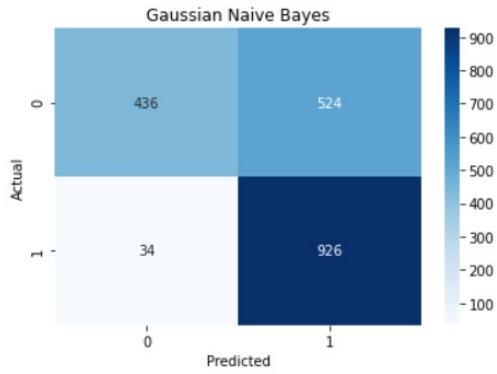
Σχήμα 3.29: ROC-AUC SVM ημέρας



mean auc SVM: 0.73
std auc 0.27

Σχήμα 3.30: ROC-AUC SVM νύχτας

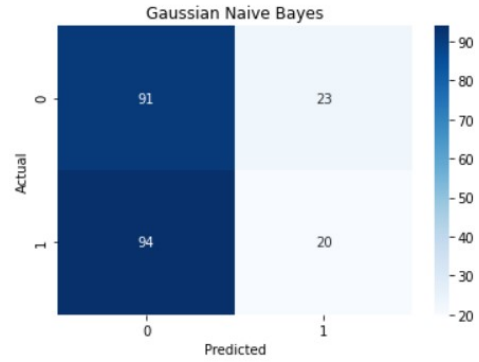
Predicted	0	1
Actual		
0	436	524
1	34	926



F1 score: 0.77

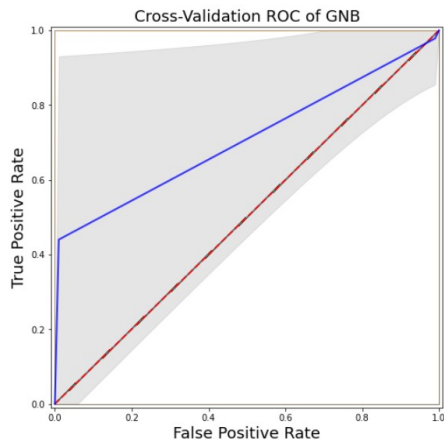
Σχήμα 3.31: GNB ημέρας

Predicted	0	1
Actual		
0	91	23
1	94	20



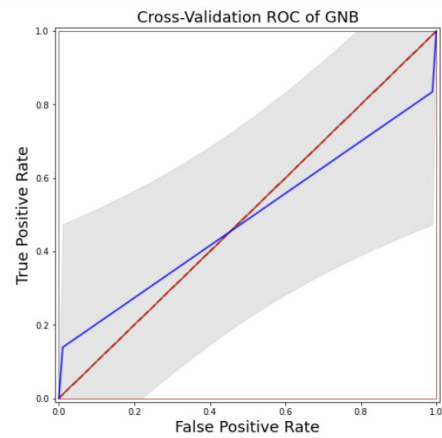
F1 score: 0.25

Σχήμα 3.32: GNB νύχτας



mean auc GNB : 0.71
std auc 0.26

Σχήμα 3.33: ROC-AUC GNB ημέρας



mean auc GNB: 0.49
std auc 0.27

Σχήμα 3.34: ROC-AUC GNB νύχτας

	Precision	Recall	F1 Score	mean AUC
DAY HRV				
KNN	0.68	0.83	0.75	0.72±0.31
Random Forest	0.73	0.81	0.71	0.76±0.29
SVM	0.77	0.9	0.83	0.81±0.26
GNB	0.64	0.96	0.77	0.71±0.26
NIGHT SLEEP HRV				
KNN	0.84	0.5	0.63	0.7±0.29
Random Forest	0.72	0.25	0.38	0.58±0.25
SVM	0.92	0.51	0.66	0.73±0.27
GNB	0.47	0.18	0.25	0.49±0.27

