



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Ταξινόμηση Σταδίων Ύπνου με Χρήση Δεδομένων  
από Φορητές Συσκευές

Sleep Stages Classification With Data Acquired From  
Wearables

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Νικολάου Σοφία

Επιβλέπων: Πέτρος Μαραγκός  
Καθηγητής Ε.Μ.Π.

ΕΡΓΑΣΤΗΡΙΟ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ, ΕΠΙΚΟΙΝΩΝΙΑΣ ΛΟΓΟΥ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ  
ΣΗΜΑΤΩΝ

Αθήνα, Ιούλιος 2022





Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Σημάτων, Ελέγχου και Ρομποτικής  
Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας  
Σημάτων

# Ταξινόμηση Σταδίων Ύπνου με Χρήση Δεδομένων από Φορητές Συσκευές

## Sleep Stages Classification With Data Acquired From Wearables

### ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

Νικολάου Σοφία

**Επιβλέπων:** Πέτρος Μαραγκός  
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 14<sup>η</sup> Ιουλίου, 2022.

.....  
Πέτρος Μαραγκός  
Καθηγητής Ε.Μ.Π.

.....  
Αθανάσιος Ροντογιάννης  
Αναπλ. Καθηγητής Ε.Μ.Π.

.....  
Γεράσιμος Ποταμιάνος  
Αναπλ. Καθηγητής Παν/μιο Θεσσαλίας

Αθήνα, Ιούλιος 2022

.....  
**ΝΙΚΟΛΑΟΣ ΣΟΦΙΑΣ**  
Φυσικός Ε.Κ.Π.Α

Copyright © – All rights reserved Νικόλαος Σοφίας, 2022.  
Με επιφύλαξη παντός δικαιώματος.

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

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



# Περίληψη

Παραδοσιακά, η παρακολούθηση του ύπνου πραγματοποιείται σε περιβάλλοντα νοσοκομείων ή κλινικών, απαιτώντας πολύπλοκη και ακριβή εγκατάσταση εξοπλισμού και βαθμολόγηση από ειδικούς ιατρούς. Στον αντίποδα, οι φορητές συσκευές παρέχουν όλο και περισσότερο μια βιώσιμη εναλλακτική λύση για την παρακολούθηση του ύπνου και είναι σε θέση να συλλέγουν δεδομένα κίνησης και καρδιακού παλμού (HR), οξυγόνου και πολλά άλλα είδη δεδομένων μέσω σύγχρονων αισθητήρων. Οπότε, αυτή η λύση ανοίγει τους ορίζοντες για τους σύγχρονους ερευνητές μηχανικούς να απασχοληθούν στον τομέα αυτό, με στόχο την καλύτερη κατανόηση της φυσιολογίας του ανθρώπινου ύπνου. Ταυτόχρονα, η ευρεία χρήση των φορητών συσκευών από τους ανθρώπους δημιουργεί μεγάλα σύνολα δεδομένων (Big Datasets) τα οποία μπορούν να αναλυθούν από τους ερευνητές, καθώς καταγράφουν δεδομένα ανθρώπινης δραστηριότητας σε πολύ σύντομα χρονικά διαστήματα (π.χ. 5sec). Γι αυτό τον λόγο, στη παρούσα μεταπτυχιακή εργασία ασχολούμαστε με το **θέμα της ταξινόμησης των δεδομένων ακτιγραφίας και καρδιακών παλμών στα 4 βασικά στάδια του ύπνου: Wake, Light Sleep, Deep Sleep και REM** με τη χρήση μοντέλων Βαθιάς Μάθησης (Deep Learning) και με διαφορετικές προσεγγίσεις προεπεξεργασίας των δεδομένων. Στόχος μας είναι η επίτευξη όσο το δυνατόν καλύτερων αποτελεσμάτων ταξινόμησης. Γενικότερα, η εφαρμογή της Τεχνητής Νοημοσύνης στον τομέα του ύπνου μπορεί να βοηθήσει τους ιατρούς να έχουν μια καλύτερη επίβλεψη των ασθενών τους, προσφέροντας διάφορες χρήσιμες πληροφορίες που προκύπτουν από την ανάλυση των ανθρώπινων "βιο-σημάτων" (data analysis) και από την μοντελοποίηση τους μέσω μοντέλων Μηχανικής Μάθησης, οπότε θα μπορούσαν να αντιμετωπιστούν διάφορες ασθένειες που σχετίζονται με τον ύπνο όπως η αϋπνία, άπνοια.

Έχοντας υπόψιν όλα τα παραπάνω, στη παρούσα μεταπτυχιακή εργασία στοχεύουμε στην ταξινόμηση των δεδομένων ακτιγραφίας και καρδιακών παλμών, που δίνονται από μια δημοφιλή βάση δεδομένων (MESA dataset), στα 4 βασικά στάδια του ύπνου με τη χρήση μοντέλων Βαθιάς Μάθησης. Έχοντας ως βάση της έρευνας μας, μια έρευνα στο τομέα ταξινόμησης των σταδίων ύπνου, προσπαθούμε να βελτιώσουμε τα αποτελέσματα, δοκιμάζοντας διάφορες τεχνικές, από την επεξεργασία των δεδομένων (pre-processing), δοκιμή διαφορετικών μοντέλων Βαθιάς Μάθησης, τεχνικές αντιμετώπισης μη ισορροπημένου συνόλου δεδομένων (imbalance dataset), μέχρι και τη μοντελοποίηση μόνο των καρδιακών παλμών. **Δηλαδή, ακολουθούμε μια ευρεία δοκιμή διαφόρων πειραμάτων τα οποία θα δώσουν κάποιες κατευθυντήριες γραμμές για την περαιτέρω βελτίωση των αποτελεσμάτων ταξινόμησης.** Συγκεκριμένα, στο πρώτο στάδιο των πειραματισμών μας, προσπαθούμε να αναπαράγουμε τα αποτελέσματα της προαναφερθείσας έρευνας και παρατηρήσαμε ότι προκύπτουν τα ίδια αποτελέσματα ακολουθώντας τη μεθοδολογία που δίνουν οι ερευνητές. Σχολιάζοντας τα αποτελέσματα και τα πιθανά μειονεκτήματα που προκύπτουν, σε επόμενο στάδιο προσπαθούμε να βελτιώσουμε τα αποτελέσματα βάσει αυτών των παρατηρήσεων. Δηλαδή, εφόσον έχουμε παρατηρήσει τις αδυναμίες της προαναφερθείσας έρευνας, προσπαθούμε να αντιμετωπίσουμε τα προβλήματα που αναδύθηκαν, χρησιμοποιώντας τα ίδια μοντέλα αλλά με διαφορετικές υπερπαραμέτρους και αρχιτεκτονική. Επίσης, δοκιμάζουμε άλλα μοντέλα Βαθιάς Μάθησης, και κάποιες διαφορετικές τεχνικές επεξεργασίας δεδομένων

(Data Augmentation) και πειραματισμός με διαφορετικό μήκος παραθύρωσης των δεδομένων (data windowing). Σχολιάζουμε τα αποτελέσματα που προκύπτουν και προτείνουμε διάφορες ερευνητικές κατευθύνσεις για μεγαλύτερη εμβάθυνση και ακόμη καλύτερα αποτελέσματα. Τέλος, ακολουθούμε μια διαφορετική προσέγγιση της ταξινόμησης των δεδομένων ύπνου της βάσης MESA. Χρησιμοποιούμε μόνο τα δεδομένα καρδιακών παλμών της βάσης (RR-Intervals) τα οποία εν συνεχεία μοντελοποιούνται χωρίς εξαγωγή χαρακτηριστικών (raw data). Οπότε, πειραματιζόμαστε με διαφορετικό είδος δεδομένων που δέχονται τα μοντέλα Βαθιάς Μάθησης με στόχο την επίτευξη όσο το δυνατόν καλύτερων αποτελεσμάτων. Τελικά, από όλους τους πειραματισμούς μας, πετυχαίνουμε ικανοποιητικά αποτελέσματα με βάσει τα δεδομένα της βάσης MESA, βελτιώνοντας τα αποτελέσματα της έρευνας [1], οπότε σε επόμενο στάδιο μπορούν να διερευνηθούν περισσότερο σε βάθος οι διάφοροι πειραματισμοί που πραγματοποιήσαμε και ίσως προκύψουν ακόμη καλύτερα αποτελέσματα ταξινόμησης. **Το σημαντικό μέρος της έρευνας μας είναι ότι εντοπίστηκαν και αντιμετωπίστηκαν διάφορα προβλήματα της ταξινόμησης των δεδομένων ύπνου, της βάσης δεδομένων MESA, στα 4 βασικά στάδια (Wake, Light, Deep, REM).**

# Abstract

Traditionally, sleep monitoring has been performed in hospital or clinic settings, requiring complex and expensive equipment installation and certification by specialist physicians. In contrast, mobile devices (wearables) are increasingly providing a viable alternative to sleep monitoring and they are able to collect motion and heart rate (HR), oxygen and many other types of data via state-of-the-art sensors. Thus, this solution opens the horizons for modern researchers of engineers to work in this field, in order to better understand the physiology of human sleep and at the same time, their widespread use by humans creates Big Datasets that can be analyzed by researchers. For this reason, in this master thesis we deal with the **issue of classifying actigraphy and heart rate data into the 4 basic stages of sleep: Wake, Light Sleep, Deep Sleep and REM** using Deep Learning models applying different data pre-processing approaches. The goal is to achieve the best possible classification results. In general, the application of Artificial Intelligence in the field of sleep analysis can help doctors to have a better supervision of their patients, providing various and useful information from the analysis of human biosignals, so that various diseases could be treated or discovered, such as insomnia and apnea. With all of the above in mind, in this master's thesis we aim at the classification of actigraphy and heart rate data, given by a popular database (MESA database), into the 4 basic stages of sleep, using Deep Learning models. Based on our research [1], we try to improve the results by testing various techniques, from data processing (pre-processing), testing of different Deep Learning models, treatment of imbalance database (imbalance database), even modeling only raw heart rate data. **That is, we follow a broad test of various experiments which will provide some guidelines for further improving the classification results.** Specifically, in the first stage of our experiments, we try to reproduce the results of the aforementioned research and finally the same results were emerged following the methodology given by the researchers. Commenting on the results and the possible disadvantages that arose, in the next step we tried to improve the results based on these observations. That is, since we have noticed the weaknesses of the aforementioned research, we try to address the possible problems, using the same Deep Learning models but with different hyperparameters and architecture. We also tested other Deep Learning models, and some different data processing techniques (Data Augmentation) and we experimented with different lengths of data windowing. We comment on the results that emerged and suggest various research directions for greater depth and even better classification results. Finally, we follow a different approach to classify MESA database sleep data. We used only the base heart rate data (RR-Intervals) which are then modeled without extracting features (raw data). So, we experimented with different types of data that the Deep Learning models accept in order to achieve the best possible results. Finally, from all our experiments, we achieve satisfactory results based on the data of the MESA database, improving the results of the research [1], so in the next stage the various experiments that can be further investigated and may produce even better classification results. **The important part of our research is that various problems were identified and addressed to some extent, regarding the classification of sleep stages into 4 stages.**



# Ευχαριστίες

Με την ολοκλήρωση της μεταπτυχιακής μου εργασίας κλείνει ένα μεγάλο και παράλληλα σύντομο κεφάλαιο της ζωής μου. Τελικά είναι αλήθεια ότι τα φοιτητικά χρόνια δημιουργούν ξεχωριστές εμπειρίες και ο χρόνος περνά χωρίς να τον αντιληφθούμε. Γι αυτό και το ταξίδι μου στον ακαδημαϊκό χώρο δεν θα ολοκληρωθεί εδώ. Κοιτώντας όλο το διάστημα ως φοιτητής, θα ήθελα να ευχαριστήσω:

- Τον επιβλέποντα Καθηγητή κ. Πέτρο Μαραγκό και όλα τα μέλη του Εργαστηρίου Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας Σημάτων για την ευκαιρία που μου έδωσαν να εκπονήσω την μεταπτυχιακή μου εργασία στο Εργαστήριο.
- Ιδιαίτερως, την μεταδιδασκτορική ερευνήτρια Δρ. Αθανασία Ζλατίντση η οποία με συμβούλευε κατά της διάρκειας εκπόνησης της μεταπτυχιακής μου εργασίας παρά το βαρύ ακαδημαϊκό πρόγραμμα στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών αλλά και γενικότερα. Θέλω να την ευχαριστήσω που με την υπομονή της και τη θέλησή της για υψηλού επιπέδου δουλειά, η συνεργασία μας με έκανε να αγαπήσω ακόμα περισσότερο τον κόσμο της έρευνας. Εύχομαι οι επερχόμενοι άνθρωποι που θα συναντήσω στη ζωή μου και θα συνεργαστώ επαγγελματικά να έχουν χαρακτηριστικά της, πρωτίστως ως άνθρωποι και δευτερευόντως ως ερευνητές.
- Τους συμφοιτητές και καθηγητές μου στο μεταπτυχιακό Μαθηματικής Προτυποποίησης σε Σύγχρονες Τεχνολογίες και τη Χρηματοοικονομική, όπου μέσω διεπιστημονικών συζητήσεων ανακάλυψα την ουσιαστική προσφορά του μεταπτυχιακού στην κοινωνία με την προώθηση της διεπιστημονικότητας. Επίσης, με βοήθησε να διευρύνω τις γνώσεις μου και σε άλλα γνωστικά πεδία πέραν της Φυσικής.
- Την οικογένειά μου και τους φίλους μου, οι οποίοι με στήριζαν όλο αυτό το καιρό εκπόνησης της μεταπτυχιακής εργασίας αλλά και σε όλο μου το ταξίδι ως φοιτητής.

Σοφίας Νικόλαος  
14/7/2022



# Περιεχόμενα

Περιεχόμενα	xī
Λίστα Σχημάτων	xii
Κατάλογος Πινάκων	xvii
<b>1 Εισαγωγή</b>	<b>1</b>
1.1 Ύπνος	2
1.1.1 Ύπνος - Βασικές ιατρικές έννοιες	2
1.1.2 Τεχνολογία Επίβλεψης ύπνου με φορητές συσκευές (wearables)	10
1.1.3 Σύνολο Δεδομένων MESA	14
1.2 Μοντέλα Βαθιάς Μάθησης (Deep Learning Models)	18
1.2.1 LSTM	18
1.2.2 Συνελικτικά Νευρωνικά Δίκτυα - CNNs	20
1.2.3 Μοντέλο Μετασχηματιστής - Transformer Model	23
1.3 Στόχοι και Οργάνωση Εργασίας	26
<b>2 Μέθοδος Ταξινόμησης Σταδίων Ύπνου με χρήση δεδομένων Ακτιγραφίας και HRV χαρακτηριστικών</b>	<b>29</b>
2.1 Βιβλιογραφική Έρευνα	30
2.2 Σύνολο δεδομένων MESA - Προεπεξεργασία	31
2.2.1 Εισαγωγή στα δεδομένα	31
2.2.2 Καθαρισμός δεδομένων - Data Cleaning	32
2.2.3 Εξαγωγή χαρακτηριστικών (Feature Extraction)	35
2.2.4 Ανάλυση χαρακτηριστικών Activity Counts και HRV (Explanatory Data Analysis)	39
2.3 Ταξινόμηση δεδομένων με αλγόριθμους Βαθιάς Μάθησης (Deep Learning)	45
2.3.1 Προετοιμασία Δεδομένων	48
2.3.2 Ταξινόμηση με Συνελικτικό Νευρωνικό Δίκτυο (CNN)	49
2.3.3 Ταξινόμηση με Νευρωνικό Δίκτυο Μακροπρόθεσμης Μνήμης (LSTM)	55
2.3.4 Ταξινόμηση με τη Μέθοδο Ensemble (Ensemble Method)	60
2.3.5 Σύνοψη Αποτελεσμάτων	62
2.4 Βελτίωση Αλγόριθμων Βαθιάς Μάθησης (Deep Learning)	64
2.4.1 Περιγραφή διαδικασίας - Προεπεξεργασία Δεδομένων	64
2.4.2 Ταξινόμηση με Συνελικτικό Νευρωνικό Δίκτυο (CNN)	69
2.4.3 Ταξινόμηση με Νευρωνικό Δίκτυο Μακροπρόθεσμης Μνήμης Δύο Κατευθύνσεων (Bidirectional LSTM)	74
2.4.4 Ταξινόμηση με διαφορετικό μήκος παραθύρωσης (Sliding Window)	80
2.4.5 Ταξινόμηση με τη Μέθοδο Ensemble (Ensemble Method)	84

2.4.6	Ταξινόμηση με Μοντέλο Μετασχηματιστή (Transformer Model)	85
2.5	Σύνοψη Αποτελεσμάτων - Συμπεράσματα	91
<b>3</b>	<b>Μέθοδος Ταξινόμησης Σταδίων Ύπνου με χρήση ακατέργαστων (raw) δεδομένων RR-Intervals</b>	<b>97</b>
3.1	Βιβλιογραφική Έρευνα	98
3.2	Δεδομένα RR- Intervals - Προεπεξεργασία	99
3.2.1	Περιγραφή διαδικασίας - Προετοιμασία δεδομένων για μοντελοποίηση	101
3.2.2	Εφαρμογή τεχνικών αύξησης συνόλου δεδομένων (Data Augmentation)	110
3.3	Μοντελοποίηση Δεδομένων RR-Intervals με αλγόριθμους Deep Learning	112
3.3.1	Μοντελοποίηση με Συνελικτικό Νευρωνικό Δίκτυο - CNN	114
3.3.2	Μοντελοποίηση με LSTM	117
3.3.3	Μοντελοποίηση με LSTM και μηχανισμό Προσοχής (Attention)	119
3.4	Ανάλυση Αποτελεσμάτων και Συμπεράσματα	122
<b>4</b>	<b>Επίλογος</b>	<b>127</b>
4.1	Σύνοψη και Μελλοντικές Επεκτάσεις	127
<b>A</b>	<b>Βιβλιογραφία</b>	<b>129</b>



# Λίστα Σχημάτων

1.1.1 Υπνόγραμμα νυχτερινού ύπνου υγιούς ενήλικα [10] . . . . .	5
1.1.2 Χαρακτηριστική δραστηριότητα EEG καθενός από τα τέσσερα στάδια του ύπνου NREM, από έναν εθελοντή γυναικείου φύλου 19 ετών [10]. Η παρουσία ατράκτων ύπνου (sleep spindles) είναι τα υπογραμμισμένα σημεία στο στάδιο 2 και το K-σύμπλεγμα φαίνεται από το βέλος. . . . .	5
1.1.3 Φυσιολογικές διαφορές μεταξύ NREM και REM ύπνου [21] . . . . .	8
1.1.4 Συσσκευή PSG Polysomnography καταγραφής δεδομένων κατά τη διάρκεια του ύπνου. Το σύστημα PSG καταγράφει δεδομένα, εγκεφαλικής δραστηριότητας (EEG), καρδιακών παλμών από τον ανθρώπινο δείκτη (heart rate), δεδομένα ροής οξυγόνου (air flow) από τη μύτη και τέλος, δεδομένα κίνησης του σώματος και χεριών μέσω μιας ελαστικής ζώνης. Όλα τα δεδομένα που καταγράφονται, μεταφέρονται σε ηλεκτρονικό υπολογιστή [43]. . . . .	12
1.1.5 Σήματα καταγραφής από συσκευή PSG Polysomnography καταγραφής δεδομένων [44] .	12
1.1.6 Συσσκευή καταγραφής δεδομένων ακτιγραφίας [47] . . . . .	13
1.1.7 Ακτιγράμματα. Διάγραμμα καταγραφής κινητικής δραστηριότητας (activity counts) για δύο στάδια ύπνου: Sleep και Wake [48] . . . . .	14
1.1.8 Δεδομένα από περικάρπιο μετρητή (Actigraph) της MESA [53]. . . . .	16
1.2.1 Επαναλαμβανόμενη κυψέλη ενός επιπέδου LSTM [72]. . . . .	19
1.2.2 Εξισώσεις μίας LSTM κυψέλης (LSTM Cell). . . . .	20
1.2.3 Παράδειγμα πράξεων Συνέλιξης στο Επίπεδο Συνέλιξης. . . . .	21
1.2.4 Αρχιτεκτονική του μοντέλου Μετασχηματιστή (Transformer model). . . . .	26
2.2.1 Σύνολο Δεδομένων ανά κλάση της βάσης MESA (30sec - segments). . . . .	32
2.2.2 Ιστόγραμμα δεδομένων Ακτιγραφίας. . . . .	34
2.2.3 Διάγραμμα δεδομένων Ακτιγραφίας για 1000 σημεία. . . . .	34
2.2.4 Ιστόγραμμα δεδομένων Ακτιγραφίας ανά κλάση. <i>Wake:0, NREM<sub>1</sub>:1, NREM<sub>2</sub>:2, NREM<sub>3</sub>:3, REM:4</i> . . . . .	34
2.2.5 Διαγράμματα RR-Intervals για 4 χρήστες. . . . .	35
2.2.6 Ιστογράμματα RR-Intervals για 4 χρήστες. . . . .	36
2.2.7 Υπνογράμματα από 4 χρήστες. . . . .	36
2.2.8 Ιστογράμματα συνόλου χαρακτηριστικών ενός χρήστη (UserId=2982). . . . .	39
2.2.9 Διαγράμματα χαρακτηριστικών ενός χρήστη (UserId=2982). . . . .	40
2.2.10 Διαγράμματα χαρακτηριστικών ενός χρήστη (UserId=2982). Για το τελευταίο διάγραμμα που παρουσιάζει το υπνόγραμμα του χρήστη UserId=2982, παραθέτουμε την αντιστοίχιση των σταδίων ύπνου: <i>Wake:0, N1:1, N2:2, Deep:3, REM:4</i> . . . . .	41
2.2.11 Πίνακας Γραμμικής Συσχέτισης των 9 χαρακτηριστικών για όλους τους χρήστες. . . . .	42
2.2.12 Πίνακας στατιστικών των 9 χαρακτηριστικών για όλους τους χρήστες. . . . .	43

2.2.13 Πίνακας στατιστικών των 9 χαρακτηριστικών για όλους τους χρήστες για το στάδιο <b>Wake</b> .	43
2.2.14 Πίνακας στατιστικών των 9 χαρακτηριστικών για όλους τους χρήστες για το στάδιο <b>Light</b> .	44
2.2.15 Πίνακας στατιστικών των 9 χαρακτηριστικών για όλους τους χρήστες για το στάδιο <b>Deep</b> .	44
2.2.16 Πίνακας στατιστικών των 9 χαρακτηριστικών για όλους τους χρήστες για το στάδιο <b>REM</b> .	44
2.3.1 Διαισθητική απεικόνιση της παραθύρωσης δεδομένων με μήκη 21, 51 και 101 (γκρι πλαίσιο) [1]. Η επικάλυψη (overlapping) μεταξύ διαδοχικών παραθύρων είναι 30 δευτερόλεπτα. Το μαύρο πλαίσιο μπορεί να αγνοηθεί. Τα μοντέλα δέχονται σαν είσοδο το γκρι πλαίσιο.	46
2.3.2 Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου. Μεταβλητή των μοντέλων αποτελεί το μήκος παραθύρωσης (sliding window) και αναφέρεται ως $k$ , $k \in (21, 51, 101)$ .	50
2.3.3 Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 21.	51
2.3.4 Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 51.	51
2.3.5 Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 101.	52
2.3.6 Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 21.	53
2.3.7 Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 51.	54
2.3.8 Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 101.	55
2.3.9 Αρχιτεκτονική Νευρωνικού Δικτύου Μακροπρόθεσμης Μνήμης. Μεταβλητή των μοντέλων αποτελεί το μήκος παραθύρωσης (sliding window) και αναφέρεται ως $k$ , $k \in (21, 51, 101)$ .	55
2.3.10 Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 21.	57
2.3.11 Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 51.	57
2.3.12 Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 101.	58
2.3.13 Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 21.	59
2.3.14 Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 51.	60
2.3.15 Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 101.	61
2.4.1 Τεχνική Jittering για την κλάση Deep για κάθε χαρακτηριστικό ακτιγραφίας και HRV.	66
2.4.2 Ιστόγραμμα συνόλου δεδομένων εκπαίδευσης (train set). Αύξηση συνόλου δεδομένων της κλάσης Deep.	67

2.4.3	Ιστόγραμμα συνόλου δεδομένων εκπαίδευσης (train set) για μήκος παραθύρωσης 101, μετά από υποδειγματοληψία της κλάσης Light.	68
2.4.4	Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου. Μεταβλητή των μοντέλων αποτελεί το μήκος παραθύρωσης (sliding window) και αναφέρεται ως $k$ , $k \in (21, 51, 101)$ .	69
2.4.5	Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 21.	70
2.4.6	Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 51.	70
2.4.7	Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 101.	71
2.4.8	Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 21.	72
2.4.9	Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 51.	73
2.4.10	Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 101.	74
2.4.11	Αρχιτεκτονική Νευρωνικού Δικτύου Μακροπρόθεσμης Μνήμης Δύο Κατευθύνσεων. Μεταβλητή των μοντέλων αποτελεί το μήκος παραθύρωσης (sliding window) και αναφέρεται ως $k$ , $k \in (21, 51, 101)$ .	74
2.4.12	Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 21.	76
2.4.13	Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 51.	76
2.4.14	Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 101.	77
2.4.15	Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 21.	77
2.4.16	Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 51.	78
2.4.17	Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 101.	79
2.4.18	Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 71. Μοντέλο CNN.	81
2.4.19	Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 71. Μοντέλο LSTM.	81
2.4.20	Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης του μοντέλου CNN. Μήκος παραθύρωσης 71.	83
2.4.21	Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης του μοντέλου LSTM. Μήκος παραθύρωσης 71.	83
2.4.22	Αρχιτεκτονική μοντέλου Μετασχηματιστή (Transformer). Μεταβλητή των μοντέλων αποτελεί το μήκος παραθύρωσης (sliding window) και αναφέρεται ως $k$ , $k \in (21, 51, 101)$ .	86