Πίνακας 3.6: Μετρικές ακρίβειας, ανάκλησης, F1 για κάθε ταξινομητή

3.10 Αποτελέσματα

Οι πίνακες σύγκρισης παρουσιάζουν το κατά πόσο οι τεχνικές ταξινόμησης καταφέρνουν να εκπαιδευτούν αποτελεσματικά σε όλο το φάσμα των συνόλων δεδομένων. Η ακρίβεια ποσοτικοποιεί τον αριθμό των θετικών προβλέψεων των κλάσεων που ανήκουν στην πραγματικότητα στη θετική τάξη, ενώ η ανάκληση αντιπροσωπεύει την ικανότητα του μοντέλου να προβλέπει σωστά τα θετικά από τα πραγματικά θετικά. Αυτό σε αντίθεση με την ακρίβεια σημαίνει ότι μετρά όσες προβλέψεις γίνονται από τα μοντέλα είναι πραγματικά θετικές από όλες τις θετικές προβλέψεις που έγιναν. Η ακρίβεια χρησιμοποιείται σε συνδυασμό με την ανάκληση για αντιστάθμιση ψευδώς θετικών και ψευδώς αρνητικών. Η ακρίβεια επηρεάζεται από την κατανομή της κλάσης. Εάν υπάρχουν περισσότερα δείγματα στην κατηγορία μειοψηφίας, τότε η ακρίβεια θα είναι μικρότερη. Η ακρίβεια μπορεί να θεωρηθεί ως μέτρο ακρίβειας ή ποιότητας. Αν θέλουμε να ελαχιστοποιήσουμε τα ψευδώς αρνητικά, θα επιλέγαμε ένα μοντέλο με υψηλή ακρίβεια. Αντίθετα, εάν θέλουμε να ελαχιστοποιήσουμε τα ψευδώς θετικά, θα επιλέγαμε ένα μοντέλο με υψηλή ανάκληση. Η ακρίβεια χρησιμοποιείται κυρίως όταν χρειάζεται να προβλέψουμε τη θετική κατηγορία και υπάρχει μεγαλύτερο κόστος που σχετίζεται με τα ψευδώς θετικά παρά με τα ψευδώς αρνητικά. Παράλληλα, παρουσιάζεται και η αμοιβαία συνεισφορά των δύο μέτρων που περιγράφεται απ' τον δείκτη F1. Οι καμπύλες ROC και οι καμπύλες ανάκλησης ακριβείας παρέχουν ένα διαγνωστικό εργαλείο για μοντέλα δυαδικής ταξινόμησης. Οι ROC AUC και το Precision-Recall AUC παρέχουν βαθμολογίες που συνοψίζουν τις καμπύλες και μπορούν να χρησιμοποιηθούν για τη σύγκριση ταξινομητών. Σε μια καμπύλη ROC αποτυπώνεται το ποσοστό των αληθινών θετικών έναντι των ψευδώς θετικών. Το AUC (συντομογραφία του εμβαδού κάτω από την καμπύλη ROC) είναι η πιθανότητα ότι ένας ταξινομητής θα κατατάξει μια τυχαία επιλεγμένη θετική παρουσία υψηλότερη από μια τυχαία επιλεγμένη αρνητική. Όσο περισσότερο "αγκαλιάζει" η καμπύλη ROC την επάνω αριστερή γωνία του πρώτου τεταρτημορίου, τόσο καλύτερα το μοντέλο ταξινομεί τα δεδομένα σε κατηγορίες.

Για να το ποσοτικοποιήσουμε αυτό, μπορούμε να υπολογίσουμε το εμβαδόν του χωρίου που προκύπτει εάν φέρουμε μια ευθεία από την αρχή έως το πέρας της καμπύλης. Οι καμπύλες ROC και η ROC AUC μπορούν να είναι βασίμες για σοβαρά προβλήματα

ταξινόμησης που έχουν imbalanced data sets, με λίγα δείγματα της κατηγορίας μειοψηφίας.