2.4.23	Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 21. . . . .	87
2.4.24	Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 51. . . . .	88
2.4.25	Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 101. . . . .	88
2.4.26	Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 21. . . . .	89
2.4.27	Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 51. . . . .	90
2.4.28	Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 101. . . . .	91
3.2.1	Σύνολο δεδομένων ανά κλάση. . . . .	100
3.2.2	Ιστογράμματα μηκών 30sec RR-Intervals για όλους τους χρήστες της βάσης δεδομένων MESA. . . . .	101
3.2.3	Κατανομή μέσου μήκους 30sec RR-Intervals όλων των χρηστών. . . . .	102
3.2.4	Ιστογράμματα 30sec RR-Intervals όλων των χρηστών με τη μεγαλύτερη πληροφορία. . . . .	103
3.2.5	Κατανομή κλάσεων πριν την μέθοδο επιλογής ακολουθιών. . . . .	103
3.2.6	Κατανομή κλάσεων μετά την μέθοδο επιλογής ακολουθιών. . . . .	104
3.2.7	Διαγράμματα RR-Intervals για 4 χρήστες. Ο άξονας time δεν απεικονίζει δευτερόλεπτα αλλά πλήθος σημείων όπου κάθε sample αποτελείται από 31 σημεία. . . . .	105
3.2.8	Διαδοχικά Διαγράμματα RR-Intervals για 1 χρήστη κατά το στάδιο Wake. . . . .	106
3.2.9	Ιστογράμματα RR-Intervals για 4 χρήστες. . . . .	106
3.2.10	Υπνογράμματα από 4 χρήστες. Ο άξονας time δεν απεικονίζει δευτερόλεπτα αλλά πλήθος σημείων/samples ανά 30 δευτερόλεπτα. Π.χ για τον User 3939 το σύνολο των σημείων RR-points είναι $400 \cdot 31 = 12400$ όπως φαίνεται στο Σχήμα 3.2.7. . . . .	107
3.2.11	Κατανομή κλάσεων από 4 χρήστες. . . . .	107
3.2.12	Υποδειγματοληψία της κλάσης Light Sleep. . . . .	110
3.2.13	Data Augmentation: (1) Reverse Deep Class, (2) Reverse + Jitter Deep Class και (3) Reverse REM Class. . . . .	112
3.2.14	Κατανομή κλάσεων μετά τις τεχνικές Data Augmentation. . . . .	113
3.3.1	Κατανομή κλάσεων στο σύνολο δεδομένων train και test. . . . .	114
3.3.2	Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου. . . . .	114
3.3.3	Υπνόγραμμα 100 διαδοχικών ακολουθιών του συνόλου αξιολόγησης και των αντίστοιχων πραγματικών κλάσεων. Σημείωση - Wake:0, Light:1, Deep:2, REM:3. . . . .	115
3.3.4	Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. . . . .	116
3.3.5	Πίνακας Σύγχυσης (Confusion Matrix) . . . . .	116
3.3.6	Αρχιτεκτονική LSTM Δικτύου. . . . .	117
3.3.7	Υπνόγραμμα 100 διαδοχικών ακολουθιών του συνόλου αξιολόγησης και των αντίστοιχων πραγματικών κλάσεων. Wake:0, Light:1, Deep:2, REM:3 . . . . .	117
3.3.8	Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. . . . .	118
3.3.9	Πίνακας Σύγχυσης (Confusion Matrix) . . . . .	119
3.3.10	Αρχιτεκτονική LSTM + Attention Δικτύου. . . . .	119
3.3.11	Υπνόγραμμα 100 διαδοχικών ακολουθιών του συνόλου αξιολόγησης και των αντίστοιχων πραγματικών κλάσεων. . . . .	120

---

3.3.12 Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. . . . .	121
3.3.13 Πίνακας Σύγχυσης (Confusion Matrix) . . . . .	122



# Κατάλογος Πινάκων

2.1	Στατιστικά των δεδομένων Ακτιγραφίας. . . . .	33
2.2	Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης. . . . .	52
2.3	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 21. . . . .	53
2.4	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 51. . . . .	53
2.5	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 101. . . . .	54
2.6	Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης. . . . .	58
2.7	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 21. . . . .	59
2.8	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 51. . . . .	59
2.9	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 101. . . . .	60
2.10	Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης. . . . .	62
2.11	Μετρική Χρονικής Απόκλισης (Time Deviation) ανα κλάση σε λεπτά (minutes). . . . .	62
2.12	Πίνακας αποτελεσμάτων που προκύπτουν από τις υλοποιήσεις μας με βάση την έρευνα [1]. . . . .	63
2.13	Πίνακας αποτελεσμάτων που δίνουν οι ερευνητές στο [1]. . . . .	63
2.14	Πίνακας μετρικών για τις μεθόδους Ensemble που δίνουν οι ερευνητές στο [1]. . . . .	64
2.15	Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης. . . . .	71
2.16	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 21. . . . .	72
2.17	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 51. . . . .	72
2.18	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 101. . . . .	73
2.19	Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης. . . . .	78
2.20	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 21. . . . .	78
2.21	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 51. . . . .	79
2.22	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 101. . . . .	79
2.23	Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά τα μοντέλα CNN και LSTM με μήκος παραθύρωσης 71. . . . .	82
2.24	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση για το μοντέλο CNN. Μήκος παραθύρωσης 71. . . . .	82
2.25	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 51. . . . .	82
2.26	Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης. . . . .	84
2.27	Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης. . . . .	87
2.28	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 21. . . . .	89
2.29	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 51. . . . .	89
2.30	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 101. . . . .	90
2.31	Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για τα καλύτερα μοντέλα της έρευνας [1] (Max Ensemble) και της δικής μας (Mean Ensemble). . . . .	92
2.32	Μετρική Χρονικής Απόκλισης (Time Deviation) ανα κλάση σε λεπτά (minutes) της έρευνας [1]. . . . .	92
2.33	Μετρική Χρονικής Απόκλισης (Time Deviation) ανα κλάση σε λεπτά (minutes) των βελτιωμένων μοντέλων (CNN και BiLSTM). . . . .	93

2.34	Πίνακας αποτελεσμάτων για όλα τα μήκη παραθύρωσης (CNN).	94
2.35	Πίνακας αποτελεσμάτων για όλα τα μήκη παραθύρωσης (BiLSTM).	94
2.36	Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για τα καλύτερα μοντέλα της έρευνας μας εκτός των μεθόδων Ensemble.	94
3.1	Στατιστικά των μηκών 30sec RR-Intervals.	102
3.2	Στατιστικά της μέσης κατανομής των μηκών 30sec RR-Intervals.	102
3.3	Στατιστικά των ακολουθιών 30sec RR-Intervals για την κλάση Wake.	108
3.4	Στατιστικά των ακολουθιών 30sec RR-Intervals για την κλάση N1.	108
3.5	Στατιστικά των ακολουθιών 30sec RR-Intervals για την κλάση N2.	109
3.6	Στατιστικά των ακολουθιών 30sec RR-Intervals για την κλάση Deep Sleep.	109
3.7	Στατιστικά των ακολουθιών 30sec RR-Intervals για την κλάση REM.	110
3.8	Μετρικές στο σύνολο δεδομένων αξιολόγησης.	115
3.9	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση.	115
3.10	Μετρικές στο σύνολο δεδομένων αξιολόγησης.	118
3.11	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανά κλάση.	118
3.12	Μετρικές στο σύνολο δεδομένων αξιολόγησης.	121
3.13	Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση.	121
3.14	Μετρικές στο σύνολο δεδομένων αξιολόγησης.	122
3.15	Μετρικές στο σύνολο δεδομένων αξιολόγησης.	123
3.16	Μετρικές στο σύνολο δεδομένων αξιολόγησης για τις δύο μεθοδολογίες που ακολουθήσαμε. 1) Εξαγωγή χαρακτηριστικών ( $BiLSTM_{101}$ ) και 2) Ακατέργαστα (raw) δεδομένα χωρίς εξαγωγή χαρακτηριστικών ( $LSTM + Attention$ ). Στη μεθοδολογία με εξαγωγή χαρακτηριστικών, η αξιολόγηση έγινε ανά χρήστη (user id), ενώ στην μεθοδολογία με χρήση raw δεδομένων RR-Intervals μόνο η αξιολόγηση έγινε ανεξαρτήτου χρήστη. Άρα Η σύγκριση δεν είναι άμεση, μόνο διαισθητική.	124



# Κεφάλαιο 1

## Εισαγωγή

---

<b>1.1</b>	<b>Ύπνος</b> . . . . .	<b>2</b>
1.1.1	Ύπνος - Βασικές ιατρικές έννοιες . . . . .	2
1.1.2	Τεχνολογία Επίβλεψης ύπνου με φορητές συσκευές (wearables) . . . . .	10
1.1.3	Σύνολο Δεδομένων MESA . . . . .	14
<b>1.2</b>	<b>Μοντέλα Βαθιάς Μάθησης (Deep Learning Models)</b> . . . . .	<b>18</b>
1.2.1	LSTM . . . . .	18
1.2.2	Συνελικτικά Νευρωνικά Δίκτυα - CNNs . . . . .	20
1.2.3	Μοντέλο Μετασχηματιστής - Transformer Model . . . . .	23
<b>1.3</b>	<b>Στόχοι και Οργάνωση Εργασίας</b> . . . . .	<b>26</b>

---

## 1.1 Ύπνος

Ο ύπνος είναι ένα σημαντικό μέρος της καθημερινότητάς μας και ξοδεύουμε περίπου το ένα τρίτο του χρόνου μας σε αυτόν (περίπου 8 ώρες). Ο ποιοτικός ύπνος είναι τόσο απαραίτητος για την επιβίωση όσο το φαγητό και το νερό. Χωρίς ύπνο δεν μπορούμε να σχηματίσουμε ή να διατηρήσουμε τις "διαδρομές" στον εγκέφαλό μας (brain pathways) που μας επιτρέπουν να μάθουμε και να δημιουργούμε νέες αναμνήσεις και είναι πιο δύσκολο να συγκεντρωθούμε και να ανταποκριθούμε γρήγορα στις απαιτήσεις της καθημερινότητας. Ο ύπνος είναι σημαντικός για μια σειρά εγκεφαλικών λειτουργιών, συμπεριλαμβανομένου του τρόπου με τον οποίο τα νευρικά κύτταρα (νευρώνες, brain cells) επικοινωνούν μεταξύ τους. Στην πραγματικότητα, ο εγκέφαλος και το σώμα παραμένουν εξαιρετικά ενεργά ενώ κοιμόμαστε [2]. Πρόσφατα ευρήματα υποδηλώνουν ότι ο ύπνος παίζει καταλυτικό ρόλο στην απομάκρυνση των τοξινών που συσσωρεύονται στον εγκέφαλό ενώ είμαστε ξύπνιοι. Επιπλέον, είναι λογικό ότι όλοι οι άνθρωποι χρειαζόμαστε ύπνο ως βιολογική ανάγκη, αλλά ο βιολογικός του σκοπός παραμένει μυστήριο. **Ο ύπνος επηρεάζει σχεδόν κάθε τύπο ιστού και συστήματος στο σώμα: από τον εγκέφαλο, την καρδιά, τους πνεύμονες μέχρι το μεταβολισμό, τη λειτουργία του ανοσοποιητικού, τη διάθεση και την αντοχή στις ασθένειες.** Έρευνες δείχνουν ότι η χρόνια έλλειψη ύπνου ή ο ύπνος κακής ποιότητας, αυξάνει τον κίνδυνο διαταραχών που περιλαμβάνουν υψηλή αρτηριακή πίεση, καρδιαγγειακές παθήσεις, διαβήτη, κατάθλιψη και παχυσαρκία.

Συμπερασματικά, ο ύπνος είναι μια σύνθετη και δυναμική διαδικασία που επηρεάζει την λειτουργία του ανθρώπινου σώματος με τρόπο που οι επιστήμονες έχουν αρχίσει να καταλαβαίνουν μέσω της εντατικής τους έρευνας, γεγονός ενθαρρυντικό για την κατανόησή του και την αντιμετώπιση ασθενειών που προκαλούνται από τον κακό ή ελλιπή ύπνο. Οπότε, στην επόμενη υποενότητα, θα γίνει μια εισαγωγή στην ανατομία του ύπνου όπου θα παρατεθούν τα ευρήματα ιατρών που θα μας βοηθήσουν να τον κατανοήσουμε περισσότερο και να δημιουργήσει το υπόβαθρο για την εισαγωγή στα στάδιά του που θα μας απασχολήσει στην παρούσα έρευνα μας.

### 1.1.1 Ύπνος - Βασικές ιατρικές έννοιες

Μια σύντομη εισαγωγή στις βασικές έννοιες του ύπνου **από τη σκοπιά της ιατρικής επιστήμης** θα σχηματίσει ένα υπόβαθρο των βασικών λειτουργιών του ανθρώπινου σώματος κατά τη διάρκεια του ύπνου, ώστε να μπορέσουμε να αναλύσουμε δεδομένα που σχετίζονται με αυτόν, έχοντας κάποιες βασικές γνώσεις. Αρχικά, γνωρίζουμε από πλευράς ιατρικής, ότι αρκετές δομές μέσα στον εγκέφαλο εμπλέκονται στον ύπνο [3]. Συγκεκριμένα:

Ο **υποθάλαμος**, μια δομή μεγέθους φιστικιού που βρίσκεται βαθιά μέσα στον εγκέφαλο, περιέχει ομάδες νευρικών κυττάρων που λειτουργούν ως κέντρα ελέγχου που επηρεάζουν τον ύπνο και τη διέγερση. Μέσα στον υποθάλαμο βρίσκεται ο υπερχεισματικός πυρήνας (SCN), δηλαδή ομάδες χιλιάδων κυττάρων που λαμβάνουν πληροφορίες σχετικά με την έκθεση στο φως απευθείας από τα μάτια και ελέγχουν τον ρυθμό συμπεριφοράς μας. Μερικοί άνθρωποι με βλάβη στο SCN κοιμούνται ασταθώς όλη την ημέρα επειδή δεν είναι σε θέση να ταιριάζουν τους κιρκαδικούς ρυθμούς (circadian rhythm) τους με τον κύκλο φωτός-σκοταδιού [4]. Οι περισσότεροι τυφλοί άνθρωποι διατηρούν κάποια ικανότητα να αντιλαμβάνονται το φως και είναι σε θέση να τροποποιήσουν τον κύκλο ύπνου/αφύπνισης (sleep/wake).

Το **στέλεχος του εγκεφάλου**, που βρίσκεται στη βάση του εγκεφάλου, επικοινωνεί με τον υποθάλαμο για να ελέγξει τις εναλλαγές μεταξύ αφύπνισης και ύπνου (sleep-wake transitions) (Το στέλεχος του εγκεφάλου περιλαμβάνει δομές που ονομάζονται γόνοι, μυελός και μεσαίος εγκέφαλος). Τα κύτταρα που προάγουν τον ύπνο στον υποθάλαμο και το στέλεχος του εγκεφάλου παράγουν μια χημική ουσία του εγκεφάλου που ονομάζεται GABA, η οποία δρα για να μειώσει τη δραστηριότητα των κέντρων διέγερσης στον υποθάλαμο και στο στέλεχος του εγκεφάλου. Το στέλεχος του εγκεφάλου (ειδικά οι γόνοι και ο μυελός) παίζει επίσης ιδιαίτερο ρόλο στον ύπνο REM. Στέλνει σήματα για να χαλαρώσουν

οι μύες που είναι απαραίτητοι για τη στάση του σώματος και τις κινήσεις των άκρων, έτσι ώστε να μην πραγματοποιήσουμε τα όνειρά μας.

Ο **θάλαμος** λειτουργεί ως διακόπτης για πληροφορίες από τις αισθήσεις στον εγκεφαλικό φλοιό (το κάλυμμα του εγκεφάλου που ερμηνεύει και επεξεργάζεται πληροφορίες από τη βραχυπρόθεσμη στη μακροπρόθεσμη μνήμη). Κατά τη διάρκεια των περισσότερων σταδίων του ύπνου, ο θάλαμος ησυχάζει, επιτρέποντας τον συντονισμό του ανθρώπου με τον εξωτερικό κόσμο. Αλλά κατά τη διάρκεια του ύπνου REM, ο θάλαμος είναι ενεργός, στέλνοντας στον φλοιό εικόνες, ήχους και άλλες αισθήσεις που γεμίζουν τα όνειρά μας.

Η **επίφυση**, που βρίσκεται μέσα στα δύο ημισφαίρια του εγκεφάλου, λαμβάνει σήματα από το SCN και αυξάνει την παραγωγή της ορμόνης μελατονίνης, η οποία μας βοηθά να κοιμηθούμε μόλις σβήσουμε τα φώτα. Τα άτομα που έχουν χάσει την όρασή τους και δεν μπορούν να συντονίσουν τον φυσικό τους κύκλο αφύπνισης-ύπνου χρησιμοποιώντας φυσικό φως, μπορούν να σταθεροποιήσουν τα πρότυπα του ύπνου τους λαμβάνοντας μικρές ποσότητες μελατονίνης την ίδια ώρα κάθε μέρα. Οι επιστήμονες πιστεύουν ότι οι κορυφές και οι κοιλάδες της μελατονίνης με την πάροδο του χρόνου είναι σημαντικές για την αντιστοίχιση του κιρκάδιου ρυθμού (circadian rhythm) του σώματος με τον εξωτερικό κύκλο του φωτός και του σκότους.

Το βασικό πρόσθιο τμήμα του εγκεφάλου, κοντά στο μπροστινό και κάτω μέρος του εγκεφάλου, προάγει επίσης τον ύπνο και την εγρήγορση, ενώ μέρος του μεσαίου τμήματος του εγκεφάλου λειτουργεί ως σύστημα διέγερσης. Η απελευθέρωση αδενοσίνης, το οποίο είναι ένα χημικό υποπροϊόν της κυτταρικής κατανάλωσης ενέργειας από κύτταρα στον βασικό πρόσθιο εγκέφαλο και πιθανώς σε άλλες περιοχές, καθοδηγεί τον ανθρώπινο ύπνο. Δηλαδή, η αδενοσίνη συμμετέχει στην εναλλαγή των σταδίων του ύπνου. Τέλος, η καφεΐνη εξουδετερώνει την υπνηλία εμποδίζοντας τις δράσεις της αδενοσίνης.

Η αμυγδαλή, μια δομή σε σχήμα αμυγδάλου που εμπλέκεται στην επεξεργασία των συναισθημάτων, γίνεται όλο και πιο ενεργή κατά τη διάρκεια του ύπνου REM.

Τελικά, με βάση τα παραπάνω συμπεραίνουμε ότι διάφορα μέρη του εγκεφάλου παίζουν πολύ σημαντικό ρόλο που επηρεάζουν καταλυτικά τον ύπνο και την εναλλαγή του ανθρώπου στα στάδια του ύπνου που θα αναλυθούν παρακάτω. Άρα ο ύπνος και η ποιότητα του καθορίζεται από την εγκεφαλική δραστηριότητα του ανθρώπου αλλά όπως θα δούμε παρακάτω στην ανάλυση των σταδίων του ύπνου, υπάρχουν και άλλοι παράγοντες που καθορίζουν το στάδιο του ύπνου στο οποίο βρίσκεται ο άνθρωπος.

## Στάδια ύπνου

Η αρχιτεκτονική του ύπνου αναφέρεται στη βασική δομική οργάνωση του κανονικού ύπνου. [3] Υπάρχουν δύο τύποι ύπνου, ο ύπνος μη γρήγορης κίνησης των ματιών (NREM) και ο ύπνος γρήγορης κίνησης των ματιών (REM). Ο ύπνος NREM χωρίζεται σε στάδια 1, 2, 3 και 4, που αντιπροσωπεύουν μια συνέχεια σχετικού βάρους. Δηλαδή, όσο αυξάνεται ο αριθμός (από 1-4), τόσο ο άνθρωπος πέφτει σε βαθύτερο ύπνο. **Το κάθε στάδιο έχει μοναδικά χαρακτηριστικά, συμπεριλαμβανομένων των διαφοροποιήσεων στα μοτίβα των εγκεφαλικών κυμάτων, των κινήσεων των ματιών και του μυϊκού τόνου.**

Οι κύκλοι και τα στάδια ύπνου αποκαλύφθηκαν με τη χρήση ηλεκτροεγκεφαλικών εγγραφών (EEG), που εντοπίζουν τα ηλεκτρικά πρότυπα της εγκεφαλικής δραστηριότητας. Λεπτομερέστερα, ο ανθρώπινος εγκέφαλος αποτελείται από εκατομμύρια νευρώνες που παίζουν σημαντικό ρόλο στον έλεγχο της απόκρισης του ανθρώπινου σώματος σε σχέση με τα εσωτερικά/εξωτερικά κινητικά/αισθητηριακά ερεθίσματα, όπως και κατά τη διάρκεια του ύπνου. Αυτοί οι νευρώνες θα λειτουργήσουν ως φορείς πληροφοριών μεταξύ του ανθρώπινου σώματος και του εγκεφάλου. Η κατανόηση της γνωστικής συμπεριφοράς του εγκεφάλου μπορεί να γίνει με την ανάλυση είτε σημάτων είτε εικόνων από τον εγκέφαλο. Η αν-

θρώπινη συμπεριφορά μπορεί να οπτικοποιηθεί ως προς τις κινητικές και αισθητηριακές καταστάσεις, όπως η κίνηση των ματιών, η κίνηση των χειλιών, η ανάμνηση, η προσοχή, το σφίξιμο των χεριών κλπ. Αυτές οι καταστάσεις σχετίζονται με συγκεκριμένη συχνότητα σήματος που βοηθά στην κατανόηση της λειτουργικής συμπεριφοράς της πολύπλοκης δομής του εγκεφάλου. Η **ηλεκτροεγκεφαλογραφία (EEG)** είναι μια αποτελεσματική μέθοδος που βοηθά στην απόκτηση εγκεφαλικών σημάτων που αντιστοιχούν σε διάφορες καταστάσεις από την επιφάνεια του τριχωτού της κεφαλής. Αυτά τα σήματα γενικά κατηγοριοποιούνται ως δέλτα, θήτα, άλφα, βήτα και γάμμα με βάση τις συχνότητες σημάτων από 0.1 Hz έως περισσότερα από 100 Hz [5].

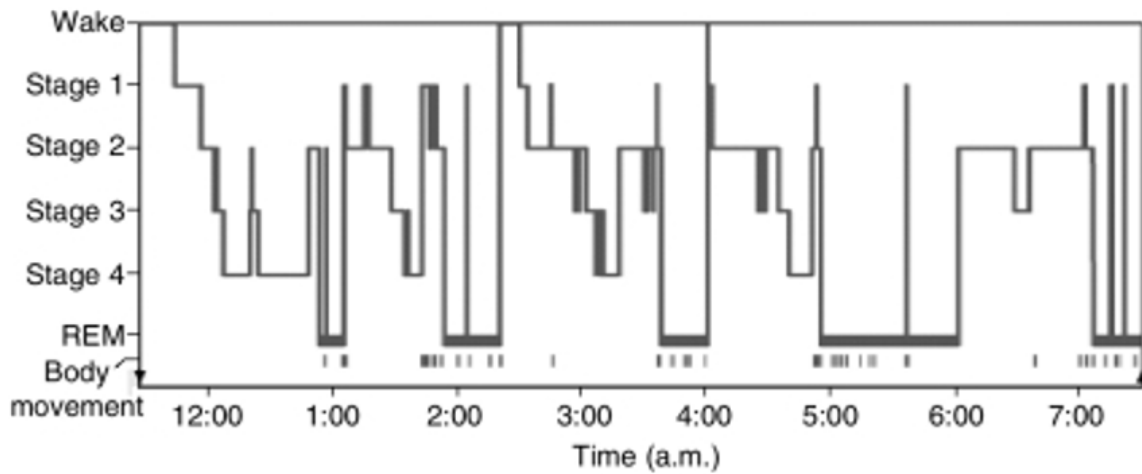
Κατά τη διάρκεια μιας περιόδου ύπνου, ο ύπνος NREM και REM εναλλάσσονται κυκλικά (Σχήμα 1.1.1). Η λειτουργία των εναλλαγών μεταξύ αυτών των δύο τύπων ύπνου δεν είναι ακόμη κατανοητή, αλλά η ακανόνιστη κυκλικότητα και/ή στάδια ύπνου που απουσιάζουν, συνδέονται με διαταραχές του ύπνου. Για παράδειγμα, αντί για ύπνο μέσω NREM, όπως είναι στον τυπικό ύπνο, τα άτομα με ναρκοληψία εισέρχονται στον ύπνο απευθείας στον ύπνο REM.