Η ανάλυση ROC δεν έχει καμία προκατάληψη προς μοντέλα που αποδίδουν καλά στην κατηγορία μειοψηφίας σε βάρος της τάξης της πλειοψηφίας - μια ιδιότητα που είναι αρκετά καλή όταν αντιμετωπίζουμε μη ισορροπημένα δεδομένα. Ο πίνακας 3.6 παρουσιάζει μία ξεκάθαρη οπτική για τον τρόπο με τον οποίο οι αλγόριθμοι αποδίδουν.

Παρατηρούμε ότι οι αρχικές μας στατιστικές εκτιμήσεις επαληθεύονται. Για τα δεδομένα ημέρας από το σύνολο των 960 πραγματικά θετικών, ο αλγόριθμος Naive Bayes πρόβλεψε σωστά τις 926 θετικές περιπτώσεις με recall=0.96 και από το σύνολο των πραγματικά 960 αρνητικών προβλέφθηκαν σωστά οι 698 από τον αλγόριθμο SVM με precision=0.77. Σύμφωνα με την καμπύλη AUC, καλύτερο μοντέλο ταξινομητή θεωρείται ο SVM με AUC=0.81. Για τα δεδομένα νυχτερινού ύπνου, καθώς πρόκειται για ένα imbalanced σετ δεδομένων με 19 μη νοσούντες της τάξης 0 και 6 νοσούντες της τάξης 1, από το σύνολο δεδομένων των 114 θετικών περιπτώσεων, 58 πρόβλεψε σωστά ο SVM και από το σύνολο των 114 αρνητικών περιπτώσεων ο SVM πρόβλεψε σωστά 109 αρνητικές περιπτώσεις, κάτι που θεωρείται λογικό αν ληφθεί υπόψη η μειοψηφία της κλάσης 1. Ο καλύτερος ταξινομητής προσδιορίζεται από την καμπύλη ROC, ως ο SVM με 0,73.

ΚΕΦΑΛΑΙΟ 4

Συμπεράσματα

Η βιβλιογραφία σχετικά με την αξιολόγηση του HRV σε ασθενείς με σχιζοφρένεια είναι ελλιπής και οι μέχρι σήμερα διαθέσιμες μελέτες περιορίζονται σε έρευνες που αφορούν τη διάρκεια της εγρήγορσης ανάμεσα σε ασθενείς με σχιζοφρένεια και σε υγιή άτομα. Σε συμφωνία με τα αποτελέσματά μας, οι Zahn et al. (1981) [56] ανέφεραν σημαντικά υψηλότερο καρδιακό ρυθμό και μειωμένο HRV σε ασθενείς με σχιζοφρένεια. Τα αποτελέσματα της παρούσας έρευνας παρέχουν κάποιες ενδείξεις ότι η σοβαρότητα των ψυχωτικών συμπτωμάτων, σχετίζεται με μειωμένο HRV. Ακόμη άτομα σε υποτροπή δεν παρουσιάζουν σημαντικές διαφορές στην HRV ημερας-νύχτας, ενώ άτομα που δεν υποτροπιάζουν παρουσιάζουν διαφορές στην HRV ημερας-νύχτας. Επίσης θα μπορούσε να αποτελέσει μελλοντικό ερευνητικό στόχο η αναλογία HRV ηρεμίας κατά τη διάρκεια της ημέρας σε σχέση με την HRV του νυχτερινού ύπνου σε νοσούντες.