Συγκεκριμένα, στο Σχήμα 1.1.1 παρουσιάζεται ένα υπνόγραμμα [6]. Το υπνόγραμμα είναι μια μορφή πολυυπνογραφίας και αναπαριστά τα στάδια του ύπνου σε συνάρτηση με το χρόνο. Αναπτύχθηκε ως ένας εύκολος τρόπος παρουσίασης των καταγραφών της δραστηριότητας των εγκεφαλικών κυμάτων από ένα ηλεκτροεγκεφαλογράφημα (EEG) κατά τη διάρκεια μιας περιόδου ύπνου. Επιτρέπει την αναγνώριση των διαφορετικών σταδίων του ύπνου: ύπνου ταχείας κίνησης των ματιών (REM) και ύπνου μη γρήγορης κίνησης των ματιών (NREM) κατά τη διάρκεια του κύκλου του ύπνου. Ο ύπνος NREM μπορεί περαιτέρω να ταξινομηθεί στο στάδιο NREM 1, 2 και 3. Το προηγούμενος θεωρημένο 4ο στάδιο του ύπνου NREM μπορεί να συμπεριληφθεί στο στάδιο 3. Αυτό το στάδιο ονομάζεται επίσης και ύπνος αργών κυμάτων (SWS) και είναι το βαθύτερο στάδιο του ύπνου. Επιπρόσθετα, τα υπνογράμματα λαμβάνονται συνήθως με οπτική βαθμολόγηση των καταγραφών από ηλεκτροεγκεφαλογράφημα (EEGs), ηλεκτροοφθαλμογραφία (EOGs) και ηλεκτρομυογραφία (EMGs) [7]. Η έξοδος από αυτές τις τρεις πηγές καταγράφεται ταυτόχρονα σε ένα γράφημα από οθόνη ή υπολογιστή ως υπνόγραμμα. Ορισμένες συχνότητες που εμφανίζονται από EEG, EOG και EMG είναι χαρακτηριστικές και καθορίζουν σε ποιο στάδιο ύπνου ή εγρήγορσης βρίσκεται ο άνθρωπος. Υπάρχει ένα πρωτόκολλο που ορίζεται από την Αμερικανική Ακαδημία Ιατρικής Ύπνου (AASM) για τη βαθμολογία ύπνου, σύμφωνα με το οποίο η κατάσταση ύπνου ή εγρήγορσης καταγράφεται σε εποχές (epochs) 30 δευτερολέπτων [8]. Πριν από αυτό, οι κανόνες Rechtschaffen και Kales (RK) χρησιμοποιούνταν για την ταξινόμηση των σταδίων ύπνου [9].

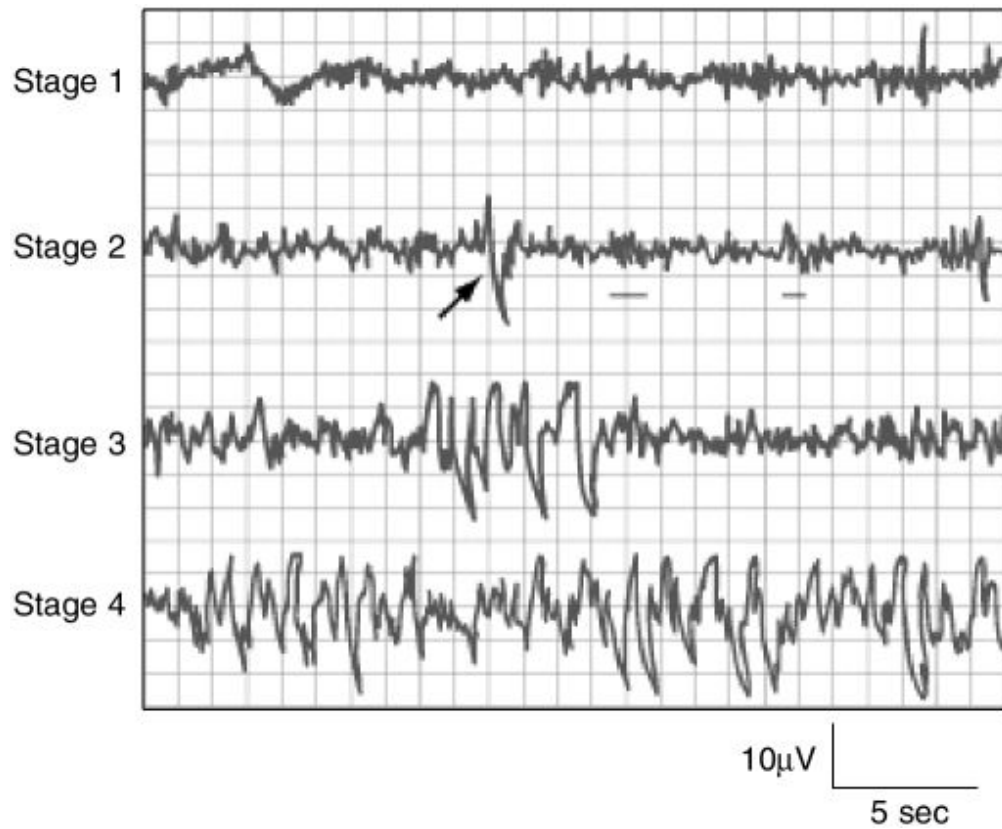
Ένα επεισόδιο ύπνου ξεκινά με μια σύντομη περίοδο του NREM σταδίου 1 που προχωρά στο στάδιο 2, ακολουθούμενη από τα στάδια 3 και 4 και τέλος σε REM. Ωστόσο, τα άτομα δεν παραμένουν σε ύπνο REM το υπόλοιπο της νύχτας αλλά, συνήθως, κάνουν κύκλους μεταξύ των σταδίων του NREM και του REM καθ' όλη τη διάρκεια της νύχτας (Σχήμα 1.1.1). Ο ύπνος NREM αποτελεί περίπου το 75 έως 80 % του συνολικού χρόνου στον ύπνο και ο ύπνος REM το υπόλοιπο 20 έως 25 %. Η μέση διάρκεια του πρώτου κύκλου ύπνου NREM-REM είναι 70 έως 100 λεπτά. Ο δεύτερος και μεταγενέστερος κύκλος διαρκεί περισσότερο - περίπου 90 έως 120 λεπτά [10]. Στους φυσιολογικούς ενήλικες, ο ύπνος REM αυξάνεται σε διάρκεια καθώς προχωρά η νύχτα και είναι μεγαλύτερος στο τελευταίο ένα τρίτο της συνολικής διάρκειας του ύπνου. Καθώς εξελίσσεται ο ύπνος (sleep event), το στάδιο 2 αρχίζει να αντιπροσωπεύει το μεγαλύτερο μέρος του ύπνου NREM και τα στάδια 3 και 4 ενδέχεται μερικές φορές να εξαφανιστούν εντελώς.

Τα τέσσερα στάδια του ύπνου NREM, που περιγράφονται παρακάτω, συνδέονται το καθένα με ξεχωριστή εγκεφαλική δραστηριότητα και φυσιολογία. Το σχήμα 1.1.2 δείχνει τα πρότυπα EEG που είναι χαρακτηριστικά των τεσσάρων σταδίων NREM.

Οπότε, έχοντας μια γενική εικόνα για την εναλλαγή των σταδίων και το υπνόγραμμα, στις επόμενες



Σχήμα 1.1.1: Υπνόγραμμα νυχτερινού ύπνου υγιούς ενήλικα [10]



Σχήμα 1.1.2: Χαρακτηριστική δραστηριότητα EEG καθενός από τα τέσσερα στάδια του ύπνου NREM, από έναν εθελοντή γυναικείου φύλου 19 ετών [10]. Η παρουσία ατράκτων ύπνου (sleep spindles) είναι τα υπογραμμισμένα σημεία στο στάδιο 2 και το K-σύμπλεγμα φαίνεται από το βέλος.

παραγράφους θα αναλύσουμε τα στάδια του ύπνου με σκοπό να καταλάβουμε περισσότερο τις διαφορές μεταξύ τους, διότι θα μας βοηθήσει στην ανάπτυξη των μοντέλων ταξινόμησης και στην κατάλληλη

προεπεξεργασία των δεδομένων ύπνου.

### **Στάδιο Εγρήγορσης (Wake):**

Το πρώτο στάδιο είναι το στάδιο εγρήγορσης ή το στάδιο Wake, το οποίο εξαρτάται περαιτέρω από το αν τα μάτια είναι ανοιχτά ή κλειστά. Κατά τη διάρκεια της εγρήγορσης με ανοιχτά μάτια, υπάρχουν κύματα άλφα και βήτα, κυρίως βήτα που φαίνονται στο εγκεφαλογράφημα EEG. Καθώς τα άτομα νυστάζουν και τα μάτια κλείνουν, ο άλφα ρυθμός είναι το κυρίαρχο μοτίβο. Μια εποχή θεωρείται στάδιο W εάν περιέχει περισσότερα από 50% κύματα άλφα και κινήσεις των ματιών που σχετίζονται με την εγρήγορση [11].

### **Στάδιο NREM 1:**

Αυτό είναι το ελαφρύτερο στάδιο του ύπνου και ξεκινά όταν περισσότερο από το 50% των κυμάτων άλφα αντικαθίστανται με δραστηριότητα μικτής συχνότητας χαμηλού πλάτους (LAMF). Υπάρχει μυϊκός τόνος στους σκελετικούς μύς και η αναπνοή τείνει να εμφανίζεται με κανονικό ρυθμό. Αυτό το στάδιο τείνει να διαρκεί 1 έως 5 λεπτά, αποτελώντας περίπου το 5% του συνολικού κύκλου [12]. Επίσης, ο ύπνος NREM σταδίου 1 εξυπηρετεί έναν μεταβατικό ρόλο στην κυκλικότητα του ύπνου. Από άλλη έρευνα [10], προκύπτει ότι το στάδιο NREM 1 διαρκεί συνήθως 1 έως 7 λεπτά στον αρχικό κύκλο, που αποτελεί το 2-5 % του συνολικού ύπνου και μπορεί να διακοπεί εύκολα από έναν θόρυβο. Η εγκεφαλική δραστηριότητα στο EEG στο στάδιο NREM 1 μεταβαίνει από την εγρήγορση (που χαρακτηρίζεται από ρυθμικά άλφα κύματα) σε κύματα μικτής τάσης και μικτής συχνότητας (low-voltage, mixed frequency). Τα κύματα άλφα (alpha waves) συνδέονται με μια κατάσταση χαλάρωσης όταν ο άνθρωπος είναι σε στάδιο που είναι ξύπνιος κατά την αρχή του ύπνου (wakeful relaxation) και χαρακτηρίζονται από συχνότητα 8 έως 13 κύκλων ανά δευτερόλεπτο. Επίσης, εκτός από τα νεογέννητα άτομα, τα άτομα με ναρκοληψία και άλλες ειδικές νευρολογικές διαταραχές, το μέσο επεισόδιο ύπνου του μέσου ατόμου ξεκινά στο στάδιο NREM 1 (mean sleep episode of average person).

### **Στάδιο NREM 2:**

Αυτό το στάδιο αντιπροσωπεύει τον βαθύτερο ύπνο καθώς πέφτουν οι καρδιακοί παλμοί και η θερμοκρασία του σώματός. Χαρακτηρίζεται από την παρουσία ατράκτων ύπνου [13], συμπλεγμάτων K [14] ή και των δύο όπως φαίνεται στο Σχήμα 1.2.1 (Stage 2). Αυτές οι άτρακτοι ύπνου θα ενεργοποιήσουν την άνω κροταφική έλικα, τον πρόσθιο τριγωνικό σωλήνα και τον θάλαμο. Επίσης, η εμφάνιση των συμπλεγμάτων K, δείχνουν την μετάβαση σε βαθύτερο ύπνο. Συνήθως είναι μεμονωμένα μεγάλα κύματα δέλτα που διαρκούν μόνο για ένα δευτερόλεπτο. Καθώς ακολουθεί βαθύτερος ύπνος και το άτομο μετακινείται στη N3. Όλα τα κύματα τους θα αντικατασταθούν με κύματα δέλτα. Το στάδιο 2 του ύπνου διαρκεί περίπου 25 λεπτά στον αρχικό κύκλο και επιμηκύνεται με κάθε διαδοχικό κύκλο, αποτελώντας τελικά περίπου το 50% του συνολικού ύπνου. Σε άλλη έρευνα [15], παρατίθεται ότι το στάδιο 2 του ύπνου διαρκεί περίπου 10 έως 25 λεπτά στον αρχικό κύκλο και παρατείνεται με κάθε διαδοχικό κύκλο, αποτελώντας τελικά το 45-55 % του συνολικού ύπνου. Ένα άτομο στο στάδιο 2 του ύπνου χρειάζεται πιο έντονα ερεθίσματα από ότι στο στάδιο 1 για να ξυπνήσει. Η εγκεφαλική δραστηριότητα σε ένα EEG δείχνει σχετικά χαμηλής τάσης, μικτής συχνότητας δραστηριότητα που χαρακτηρίζεται από την παρουσία ατράκτων ύπνου (sleep spindles) και συμπλεγμάτων K (K-complex) (Σχήμα 1.1.2). Υποτίθεται ότι οι άτρακτοι του ύπνου είναι σημαντικοί για την εδραίωση της μνήμης. Έχει παρατηρηθεί από ιατρούς ερευνητές ότι τα άτομα που μαθαίνουν μια νέα εργασία έχουν σημαντικά υψηλή πυκνότητα ατράκτων ύπνου. Θεωρείται ότι παίζουν ρόλο στην πλαστικότητα του εγκεφάλου ή στη διαδικασία μάθησης και ενσωμάτωσης νέων αναμνήσεων. Οι άτρακτοι ύπνου φαίνεται επίσης να μειώνουν την απόκρισή μας σε εξωτερικά ερεθίσματα ενώ κοιμόμαστε.

### **Στάδια NREM 3 και 4, Slow-Wave Sleep:**

Αυτό θεωρείται το βαθύτερο στάδιο του ύπνου και χαρακτηρίζεται από πολύ πιο αργή συχνότητα με



σήματα υψηλού πλάτους γνωστά ως κύματα δέλτα. Το κύμα δέλτα είναι ένας τύπος εγκεφαλικού κύματος υψηλού πλάτους που βρίσκεται στους ανθρώπους και σχετίζεται με βαθύ ύπνο. Τα κύματα Δέλτα έχουν συχνότητα από ένα έως τρία Hertz (Hz) και μετρώνται με ηλεκτροεγκεφαλογράφημα (EEG) [16]. Συγκεκριμένα, το στάδιο ύπνου NREM 3 και 4, είναι το πιο δύσκολο να ξυπνήσει κανείς, και για μερικούς ανθρώπους, ακόμη και δυνατοί θόρυβοι (πάνω από 100 dB) δεν θα τους ξυπνήσουν. Καθώς οι άνθρωποι μεγαλώνουν, τείνουν να περνούν λιγότερο χρόνο σε αυτόν τον αργό ύπνο, όπου εμφανίζονται στο ηλεκτροεγκεφαλογράφημα τα κύματα δέλτα, και περισσότερο χρόνο ύπνου στο στάδιο N2. Αν και αυτό το στάδιο έχει το μεγαλύτερο όριο διέγερσης, αν κάποιος ξυπνήσει κατά τη διάρκεια αυτού του σταδίου, θα έχει μια παροδική φάση αδράνειας. Αυτό είναι γνωστό ως αδράνεια ύπνου. Τα γνωστικά τεστ δείχνουν ότι τα άτομα που ξυπνούν σε αυτό το στάδιο τείνουν να έχουν μέτρια διαταραχή της νοητικής απόδοσης για 30 λεπτά έως μία ώρα. Αυτό είναι το στάδιο όπου το σώμα επισκευάζει και αναγεννά τους ιστούς του, χτίζει οστά και μυς και ενισχύει το ανοσοποιητικό σύστημα [12]. Επίσης, όπως αναφέραμε προηγουμένως, τα στάδια ύπνου 3 και 4 αναφέρονται συλλογικά ως ύπνος αργού κύματος (Slow-Wave Sleep), τα περισσότερα από τα οποία συμβαίνουν κατά το πρώτο τρίτο της νύχτας. Το καθένα έχει συγκεκριμένα χαρακτηριστικά. Το στάδιο 3 διαρκεί μόνο λίγα λεπτά και αποτελεί περίπου το 3 έως 8 % της συνολικής διάρκειας του ύπνου. Το EEG παρουσιάζει αυξημένη δραστηριότητα υψηλής τάσης, αργού κύματος (Εικόνα 1.1.2). Το τελευταίο στάδιο NREM είναι το στάδιο 4, το οποίο διαρκεί περίπου 20 έως 40 λεπτά στον πρώτο κύκλο και αποτελεί περίπου το 10 έως 15 % του ύπνου. Το κατώφλι διέγερσης είναι υψηλότερο για όλα τα στάδια NREM στο στάδιο 4. Αυτό το στάδιο χαρακτηρίζεται από αυξημένες ποσότητες δραστηριότητας υψηλής τάσης, αργών κυμάτων στο EEG.

### Στάδιο REM:

Το στάδιο ύπνου REM, είναι το στάδιο που συνδέεται με το όνειρο. Είναι ενδιαφέρον ότι το EEG είναι παρόμοιο με ένα ξύπνιο άτομο, αλλά οι σκελετικοί μύες είναι άτονοι και χωρίς κίνηση. Εξαιρέση αποτελούν οι μύες του ματιού και του διαφραγματικού αναπνευστικού, οι οποίοι παραμένουν ενεργοί. Ο ρυθμός αναπνοής αλλάζει, ωστόσο, καθώς είναι πιο ασταθής και ακανόνιστος. Αυτό το στάδιο ξεκινά συνήθως 90 λεπτά αφότου έχει αρχίσει ο ύπνος και κάθε ένας από τους κύκλους REM διαρκεί περισσότερο από 10 λεπτά κατά τη διάρκεια της νύχτας. Η πρώτη περίοδος διαρκεί συνήθως 10 λεπτά και η τελευταία μπορεί να διαρκέσει έως και μία ώρα [17] [18]. Σε άλλη έρευνα [10], ο ύπνος REM ορίζεται από την παρουσία μη συγχρονισμένης (χαμηλής τάσης, μικτής συχνότητας) δραστηριότητας εγκεφαλικού κύματος, μυϊκής ατονίας και εκρήξεων γρήγορων κινήσεων των ματιών. Οι μορφές κύματος "Sawtooth", η δραστηριότητα θήτα (3 έως 7 μετρήσεις ανά δευτερόλεπτο) και η αργή δραστηριότητα άλφα χαρακτηρίζουν επίσης τον ύπνο REM. Κατά τη διάρκεια του αρχικού κύκλου, η περίοδος REM μπορεί να διαρκέσει μόνο 1 έως 5 λεπτά. Ωστόσο, παρατείνεται προοδευτικά καθώς εξελίσσεται η διαδικασία του ύπνου. Υπάρχουν πολλές φυσιολογικές διαφορές μεταξύ του ύπνου NREM και REM (Σχήμα 1.1.3). Επίσης, όπως αναφέραμε αρχικά, τα όνειρα που βλέπει ο άνθρωπος σχετίζεται συχνότερα με το στάδιο ύπνου REM. Η απώλεια μυϊκού τόνου και αντανακλαστικών εξυπηρετεί πιθανώς μια σημαντική λειτουργία, επειδή εμποδίζει ένα άτομο να «κάνει πράξη» τα όνειρά του ή τους εφιάλτες του ενώ κοιμάται. Περίπου το 80 % των ζωντανών ανακλήσεων στα όνειρα προκύπτουν μετά από διέγερση από αυτό το στάδιο του ύπνου. Ο ύπνος REM μπορεί επίσης να είναι σημαντικός για τη σταθεροποίηση της μνήμης [19][20].

Εφόσον, έχουμε μια πιο λεπτομερή εικόνα για κάθε στάδιο του ύπνου, στις επόμενες παραγράφους θα περιγράψουμε τη φυσιολογία του ύπνου, δηλαδή τι διαδικασίες/αλλαγές περνά το ανθρώπινο σώμα κατά τη διάρκεια του ύπνου και την εναλλαγή των σταδίων του ύπνου.

TABLE 2-1 Physiological Changes During NREM and REM Sleep

Physiological Process	<u>NREM</u>	<u>REM</u>
Brain activity	Decreases from wakefulness	Increases in motor and sensory areas, while other areas are similar to <u>NREM</u>
Heart rate	Slows from wakefulness	Increases and varies compared to <u>NREM</u>
Blood pressure	Decreases from wakefulness	Increases (up to 30 percent) and varies from <u>NREM</u>
Sympathetic nerve activity	Decreases from wakefulness	Increases significantly from wakefulness
Muscle tone	Similar to wakefulness	Absent
Blood flow to brain	Decreases from wakefulness	Increases from <u>NREM</u> , depending on brain region
Respiration	Decreases from wakefulness	Increases and varies from <u>NREM</u> , but may show brief stoppages; coughing suppressed
Airway resistance	Increases from wakefulness	Increases and varies from wakefulness
Body temperature	Is regulated at lower set point than wakefulness; shivering initiated at lower temperature than during wakefulness	Is not regulated; no shivering or sweating; temperature drifts toward that of the local environment
Sexual arousal	Occurs infrequently	Greater than <u>NREM</u>

SOURCES: [NHLBI \(2003\)](#), [Somers et al. \(1993\)](#), [Madsen et al. \(1991b\)](#).

Σχήμα 1.1.3: Φυσιολογικές διαφορές μεταξύ NREM και REM ύπνου [21]

### Φυσιολογία του ύπνου

Εκτός από τις φυσιολογικές αλλαγές που αναφέρονται στον Σχήμα 1.1.3, υπάρχουν και άλλες αλλαγές στο ανθρώπινο σώμα που συμβαίνουν κατά τη διάρκεια του ύπνου. Γενικά, αυτές οι αλλαγές είναι ανεκτές σε υγιή άτομα, αλλά μπορεί να θέσουν σε κίνδυνο την μερικές φορές εύθραυστη ισορροπία ευάλωτων σωματικά ατόμων, όπως αυτά με καρδιαγγειακές παθήσεις. Οι φυσιολογικές αλλαγές συμβαίνουν επίσης στα ακόλουθα συστήματα:

#### Καρδιαγγειακά:

Όσον αφορά την καρδιαγγειακή κυκλοφορία του αίματος, παρατηρούνται μεταβολές στην αρτηριακή πίεση και στον καρδιακό ρυθμό κατά τη διάρκεια του ύπνου και καθορίζονται κυρίως από τη δραστηριότητα του αυτόνομου νευρικού συστήματος. Για παράδειγμα, παρατηρούνται σύντομες αυξήσεις της αρτηριακής πίεσης και των καρδιακών παλμών κατά το Στάδιο NREM 2 καθώς και η διέγερση και οι μεγάλες κινήσεις του σώματος [22]. Επιπλέον, υπάρχει αυξημένος κίνδυνος εμφράγματος του μυοκαρδίου το πρωί λόγω των απότομων αυξήσεων του καρδιακού ρυθμού και της αρτηριακής πίεσης που συνοδεύουν το ξύπνημα [23].

#### Δραστηριότητα συμπαθητικού νεύρου:

Η δραστηριότητα του συμπαθητικού νεύρου μειώνεται καθώς ο ύπνος NREM "βαθαίνει". Ωστόσο, υπάρχει μια έκρηξη δραστηριότητας του συμπαθητικού νεύρου κατά τη διάρκεια του ύπνου NREM λόγω της σύντομης αύξησης της αρτηριακής πίεσης και του καρδιακού ρυθμού που ακολουθεί τα συμπλέγματα K, τα οποία παρατηρούνται στο EEG. Τέλος, σε σύγκριση με την εγρήγορση (στάδιο ύπνου Wake), υπάρχει αύξηση της δραστηριότητας κατά τη διάρκεια του ύπνου REM [19].



### Αναπνευστικό:

Η αναπνευστική ροή αλλάζει κατά τη διάρκεια του ύπνου και γίνεται όλο και πιο γρήγορη και ακανόνιστη, συγκεκριμένα κατά τη διάρκεια του ύπνου REM. Τα δεδομένα αναπνοής κατά τη διάρκεια του ύπνου REM είναι κάπως ασαφή, αλλά υποδηλώνουν ότι ο υποαερισμός (ανεπαρκής αερισμός των πνευμόνων που οδηγεί σε μείωση της περιεκτικότητας σε οξυγόνο ή αύξηση της περιεκτικότητας σε διοξείδιο του άνθρακα στο αίμα ή και στα δύο) συμβαίνει με παρόμοιο τρόπο όπως κατά τη διάρκεια του ύπνου NREM. Διάφοροι παράγοντες συμβάλλουν στον υποαερισμό κατά τη διάρκεια του ύπνου NREM, και πιθανώς REM, όπως ο μειωμένος μυϊκός τόνος του φάρυγγα. Επιπλέον, κατά τη διάρκεια του ύπνου REM, υπάρχει μειωμένη κίνηση του κλωβού των πλευρών και αυξημένη αντίσταση των ανώτερων αεραγωγών λόγω της απώλειας τόνου στους μεσοπλεύριους και τους άνω μυς των αεραγωγών. Γενικότερα, η αναπνευστική ροή παρουσιάζει λιγότερο αποτελεσματικές προσαρμοστικές αντιδράσεις κατά τη διάρκεια του ύπνου. Το αντανακλαστικό του βήχα, το οποίο συνήθως αντιδρά σε ερεθιστικούς παράγοντες στον αεραγωγό, καταστέλλεται κατά τη διάρκεια του ύπνου REM και NREM. Η υποξική αναπνευστική απόκριση είναι επίσης χαμηλότερη στον ύπνο NREM από ό,τι κατά τη διάρκεια της εγρήγορσης και μειώνεται περαιτέρω κατά τον ύπνο REM. Ομοίως, η απόκριση διέγερσης στην αναπνευστική αντίσταση (για παράδειγμα, αντίσταση στην εισπνοή ή στην εκπνοή) είναι χαμηλότερη στο στάδιο 3 και στο στάδιο 4 του ύπνου [24].

Στο τελευταίο μέρος της ανάλυσης του ύπνου από ιατρικής πλευράς, παραθέτουμε την Φυσιοπαθολογία του ύπνου, δηλαδή την επίδραση του ύπνου στον άνθρωπο και τις διαταραχές που προκαλεί ο μη φυσιολογικός ύπνος.