- [1] Daniel A Adler **and others**. “Predicting early warning signs of psychotic relapse from passive sensing data: an approach using encoder-decoder neural networks”. **in** *JMIR mHealth and uHealth*: 8.8 (2020), e19962.
- [2] Richard A Armstrong. “Should Pearson’s correlation coefficient be avoided?” **in** *Ophthalmic and Physiological Optics*: 39.5 (2019), **pages** 316–327.
- [3] American Psychiatric Association **and others**. “American Psychiatric Association Task Force on DSM-IV.(2000)”. **in** *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*: ().
- [4] Karl-Jürgen Bär **and others**. “Non-linear complexity measures of heart rate variability in acute schizophrenia”. **in** *Clinical neurophysiology*: 118.9 (2007), **pages** 2009–2015.
- [5] Karl-Jürgen Bär **and others**. “Relationship between cardiovagal modulation and psychotic state in patients with paranoid schizophrenia”. **in** *Psychiatry research*: 157.1-3 (2008), **pages** 255–257.
- [6] Ian Barnett **and others**. “Relapse prediction in schizophrenia through digital phenotyping: a pilot study”. **in** *Neuropsychopharmacology*: 43.8 (2018), **pages** 1660–1666.
- [7] Michael Brennan, Marimuthu Palaniswami **and** Peter Kamen. “Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability?” **in** *IEEE transactions on biomedical engineering*: 48.11 (2001), **pages** 1342–1347.
- [9] Cheng-Ding Chang, Chien-Chih Wang **and** Bernard C Jiang. “Singular value decomposition based feature extraction technique for physiological signal analysis”. **in** *Journal of medical systems*: 36.3 (2012), **pages** 1769–1777.
- [10] Joshua Chiappelli **and others**. “Stress-induced increase in kynurenic acid as a potential biomarker for patients with schizophrenia and distress intolerance”. **in** *JAMA psychiatry*: 71.7 (2014), **pages** 761–768.

- [11] Anthony B Ciccone **and others**. “Reminder: RMSSD and SD1 are identical heart rate variability metrics”. **in** *Muscle & nerve*: 56.4 (2017), **pages** 674–678.
- [12] Ben Derrick, Deirdre Toher **and** Paul White. “Why Welch’s test is Type I error robust”. **in** *The Quantitative Methods in Psychology*: 12.1 (2016).
- [13] Panagiotis P Filntisis **and others**. “Identifying differences in physical activity and autonomic function patterns between psychotic patients and controls over a long period of continuous monitoring using wearable sensors”. **in** *arXiv preprint arXiv:2011.02285*: (2020).
- [14] Fay CM Geisler **and others**. “The impact of heart rate variability on subjective well-being is mediated by emotion regulation”. **in** *Personality and individual differences*: 49.7 (2010), **pages** 723–728.
- [16] George D Gulevich, William C Dement **and** Vincent P Zarcone. “All-night sleep recordings of chronic schizophrenics in remission”. **in** *Comprehensive psychiatry*: 8.3 (1967), **pages** 141–149.
- [17] Sunil R Gupta **and** MS Ali. “A New Transaction Scheduling Scheme for Active Real-time Database Systems: Space-Filling Curve”. **in** *International Journal Of Computer Science And Applications*: 6.2 (2013).
- [19] Teresa Henriques **and others**. “Nonlinear methods most applied to heart-rate time series: A review”. **in** *Entropy*: 22.3 (2020), **page** 309.
- [20] Miguel Henry **and** George Judge. “Permutation entropy and information recovery in nonlinear dynamic economic time series”. **in** *Econometrics*: 7.1 (2019), **page** 10.
- [21] T Higuchi. “Relationship between the fractal dimension and the power law index for a time series: a numerical investigation”. **in** *Physica D: Nonlinear Phenomena*: 46.2 (1990), **pages** 254–264.
- [22] Tomoyuki Higuchi. “Approach to an irregular time series on the basis of the fractal theory”. **in** *Physica D: Nonlinear Phenomena*: 31.2 (1988), **pages** 277–283.
- [24] Michael J Katz. “Fractals and the analysis of waveforms”. **in** *Computers in biology and medicine*: 18.3 (1988), **pages** 145–156.
- [25] Stephen H Kellert **and** Lawrence Sklar. “In the wake of chaos: Unpredictable order in dynamical systems”. **in** *Philosophy of Science*: 64.1 (1997).
- [26] MS Keshavan, Charles F Reynolds **and** David J Kupfer. “Electroencephalographic sleep in schizophrenia: a critical review”. **in** *Comprehensive Psychiatry*: 31.1 (1990), **pages** 34–47.
- [27] Hye-Geum Kim **and others**. “Stress and heart rate variability: A meta-analysis and review of the literature”. **in** *Psychiatry investigation*: 15.3 (2018), **page** 235.
- [28] Jong-Hoon Kim, Jun-Hyung Ann **and** Jinyoung Lee. “Relationship between heart rate variability and the severity of psychotic symptoms in schizophrenia”. **in** *Acta Neuropsychiatrica*: 23.4 (2011), **pages** 161–166. DOI: 10.1111/j.1601-5215.2011.00549.x.

- [29] Robert E Kleiger **and others**. “Time domain measurements of heart rate variability”. *in* *Cardiology clinics*: 10.3 (1992), **pages** 487–498.
- [30] C Kneisl, H Wilson **and** Eileen Trigoboff. “Σύγχρονη νοσηλευτική ψυχικής υγείας”. *in* *Εκδόσεις Έλληνα*, Αθήνα: 208 (2009).
- [31] Shi-yang Li **and others**. “Analysis of heart rate variability based on singular value decomposition entropy”. *in* *Journal of Shanghai University (English Edition)*: 12.5 (2008), **pages** 433–437.
- [32] Ya Liu **and others**. “Altered Heart Rate Variability in Patients With Schizophrenia During an Autonomic Nervous Test”. *in* *Frontiers in Psychiatry*: 12 (2021).
- [33] Dominique Makowski **and others**. “NeuroKit2: A Python toolbox for neurophysiological signal processing”. *in* *Behavior Research Methods*: 53.4 (february 2021), **pages** 1689–1696. DOI: 10.3758/s13428-020-01516-y. URL: <https://doi.org/10.3758/2Fs13428-020-01516-y>.
- [34] Marek Malik. “Time-domain measurement of heart rate variability”. *in* *Cardiac Electrophysiology Review*: 1.3 (1997), **pages** 329–334.
- [35] Klaus Mann **and others**. “Heart rate variability during sleep in patients with schizophrenia treated with olanzapine”. *in* *International clinical psychopharmacology*: 19.6 (2004), **pages** 325–330.
- [36] Laia Mas-Expósito **and others**. “The World Health Organization quality of life scale brief version: a validation study in patients with schizophrenia”. *in* *Quality of life Research*: 20.7 (2011), **pages** 1079–1089.
- [38] Dib Nabil **and** F Bereksi Reguig. “Ectopic beats detection and correction methods: A review”. *in* *Biomedical Signal Processing and Control*: 18 (2015), **pages** 228–244.
- [39] Keith H Nuechterlein **and others**. “The vulnerability/stress model of schizophrenic relapse: a longitudinal study”. *in* *Acta Psychiatrica Scandinavica*: 89 (1994), **pages** 58–64.
- [40] Mohamed I Owis **and others**. “Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification”. *in* *IEEE transactions on Biomedical Engineering*: 49.7 (2002), **pages** 733–736.
- [41] Steve Pincus. “Approximate entropy (ApEn) as a complexity measure”. *in* *Chaos: An Interdisciplinary Journal of Nonlinear Science*: 5.1 (1995), **pages** 110–117.
- [42] A Provenzale **and others**. “Distinguishing between low-dimensional dynamics and randomness in measured time series”. *in* *Physica D: nonlinear phenomena*: 58.1-4 (1992), **pages** 31–49.
- [43] Daniel S Quintana **and others**. “Reduced heart rate variability in schizophrenia and bipolar disorder compared to healthy controls”. *in* *Acta Psychiatrica Scandinavica*: 133.1 (2016), **pages** 44–52.
- [45] Md Mashiur Rahman **and others**. “Detection of Abnormality in Electrocardiogram (ECG) Signals Based on Katz’s and Higuchi’s Method Under Fractal Dimensions”. *in* *Comput. Biol. Bioinform*: 4 (2016), **pages** 27–36.