### Φυσιοπαθολογία του ύπνου

Τα άτομα με διαταραχές που σχετίζονται με τον ύπνο, υποφέρουν από κατακερματισμό του ύπνου και άπνοια κατά τη διάρκεια του ύπνου τους. Καθώς αρχίζουν να πέφτουν σε βαθύτερα στάδια ύπνου (στάδια NREM 3 και 4), ο ανώτερος αεραγωγός τους καταρρέει και παρεμποδίζει την κανονική τους αναπνοή. Αυτή η παρεμβολή αναγκάζει το σώμα να επιστρέψει στα ελαφρύτερα στάδια ύπνου για να συνεχίσει την καλύτερη αναπνοή. Τα στάδια στον βαθύ ύπνο είναι σημαντικά για την κανονική λειτουργία για την αναπλήρωση του ανοσοποιητικού συστήματος και τον σωστό μεταβολισμό και ανάπτυξη. Τα άτομα με άπνοια ύπνου, δεν περνούν από τα φυσιολογικά στάδια του κύκλου του ύπνου. Έχουν μειωμένο στάδιο N3 και REM ύπνου καθώς η κατάρρευση των αεραγωγών τους εμποδίζει να αναπνεύσουν σε αυτές τις βαθύτερες καταστάσεις ύπνου. Αυτό οδηγεί σε υπερβολική υπνηλία κατά τη διάρκεια της ημέρας, καθώς ο σωστός, αποτελεσματικός ύπνος δεν επιτυγχάνεται όλη τη νύχτα. Υπάρχουν δύο τύποι υπνικής άπνοιας: η κεντρική και η αποφρακτική. Η κεντρική άπνοια ύπνου εμφανίζεται όταν ο εγκέφαλος αποτυγχάνει να σηματοδοτήσει σωστά τους αναπνευστικούς μύες κατά τη διάρκεια του ύπνου. Από την άλλη πλευρά, η αποφρακτική άπνοια ύπνου είναι ένα μηχανικό πρόβλημα στο οποίο υπάρχει μερική ή πλήρης απόφραξη του ανώτερου αεραγωγού [25] [26].

Κατά τη διάρκεια του ύπνου REM, συνήθως δεν κινούμαστε καθώς οι μύες μας είναι παράλυτοι. Εάν διαταραχθεί η προσωρινή παράλυση (ατονία) του ύπνου REM, μπορεί κανείς να πραγματοποιήσει όνειρα φυσιολογικά. Αυτό ονομάζεται διαταραχή ύπνου ταχείας κίνησης των ματιών (REM). Με αυτή τη διαταραχή, δεν επιτυγχάνεται φυσιολογική παράλυση και υπάρχουν επεισόδια μη φυσιολογικής κίνησης όλη τη νύχτα. Αυτά τα επεισόδια συμπίπτουν με το στάδιο του ύπνου REM. Η αιτία αυτής της διαταραχής δεν είναι πλήρως γνωστή, αλλά μπορεί να σχετίζεται με ορισμένες εκφυλιστικές νευρολογικές καταστάσεις, όπως η νόσος του Πάρκινσον ή η άνοια του σώματος Lewy [27] [28].

Η ναρκοληψία είναι μια άλλη ανωμαλία του κύκλου ύπνου στην οποία τα άτομα παρουσιάζουν συνήθως επίμονη υπνηλία κατά τη διάρκεια της ημέρας και σύντομα επεισόδια μυϊκής αδυναμίας γνωστά ως καταπληξία. Στη ναρκοληψία, η ρύθμιση του ύπνου διαταράσσεται και τα άτομα τείνουν να παραλείπουν

τις αρχικές φάσεις του ύπνου και να πηγαίνουν απευθείας στον ύπνο REM. Ακόμη και κατά τη διάρκεια σύντομων μεσημεριανών ύπνων, αυτά τα άτομα μπορούν να εισέλθουν στη φάση REM και να έχουν σύντομα όνειρα. Αυτό περιορίζει την διάρκεια του ύπνου τους στο στάδιο του βαθύ ύπνου N3 και έτσι προκαλείται ακανόνιστο μοτίβο ύπνου. Αυτά τα άτομα βιώνουν επίσης ξαφνικές απώλειες μυϊκής δύναμης καθώς οι μύες του σώματος είναι ατονικοί και παραλύουν στη φάση του ύπνου REM. Αυτά τα κενά στον ύπνο REM συμβαίνουν οποιαδήποτε στιγμή κατά τη διάρκεια της ημέρας και συνήθως διαρκούν από δευτερόλεπτα έως λεπτά [29] [30].

Η υπονοβασία είναι συχνό φαινόμενο σε παιδιά σχολικής ηλικίας. Αυτά τα άτομα τείνουν να κάνουν κινήσεις που φαίνονται σκόπιμες και είναι σημαντικό να καταλάβουμε ότι δεν πραγματοποιούν τα όνειρά τους. Τα όνειρα συμβαίνουν κατά τη φάση της γρήγορης κίνησης των ματιών του κύκλου, κατά την οποία το σώμα είναι πλήρως παράλυτο. Η υπονοβασία τείνει να συμβαίνει επειδή ο κύκλος ύπνου βρίσκεται ακόμα στη φάση ωρίμανσης και οι σωστοί κύκλοι ύπνου/αφύπνισης δεν έχουν ακόμη ρυθμιστεί. Η υπονοβασία συνδέεται με κοινές συμπεριφορές, όπως το ντύσιμο, το φαγητό και η ούρηση. Επομένως, η υπονοβασία λαμβάνει χώρα στις φάσεις της μη γρήγορης κίνησης των ματιών, συνήθως στη N3 [31].

Στην επόμενη ενότητα, θα κάνουμε μια εισαγωγή στις τεχνολογίες των φορητών συσκευών, εφόσον στη παρούσα ενότητα έχουμε αποκτήσει βασικές γνώσεις για τον ύπνο, τα στάδια του, τη φυσιολογία του και την επίδραση του στον άνθρωπο.

### 1.1.2 Τεχνολογία Επίβλεψης ύπνου με φορητές συσκευές (wearables)

Η ακριβής εκτίμηση του ύπνου είναι ζωτικής σημασίας για την καλύτερη κατανόηση και αξιολόγηση του ρόλου του στην υγεία και τις ασθένειες. Η έκρηξη της τεχνολογίας φορητών συσκευών είναι μέρος της ψηφιακής επανάστασης στην υγεία και παράγει πολλές νέες, εξαιρετικά εξελιγμένες και σχετικά φθηνές συσκευές καταναλωτών που συλλέγουν δεδομένα από πολλαπλούς αισθητήρες και εξάγουν πληροφορίες σχετικά με τις συμπεριφορές των χρηστών, συμπεριλαμβανομένου του ύπνου. Αυτές οι συσκευές είναι πλέον σε θέση να συλλέξουν διαφορετικά βιοσήματα για τον προσδιορισμό, για παράδειγμα, του καρδιακού ρυθμού και της μεταβλητότητάς του, της αγωγιμότητας του δέρματος και της θερμοκρασίας, επιπλέον της κινητικής δραστηριότητας. Καταγράφουν συνεχόμενα, δημιουργώντας συντριπτικά μεγάλα σύνολα δεδομένων (Big Data), με τη δυνατότητα να προσφέρουν μια άνευ προηγουμένου παρακολούθηση και επίβλεψη (monitoring) της υγείας των χρηστών. Δυστυχώς, υπάρχουν λίγες οδηγίες εντός και εκτός της επιστημονικής κοινότητας, που ασχολείται ερευνητικά με τον ύπνο, για τη σωστή χρήση τους, οδηγώντας σε σύγχυση και αντιπαράθεσεις σχετικά με την εγκυρότητα και την εφαρμογή τους. Η τρέχουσα τελευταία τεχνολογία των "φορητών" συσκευών στοχεύει να αναδείξει τη χρήση, την αξιοποίηση και τη χρησιμότητα τους από τους καταναλωτές, στην κλινική πρακτική και έρευνα. Οι οδηγίες για τυποποιημένη αξιολόγηση της απόδοσης της συσκευής, κρίνονται απαραίτητες και πρέπει να ληφθούν υπόψη αρκετοί κρίσιμοι παράγοντες πριν από τη χρήση αυτών των συσκευών σε κλινικά και ερευνητικά πρωτόκολλα ύπνου. Τελικά, η "φορητή" τεχνολογία ύπνου υπόσχεται για την προώθηση της κατανόησης της υγείας του ύπνου, ωστόσο, χρειάζεται μια προσεκτική πορεία προς τα εμπρός, κατανοώντας τα οφέλη και τις παγίδες αυτής της τεχνολογίας όπως εφαρμόζεται στην έρευνα ύπνου και την κλινική ιατρική ύπνου [32][33].

Από πλευράς καταναλωτών ή μη ειδικών στον τομέα του ύπνου, έχει παρατηρηθεί μια τεράστια αύξηση στην ενημέρωσή τους την τελευταία δεκαετία πάνω στο θέμα του ύπνου. Αν και οι επαγγελματίες του ύπνου λαμβάνουν συχνά κάποια εύσημα για αυτήν την αυξημένη ευαισθητοποίηση, η πραγματικότητα είναι ότι η ευαισθητοποίηση του κοινού οφείλεται κυρίως στην έκρηξη των συσκευών "αυτοπαρακολούθησης" και στην αυξημένη προσοχή των μέσων ενημέρωσης σχετικά με τον ύπνο. Ως αποτέλεσμα αυτών των αλλαγών, οι καταναλωτές έχουν μεγαλύτερη επίγνωση των προτύπων ύπνου τους. Ορισμένοι καταναλωτές επενδύουν σε συσκευές παρακολούθησης για να αξιολογήσουν τις ανησυχίες τους για τον ύπνο. Πολλοί άλλοι εύκολα μαθαίνουν για τον ύπνο από ένα πρόγραμμα παρακολούθησης φυσικής κατάστασης που

αγοράστηκε για άλλους λόγους [34].

Επιπρόσθετα, οι "φορητές συσκευές" παρακολούθησης ύπνου (π.χ. βραχιολάκια, smartwatches, headbands, δαχτυλίδια) αποτελούν μέρος μιας μεγαλύτερης οικογένειας καταναλωτικών τεχνολογιών ύπνου (Consumer Sleep Technology (CST)). Το CST περιλαμβάνει smartphones, αισθητήρες στο κρεβάτι και αισθητήρες χωρίς επαφή, καθώς και άλλες συσκευές που έχουν σχεδιαστεί για να βελτιώνουν τον ύπνο ή/και να βελτιώνουν τις αποκρίσεις/αντιδράσεις του ανθρώπου κατά τον ύπνο, όπως νευροδιεγέρτες, συσκευές βιοανάδρασης και συστήματα παραπλάνησης εγκεφαλικών κυμάτων [35][34].

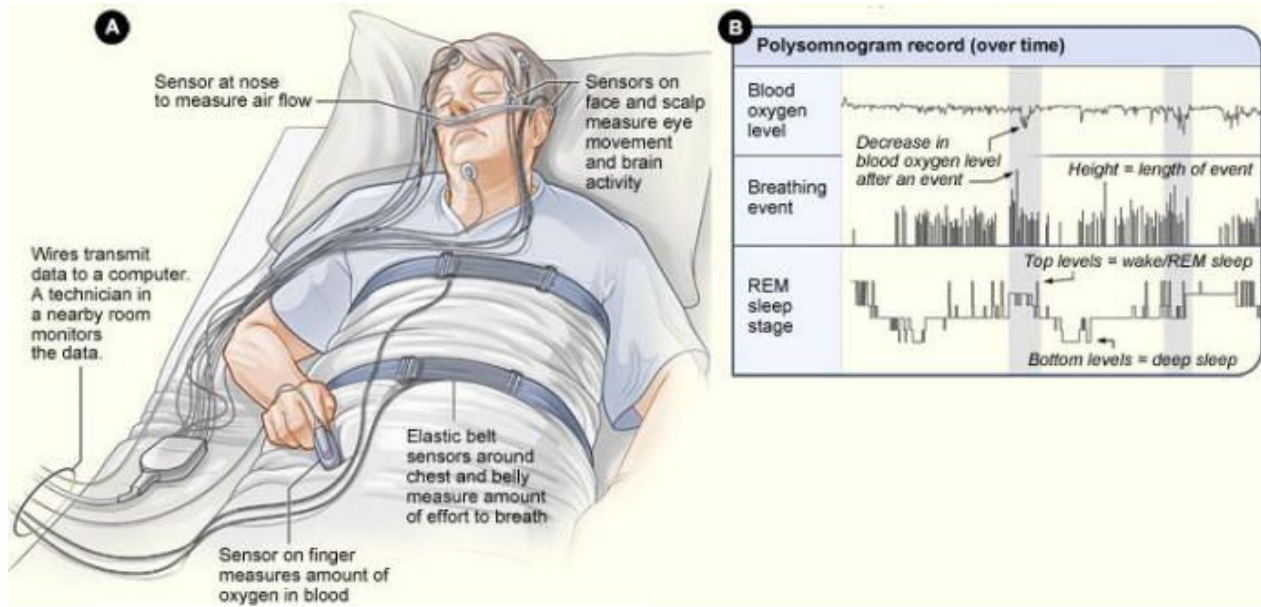
Θεωρούμε τις "φορητές" συσκευές καταγραφής δεδομένων ύπνου, αυτές τις συσκευές που είναι σχετικά χαμηλού κόστους και διαθέσιμες χωρίς ιατρική συνταγή ή κλινικές συστάσεις. Είναι επίσης σχεδιασμένες για την καταγραφή της ανθρώπινης δραστηριότητας μετρώντας πολλά βιοσήματα (π.χ. καρδιακό ρυθμό και μεταβλητότητα, αγωγιμότητα του δέρματος, θερμοκρασία), εκτός από την κίνηση, από την οποία μπορούν να αντληθούν πληροφορίες για διάφορες ανθρώπινες συμπεριφορές, συμπεριλαμβανομένου του ύπνου που εξάγεται. Η εύκολη πρόσβαση στα καταγεγραμμένα δεδομένα σε νεοφοϋπολογιστικές υποδομές (Cloud), η χρηστικότητα (διεπαφές χρηστών για κινητά), η καινοτομία και η προσιτή τιμή οδήγησαν στην ευρεία χρήση τους και συνέβαλαν στην αύξηση της ευαισθητοποίησης σχετικά με τη σημασία του ύπνου στον γενικό πληθυσμό [36][37].

Μέσα στις ερευνητικές και κλινικές κοινότητες ύπνου, υπάρχει αυξανόμενη αναγνώριση των πιθανών οφελών από τη χρήση φορητών συσκευών ύπνου. Τα οφέλη περιλαμβάνουν την εύκολη πρόσβαση σε μια απίστευτη και άνευ προηγουμένου ποσότητα πληροφοριών σχετικά με τον ύπνο, που συλλέγονται στο φυσικό περιβάλλον των ανθρώπων για εκτεταμένες περιόδους. Τα δεδομένα μπορούν να συλλεχθούν ανά πάσα στιγμή χωρίς ενεργή συμμετοχή των χρηστών (που φορούν απλώς μια συσκευή) και χωρίς την ανάγκη εξειδικευμένων τεχνικών που επεξεργάζονται τα δεδομένα (τα οποία παρέχονται συνήθως σε συνοπτική μορφή, όπως τα συνολικά λεπτά ύπνου). Ωστόσο, παρά αυτά τα πιθανά πλεονεκτήματα, ένα θεμελιώδες ζήτημα παραμένει άλυτο. Για πολλές από τις συσκευές και τα σχετικά συστήματα, **υπάρχουν ανεπαρκή διαθέσιμα δεδομένα σχετικά με την εγκυρότητα, την ακρίβεια και την αξιοπιστία τους στη μέτρηση των διαφόρων παραμέτρων ύπνου και άλλων δεικτών, όπως εκείνων που αντανακλούν την καρδιακή λειτουργία.** Επίσης, δεν υπάρχει συναίνεση μεταξύ των κλινικών ιατρών ύπνου και των επιστημόνων έρευνας σχετικά με τον τρόπο αντιμετώπισης της έκρηξης της "φορητής" τεχνολογίας και δεν υπάρχουν ευρέως αποδεκτά πρότυπα ως προς τον τρόπο εφαρμογής της χρήσης αυτών των συσκευών στην έρευνα και τις κλινικές ρυθμίσεις ύπνου [38][39].

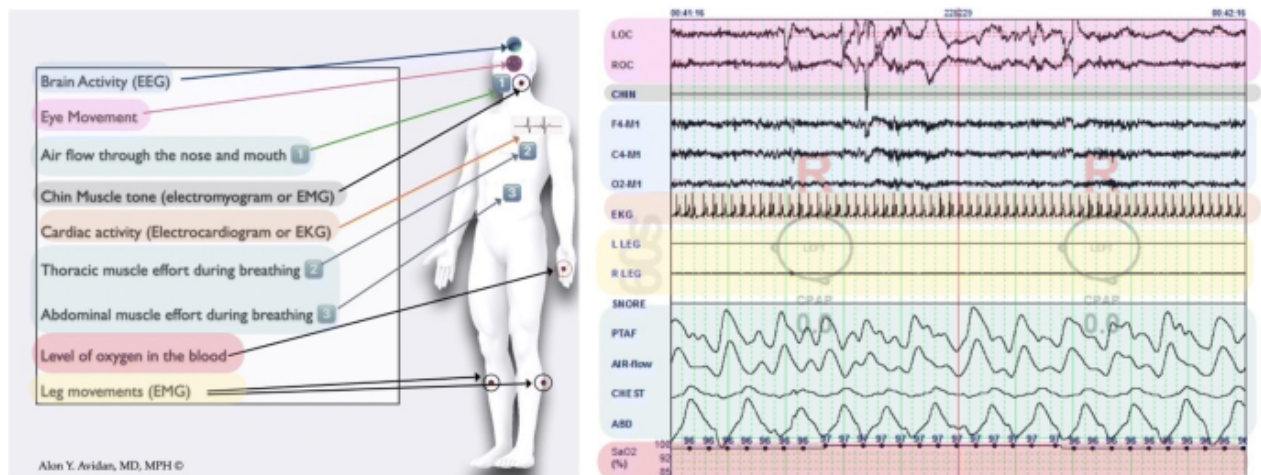
## Μέθοδοι καταγραφής ύπνου: Πολυυπνογραφία και ακτιγραφία (PSG και actigraphy)

Το PSG (Polysomnography) όπως φαίνεται στο Σχήμα 1.1.4 και 1.1.5, είναι η καθιερωμένη τυπική μέθοδος για την αξιολόγηση του ύπνου και αποτελεί τη κύρια αναφορά ανάμεσα σε πολλές μεθόδους καταγραφής δεδομένων ανθρώπινης δραστηριότητας. Το PSG είναι ένας ολοκληρωμένος μετρητής ύπνου, βασισμένος στην ταυτόχρονη καταγραφή φλοιώδους (ηλεκτροεγκεφαλογράφημα [EEG]), μυϊκής (ηλεκτρομυογράφημα EMG [40]) και ηλεκτροφθαλμικής δραστηριότητας, μέσω της τυποποιημένης τοποθέτησης (διεθνές σύστημα 10/20 EEG) ηλεκτροδίων επιφανείας του τριχωτού της κεφαλής. Στο πλαίσιο της αξιολόγησης του PSG, μια σειρά επιπλέον φυσιολογικών σημάτων (π.χ. ηλεκτροκαρδιογράφημα [ECG] [41], αναπνοή, κινήσεις ποδιών, ρινική πίεση, αποκορεσμός οξυγόνου και θέση σώματος) αξιολογούνται τακτικά και βοηθούν στον χαρακτηρισμό της σύνθετης φύσης του ύπνου και της πιθανής παρουσίας των διαταραχών του ύπνου. Ακολουθώντας τυποποιημένους οπτικούς κανόνες βασισμένους στις συστάσεις της Αμερικανικής Ακαδημίας Ιατρικής του ύπνου (AASM), ο ύπνος βαθμολογείται χειροκίνητα σε διαστήματα **30 δευτερολέπτων** [42] με οπτική αναγνώριση συγκεκριμένων φάσεων (π.χ. διεγερση, σύμπλεγμα K, ατράκτους) και τονωτικού (π.χ. ποσοστό "αργού" κύματος (slow wave)

ύπνου σε μια εποχή). Χαρακτηριστικά από το πολλαπλό EEG και τα υπόλοιπα κανάλια γίνεται αντιστοίχιση κάθε εποχής ως εξής: ύπνος αφύπνισης (Wake), N1, N2, N3 ή REM. Το PSG συνήθως περιορίζεται σε εργαστηριακές έρευνες ύπνου και κλινικές ρυθμίσεις καθώς απαιτεί εξειδικευμένο εξοπλισμό (ειδικό σύστημα απόκτησης PSG) και εξειδίκευση (επαγγελματικά εκπαιδευμένο προσωπικό) για την καταγραφή, βαθμολόγηση και ερμηνεία δεδομένων PSG. Αν και υπάρχουν φορητά συστήματα PSG, η χρήση του PSG είναι πολύ δαπανηρή και μη πρακτική για να είναι εφικτή για τη μέτρηση του ύπνου για παρατεταμένες περιόδους εκτός ερευνητικών μελετών.



Σχήμα 1.1.4: Συσκευή PSG Polysomnography καταγραφής δεδομένων κατά τη διάρκεια του ύπνου. Το σύστημα PSG καταγράφει δεδομένα, εγκεφαλικής δραστηριότητας (EEG), καρδιακών παλμών από τον ανθρώπινο δείκτη (heart rate), δεδομένα ροής οξυγόνου (air flow) από τη μύτη και τέλος, δεδομένα κίνησης του σώματος και χεριών μέσω μιας ελαστικής ζώνης. Όλα τα δεδομένα που καταγράφονται, μεταφέρονται σε ηλεκτρονικό υπολογιστή [43].



Σχήμα 1.1.5: Σήματα καταγραφής από συσκευή PSG Polysomnography καταγραφής δεδομένων [44]

Η αποδεκτή εναλλακτική λύση του PSG για μη εργαστηριακές ρυθμίσεις είναι η **ακτιγραφία (actigra-**

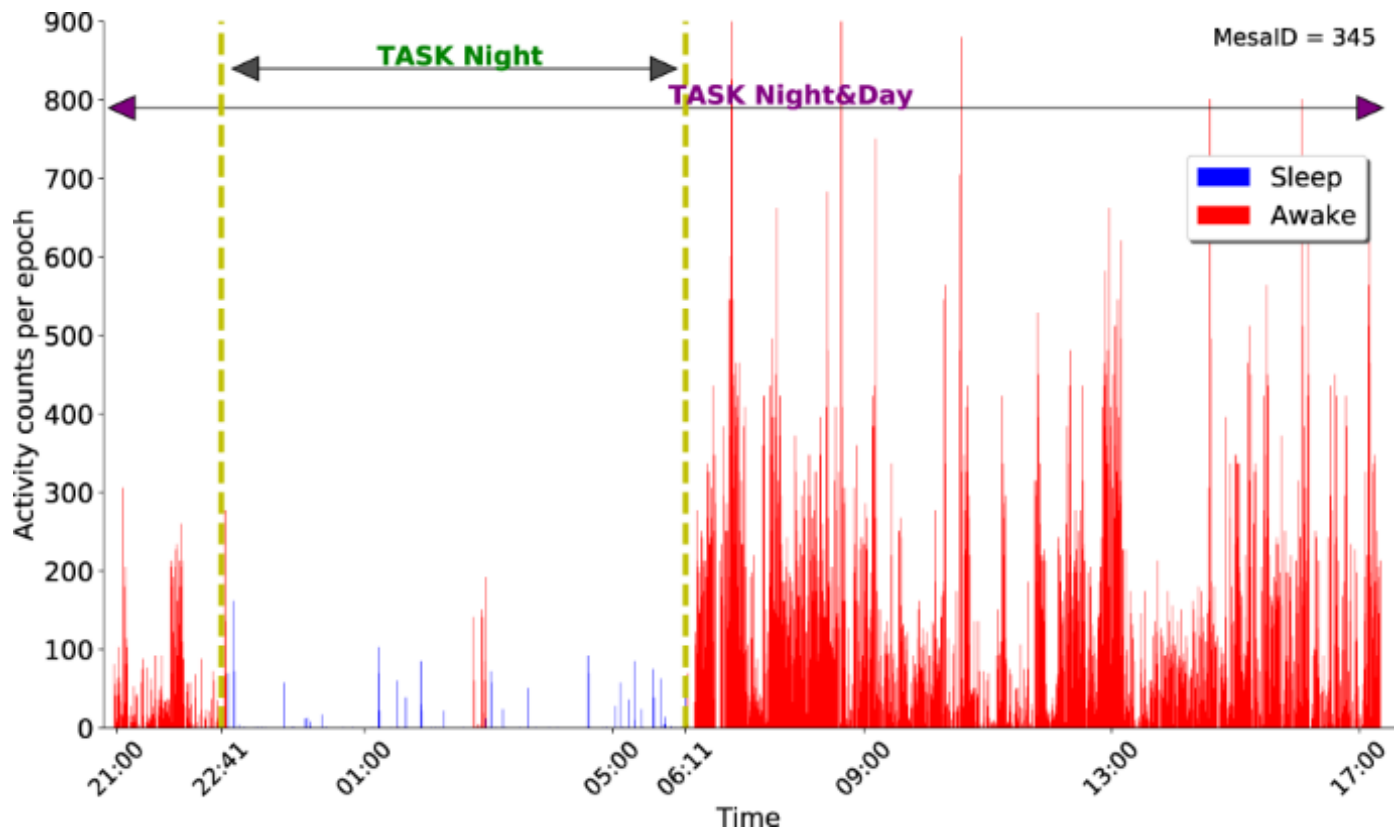


phy) [45] όπως φαίνεται στο Σχήμα 1.1.6. Οι συσκευές ακτιγραφίας (κυρίως συσκευές που φοριούνται στον καρπό) βασίζονται σε ένα επιταχυνσιόμετρο [46] που μας επιτρέπει να μετρήσουμε την επιτάχυνση και να έχουμε την γραφική απεικόνισή της σε σχέση με τον χρόνο ή με τις τιμές που μας δίνουν άλλοι αισθητήρες που χρησιμοποιούμε για τη μέτρηση μοτίβων δραστηριότητας (κίνηση) και την εκτίμηση των καταστάσεων ύπνου/αφύπνισης αποδεχόμενη την απλή υπόθεση ότι η κίνηση συνεπάγεται αφύπνιση και η μη κίνηση συνεπάγεται ύπνο, όπως φαίνεται στο Σχήμα 1.1.7. Συγκεκριμένα, στο Σχήμα 1.1.7 παρατηρούμε ότι όταν ο χρήστης του ρολογιού ακτιγραφίας κινείται και δεν κοιμάται, οι τιμές που καταγράφονται είναι υψηλές, ενώ όταν κοιμάται είναι χαμηλές έως και μηδενικές, συγκριτικά με όταν κινείται. Επίσης, λόγω του μικρού μεγέθους, της άνεσης και των αδιάβροχων ιδιοτήτων τους, οι συσκευές ακτιγραφίας έχουν σχεδιαστεί για να φοριούνται όλο το 24ωρο και επομένως είναι κατάλληλες για παρατεταμένες εγγραφές σε μη εργαστηριακές ρυθμίσεις. Το επιταχυνσιόμετρο της συσκευής ανιχνεύει την εμφάνιση και τον βαθμό κίνησης σε πολλαπλές κατευθύνσεις (π.χ., 3-άξονες), το οποίο μετατρέπεται σε ψηφιακό σήμα για να εξαχθεί ένας αριθμός δραστηριοτήτων (activity counts). Στη συνέχεια, αν οι τιμές δραστηριότητας υπερβούν ένα προκαθορισμένο κατώφλι που συνήθως καθορίζεται από κάποιον αλγόριθμο, τότε μια εποχή/χρονική περίοδος ύπνου αναγνωρίζεται ως αφύπνιση (wake) αλλιώς αναγνωρίζεται ως κατάσταση ύπνου, δηλαδή εάν οι τιμές δραστηριότητας είναι κάτω από το προκαθορισμένο κατώφλι. Επίσης, τα δεδομένα μπορούν να αποθηκευτούν με διαφορετικούς ρυθμούς, γεγονός που συμβάλλει στο πόσο καιρό μια συσκευή μπορεί να αποθηκεύσει συνεχή δεδομένα. Λόγω περιορισμών στην αποθήκευση δεδομένων, η πλειοψηφία της βιβλιογραφίας που χρησιμοποιεί ακτιγραφία βασίζεται σε ανάλυση 1 λεπτού για τη συλλογή δεδομένων.



Σχήμα 1.1.6: Συσκευή καταγραφής δεδομένων ακτιγραφίας [47]

Αν και η ακτιγραφία έχει πολλά πλεονεκτήματα, υπάρχουν περιορισμοί που πρέπει να ληφθούν υπόψη. Ενώ είναι **λιγότερο δαπανηρό από ένα σύστημα PSG**, δηλαδή οι κλινικές συσκευές είναι συχνά άνω των \$ 1000 η καθεμία, το οποίο παραμένει ένας περιοριστικός παράγοντας, όταν ο ύπνος πρέπει να καταγράφεται σε μεγάλα σύνολα δεδομένων σε πληθυσμούς, μπορεί η χωρητικότητα της συσκευής ακτιγραφίας να είναι περιορισμένη. Επιπλέον, αν και η ακτιγραφία δεν απαιτεί από κάποιον ειδικό να βαθμολογεί χειροκίνητα τις εγγραφές ύπνου ή να παρακολουθεί εγγραφές κατά τη διάρκεια της νύχτας, ένα έμπειρο μέλος του προσωπικού με εμπειρία στην ανάλυση ύπνου εξακολουθεί να απαιτείται για να



Σχήμα 1.1.7: Ακτίγραμμα. Διάγραμμα καταγραφής κινητικής δραστηριότητας (activity counts) για δύο στάδια ύπνου: Sleep και Wake [48]

εντοπίσει τυχόν προβλήματα με το ακτίγραμμα, όπως ελλείποντα δεδομένα.

Παρά τους αρκετούς περιορισμούς του πεδίου της ακτιγραφίας, πιθανώς το κόστος της ακτιγραφίας, η απαίτηση τεχνικού προσωπικού και του χρόνου για την επεξεργασία των δεδομένων, είναι μεταξύ των κύριων παραγόντων που οδηγούν τους ερευνητές και τους κλινικούς ιατρούς να θεωρούν τις καταναλωτικές φορητές συσκευές (wearables) ως εναλλακτική λύση η οποίες συλλέγουν εύκολα δεδομένα ύπνου σε μη εργαστηριακές ρυθμίσεις και συνθήκες [49].

### 1.1.3 Σύνολο Δεδομένων MESA

Η Πολυεθνική Μελέτη Αθληροσκλήρωσης MESA [50] [51] είναι μια μελέτη των χαρακτηριστικών της υποκλινικής καρδιαγγειακής νόσου (ασθένεια που ανιχνεύεται μη επεμβατικά πριν προκαλέσει κλινικά σημεία και συμπτώματα) και των παραγόντων κινδύνου που προβλέπουν την εξέλιξη σε κλινικά εμφανή καρδιαγγειακή νόσο ή την εξέλιξη της υποκλινικής νόσου. Οι ερευνητές του MESA μελέτησαν ένα ποικίλο, πληθυσμιακό δείγμα 6.814 ασυμπτωματικών ανδρών και γυναικών ηλικίας 45-84 ετών. Το 38 % των συμμετεχόντων είναι λευκοί, το 28 % Αφροαμερικανοί, το 22 % Ισπανόφωνοι και το 12 % Ασιάτες, κυρίως κινεζικής καταγωγής.

Οι συμμετέχοντες προέρχονται από έξι κέντρα σε όλες τις Ηνωμένες Πολιτείες: Πανεπιστήμιο Wake Forest, Πανεπιστήμιο Columbia, Πανεπιστήμιο Johns Hopkins, Πανεπιστήμιο της Μινεσότα, Πανεπιστήμιο Northwestern και Πανεπιστήμιο της Καλιφόρνια στο Λος Άντζελες. Κάθε συμμετέχων έλαβε εκτεταμένη φυσική εξέταση για τον προσδιορισμό της στεφανιαίας αβεστοποιήσεως, της κοιλιακής μάζας και λειτουργίας, της ενδοθηλιακής αγγειοδιαστολής με τη μεσολάβηση ροής, του πάχους του έσω-έσω

τοιχώματος της καρωτίδας και της παρουσίας ηχογενών διακρίσεων στην καρωτιδική αρτηρία, αγγειακή ανεπάρκεια κάτω άκρων, μορφές αρτηριακών κυμάτων (ηλεκτροκαρδιογραφικές μορφές, ECG) μέτρα, τυπικοί παράγοντες στεφανιαίου κινδύνου, κοινωνικοδημογραφικοί παράγοντες, παράγοντες τρόπου ζωής και ψυχοκοινωνικοί παράγοντες. Η επιλεγμένη επανάληψη των μέτρων υποκλινικής νόσου και των παραγόντων κινδύνου στις επισκέψεις παρακολούθησης, επιτρέπει τη μελέτη της εξέλιξης της νόσου. Τα δείγματα αίματος έχουν αναλυθεί για πιθανούς βιοχημικούς παράγοντες κινδύνου και έχουν αποθηκευτεί για μελέτες περιπτώσεων ελέγχου. Έχει εξαχθεί DNA και τα λεμφοκύτταρα συντηρήθηκαν στο κρύο για μελέτη υποψήφιων γονιδίων και πιθανώς, σάρωση σε όλο το γονιδίωμα, έκφραση και άλλες γενετικές τεχνικές. Οι συμμετέχοντες παρακολουθούνται για αναγνώριση και χαρακτηρισμό συμβάντων καρδιαγγειακής νόσου, συμπεριλαμβανομένου του οξέος εμφράγματος του μυοκαρδίου και άλλων μορφών στεφανιαίας νόσου (CHD), εγκεφαλικού επεισοδίου και συμφορητικής καρδιακής ανεπάρκειας, για παρεμβάσεις καρδιαγγειακών παθήσεων και για τη θνησιμότητα.

Εκτός από τα έξι κέντρα, το MESA περιλαμβάνει ένα Συντονιστικό Κέντρο, ένα Κεντρικό Εργαστήριο και Κεντρικά Κέντρα Ανάγνωσης για Υπολογιστική Τομογραφία (CT), Απεικόνιση Μαγνητικού Συντονισμού (MRI), Υπερηχογράφημα και Ηλεκτροκαρδιογραφία (ECG). Η ανάπτυξη πρωτοκόλλου, η εκπαίδευση του προσωπικού και οι πιλοτικές δοκιμές πραγματοποιήθηκαν τους πρώτους 18 μήνες της μελέτης. Η πρώτη εξέταση πραγματοποιήθηκε σε διάστημα δύο ετών, από τον Ιούλιο του 2000 έως τον Ιούλιο του 2002. Ακολούθησαν τρεις περίοδοι εξέτασης διάρκειας 17-20 μηνών. Πραγματοποιήθηκε επικοινωνία με τους συμμετέχοντες κάθε 9 έως 12 μήνες καθ' όλη τη διάρκεια της μελέτης για την αξιολόγηση της κλινικής νοσηρότητας και θνησιμότητας [52].

### Σύνολο Δεδομένων:

- **Ακτιγραφία (activity counts):** Στο πείραμα της MESA Sleep συμμετείχαν συνολικά 2.237 άνθρωποι οι οποίοι φορούν συσκευές ακτιγραφίας [53] που φοριούνται στον καρπό (Actiwatch Spectrum, Philips Respironics [54]) για το χρονικό διάστημα μεταξύ 2010 και 2013. Οι συμμετέχοντες έλαβαν οδηγίες να φορούν το ρολόι για μια εβδομάδα. Οι καταγραφές σημειώθηκαν από εκπαιδευμένο τεχνικό στο Boston Sleep Reading Center. Έχουν δημιουργηθεί αρχεία δεδομένων (.csv) για 2.159 συμμετέχοντες με δεδομένα ακτιγραφίας. Κάθε σειρά σε αυτά τα αρχεία αντιπροσωπεύει συνοπτικά δεδομένα διάρκειας 30 δευτερολέπτων από τη συσκευή ακτιγραφίας. Σε μια εποχή αντιστοιχούν δηλαδή 30s (οι ημερομηνίες έχουν αφαιρεθεί). Επίσης, εποχές χωρίς μετρήσεις δραστηριότητας ή/και φωτεινές ενδείξεις συνήθως υποδεικνύουν την αποτυχία του ρολογιού να παρέχει έγκυρη μέτρηση για αυτήν την εποχή. Τέλος, στο Σχήμα 1.1.8 παρουσιάζονται τα δεδομένα ακτιγραφίας της βάσης δεδομένων MESA, η ανάλυση των οποίων θα γίνει σε παρακάτω Κεφάλαια με την προεπεξεργασία των δεδομένων.
- **Πολυυπνογραφία (PSG data)** Η πολυυπνογραφία (PSG) για τους συμμετέχοντες [55], διεξήχθη χρησιμοποιώντας το σύστημα Compumedics Somte (Compumedics Ltd., Abbotsford, Αυστραλία) [56]. Οι αισθητήρες και το προφίλ καταγραφής αποτελούνταν από ηλεκτροεγκεφαλογραφήματα φλοιού (κεντρικά C4-M1, ινιακά Oz-Cz και μετωπικές απαγωγές Fz-Cz), αμφίπλευρη ηλεκτροφθαλμογραφία (EOG) [57], ηλεκτρομυογράφημα (EMG) του πηγουνιού, θωρακική και κοιλιακή αναπνευστική επαγωγή πληθυσμογραφίας (με αυτοβαθμονόμηση ζώνης), ροή αέρα (με ρινικό-στοματικό θερμοστοιχείο και καταγραφή πίεσης από ρινικό σωληνίσκο), ηλεκτροκαρδιογράφημα ECG, κινήσεις των ποδιών και παλμική οξυμετρία των δακτύλων. Εκτός από τη σύνδεση αισθητήρων και ηλεκτροδίων, τα μέλη του προσωπικού πραγματοποίησαν βαθμονομήσεις σήματος. Οι νυχτερινές ηχογραφήσεις διαβιβάστηκαν στο κέντρο ανάγνωσης στο Brigham and Women's Hospital και τα δεδομένα βαθμολογήθηκαν από εκπαιδευμένους τεχνικούς χρησιμοποιώντας τις τρέχουσες οδηγίες. Όπως και στην ακτιγραφία, τα ακατέργαστα δεδομένα πολυυπνογραφίας (raw data) είναι διαθέσιμα για τους 2.056 συμμετέχοντες στο MESA Sleep. Κάθε ηχογράφηση έχει ένα αρχείο σήματος (.EDF) [58] και δύο εκδόσεις σχολιασμών βαθμολογίας συμβάντος και

Name	Label	Units / Categories / Notes
mesaid	Random BioLINCC ID	
line	Epoch line number	
linetime	Clock time	HH:MM:SS
offwrist	Off wrist indicator	0 = On wrist / 1 = Off wrist
activity	Activity count	
marker	Event marker indicator	0 = Marker not pressed / 1 = Marker pressed
whitelight	White light	Lux
redlight	Red light	Microwatts per square centimeter
greenlight	Green light	Microwatts per square centimeter
bluelight	Blue light	Microwatts per square centimeter
wake	Awake indicator*	0 = Asleep / 1 = Awake
interval	Interval type*	
dayofweek	Day of the week	1 = Sunday / 2 = Monday / etc.
daybymidnight	Incrementing day number (midnight)	Counter turns over at midnight
daybynoon	Incrementing day number (noon)	Counter turns over at noon

Σχήμα 1.1.8: Δεδομένα από περικάρπιο μετρητή (Actigraph) της MESA [53].

αναγνώριση-καταγραφή εποχής (.XML) (sleep epoch).

## Σκοπός χρήσης του συνόλου δεδομένων MESA

Η βάση δεδομένων MESA είναι ευρέως διαδεδομένη στην επιστημονική κοινότητα που ασχολείται με την επίβλεψη του ύπνου, περιέχει μεγάλο όγκο δεδομένων για ευρεία ανάλυση και εξαγωγή αποτελεσμάτων σε αυτόν τον αναδυόμενο κλάδο της ψηφιακής επίβλεψης ύπνου. Η ευρεία ανάπτυξη γύρω από τον κλάδο e-health/e-prevention [59] έχει οδηγήσει επίσης και στην αντίστοιχη ανάπτυξη επιστημονικών ερευνών στον τομέα της κατηγοριοποίησης, σε κλάσεις, των σταδίων του ύπνου του ανθρώπου αλλά και γενικότερα στην επίβλεψη της ανθρώπινης δραστηριότητας με μη επεμβατικούς τρόπους. Επομένως, η ενασχόληση με τη συγκεκριμένη, αλλά και άλλες διαδεδομένες βάσεις, αποτελεί κίνητρο για τους ερευνητές να δημιουργήσουν νέες τεχνολογίες γύρω από την επίβλεψη ύπνου. Επίσης, η βάση MESA έχει αναλυθεί από αρκετούς επιστήμονες και είναι μια εγκεκριμένη βάση με σωστή καταγραφή δεδομένων. Η **ποικιλομορφία** σε δεδομένα, είναι βασικός παράγοντας όπου επιλέχθηκε η συγκεκριμένη βάση, διότι αυτό βοηθά πολύ στην γενίκευση των αποτελεσμάτων ταξινόμησης του ύπνου του ανθρώπου, δηλαδή όπως αναφέραμε και η προηγουμένως, η βάση MESA περιλαμβάνει πολλών ειδών διαφορετικών ανθρώπων με διαφορετικά χαρακτηριστικά (π.χ φυλετικά, ηλικία κλπ). Επίσης, ο **μεγάλος όγκος** των δεδομένων επιτρέπει την εκτεταμένη χρήση της μοντέρνας τεχνολογίας του **Deep Learning** και άλλων



σύγχρονων μοντέλων. Άρα αυξάνεται ο αριθμός των διαφορετικών μοντέλων όπου ένας ερευνητής μπορεί να χρησιμοποιήσει και να δοκιμάσει. Οπότε μπορεί να οδηγηθεί σε ασφαλέστερα αποτελέσματα και τελικά να προωθήσει τις νέες τεχνολογίες. Τέλος, η βάση δεδομένων MESA εξαιτίας του μεγάλου όγκου δεδομένων και της ποικιλομορφίας της, μπορεί να βοηθήσει και την τεχνολογική βιομηχανία στην ανάπτυξη εφαρμογών επίβλεψης της ανθρώπινης δραστηριότητας, όπως για παράδειγμα, μεγάλες εταιρίες ήδη έχουν αναπτύξει τέτοιες εφαρμογές (FitBit, AppleWatch, Samsung Galaxy). Για όλους τους παραπάνω λόγους επιλέχθηκε η συγκεκριμένη βάση για ανάλυση και εξαγωγή αποτελεσμάτων, με απώτερο σκοπό την **παροχή ψηφιακής βοήθειας στους ιατρούς** οι οποίοι ασχολούνται με τον ύπνο και τις διαταραχές του. Η ψηφιακή επίβλεψη του ύπνου μπορεί να βοηθήσει τους ιατρούς να βγάλουν έγκυρα συμπεράσματα με τον μικρότερο δυνατό κόστος και με σχεδόν μηδαμινό επεμβατικό τρόπο.

## Εφαρμογές του συνόλου δεδομένων MESA

Σε αυτή την υποενότητα θα παραθέσουμε μερικές έρευνες οι οποίες χρησιμοποιούν τη βάση δεδομένων MESA στον τομέα Μηχανικής Μάθησης και ταξινόμησης δεδομένων και πιο συγκεκριμένα στις εφαρμογές αναγνώρισης ανθρώπινης δραστηριότητας. Αρχικά, στην έρευνα [60] χρησιμοποιείται η βάση MESA για την πρόβλεψη επικίνδυνων καρδιαγγειακών εκβάσεων/συμβάντων (cardiovascular outcomes) για 12 χρόνια παρακολούθησης των συμμετεχόντων οι οποίοι δεν είχαν κάποια καρδιαγγειακή νόσο. Στην έρευνα συμπεριλήφθηκαν όλοι οι 6.814 συμμετέχοντες και χρησιμοποιήθηκαν δέντρα τυχαίας επιβίωσης (Random Survival Forests) για να προσδιοριστούν οι 20 πιο πιθανοί παράγοντες κάθε καρδιαγγειακού συμβάντος. Τελικά, τα μοντέλα Random Forests έδωσαν πολύ καλά αποτελέσματα πρόβλεψης, αυξημένα κατά 10-15% συγκριτικά με άλλες έρευνες. Αντίστοιχα, σε άλλες δύο έρευνες [61] [62], χρησιμοποιήθηκε η βάση MESA για την πρόβλεψη επικίνδυνων καρδιαγγειακών εκβάσεων/συμβάντων με μοντέλα Μηχανικής Μάθησης (Machine Learning) και Βαθιάς Μάθησης (Deep Learning) αντίστοιχα. Στη πρώτη έρευνα χρησιμοποιήθηκαν μοντέλα όπως XGBoost [63] και Random Forests [64], και στη δεύτερη έρευνα χρησιμοποιήθηκαν Βαθιά Νευρωνικά Δίκτυα (Deep Neural Networks). Και στις δύο έρευνες ο στόχος ήταν ο ίδιος - cardiovascular risk prediction, πετυχαίνοντας καλά αποτελέσματα, απλά με διαφορετικές προσεγγίσεις ως προς τα μοντέλα και την προεπεξεργασία των δεδομένων. Μια άλλη εφαρμογή της βάσης MESA βρίσκεται στη έρευνα [65], η οποία προσπαθεί να κάνει ταξινόμηση των δεδομένων στα 5 στάδια του ύπνου (Wake, NREM1, NREM2, NREM3 και REM). Συγκεκριμένα, χρησιμοποιεί τα δεδομένα πολυυπνογραφίας (PSG) των συμμετεχόντων, και για τη ταξινόμηση εφαρμόζει μοντέλα Βαθιάς Μάθησης που χρησιμοποιήθηκαν σε άλλες έρευνες [66] και [67] για τον ίδιο σκοπό. Δηλαδή, τα μοντέλα που χρησιμοποιούνται είναι Συνελικτικά Νευρωνικά Δίκτυα (CNN) και απλά Νευρωνικά Δίκτυα (FNN) τα οποία θα αναλυθούν σε παρακάτω ενότητα. Τελικά, χρησιμοποιώντας τα δεδομένα PSG, τα μοντέλα πετυχαίνουν πολύ καλά αποτελέσματα. Πρέπει να αναφερθεί όμως ότι σε αυτή την έρευνα [65], χρησιμοποιήθηκαν και άλλες βάσεις δεδομένων για την πληρότητα και γενίκευση των αποτελεσμάτων σε πολλούς διαφορετικούς πληθυσμούς. Επιπρόσθετα, στην έρευνα [67], οι ερευνητές προσπαθούν να αναλύσουν τα δεδομένα ύπνου της βάσης MESA και να διαγνώσουν-αναγνωρίσουν μέσα από αυτά, φαινόμενα που σχετίζονται με τη ναρκοληψία. Συγκεκριμένα, η ανάλυση του ύπνου για τη διάγνωση διαταραχών ύπνου όπως η Ναρκοληψία Τύπου 1 (T1N) απαιτεί επί του παρόντος οπτική επιθεώρηση των αρχείων πολυυπνογραφίας από εκπαιδευμένους τεχνικούς βαθμολόγησης. Οι ερευνητές, χρησιμοποιούν νευρωνικά δίκτυα σε περίπου 3.000 εγγραφές κανονικού και μη φυσιολογικού ύπνου για να αυτοματοποιήσουν τη διαδικασία ταξινόμησης των δεδομένων στα στάδια ύπνου (sleep staging classification), δημιουργώντας ένα γράφημα "υπνοδυσκνότητας", δηλαδή μια κατανομή πιθανότητας που μεταφέρει περισσότερες πληροφορίες από τα κλασικά υπνογράμματα. Τέλος, στην έρευνα [68], χρησιμοποιούνται μόνο τα δεδομένα καρδιακών παλμών της βάσης MESA ως είσοδο ενός Συνελικτικών Νευρωνικού Δικτύου (CNN) για την ταξινόμηση τους στα 4 στάδια του ύπνου. Το CNN μοντέλο πέτυχε πολύ καλά αποτελέσματα στο σύνολο αξιολόγησης το οποίο περιείχε και δεδομένα από άλλη

βάση δεδομένων, την Sleep Heart Health Study (SHHS). Η έρευνα αυτή είναι πολύ σημαντική για την δικιά μας έρευνα, επειδή σε επόμενο Κεφάλαιο θα χρησιμοποιούμε μόνο δεδομένα καρδιακών παλμών της βάσης MESA και θα συγκρίνουμε τα αποτελέσματα.

Συμπερασματικά, από τις παραπάνω έρευνες, παρατηρούμε ότι η βάση δεδομένων MESA έχει διάφορες εφαρμογές όπως σε θέματα καρδιαγγειακά (πρόβλεψη επικίνδυνων συμβάντων), θέματα ταξινόμησης δεδομένων ύπνου στα στάδια του και θέματα αναγνώρισης ασθενειών του ύπνου όπως η ναρκοληψία. Προφανώς, μπορεί κάποιος άλλος ερευνητής να βρει και άλλες εφαρμογές της, αλλά σε αυτό το στάδιο θέλουμε να αναδείξουμε τη σημασία της βάσης MESA σε διάφορες θεματολογίες της ιατρικής. Επίσης, διάφορα μοντέλα που προαναφέρθηκαν, θα εξηγηθούν αναλυτικότερα στην επόμενη ενότητα για τη λειτουργία τους και την εφαρμογή τους.

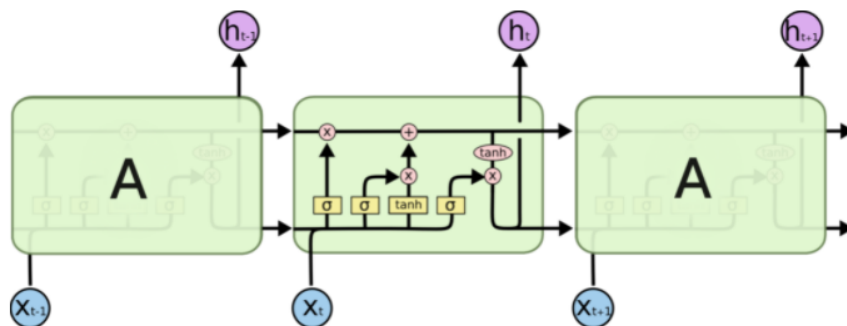
## 1.2 Μοντέλα Βαθιάς Μάθησης (Deep Learning Models)

Τα μοντέλα Βαθιάς Μάθησης αντιπροσωπεύουν ένα νέο πρότυπο μάθησης στην τεχνητή νοημοσύνη (AI) και τη μηχανική μάθηση. Πρόσφατα καινοτόμα αποτελέσματα στην ανάλυση εικόνων και την αναγνώριση ομιλίας έχουν δημιουργήσει τεράστιο ενδιαφέρον σε αυτόν τον τομέα, επειδή επίσης φαίνονται δυνατές εφαρμογές σε πολλούς άλλους τομείς που παρέχουν μεγάλα δεδομένα [69]. Από την άλλη πλευρά, η μαθηματική και η υπολογιστική μεθοδολογία που διέπει τα μοντέλα Βαθιάς Μάθησης είναι απαιτητική, ειδικά για τους διεπιστημονικούς επιστήμονες όπως τους ιατρούς του ύπνου που μας ενδιαφέρει για τη παρούσα μεταπτυχιακή εργασία. Συγκεκριμένα, τα μοντέλα Deep Learning έχουν φέρει επανάσταση στην έρευνα της Μηχανικής Μάθησης και τις εφαρμογές της, κυρίως λόγω της ικανότητας πολλαπλών επιπέδων δικτύου να εξάγουν χαρακτηριστικά (feature extraction) από δεδομένα εισόδου. Ως εκ τούτου, αναδεικνύονται ως αποτελεσματικοί "εξαγωγείς" χαρακτηριστικών και μας βοηθούν όχι μόνο αποφεύγοντας την δαπανηρή διαδικασία εξαγωγής χειροποίητων (hand crafted) χαρακτηριστικών, αλλά επίσης παρέχοντάς μας χαρακτηριστικά ακόμα καλύτερης ποιότητας. Αυτή η διαδικασία εξαγωγής χαρακτηριστικών μερικές φορές αναφέρεται ως εκμάθηση αναπαράστασης (Representation Learning). Στη συνέχεια, μετά την εξαγωγή χαρακτηριστικών, τα επόμενα επίπεδα και η έξοδος, μετατρέπουν αυτά τα χαρακτηριστικά σε προβλέψεις κλάσεων, εκτελώντας το στάδιο της ταξινόμησης. Οπότε στη παρούσα ενότητα θα αναλύσουμε τα μοντέλα Βαθιάς Μάθησης που θα χρησιμοποιηθούν στην έρευνα μας.