- [46] Graeme D Ruxton. “The unequal variance t-test is an underused alternative to Student’s t-test and the Mann–Whitney U test”. **in***Behavioral Ecology*: 17.4 (2006), **pages** 688–690.
- [48] Fred Shaffer **and** Jay P Ginsberg. “An overview of heart rate variability metrics and norms”. **in***Frontiers in public health*: (2017), **page** 258.
- [49] CE Shannon. “Une théorie mathématique des télécommunications”. **in***Bell. Syst. Techn. J.*: 27.3 (1948), **pages** 379–423.
- [50] AS Sokolenko. “FRACTAL GEOMERTY”. **in**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ВН КАРАЗІНА ФАКУЛЬТЕТ ІНОЗЕМНИХ МОВ КАФЕДРА АНГЛІЙСЬКОЇ МОВИ: ().
- [51] Phyllis K Stein. “Assessing heart rate variability from real-world Holter reports”. **in***Cardiac Electrophysiology Review*: 6.3 (2002), **pages** 239–244.
- [52] Phyllis K Stein **and others**. “Sometimes higher heart rate variability is not better heart rate variability: results of graphical and nonlinear analyses”. **in***Journal of cardiovascular electrophysiology*: 16.9 (2005), **pages** 954–959.
- [53] Juan Sztajzel **and others**. “Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system”. **in***Swiss medical weekly*: 134.35-36 (2004), **pages** 514–522.
- [54] Μάρκος Τσιπούρας **and others**. “Διακύμανση Καρδιακού Ρυθμού”. **in**(2015).
- [55] PE Tikkanen. “Nonlinear wavelet and wavelet packet denoising of electrocardiogram signal”. **in***Biological cybernetics*: 80.4 (1999), **pages** 259–267.
- [56] Motomi Toichi **and others**. “The influence of psychotic states on the autonomic nervous system in schizophrenia”. **in***International Journal of Psychophysiology*: 31.2 (1999), **pages** 147–154.
- [57] Gaetano Valenza **and others**. “Predicting mood changes in bipolar disorder through heartbeat nonlinear dynamics”. **in***IEEE journal of biomedical and health informatics*: 20.4 (2016), **pages** 1034–1043.
- [58] Ruud Van Winkel, Nicholas C Stefanis **and** Inez Myin-Germeys. “Psychosocial stress and psychosis. A review of the neurobiological mechanisms and the evidence for gene-stress interaction”. **in***Schizophrenia bulletin*: 34.6 (2008), **pages** 1095–1105.
- [59] Dawn I Velligan **and** Larry D Alphs. “Negative symptoms in schizophrenia: the importance of identification and treatment”. **in***Psychiatric Times*: 25.3 (2008), **pages** 39–39.
- [61] Bernard L Welch. “The generalization of ‘STUDENT’S’ problem when several different population variances are involved”. **in***Biometrika*: 34.1-2 (1947), **pages** 28–35.
- [62] Jie Wu **and others**. “Determination of weights for ultimate cross efficiency using Shannon entropy”. **in***Expert Systems with Applications*: 38.5 (2011), **pages** 5162–5165.

- [8] V.B. Carson. *Mental Health Nursing: The Nurse-patient Journey*. W.B. Saunders, 2000. ISBN: 9780721680538. URL: <https://books.google.gr/books?id=QM5rAAAAMAAJ>.
- [18] John E Hall **and** Michael E Hall. *Guyton and Hall textbook of medical physiology e-Book*. Elsevier Health Sciences, 2020.
- [23] John Hughlings Jackson **and** James Taylor. *Selected Writings of John Hughlings Jackson: Vol. 2: Evolution and Dissolution of the Nervous System; Speech; Various Papers, Addresses and Lectures*. Staples Press, 1958.
- [37] Tom M Mitchell **and** Tom M Mitchell. *Machine learning. volume 1. 9*. McGraw-hill New York, 1997.