### 1.2.1 LSTM

Τα δίκτυα μακράς βραχυπρόθεσμης μνήμης – που συνήθως ονομάζονται απλώς LSTMs – είναι ένα ειδικό είδος Recurrent Neural Network (RNN), ικανό να μάθει μακροπρόθεσμες εξαρτήσεις. Εισήχθησαν από τους Hochreiter Schmidhuber (1997) [70] και βελτιώθηκαν και διαδόθηκαν από πολλούς ερευνητές. Λειτουργούν εξαιρετικά καλά σε μια μεγάλη ποικιλία προβλημάτων και χρησιμοποιούνται πλέον ευρέως. Τα LSTM έχουν σχεδιαστεί ρητά για την αποφυγή του προβλήματος της μακροπρόθεσμης εξάρτησης. Το να "θυμούνται" πληροφορίες για μεγάλες χρονικές περιόδους είναι πρακτικά η προεπιλεγμένη συμπεριφορά τους, όχι κάτι που δυσκολεύονται να μάθουν. Όλα τα επαναλαμβανόμενα νευρωνικά δίκτυα έχουν τη μορφή μιας αλυσίδας επαναλαμβανόμενων μονάδων νευρωνικού δικτύου. Στα τυπικά RNN [71], αυτή η επαναλαμβανόμενη μονάδα θα έχει μια πολύ απλή δομή, όπως ένα στρώμα *tanh*. Τα LSTM έχουν επίσης αυτή τη δομή σαν αλυσίδα, αλλά η επαναλαμβανόμενη μονάδα έχει διαφορετική δομή. Αντί να έχουμε ένα ενιαίο επίπεδο νευρωνικού δικτύου, υπάρχουν τέσσερα που αλληλεπιδρούν με έναν πολύ ιδιαίτερο τρόπο.

Στο Σχήμα 1.2.1, **κάθε γραμμή** φέρει ένα ολόκληρο διάνυσμα, από την έξοδο ενός κόμβου έως τις εισόδους άλλων. Οι **ροζ κύκλοι** αντιπροσωπεύουν σημειακές (point-wise) πράξεις, όπως η πρόσθεση διανυσμάτων, ενώ τα **κίτρινα πλαίσια** είναι στρώματα νευρωνικού δικτύου. Η **συγχώνευση**



The repeating module in an LSTM contains four interacting layers.

Σχήμα 1.2.1: Επαναλαμβανόμενη κυψέλη ενός επιπέδου LSTM [72].

γραμμών υποδηλώνει την συνένωση, ενώ μια **διχάλα γραμμών** υποδηλώνει ότι το περιεχόμενό της αντιγράφεται και τα αντίγραφα πηγαίνουν σε διαφορετικές τοποθεσίες. **Το κλειδί για τα LSTM είναι η κατάσταση κελιού (cell state), δηλαδή η οριζόντια γραμμή που διατρέχει την κορυφή του διαγράμματος.** Η κατάσταση του κελιού μοιάζει με "μεταφορική ταινία". Τρέχει κατευθείαν σε ολόκληρη την αλυσίδα, με μερικές μόνο μικρές γραμμικές αλληλεπιδράσεις. Είναι πολύ εύκολο για τις πληροφορίες να ρέουν κατά μήκος της αμετάβλητης. Το LSTM έχει την ικανότητα να αφαιρεί ή να προσθέτει πληροφορίες στην κατάσταση του κελιού, που ρυθμίζεται προσεκτικά από δομές που ονομάζονται πύλες (gates). Οι πύλες είναι ένας τρόπος για να περάσουν προαιρετικά πληροφορίες. Αποτελούνται από ένα στρώμα σιγμοειδούς (sigmoid) νευρικού δικτύου και μια πράξη πολλαπλασιασμού κατά σημείο. Το σιγμοειδές στρώμα εξάγει αριθμούς μεταξύ 0 και 1, περιγράφοντας την ποσότητα της πληροφορίας που πρέπει να περάσει από κάθε στοιχείο. Ένα LSTM έχει τρεις από αυτές τις πύλες, για την προστασία και τον έλεγχο της κατάστασης των κελιών. Το πρώτο βήμα στο LSTM είναι να αποφασιστεί ποιες πληροφορίες θα πεταχτούν από την κατάσταση του κελιού (cell state  $C_t$ ). Αυτή η απόφαση λαμβάνεται από ένα σιγμοειδές στρώμα που ονομάζεται «στρώμα της πύλης απόρριψης» (forget gate  $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$ ). Η πύλη απόρριψης δέχεται ως είσοδο τα  $h_{t-1}$  και  $x_t$  και επιστρέφει έναν αριθμό μεταξύ 0 και 1 για κάθε αριθμό στην κατάσταση της προηγούμενης κυψέλης  $C_{t-1}$ . Η τιμή 1 επιτρέπει όλη την πληροφορία να "περάσει" από το σιγμοειδές στρώμα, ενώ η τιμή 0 δεν επιτρέπει καμία πληροφορία να "περάσει". Ας θεωρήσουμε το παράδειγμα ενός γλωσσικού μοντέλου που προσπαθεί να προβλέψει την επόμενη λέξη με βάση όλες τις προηγούμενες. Σε ένα τέτοιο πρόβλημα, η κατάσταση του κελιού μπορεί να περιλαμβάνει το γένος του παρόντος υποκειμένου, έτσι ώστε να μπορούν να χρησιμοποιηθούν οι σωστές αντωνυμίες. Όταν βλέπουμε ένα νέο υποκείμενο, θέλουμε να ξεχάσουμε το γένος του παλιού υποκειμένου. Το επόμενο βήμα είναι να αποφασίσουμε ποιες νέες πληροφορίες πρόκειται να αποθηκεύσουμε στην κατάσταση του κελιού. Η διαδικασία αυτή αποτελείται από δύο μέρη. Πρώτον, ένα σιγμοειδές στρώμα που ονομάζεται "στρώμα πύλης εισόδου" (input gate  $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$ ), αποφασίζει ποιες τιμές θα ενημερωθούν. Στη συνέχεια, ένα επίπεδο **tanh** δημιουργεί ένα διάνυσμα νέων υποψήφιας τιμών,  $\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$ , που θα μπορούσαν να προστεθούν στην κατάσταση του κελιού. Στο επόμενο βήμα, συνδυάζονται αυτά τα δύο για να δημιουργηθεί μια ενημέρωση για την κατάσταση του κελιού. Στο παράδειγμα του γλωσσικού μοντέλου, θα θέλαμε να προσθέσουμε το γένος του νέου υποκειμένου στην κατάσταση κελιού, για να αντικαταστήσουμε το παλιό που ξεχνάμε. Επόμενο βήμα είναι να ενημερωθεί η παλιά κατάσταση κελιού,  $C_{t-1}$ , με νέα κατάσταση κελιού  $C_t$ . Οπότε, πολλαπλασιάζεται η παλιά κατάσταση με  $f_t$ , ξεχνώντας τα πράγματα που αποφασίσαμε να ξεχαστούν νωρίτερα. Μετά γίνεται πρόσθεση με  $i_t * \tilde{C}_t$ . Οπότε, οι νέες υποψήφιας τιμές είναι ( $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$ ), που προσαρμόζονται με το πόσο αποφασίσαμε να ενημερώσουμε κάθε τιμή της κατάστασης κελιού. Στην περίπτωση του γλωσσικού μοντέλου, θα

ξεχνούσαμε στην πραγματικότητα τις πληροφορίες σχετικά με το γένος του παλιού υποκειμένου και θα προσθέταμε τις νέες πληροφορίες, όπως αποφασίσαμε στα προηγούμενα βήματα. Τέλος, πρέπει να αποφασίσουμε τι θα βγάλουμε ως έξοδο. Αυτή η έξοδος θα βασίζεται στην κατάσταση του κελιού, αλλά θα είναι μια φιλτραρισμένη έκδοση. Αρχικά, εκτελείται ένα σιγμοειδές στρώμα το οποίο αποφασίζει ποια μέρη της κατάστασης κελιού θα εξαχθεί (output  $\mathbf{o}_t = \sigma(\mathbf{W}_O \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_O)$ ). Στη συνέχεια, εισάγεται η κατάσταση του κελιού στη συνάρτηση  $\tanh$  (για να ωθήσουμε τις τιμές να είναι μεταξύ 1 και 1) και το αποτέλεσμα το πολλαπλασιάζουμε με την έξοδο της σιγμοειδούς πύλης, έτσι ώστε να προκύπτουν μόνο τα μέρη που αποφασίσαμε ( $\mathbf{h}_t = \mathbf{o}_t * \tanh(\mathbf{C}_t)$ ). Στο παράδειγμα του γλωσσικού μοντέλου, αφού το LSTM κελί μόλις είδε ένα υποκείμενο, μπορεί να θέλει να εξάγει πληροφορίες σχετικές με ένα ρήμα, σε περίπτωση που αυτό είναι το επόμενο. Για παράδειγμα, μπορεί να βγάζει αν το υποκείμενο είναι σε ενικό ή πληθυντικό, έτσι ώστε να γνωρίζουμε σε ποια μορφή θα πρέπει να συζευχθεί ένα ρήμα εάν αυτό ακολουθεί στη συνέχεια [72]. Στο Σχήμα 1.2.2 απεικονίζονται συγκεντρωμένες οι εξισώσεις που περιλαμβάνονται για την εκμάθηση μια κυψέλης ενός LSTM, όπως τις αναλύσαμε προηγουμένως.

$$\begin{aligned} i_t &= \sigma(\mathbf{x}_t U^i + \mathbf{h}_{t-1} W^i) \\ f_t &= \sigma(\mathbf{x}_t U^f + \mathbf{h}_{t-1} W^f) \\ o_t &= \sigma(\mathbf{x}_t U^o + \mathbf{h}_{t-1} W^o) \\ \tilde{C}_t &= \tanh(\mathbf{x}_t U^g + \mathbf{h}_{t-1} W^g) \\ C_t &= \sigma(f_t * C_{t-1} + i_t * \tilde{C}_t) \\ h_t &= \tanh(C_t) * o_t \end{aligned}$$

Σχήμα 1.2.2: Εξισώσεις μίας LSTM κυψέλης (LSTM Cell).

## 1.2.2 Συνελικτικά Νευρωνικά Δίκτυα - CNNs

Στον τομέα της Βαθιάς Μάθησης, το Συνελικτικό Νευρωνικό Δίκτυο (CNN) είναι ο πιο διάσημος και κοινώς χρησιμοποιούμενος αλγόριθμος. Το κύριο πλεονέκτημα του CNN σε σύγκριση με τους προκατόχους του είναι ότι αυτόματα προσδιορίζει τα σχετικά χαρακτηριστικά χωρίς ανθρώπινη επίβλεψη. Τα CNN εφαρμόζονται εκτενώς σε μια σειρά διαφορετικών πεδίων, συμπεριλαμβανομένης της όρασης υπολογιστών (Computer Vision), της επεξεργασίας ομιλίας (Speech Processing) και της αναγνώρισης προσώπων (Face Recognition) κλπ. Η δομή των CNN εμπνεύστηκε από νευρώνες στον εγκέφαλο ανθρώπων και ζώων, παρόμοια με ένα συμβατικό νευρωνικό δίκτυο [73].

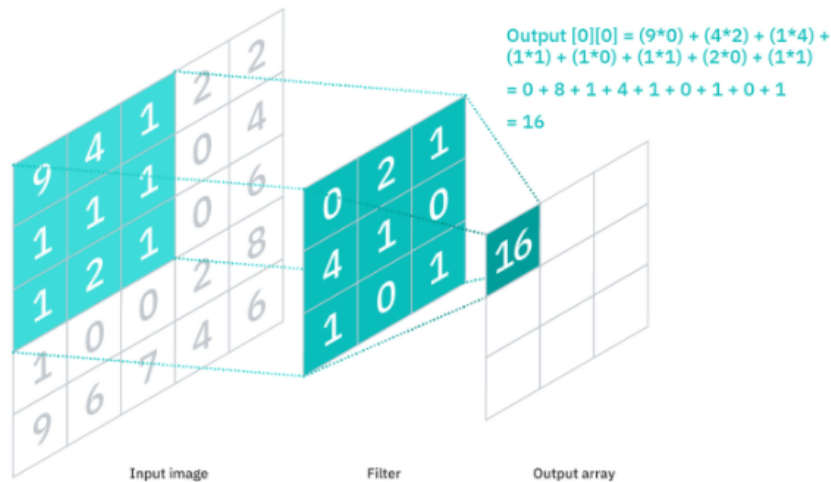
Τα συνελικτικά νευρωνικά δίκτυα διακρίνονται από άλλα νευρωνικά δίκτυα για την ανώτερη απόδοσή τους με εισόδους σημάτων εικόνας, ομιλίας ή ήχου. Έχουν τρεις κύριους τύπους επιπέδων, τα οποία είναι:

- **Επίπεδο Συνέλιξης (Convolutional Layer)**
- **Επίπεδο Pooling (Pooling Layer)**
- **Επίπεδο Πλήρως συνδεδεμένο (Fully Connected Layer)**

Το συνελικτικό επίπεδο είναι το πρώτο επίπεδο ενός συνελικτικού δικτύου. Ενώ τα συνελικτικά στρώματα μπορούν να ακολουθηθούν από πρόσθετα συνελικτικά στρώματα ή στρώματα συγκέντρωσης, το πλήρως συνδεδεμένο στρώμα είναι το τελικό στρώμα. Με κάθε επίπεδο, το CNN αυξάνει την πολυπλοκότητά του, εντοπίζοντας μεγαλύτερα τμήματα της εικόνας. Τα προηγούμενα επίπεδα επικεντρώνονται σε απλά χαρακτηριστικά, όπως χρώματα και άκρες. Καθώς τα δεδομένα εικόνας προχωρούν μέσω

των επιπέδων του CNN, αρχίζει να αναγνωρίζει μεγαλύτερα στοιχεία ή σχήματα του αντικειμένου μέχρι να αναγνωρίσει τελικά το αντικείμενο που προορίζεται.

**Επίπεδο Συνέλιξης:** Το συνελικτικό στρώμα είναι το βασικό δομικό στοιχείο ενός CNN, και εκεί συμβαίνει η πλειοψηφία των υπολογισμών. Απαιτεί μερικά στοιχεία, τα οποία είναι δεδομένα εισόδου, ένα φίλτρο και ένας χάρτης χαρακτηριστικών. Ας υποθέσουμε ότι η είσοδος θα είναι μια έγχρωμη εικόνα, η οποία αποτελείται από έναν πίνακα με pixels σε 3D. Αυτό σημαίνει ότι η **είσοδος θα έχει τρεις διαστάσεις: ύψος, πλάτος και βάθος**, που αντιστοιχούν στο RGB σε μια εικόνα. Έχουμε επίσης έναν ανιχνευτή χαρακτηριστικών, γνωστό και ως **πυρήνας ή φίλτρο (kernel/filter)**, ο οποίος θα "κινείται" στην εικόνα, ελέγχοντας αν υπάρχει το χαρακτηριστικό. Αυτή η διαδικασία είναι γνωστή ως συνέλιξη. Ο ανιχνευτής χαρακτηριστικών είναι μια διδιάστατη (2-D) διάταξη βαρών (weights), η οποία αντιπροσωπεύει μέρος της εικόνας. Ενώ μπορεί να διαφέρουν σε μέγεθος, το μέγεθος του φίλτρου είναι συνήθως ένας πίνακας  $3 * 3$ . Το μέγεθος του φίλτρου καθορίζεται συνήθως από τις διαστάσεις της εικόνας. Στη συνέχεια, το φίλτρο εφαρμόζεται σε μια περιοχή της εικόνας και υπολογίζεται ένα γινόμενο μεταξύ των εικονοστοιχείων εισόδου και του φίλτρου. Το αποτέλεσμα του γινομένου τροφοδοτείται στη συνέχεια σε έναν πίνακα εξόδου. Στη συνέχεια, το φίλτρο μετατοπίζεται κατά ένα βήμα (stride), επαναλαμβάνοντας τη διαδικασία μέχρι να σαρώσει ολόκληρη την εικόνα. Η τελική έξοδος από τη σειρά διαδοχικών γινομένων από την είσοδο και το φίλτρο είναι γνωστή ως χάρτης χαρακτηριστικών, χάρτης ενεργοποίησης ή σύνθετο χαρακτηριστικό (Feature Maps).



Σχήμα 1.2.3: Παράδειγμα πράξεων Συνέλιξης στο Επίπεδο Συνέλιξης.

Όπως παρατηρούμε στο Σχήμα 1.2.3, κάθε τιμή εξόδου στον χάρτη χαρακτηριστικών δεν χρειάζεται να συνδέεται με κάθε τιμή pixel της εικόνας εισόδου. Χρειάζεται μόνο να συνδεθεί στην εικόνα, όπου εφαρμόζεται το φίλτρο. Δεδομένου ότι ο πίνακας εξόδου δεν χρειάζεται να αντιστοιχιστεί απευθείας σε κάθε τιμή εισόδου, τα συνελικτικά επίπεδα αναφέρονται συνήθως ως επίπεδα «μερικώς συνδεδεμένα». Ωστόσο, αυτό το χαρακτηριστικό μπορεί επίσης να περιγραφεί ως τοπική συνδεσιμότητα [74].

Σημειώνουμε ότι τα βάρη στον ανιχνευτή χαρακτηριστικών παραμένουν σταθερά καθώς το φίλτρο κινείται στην εικόνα, κάτι που είναι επίσης γνωστό ως κοινή χρήση παραμέτρων (parameter sharing). Ορισμένες παράμετροι, όπως οι τιμές του βάρους, προσαρμόζονται κατά τη διάρκεια της εκπαίδευσης μέσω της διαδικασίας οπίσθιου πολλαπλασιασμού (backpropagation) [75] και κλίσης κατάβασης (gradient descent) [76]. Ωστόσο, υπάρχουν **τρεις υπερπαραμέτροι που επηρεάζουν το μέγεθος του όγκου της εξόδου** που πρέπει να ρυθμιστεί πριν ξεκινήσει η εκπαίδευση του νευρωνικού δικτύου. Αυτά περιλαμβάνουν:



- **Ο αριθμός των φίλτρων** επηρεάζει το βάθος της εξόδου. Για παράδειγμα, τρία διαφορετικά φίλτρα θα έδιναν τρεις διαφορετικούς χάρτες χαρακτηριστικών, δημιουργώντας ένα βάθος τριών.
- **Διασκελισμός (Stride)** είναι η απόσταση ή ο αριθμός των pixel που κινείται το φίλτρο πάνω από την εικόνα εισόδου. Ενώ οι τιμές διασκελισμού δύο ή μεγαλύτερες είναι σπάνιες, ένας μεγαλύτερος διασκελισμός αποδίδει μικρότερη απόδοση.
- Το **Zero-padding** χρησιμοποιείται συνήθως όταν τα φίλτρα δεν ταιριάζουν στην εικόνα εισόδου. Αυτό θέτει όλα τα στοιχεία που εμπίπτουν εκτός του πίνακα εισόδου στο μηδέν, παράγοντας μια μεγαλύτερη ή ίσου μεγέθους έξοδο. Υπάρχουν τρεις τύποι padding: **Έγκυρο padding (Valid Padding)**: Αυτό είναι επίσης γνωστό ως no padding. Σε αυτήν την περίπτωση, η τελευταία συνέλιξη απορρίπτεται εάν οι διαστάσεις δεν ευθυγραμμίζονται. **Ίδιο padding (Same Padding)**: Αυτή η "επέκταση" στο συνελκτικό δίκτυο διασφαλίζει ότι το επίπεδο εξόδου έχει το ίδιο μέγεθος με το στρώμα εισόδου. **Πλήρες Padding (Full Padding)**: Αυτός ο τύπος padding αυξάνει το μέγεθος της εξόδου προσθέτοντας μηδενικά στο όριο της εισόδου.

Μετά από τη διαδικασία της συνέλιξης, ένα CNN εφαρμόζει έναν μετασχηματισμό Rectified Linear Unit (ReLU) στον χάρτη χαρακτηριστικών, εισάγοντας μη γραμμικότητα στο μοντέλο.

Όπως αναφέραμε προηγουμένως, ένα άλλο επίπεδο συνέλιξης μπορεί να ακολουθήσει το αρχικό επίπεδο συνέλιξης. Όταν συμβεί αυτό, η δομή του CNN μπορεί να γίνει ιεραρχική καθώς τα μεταγενέστερα στρώματα μπορούν να δουν τα εικονοστοιχεία των προηγούμενων επιπέδων. Για παράδειγμα, ας υποθέσουμε ότι προσπαθούμε να προσδιορίσουμε εάν μια εικόνα περιέχει ένα ποδήλατο. Μπορούμε να σκεφτούμε το ποδήλατο ως ένα άθροισμα εξαρτημάτων. Αποτελείται από πλαίσιο, τιμόνι, τροχούς, πεντάλ κλπ. Κάθε μεμονωμένο μέρος του ποδηλάτου δημιουργεί ένα μοτίβο χαμηλότερου επιπέδου στο νευρωνικό δίκτυο και ο συνδυασμός των μερών του αντιπροσωπεύει ένα μοτίβο υψηλότερου επιπέδου, δημιουργώντας μια ιεραρχία χαρακτηριστικών εντός του CNN. Τελικά, το συνελκτικό στρώμα μετατρέπει την εικόνα σε αριθμητικές τιμές, επιτρέποντας στο νευρωνικό δίκτυο να ερμηνεύει και να εξάγει σχετικά μοτίβα.

**Επίπεδο Pooling:** Η διαδικασία pooling, γνωστή και ως υποδειγματοληψία (downsampling), πραγματοποιεί μείωση διαστάσεων, μειώνοντας τον αριθμό των παραμέτρων στην είσοδο. Παρόμοια με το συνελκτικό στρώμα, η διαδικασία padding σαρώνει ένα φίλτρο σε ολόκληρη την είσοδο, αλλά η διαφορά είναι ότι αυτό το φίλτρο δεν έχει βάρη. Αντίθετα, ο πυρήνας εφαρμόζει μια συνάρτηση συνάθροισης (aggregation function) στις τιμές της εικόνας, συμπληρώνοντας τον πίνακα εξόδου. Υπάρχουν δύο κύριοι τύποι pooling:

- **Μέγιστο Pooling (Max Pooling)**: Καθώς το φίλτρο κινείται κατά μήκος της εισόδου, επιλέγει το εικονοστοιχείο με τη μέγιστη τιμή για να στείλει στον πίνακα εξόδου. Επιπλέον, αυτή η προσέγγιση τείνει να χρησιμοποιείται πιο συχνά σε σύγκριση με τη μέση συγκέντρωση.
- **Μέσο Pooling (Average Pooling)**: Καθώς το φίλτρο μετακινείται κατά μήκος της εισόδου, υπολογίζει τη μέση τιμή για να στείλει στον πίνακα εξόδου.

Ενώ πολλές πληροφορίες χάνονται στο επίπεδο pooling, έχει επίσης μια σειρά από οφέλη για το CNN. Βοηθούν στη μείωση της πολυπλοκότητας, στη βελτίωση της αποτελεσματικότητας και στον περιορισμό του κινδύνου υπερβολικής προσαρμογής.

**Πλήρες Συνδεδεμένο Επίπεδο:** Το όνομα του πλήρως συνδεδεμένου επιπέδου περιγράφει εύστοχα τον εαυτό του. Όπως αναφέρθηκε προηγουμένως, οι τιμές των pixel της εικόνας εισόδου δεν συνδέονται απευθείας με το επίπεδο εξόδου σε μερικώς συνδεδεμένα στρώματα. Ωστόσο, στο πλήρως συνδεδεμένο επίπεδο, κάθε κόμβος στο επίπεδο εξόδου συνδέεται απευθείας με έναν κόμβο στο προηγούμενο επίπεδο. Αυτό το επίπεδο εκτελεί το έργο της ταξινόμησης με βάση τα χαρακτηριστικά που

εξάγονται μέσω των προηγούμενων επιπέδων και των διαφορετικών φίλτρων τους. Ενώ τα επίπεδα συνέλιξης και pooling τείνουν να χρησιμοποιούν συναρτήσεις ReLu, τα πλήρως συνδεδεμένα επίπεδα (Fully connected layers (FC)) συνήθως αξιοποιούν μια συνάρτηση ενεργοποίησης *softmax* για την κατάλληλη ταξινόμηση των εισόδων, παράγοντας μια πιθανότητα από 0 έως 1.

**Τύποι Συνελικτικών Νευρωνικών Δικτύων:** Οι Kunihiko Fukushima και Yann LeCun έθεσαν τα θεμέλια της έρευνας γύρω από τα συνελικτικά νευρωνικά δίκτυα στην εργασία τους το 1980 [77] και το 1989 [78], αντίστοιχα. Σε μια πολύ διάσημη έρευνα, ο Yann LeCun εφάρμοσε επιτυχώς backpropagation για να εκπαιδεύσει τα νευρωνικά δίκτυα να αναγνωρίζουν μοτίβα μέσα σε μια σειρά χειρόγραφων ταχυδρομικών κωδίκων. Συνέχισε την έρευνά του με την ομάδα του κατά τη διάρκεια της δεκαετίας του 1990, με αποκορύφωμα το "LeNet-5" [79], το οποίο εφάρμοζε τις ίδιες αρχές της προηγούμενης έρευνας για την αναγνώριση εγγράφων. Έκτοτε, ένας αριθμός διαφορετικών αρχιτεκτονικών CNN εμφανίστηκε με την εισαγωγή νέων συνόλων δεδομένων, όπως το MNIST και το CIFAR-10, και διαγωνισμούς, όπως το ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Μερικές από αυτές τις άλλες αρχιτεκτονικές είναι τα δίκτυα: AlexNet [80], VGGNet [81], GoogLeNet [82], ResNet [83], ZFNet. Ωστόσο, το LeNet-5 είναι γνωστό ως η κλασική αρχιτεκτονική του CNN.

### 1.2.3 Μοντέλο Μετασχηματιστής - Transformer Model

Ένας Μετασχηματιστής είναι ένα μοντέλο Βαθιάς Μάθησης που υιοθετεί τον μηχανισμό της "αυτοπροσοχής" (self-attention), σταθμίζοντας διαφορετικά τη "σημαντικότητα" (importance) κάθε μέρους των δεδομένων εισόδου. Χρησιμοποιείται κυρίως στους τομείς της επεξεργασίας φυσικής γλώσσας (NLP) και της όρασης υπολογιστή (CV). Όπως τα νευρωνικά δίκτυα RNN, οι μετασχηματιστές έχουν σχεδιαστεί για να χειρίζονται διαδοχικά δεδομένα εισόδου, όπως η φυσική γλώσσα, για εργασίες όπως η μετάφραση και η σύνοψη κειμένου. Ωστόσο, σε αντίθεση με τα RNN, οι μετασχηματιστές δεν επεξεργάζονται απαραίτητα τα δεδομένα με τη σειρά. Αντίθετα, ο μηχανισμός Προσοχής (Attention Mechanism) παρέχει το πλαίσιο για οποιαδήποτε θέση στην ακολουθία εισόδου. Για παράδειγμα, εάν τα δεδομένα εισόδου είναι μια πρόταση φυσικής γλώσσας, ο μετασχηματιστής δεν χρειάζεται να επεξεργαστεί την αρχή της πρότασης πριν από το τέλος αλλά προσδιορίζει το πλαίσιο που προσδίδει νόημα σε κάθε λέξη της πρότασης. Αυτή η δυνατότητα επιτρέπει μεγαλύτερη "παράλληλοποίηση" από τα RNN και επομένως μειώνει τους χρόνους εκπαίδευσης [84].

Πριν από τους μετασχηματιστές, τα περισσότερα συστήματα NLP (Natural Language Processing) τελευταίας τεχνολογίας βασίζονταν σε κλειστά RNN, όπως το LSTM και οι πυλωτές επαναλαμβανόμενες μονάδες (Gated Recurrent Units GRU) [85], με πρόσθετους μηχανισμούς Προσοχής. Οι μετασχηματιστές είναι βασισμένοι σε αυτές τις τεχνολογίες Προσοχής χωρίς τη χρήση δομής RNN, υπογραμμίζοντας το γεγονός ότι μόνο οι μηχανισμοί Προσοχής μπορούν να ταιριάζουν την απόδοση των RNN με Προσοχή.

**Διαδοχική επεξεργασία (Sequential Processing):** Τα Gated RNN επεξεργάζονται tokens διαδοχικά, διατηρώντας ένα διάλυμα κατάστασης που περιέχει μια αναπαράσταση των δεδομένων που εμφανίζονται μετά από κάθε token. Για την επεξεργασία του n-οστού token, το μοντέλο συνδυάζει την κατάσταση που αντιπροσωπεύει την ακολουθία (sequence) μέχρι το token n-1 με τις πληροφορίες του νέου διακριτικού για τη δημιουργία μιας νέας κατάστασης, που αντιπροσωπεύει την ακολουθία μέχρι το διακριτικό n. Θεωρητικά, οι πληροφορίες από ένα token μπορούν να διαδοθούν αυθαίρετα πολύ κάτω από την ακολουθία, εάν σε κάθε σημείο η κατάσταση συνεχίζει να κωδικοποιεί πληροφορίες "περιεχομένου" (contextual information) για το token. Στην πράξη, αυτός ο μηχανισμός είναι ελαττωματικός: το πρόβλημα του "εξαφανισμού" των παραγώγων (vanishing gradients), αφήνει την κατάσταση του μοντέλου στο τέλος μιας μεγάλης ακολουθίας χωρίς ακριβείς, εξαγωγικές πληροφορίες σχετικά με τα προηγούμενα token.

**Εστίαση Προσοχής (Attention):** Αυτό το πρόβλημα αντιμετωπίστηκε με μηχανισμούς εστίασης προσοχής. Οι μηχανισμοί εστίασης προσοχής επιτρέπουν σε ένα μοντέλο να αντλεί πληροφορίες από την κατάσταση σε οποιοδήποτε προηγούμενο σημείο κατά μήκος της ακολουθίας. Το επίπεδο προσοχής (attention layer) μπορεί να έχει πρόσβαση σε όλες τις προηγούμενες καταστάσεις και να τις σταθμίζει σύμφωνα με ένα μέτρο συνάφειας (measure of relevancy), παρέχοντας σχετικές πληροφορίες σχετικά με τα μακρινά token.

Ένα σαφές παράδειγμα της τιμής της προσοχής (attention value) είναι στη μετάφραση της γλώσσας, όπου το πλαίσιο είναι απαραίτητο για την εκχώρηση της σημασίας μιας λέξης σε μια πρόταση. Σε ένα σύστημα μετάφρασης από Αγγλικά προς Γαλλικά, η πρώτη λέξη της γαλλικής εξόδου εξαρτάται κατά πάσα πιθανότητα σε μεγάλο βαθμό από τις πρώτες λίγες λέξεις της αγγλικής εισόδου. Ωστόσο, σε ένα κλασικό μοντέλο LSTM, για να παραχθεί η πρώτη λέξη της γαλλικής εξόδου, δίνεται στο μοντέλο μόνο το διάνυσμα κατάστασης της τελευταίας αγγλικής λέξης. Θεωρητικά, αυτό το διάνυσμα μπορεί να κωδικοποιήσει πληροφορίες για ολόκληρη την αγγλική πρόταση, δίνοντας στο μοντέλο όλες τις απαραίτητες γνώσεις. Στην πράξη αυτές οι πληροφορίες συχνά διατηρούνται ελάχιστα από το LSTM. Ένας μηχανισμός εστίασης προσοχής μπορεί να προστεθεί για την αντιμετώπιση αυτού του προβλήματος: ο αποκωδικοποιητής (Decoder) έχει πρόσβαση στα διανύσματα κατάστασης κάθε αγγλικής λέξης εισόδου, όχι μόνο της τελευταίας, και μπορεί να μάθει βάρη προσοχής (attention weights) που υπαγορεύουν πόσο να παρακολουθεί κάθε διάνυσμα κατάστασης εισαγωγής στα αγγλικά. Όταν προστίθενται σε RNN, οι μηχανισμοί εστίασης προσοχής αυξάνουν την απόδοση. **Η ανάπτυξη της αρχιτεκτονικής του Transformer αποκάλυψε ότι οι μηχανισμοί προσοχής ήταν από μόνοι τους ισχυροί και ότι η διαδοχική επαναλαμβανόμενη επεξεργασία δεδομένων δεν ήταν απαραίτητη για να επιτευχθούν τα κέρδη απόδοσης των RNN με εστίαση προσοχής. Οι μετασχηματιστές χρησιμοποιούν έναν μηχανισμό εστίασης προσοχής χωρίς RNN, επεξεργάζοντας όλα τα token ταυτόχρονα και υπολογίζοντας τα βάρη προσοχής μεταξύ τους σε διαδοχικά επίπεδα.**

**Αρχιτεκτονική δικτύου:** Όπως τα προηγούμενα μοντέλα, ο Μετασχηματιστής υιοθετεί μια αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή (Encoder-Decoder). Ο κωδικοποιητής αποτελείται από επίπεδα κωδικοποίησης που επεξεργάζονται την είσοδο επαναληπτικά το ένα στρώμα μετά το άλλο, ενώ ο αποκωδικοποιητής αποτελείται από στρώματα αποκωδικοποίησης που κάνουν το ίδιο πράγμα στην έξοδο του κωδικοποιητή. Η λειτουργία κάθε επιπέδου κωδικοποιητή είναι να δημιουργεί κωδικοποιήσεις που περιέχουν πληροφορίες σχετικά με τα μέρη των εισόδων που σχετίζονται μεταξύ τους. Πέραν τις κωδικοποιήσεις του στο επόμενο επίπεδο κωδικοποιητή ως είσοδο. Κάθε επίπεδο αποκωδικοποιητή κάνει το αντίθετο, λαμβάνοντας όλες τις κωδικοποιήσεις και χρησιμοποιώντας τις ενσωματωμένες πληροφορίες "περιεχομένου" για να δημιουργήσει μια ακολουθία εξόδου. Για να επιτευχθεί αυτό, κάθε επίπεδο κωδικοποιητή και αποκωδικοποιητή χρησιμοποιεί έναν μηχανισμό εστίασης προσοχής. Για κάθε είσοδο, η εστίαση προσοχής σταθμίζει τη συνάφεια κάθε άλλης εισόδου και αντλεί από αυτές για να παράγει την έξοδο. Κάθε επίπεδο αποκωδικοποιητή έχει έναν πρόσθετο μηχανισμό εστίασης προσοχής που αντλεί πληροφορίες από τις εξόδους των προηγούμενων αποκωδικοποιητών, πριν το επίπεδο του αποκωδικοποιητή αντλήσει πληροφορίες από τις κωδικοποιήσεις. Τόσο το επίπεδο κωδικοποιητή όσο και του αποκωδικοποιητή έχουν: ένα νευρωνικό δίκτυο για πρόσθετη επεξεργασία των εξόδων, υπολειπόμενες συνδέσεις (residual connections) και βήματα κανονικοποίησης στρώματος (layer normalization steps).

**Εστίαση προσοχής διαβαθμισμένου εσωτερικού γινομένου (Scaled dot-product attention):**

Τα δομικά στοιχεία του Μετασχηματιστή είναι μονάδες προσοχής διαβαθμισμένου εσωτερικού γινομένου. Όταν μια πρόταση μεταβιβάζεται σε ένα μοντέλο Μετασχηματιστή, τα βάρη εστίασης προσοχής υπολογίζονται μεταξύ κάθε token ταυτόχρονα. Η μονάδα προσοχής (attention unit), παράγει em-



beddings για κάθε token στο πλαίσιο που περιέχουν πληροφορίες για το ίδιο το token μαζί με έναν σταθμισμένο συνδυασμό άλλων σχετικών token, το καθένα σταθμισμένο με το βάρος προσοχής του. Για κάθε μονάδα προσοχής το μοντέλο Μετασχηματιστή μαθαίνει τρεις πίνακες βάρους. Τα βάρη query  $\mathbf{W}_Q$ , τα keys βάρη  $\mathbf{W}_K$  και τα βάρη value  $\mathbf{W}_V$ . Για κάθε token  $i$ , η λέξη εισόδου  $\mathbf{x}_i$  πολλαπλασιάζεται με κάθε έναν από τους τρεις πίνακες βάρους για να παραχθεί ένα διάνυσμα query  $\mathbf{q}_i = \mathbf{x}_i \mathbf{W}_Q$ , ένα διάνυσμα key  $\mathbf{k}_i = \mathbf{x}_i \mathbf{W}_K$  και ένα διάνυσμα value  $\mathbf{v}_i = \mathbf{x}_i \mathbf{W}_V$ . Τα βάρη προσοχής υπολογίζονται χρησιμοποιώντας το query και τα διανύσματα keys: το βάρος προσοχής  $\mathbf{a}_{i,j}$  από το token  $i$  στο token  $j$  είναι εσωτερικό γινόμενο μεταξύ του  $\mathbf{q}_i$  και  $\mathbf{k}_j$ . Τα βάρη προσοχής διαιρούνται με την τετραγωνική ρίζα της διάστασης των key διανυσμάτων,  $\sqrt{d_k}$ , γεγονός που σταθεροποιεί τις παραγώγους (gradients) κατά την εκπαίδευση του μοντέλου, και ομαλοποιεί τα βάρη μέσα από την συνάρτηση softmax. Το γεγονός ότι οι  $\mathbf{W}_Q$  και  $\mathbf{W}_K$  είναι διαφορετικοί πίνακες, επιτρέπει την προσοχή να είναι μη συμμετρική: εάν το token  $i$  εστιάζει στο token  $j$  (δηλαδή το  $\mathbf{q}_i * \mathbf{k}_j$  είναι μεγάλο), αυτό δεν σημαίνει απαραίτητα ότι το token  $j$  θα εστιάζει στο token  $i$  (δηλ.  $\mathbf{q}_j * \mathbf{k}_i$  μπορεί να είναι μικρό). Η έξοδος της μονάδας προσοχής για το token  $i$  είναι το σταθμισμένο άθροισμα των διανυσμάτων value όλων των token, σταθμισμένο με  $\mathbf{a}_{i,j}$ , την προσοχή από το token  $i$  σε κάθε token. Ο υπολογισμός της εστίασης προσοχής για όλα τα token μπορεί να εκφραστεί ως ένας υπολογισμός μεγάλου πίνακα χρησιμοποιώντας τη συνάρτηση softmax, η οποία είναι χρήσιμη για εκπαίδευση λόγω βελτιστοποιήσεων λειτουργίας υπολογιστικών πινάκων που υπολογίζουν γρήγορα πράξεις πινάκων. Οι πίνακες  $\mathbf{Q}$ ,  $\mathbf{K}$  και  $\mathbf{V}$  ορίζονται ως οι πίνακες όπου οι  $i$ -οστές σειρές είναι διανύσματα  $\mathbf{q}_i$ ,  $\mathbf{k}_i$  και  $\mathbf{v}_i$  αντίστοιχα.

Η εξίσωση εστίασης προσοχής είναι η ακόλουθη:  $\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}$ .

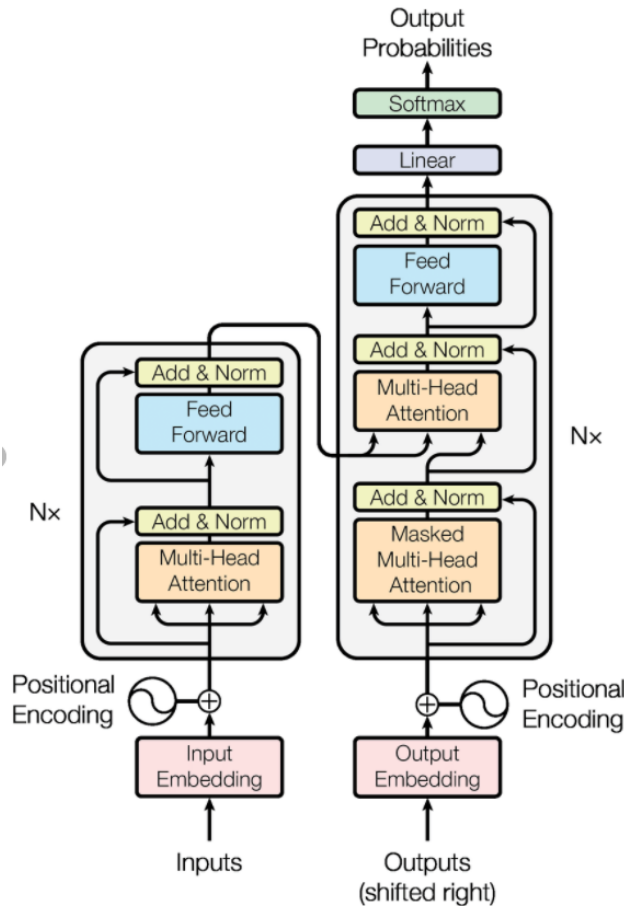
### Πολυμέτωση εστίαση προσοχής (Multi-head attention):

Ένα σύνολο πινάκων ( $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V$ ) ονομάζεται κεφαλή προσοχής (attention head) και κάθε στρώμα σε ένα μοντέλο Μετασχηματιστή έχει πολλαπλές κεφαλές προσοχής. Ενώ κάθε κεφαλή προσοχής παρακολουθεί τα token που σχετίζονται με κάθε token, με πολλαπλές κεφαλές προσοχής το μοντέλο μπορεί να το κάνει αυτό για διαφορετικούς ορισμούς της "συνάφειας" (relevance). Επιπλέον, το πεδίο επιρροής που αντιπροσωπεύει τη "συνάφεια" μπορεί να διευρυνθεί προοδευτικά σε διαδοχικά στρώματα. Πολλές κεφαλές προσοχής μετασχηματιστών κωδικοποιούν σχέσεις συνάφειας που έχουν νόημα για τον άνθρωπο. Για παράδειγμα, οι κεφαλές προσοχής μπορούν να παρακολουθούν κυρίως την επόμενη λέξη, ενώ άλλοι παρακολουθούν κυρίως από τα ρήματα στα άμεσα αντικείμενά τους. Οι υπολογισμοί για κάθε κεφαλή προσοχής μπορούν να εκτελεστούν παράλληλα, γεγονός που επιτρέπει τη γρήγορη επεξεργασία. Οι έξοδοι για το επίπεδο προσοχής, συνενώνονται για να περάσουν στα στρώματα του νευρωνικού δικτύου.

**Κωδικοποιητής (Encoder):** Κάθε κωδικοποιητής αποτελείται από δύο κύρια στοιχεία: έναν μηχανισμό αυτοπροσοχής (self-attention) και ένα νευρωνικό δίκτυο. Ο μηχανισμός αυτοπροσοχής δέχεται κωδικοποιήσεις εισόδου από τον προηγούμενο κωδικοποιητή και σταθμίζει τη συνάφειά τους μεταξύ τους για να δημιουργήσει κωδικοποιήσεις εξόδου. Το νευρωνικό δίκτυο επεξεργάζεται περαιτέρω κάθε κωδικοποίηση εξόδου ξεχωριστά. Η κωδικοποιημένη έξοδος περνά στη συνέχεια στον επόμενο κωδικοποιητή ως είσοδος, καθώς και στους αποκωδικοποιητές. Ο πρώτος κωδικοποιητής δέχεται ως είσοδο πληροφορίες "θέσης και ενσωμάτωσης" (embeddings) των ακολουθιών εισόδου, αντί για κωδικοποιήσεις. Οι πληροφορίες θέσης (positional information) είναι απαραίτητες ώστε ο μετασχηματιστής να κάνει χρήση της σειράς της ακολουθίας, επειδή κανένα άλλο μέρος του μετασχηματιστή δεν το χρησιμοποιεί.

**Αποκωδικοποιητής (Decoder):** Κάθε αποκωδικοποιητής αποτελείται από τρία κύρια στοιχεία: έναν μηχανισμό αυτοπροσοχής, έναν μηχανισμό προσοχής μετά τις κωδικοποιήσεις και ένα νευρωνικό δίκτυο. Ο αποκωδικοποιητής λειτουργεί με παρόμοιο τρόπο με τον κωδικοποιητή, αλλά εισάγεται ένας πρόσθετος μηχανισμός προσοχής, ο οποίος αντλεί σχετικές πληροφορίες από τις κωδικοποιήσεις που

δημιουργούνται από τους κωδικοποιητές. Όπως ο πρώτος κωδικοποιητής, ο πρώτος αποκωδικοποιητής λαμβάνει πληροφορίες "θέσης και ενσωμάτωσης" (embeddings) της ακολουθίας εξόδου ως είσοδο, αντί για κωδικοποιήσεις. Ο μετασχηματιστής δεν πρέπει να χρησιμοποιεί την τρέχουσα ή τη μελλοντική έξοδο για να προβλέψει μια έξοδο, επομένως η ακολουθία εξόδου πρέπει να καλύπτεται μερικώς για να αποτραπεί αυτή η αντίστροφη ροή πληροφοριών. Ο τελευταίος αποκωδικοποιητής ακολουθείται από έναν τελικό γραμμικό μετασχηματισμό και στρώμα softmax, για να παραχθούν οι πιθανότητες εξόδου της ακολουθίας.



Σχήμα 1.2.4: Αρχιτεκτονική του μοντέλου Μετασχηματιστή (Transformer model).

**Εναλλακτικές-Παραλλαγές:** Η εκπαίδευση αρχιτεκτονικών που βασίζονται σε μετασχηματιστές μπορεί να είναι υπολογιστικά ακριβή, ειδικά για μεγάλες ακολουθίες. Οι εναλλακτικές αρχιτεκτονικές περιλαμβάνουν το Reformer (το οποίο μειώνει τον υπολογιστικό φορτίο από  $O(N^2)$  σε  $O(N \ln N)$ ), ή μοντέλα όπως το ETC/BigBird (που μπορεί να το μειώσει σε  $O(N)$  όπου  $N$  είναι το μήκος της ακολουθίας. Αυτό γίνεται με χρήση locality-sensitive hashing και αναστρέψιμων επιπέδων (reversible layers). Ένα σημείο αναφοράς για τη σύγκριση αρχιτεκτονικών μετασχηματιστών εισήχθη στα τέλη του 2020.

### 1.3 Στόχοι και Οργάνωση Εργασίας

Η παρούσα μεταπτυχιακή εργασία έχει ως στόχο: 1) την διερεύνηση των δεδομένων ύπνου που καταγράφονται από φορητές συσκευές, ευρέως γνωστές ως smartwatches και 2) την ανάλυση τους

για μεγαλύτερη εμβάθυνση και κατανόηση μιας πολύ αναγκαίας, πολύπλοκης και καθημερινής ανθρώπινης δραστηριότητας όπως είναι ο ύπνος. Όμως, ο **βασικός μας στόχος** είναι η ταξινόμηση των δεδομένων φορητών συσκευών στα στάδια του ύπνου (sleep stages), με σκοπό την εξέλιξη του κλάδου της "ανάλυσης δεδομένων της ανθρώπινης δραστηριότητας" με τη χρήση μοντέλων Βαθιάς Μηχανικής Μάθησης που αποτελεί κλάδο της Τεχνητής Νοημοσύνης. Επίσης, η γρήγορα αναπτυσσόμενη τεχνολογία των φορητών συσκευών αποτελεί ευκαιρία για τους ερευνητές να εξετάσουν αυτόν τον τομέα. Λεπτομερέστερα, η παρούσα μεταπτυχιακή εργασία αποσκοπεί στην ταξινόμηση των δεδομένων ύπνου μιας πολύ διαδεδομένης βάσης δεδομένων (MESA database), με τη χρήση διαφορετικών μοντέλων Βαθιάς Μάθησης και τεχνικών κατάλληλης επεξεργασίας δεδομένων για την επίτευξη όσο το δυνατόν καλύτερων αποτελεσμάτων ταξινόμησης. Η κεντρική ιδέα της εργασίας μας είναι ότι χρησιμοποιείται ως βάση η σύγχρονη έρευνα [1] ώστε να βελτιώσουμε τα αποτελέσματα της ταξινόμησης στα 4 βασικά στάδια ύπνου με τη χρήση διάφορων τεχνικών είτε από πλευράς μοντέλων είτε από πλευράς τεχνικών προεπεξεργασίας δεδομένων και να προτείνουμε κατευθύνσεις για περαιτέρω εμβάθυνση. Δηλαδή, πειραματιζόμαστε α) με διαφορετικά μοντέλα Βαθιάς Μάθησης όπως Συνελικτικά Νευρωνικά Δίκτυα (CNN), Νευρωνικά Δίκτυα Μακροπρόθεσμης Μνήμης (LSTM), Νευρωνικά Δίκτυα Μακροπρόθεσμης Μνήμης με μηχανισμό Προσοχής (LSTM + Attention), Μετασχηματιστές (Transformers) και β) με τεχνικές προεπεξεργασίας δεδομένων όπως η τεχνική αύξησης συνόλου δεδομένων με τεχνητά δεδομένα (Data Augmentation), βελτιώνοντας τα ήδη υπάρχοντα αποτελέσματα της έρευνας [1]. Οπότε, με βάση όλα τα παραπάνω, μέσω της εργασίας μας βάζουμε ένα λιθαράκι στον τομέα της ανάλυσης ιατρικών δεδομένων.

Επιπλέον, η μεταπτυχιακή εργασία χωρίζεται σε 4 κεφάλαια και ακολουθεί μία επιγραμματική παρουσίαση των περιεχομένων κάθε κεφαλαίου:

- Στο κεφάλαιο 1 παραθέτουμε μια εισαγωγή για τον ύπνο από ιατρικής πλευράς, τα στάδια που αποτελείται δίνοντας μια γενική ιδέα στον αναγνώστη ώστε να κατανοήσει αρχικά στο θέμα. Στη συνέχεια αναφέρουμε τις υπάρχουσες τεχνολογίες καταγραφής δεδομένων της ανθρώπινης δραστηριότητας και την αναγκαιότητα των φορητών συσκευών ως μέσο εξοικονόμησης και σύγχρονης καταγραφής δεδομένων. Τέλος, αναφέρουμε το θεωρητικό υπόβαθρο των μοντέλων Βαθιάς Μάθησης και γενικών σημαντικών τεχνικών επεξεργασίας δεδομένων που χρησιμοποιούνται στη παρούσα εργασία, ώστε ο αναγνώστης να κατανοήσει τα μοντέλα και τις τεχνικές που διεξάγουμε τα πειράματα ταξινόμησης.
- Στο κεφάλαιο 2 αρχίζουμε με το πρώτο στάδιο των πειραματισμών μας, όπου προσπαθούμε να αναπαράγουμε τα αποτελέσματα της έρευνας [1] η οποία χρησιμοποιεί εξαγόμενα χαρακτηριστικά από τα δεδομένα της βάσης MESA. Σχολιάζοντας τα αποτελέσματα και τα πιθανά ελαττώματα που προκύπτουν, σε επόμενο στάδιο προσπαθήσουμε να βελτιώσουμε τα αποτελέσματα βάσει κάποιων δημοφιλών μετρικών ταξινόμησης αλλά και γενικά με διαφορετικούς τρόπους αξιολόγησης των αποτελεσμάτων. Οπότε ο στόχος μας, σε αυτό το πρώτο μέρος είναι η αναγνώριση των πιθανών αδυναμιών της προαναφερθείσας έρευνας.
- Στο δεύτερο μέρος του κεφαλαίου 2, εφόσον έχουμε παρατηρήσει τις αδυναμίες της προαναφερθείσας έρευνας, προσπαθούμε να αντιμετωπίσουμε τα προβλήματα που αναδύθηκαν, χρησιμοποιώντας τα ίδια μοντέλα αλλά με διαφορετικές υπερπαραμέτρους και διαφορετική αρχιτεκτονική. Επίσης, δοκιμάζουμε άλλα μοντέλα Βαθιάς Μάθησης (πχ. Μοντέλο Μετασχηματιστή), και κάποιες διαφορετικές τεχνικές επεξεργασίας δεδομένων προς μοντελοποίηση όπως για παράδειγμα: Data Augmentation και πειραματισμός με διαφορετικό μήκος παραθύρωσης των δεδομένων. Σχολιάζουμε τα αποτελέσματα που προκύπτουν και προτείνουμε διάφορες ερευνητικές κατευθύνσεις για μεγαλύτερη εμβάθυνση και ακόμη καλύτερα αποτελέσματα.
- Στο Κεφάλαιο 3 ακολουθούμε μια διαφορετική προσέγγιση της ταξινόμησης των δεδομένων ύπνου της βάσης MESA. Χρησιμοποιούμε μόνο τα δεδομένα καρδιακών παλμών της βάσης (RR-

Intervals) τα οποία εν συνεχεία μοντελοποιούνται χωρίς εξαγωγή χαρακτηριστικών, δηλαδή ακατέργαστα (raw). Οπότε, πειραματιζόμαστε με διαφορετικό είδος δεδομένων που δέχονται τα μοντέλα Βαθιάς Μάθησης με στόχο την επίτευξη όσο το δυνατόν καλύτερων αποτελεσμάτων. Σημαντικό να αναφερθεί είναι ότι η προσέγγιση που ακολουθήσαμε στο Κεφάλαιο 3 είναι η ταξινόμηση των δεδομένων ύπνου με τη χρήση μόνο των καρδιακών παλμών επειδή οι συνηθισμένες οικονομικές φορητές συσκευές είναι συνήθως εξοπλισμένες μόνο με καταγραφή καρδιακών παλμών. Άρα, η επίτευξη πολύ αξιόλογων αποτελεσμάτων μόνο με τη χρήση καρδιακών παλμών είναι πολύ σημαντική για τους παραπάνω λόγους.

- Τέλος, συνοψίζουμε τα αποτελέσματα μας, κρατώντας τη χρήσιμη πληροφορία από κάθε πείραμα και συλλέγουμε όλες τις πιθανές προτάσεις για την περαιτέρω εμβάθυνση στο θέμα και καλύτερα αποτελέσματα.

## Κεφάλαιο 2

# Μέθοδος Ταξινόμησης Σταδίων Ύπνου με χρήση δεδομένων Ακτιγραφίας και HRV χαρακτηριστικών

---

<b>2.1</b>	<b>Βιβλιογραφική Έρευνα</b> . . . . .	<b>30</b>
<b>2.2</b>	<b>Σύνολο δεδομένων MESA - Προεπεξεργασία</b> . . . . .	<b>31</b>
2.2.1	Εισαγωγή στα δεδομένα . . . . .	31
2.2.2	Καθαρισμός δεδομένων - Data Cleaning . . . . .	32
2.2.3	Εξαγωγή χαρακτηριστικών (Feature Extraction) . . . . .	35
2.2.4	Ανάλυση χαρακτηριστικών Activity Counts και HRV (Explanatory Data Analysis) . . . . .	39
<b>2.3</b>	<b>Ταξινόμηση δεδομένων με αλγόριθμους Βαθιάς Μάθησης (Deep Learning)</b> . . . . .	<b>45</b>
2.3.1	Προετοιμασία Δεδομένων . . . . .	48
2.3.2	Ταξινόμηση με Συνελικτικό Νευρωνικό Δίκτυο (CNN) . . . . .	49
2.3.3	Ταξινόμηση με Νευρωνικό Δίκτυο Μακροπρόθεσμης Μνήμης (LSTM) . . . . .	55
2.3.4	Ταξινόμηση με τη Μέθοδο Ensemble (Ensemble Method) . . . . .	60
2.3.5	Σύνοψη Αποτελεσμάτων . . . . .	62
<b>2.4</b>	<b>Βελτίωση Αλγόριθμων Βαθιάς Μάθησης (Deep Learning)</b> . . . . .	<b>64</b>
2.4.1	Περιγραφή διαδικασίας - Προεπεξεργασία Δεδομένων . . . . .	64
2.4.2	Ταξινόμηση με Συνελικτικό Νευρωνικό Δίκτυο (CNN) . . . . .	69
2.4.3	Ταξινόμηση με Νευρωνικό Δίκτυο Μακροπρόθεσμης Μνήμης Δύο Κατευθύνσεων (Bidirectional LSTM) . . . . .	74
2.4.4	Ταξινόμηση με διαφορετικό μήκος παραθύρωσης (Sliding Window) . . . . .	80
2.4.5	Ταξινόμηση με τη Μέθοδο Ensemble (Ensemble Method) . . . . .	84
2.4.6	Ταξινόμηση με Μοντέλο Μετασχηματιστή (Transformer Model) . . . . .	85
<b>2.5</b>	<b>Σύνοψη Αποτελεσμάτων - Συμπεράσματα</b> . . . . .	<b>91</b>

---

## 2.1 Βιβλιογραφική Έρευνα

Τα τελευταία χρόνια η τεχνολογία των φορετών συσκευών (wearables) έχει κυριαρχήσει στον τομέα ανάλυσης δεδομένων ύπνου εξαιτίας της εύκολης καταγραφής των δεδομένων και του χαμηλού κόστους των συσκευών. Ο συνδυασμός των δύο αιτιών οδηγεί στην έκρηξη ερευνητικών δημοσιεύσεων στον τομέα ταξινόμησης των σταδίων του ύπνου αλλά και γενικότερα στον τομέα αναγνώρισης της ανθρώπινης δραστηριότητας. Στο παρόν κεφάλαιο εστιάζουμε την έρευνα μας στην ταξινόμηση του ύπνου σε 4 στάδια (Wake/Light/Deep/REM) με Deep Learning αλγορίθμους. Εστιάζουμε σε αυτόν τον αριθμό σταδίων διότι χρησιμοποιείται ευρέως από εφαρμογές των wearables όπως στο Samsung Smart Watch App. Συγκεκριμένα, υπάρχουν πολλές δημοσιεύσεις που εστιάζουν στην ταξινόμηση 4 σταδίων ύπνου.

Αρχικά, στο [86] οι ερευνητές χρησιμοποίησαν δεδομένα κίνησης (3-axis accelerometer data) και δεδομένα φωτοπληθυσμογραφίας (PPG) από 50 υγιείς χρήστες και τροφοδοτούν τα Machine Learning μοντέλα: Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) και Random Forests. Εφαρμόζουν μια ειδική κανονικοποίηση του Z-Score και η εκθετική εξομάλυνση ώστε να μειωθεί η μεταβλητότητα εντός και μεταξύ των χρηστών. Επίσης, από τα δεδομένα, εξάγουν χαρακτηριστικά HRV, στατιστικά κίνησης και αναπνευστικά (movement and respiratory features). Το καλύτερο μοντέλο (LDA) πετυχαίνει στο σύνολο δεδομένων αξιολόγησης **accuracy 77 %** και **Cohen-Kappa 58%** για το πρόβλημα ταξινόμησης ύπνου σε 4 στάδια.

Σε μια άλλη έρευνα, στο [1], παρουσιάζεται ένα σύνολο αλγορίθμων για ταξινόμηση σταδίων του ύπνου με βάση την ακτιγραφία και τους καρδιακούς παλμούς μεταξύ 1.743 συμμετεχόντων της βάσης MESA. Οι ερευνητές χρησιμοποίησαν κινητικά σήματα (Movement data) και καρδιακά χαρακτηριστικά (HRV features) από φορητούς αισθητήρες (Fitbit), εκπαίδευσαν Machine Learning και Deep Learning μοντέλα και αξιολόγησαν την απόδοσή τους. Τα συνελκτικά νευρωνικά δίκτυα (CNN) και τα δίκτυα μακροπρόθεσμης βραχυπρόθεσμης μνήμης (LSTM) είχαν τις καλύτερες επιδόσεις ταξινόμησης. Τέλος, το καλύτερο μοντέλο, πέτυχε **accuracy 78.2%** και **F1- Score 69.8%** για το πρόβλημα ταξινόμησης τριών σταδίων ύπνου. Στο πρόβλημα 4 σταδίων ύπνου, ο καλύτερος αλγόριθμος πέτυχε **accuracy 71.3%**, **F1- Score 53.5%** και **Cohen-Kappa 58.8%**. Συγκεκριμένα, στο κεφάλαιο αυτό θα εστιάζουμε στην βελτίωση των αποτελεσμάτων της συγκεκριμένης δημοσίευσης.

Επιπρόσθετα, στο [87] οι ερευνητές προτείνουν ένα δίκτυο μακράς βραχυπρόθεσμης μνήμης (LSTM) για τη μοντελοποίηση των RR-Intervals που υπολογίστηκαν από ηλεκτροκαρδιογραφήματα (ECG). Το σετ δεδομένων RR-Intervals αποτελείται από 292 συμμετέχοντες, δηλαδή σύνολο 541.214 τμήματα ύπνου 30 δευτερολέπτων, και εξήχθησαν 132 HRV χαρακτηριστικά (time, frequency, entropy, regularity and miscellaneous features). Σημαντικό να αναφερθεί είναι ότι τα χαρακτηριστικά υπολογίστηκαν για κάθε περίοδο 30 δευτερολέπτων ύπνου χρησιμοποιώντας "παράθυρο" 4,5 λεπτών με επίκεντρο την εποχή αξιολόγησης (by epoch labeling). Τελικά, αποδεικνύεται ότι το μοντέλο επιτυγχάνει **accuracy 77%** και **Cohen-Kappa 61%** σε ολόκληρη τη βάση δεδομένων.

Στο [88], οι ερευνητές μοντελοποίησαν δεδομένα RR-Intervals και Breathing σήματα προκειμένου να επιτύχουν την ταξινόμηση των 4 σταδίων του ύπνου. Από τα RR-Intervals εξήγαγαν χαρακτηριστικά από το πεδίο του χρόνου και της συχνότητας (time, frequency domain). Όμως, επειδή τα RR-Intervals σήματα εμπεριέχουν απότομες αλλαγές, οι ερευνητές πρότειναν 3 επιπλέον χαρακτηριστικά από το πεδίο του χρόνου που αντιμετωπίζουν το πρόβλημα της απότομης αλλαγής των RR-Intervals. Από τα breathing σήματα, εξήγαγαν χαρακτηριστικά μόνο από το πεδίο του χρόνου. Τελικά, το προτεινόμενο Bidirectional LSTM μοντέλο πέτυχε **accuracy 80.7%** και **Cohen-Kappa 71%** στη βάση δεδομένων SHHS.

Σε μια ενδιαφέρουσα δημοσίευση, [89] η προσοχή εστιάζεται στη μοντελοποίηση δεδομένων από δι-

αφορετικούς αισθητήρες με διαφορετικά χαρακτηριστικά όπως η συχνότητα δειγματοληψίας (sampling rate), χαρακτηριστικά τα οποία δυσκολεύουν την μοντελοποίηση των διαφορετικών δεδομένων. Επίσης, οι ερευνητές εστιάζουν την προσοχή τους και εξετάζουν την ταξινόμηση των δεδομένων χρησιμοποιώντας είτε εξαγόμενα χαρακτηριστικά είτε ακατέργαστα (raw) δεδομένα. Για το πρόβλημα ταξινόμησης ύπνου κατέληξαν ότι το CNN μοντέλο που χρησιμοποίησαν, πέτυχε απόδοση **F1-Score 68%** και επίσης συμπεράναν ότι η μοντελοποίηση με ακατέργαστα raw δεδομένα δίνει καλύτερα αποτελέσματα έναντι της μοντελοποίησης με εξαγόμενα χαρακτηριστικά, εξαιτίας της πολυπλοκότητας του προβλήματος του ύπνου. Οπότε, επεξεργάζοντας raw δεδομένα, το μοντέλο μπορεί να ανακαλύψει σχέσεις και χαρακτηριστικά όπου δεν μπορούν να αποδοθούν με handcrafted χαρακτηριστικά που εξάγει ο άνθρωπος. Όμως, αυτό το αποτέλεσμα δεν μπορεί να γενικευθεί διότι η έρευνα τοποθετήθηκε σε συγκεκριμένο πλαίσιο.

Εν κατακλείδι, από την διεξαχθείσα έρευνα γύρω από τον τομέα ταξινόμησης του ύπνου σε 4 στάδια στη βάση MESA, θα αναλύσουμε την έρευνα [1] και θα προσπαθήσουμε να βελτιώσουμε τα υπάρχοντα αποτελέσματα για τους Deep Learning αλγόριθμους που δίνουν και τα καλύτερα αποτελέσματα έναντι των Machine Learning μοντέλων. Σημαντικό να αναφερθεί είναι το γεγονός ότι η ταξινόμηση του ύπνου σε στάδια με δεδομένα από φορητές συσκευές (wearables) με Deep Learning μοντέλα έχουν χαμηλότερη απόδοση από την αντίστοιχη ταξινόμηση με δεδομένα από ECG [87]. Οπότε, συγκρίνουμε τις αποδόσεις των μοντέλων που υλοποιήσαμε με τις αντίστοιχες έρευνες σε παρόμοια σετ δεδομένων ώστε να έχουμε μια βάση πάνω στην οποία θα τοποθετήσουμε την αξία της έρευνας που κάνουμε.

## 2.2 Σύνολο δεδομένων MESA - Προεπεξεργασία

### 2.2.1 Εισαγωγή στα δεδομένα

Στο παρόν κεφάλαιο θα αναλύσουμε την διαδικασία που ακολουθήθηκε για την προεπεξεργασία των δεδομένων MESA, πριν την μοντελοποίησή τους. Αρχικά, όπως αναφέραμε και προηγουμένως, η έρευνά μας επικεντρώνεται στην ταξινόμηση των δεδομένων MESA (actigraphy+HRV) με Deep Learning αλγόριθμους, οι οποίοι πετυχαίνουν καλύτερα αποτελέσματα στην ταξινόμηση του ύπνου σε 4 στάδια [1]. Τα δεδομένα που χρησιμοποιήσαμε είναι δεδομένα ακτιγραφίας και RR-Intervals (QRS Complexes) από 1.743 χρήστες. Σημαντικό να αναφερθεί είναι ότι το σύνολο δεδομένων περιλαμβάνει υγιείς και ασθενείς ανθρώπους, πληροφορία την οποία δεν μας δίνεται. Τελικά, με βάση το καλύτερο μοντέλο που πετυχαίνουν οι ερευνητές στο [1], θα δομήσουμε την έρευνα μας και θα προσπαθήσουμε να βελτιώσουμε τα αποτελέσματα για τη βάση δεδομένων MESA.

Συγκεκριμένα, επεξεργαστήκαμε τα δεδομένα για κάθε χρήστη ώστε να καθαριστούν και να είναι σε μορφή αξιοποιήσιμη από τους επιλεγμένους Deep Learning αλγόριθμους που θα παρουσιάσουμε παρακάτω. Για τη συλλογή των δεδομένων γνωρίζουμε ότι οι χρήστες φορούσαν ένα actigraphy wearable για μια εβδομάδα, και για μια νύχτα υποβλήθηκαν σε καταγραφή δεδομένων PSG το οποίο περιλαμβάνει και καταγραφή ECG. Από τις καταγραφές ECG προκύπτουν και τα δεδομένα R-points. Επίσης, γνωρίζουμε ότι τα δεδομένα ακτιγραφίας καταγράφουν "μετρήσεις δραστηριότητας" (activity counts) οι οποίες μετρήσεις ανά 30 δευτερόλεπτα αθροίζονται και προκύπτει μια τιμή "δραστηριότητας" που αντιστοιχεί στα 30 δευτερόλεπτα. Οπότε προκύπτουν τμήματα δεδομένων "δραστηριότητας" 30 δευτερολέπτων, τα οποία συνοδεύονται από την αντίστοιχη ετικέτα (label) που δείχνουν σε ποια κλάση (στάδιο ύπνου) ανήκουν. Όσον αφορά τα δεδομένα "R σημεία" (Rpoints), προκύπτουν από το ηλεκτροκαρδιογράφημα ECG που υποβάλλεται κάθε χρήστης με την τεχνική PSG. Η ταξινόμηση των δεδομένων ECG σε κλάσεις, γίνεται από ειδικούς ανά 30 δευτερόλεπτα, οπότε κάθε τμήμα (segment) 30 δευτερολέπτων συνοδεύεται από την αντίστοιχη ετικέτα (label) που δείχνει σε ποια κλάση (στάδιο ύπνου) ανήκει. Τελικά, για όλους τους χρήστες προκύπτει ένα σύνολο δεδομένων όπως φαίνεται στο Σχήμα 2.2.1.



Task 3		
Sleep Stages	# Epochs	%
Wake	652,509(314,784)	34%(20%)
Light	893,472	47%(57%)
Deep	128,874	7%(8%)
REM	229,045	12%(15%)
<b>Total</b>	<b>1,903,900(1,566,175)</b>	<b>100%</b>

Σχήμα 2.2.1: Σύνολο Δεδομένων ανά κλάση της βάσης MESA (30sec - segments).

Στο Σχήμα 2.2.1 παρατηρούμε ότι το σύνολο δεδομένων είναι **μη ισορροπημένο (imbalanced)**, δηλαδή, το σύνολο δεδομένων MESA δεν περιέχει το ίδιο αριθμό ετικετών (label) για κάθε κατηγορία/κλάση. Συγκεκριμένα, η κλάση Light έχει τα περισσότερα δεδομένα 30 δευτερολέπτων με ποσοστό 47%, ακολουθεί η κλάση Wake με ποσοστό 34% και τέλος οι κλάσεις Deep και REM με ποσοστό 7% και 12% αντίστοιχα. Η ανισορροπία αυτή εξηγείται, καθώς όπως αναφέραμε και στην εισαγωγή, κατά τη διάρκεια του ύπνου, τα στάδια REM και Deep πραγματοποιούνται για μικρό χρονικό διάστημα έως και καθόλου. Επίσης, στην ανισορροπία των δεδομένων υπεισέρχεται και το σφάλμα ταξινόμησης των δεδομένων από τους ιατρούς στις σωστές κλάσεις/στάδια ύπνου (sleep stages), δηλαδή υπεισέρχεται λάθος από ανθρώπινο παράγοντα. Από ιατρικής πλευράς οι κλάσεις Light και Deep είναι παρόμοιες μεταξύ τους, καθώς η κλάση Deep ανήκει στο στάδιο Light ως κατηγορία N3 [3]. Τέλος, πρέπει να αναφερθεί ότι οι χρήστες δεν έχουν την ίδια διάρκεια ύπνου οπότε μπορεί και αυτό να αποτελεί αιτία στην ανισοκατανομή των κλάσεων.

Στη συνέχεια, εφόσον έχουμε το σύνολο δεδομένων που αποτελείται από τμήματα των 30 δευτερολέπτων, με δεδομένα ακτιγραφίας και "R σημεία" (Rpoints) για κάθε χρήστη, στοχεύουμε στην **χρονική ευθυγράμμιση των δεδομένων (alignment)** ακτιγραφίας και "R σημείων" (Rpoints) μεταξύ τους, αφού: 1) Η καταγραφή των δεδομένων γίνεται από διαφορετικό εξοπλισμό. 2) Η συχνότητα δειγματοληψίας δεν είναι η ίδια. Δηλαδή, τα δεδομένα ακτιγραφίας καταγράφονται ανα 30 δευτερόλεπτα, ενώ τα "R σημεία" (Rpoints) από το ηλεκτροκαρδιογράφημα δεν έχουν σταθερή συχνότητα δειγματοληψίας. Δηλαδή, ανα 30 δευτερόλεπτα δεν έχουμε σταθερό αριθμό "R σημείων" (Rpoints). Αυτή η διαδικασία της χρονικής ευθυγράμμισης των δεδομένων γίνεται από το αρχείο overlap.xlsx<sup>1</sup>, που προσφέρεται από τη βάση MESA, και εξάγουμε το χρονικό διάστημα που πραγματοποιείται ταυτόχρονη καταγραφή δεδομένων ακτιγραφίας και "R σημείων" (Rpoints) για κάθε εποχή 30 δευτερολέπτων.

## 2.2.2 Καθαρισμός δεδομένων - Data Cleaning

Μετά την χρονική ευθυγράμμιση των δεδομένων, καθαρίζουμε τα δεδομένα (**Data Cleaning**) ώστε να αφαιρέσουμε την λανθασμένη πληροφορία που καταγράφηκε από τους αισθητήρες και να είναι σωστά για την ταξινόμησή τους. Συγκεκριμένα, το σύνολο δεδομένων MESA έχει 30 χρήστες οι οποίοι δεν έχουν καθόλου κλάση REM (συνολικά 2% των δεδομένων) αλλά δεν αφαιρέθηκαν από την μοντελοποίηση επειδή η τελευταία δεν γίνεται ανά χρήστη. Δηλαδή στη μοντελοποίηση δεν λαμβάνουμε υπόψιν τους διαφορετικούς χρήστες, όμως η προεπεξεργασία των δεδομένων γίνεται ανα χρήστη ώστε τα δεδομένα να προετοιμαστούν σωστά για την επακόλουθη ταξινόμησή τους.

Όσον αφορά τα "R σημεία" (Rpoints), αφαιρούμε: Πρώτον, τα διαδοχικά "R σημεία" (Rpoints) που αλληλοκαλύπτονται και δεύτερον τα σημεία τα οποία δεν είναι καλά φορεμένη η συσκευή του PSG από

<sup>1</sup>Overlap files: <https://sleepdata.org/datasets/mesa/files/overlap>

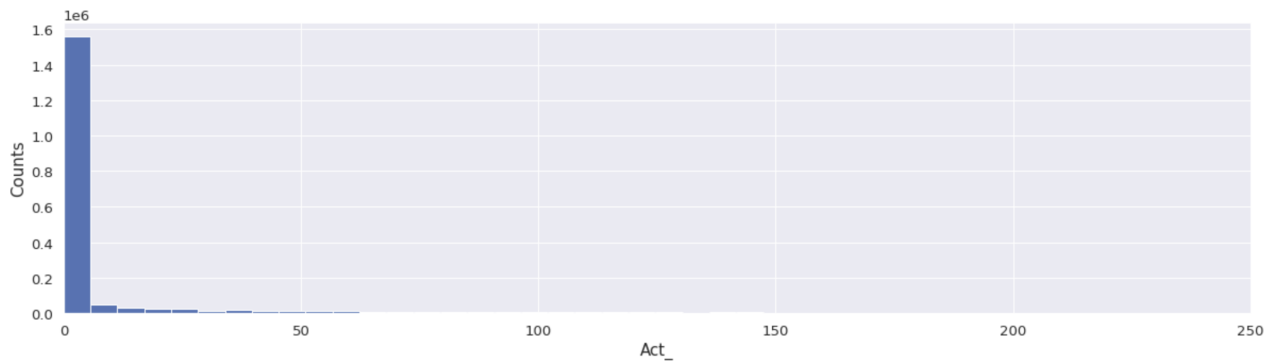
τον χρήστη και έχουν λάθος τιμές. Αυτή η πληροφορία δίνεται από το σύνολο δεδομένων MESA μέσω της μεταβλητής "TPoint" η οποία περιλαμβάνεται στο σύνολο δεδομένων των "R σημείων" (Rpoints). Η μεταβλητή "Tpoint" περιέχει τιμές μεγαλύτερες του μηδενός αλλά και την τιμή -1 αν δεν γίνεται σωστή καταγραφή δεδομένων. Οπότε για την μοντελοποίηση κρατάμε μόνο τις θετικές τιμές και με αυτόν τον τρόπο αφαιρούνται τα σημεία για τα οποία δεν έγινε καλή καταγραφή δεδομένων. Εν συνεχεία, τα "R σημεία" (Rpoints) μετατρέπονται σε RR-Intervals με βάση την  $1^{\text{η}}$  παράγωγο των millisecond και πολλαπλασιάζονται επί 1000 ώστε να μετατραπούν σε δευτερόλεπτα. Επίσης, οι κενές τιμές (missing values) που υπάρχουν μέσα στο σύνολο δεδομένων RR-Intervals αντικαθίστανται από την μέση τιμή των δεδομένων. Στο τελικό στάδιο του data cleaning, τα RR-Intervals φιλτράρονται μεταξύ των τιμών 300msec και 2000msec οπότε αφαιρούνται οι ακραίες τιμές έξω από το εύρος αυτό (outliers). Στη συνέχεια, οι κενές τιμές (missing values) αντικαθίστανται με την μέθοδο γραμμικής παρεμβολής (linear interpolation). Τέλος, αφαιρούνται οι ectopic beats με την μέθοδο Malik [90] και επαναλαμβάνεται η αντικατάσταση των missing values με την μέθοδο της γραμμικής παρεμβολής. Με αυτόν τον τρόπο ολοκληρώνεται η προεπεξεργασία/καθαρισμός των δεδομένων RR-Intervals με τη βοήθεια της βιβλιοθήκης HRVanalysis (Python) [91]. Η διαδικασία που ακολουθήσαμε για τον καθαρισμό των RR-Intervals είναι ο ίδιος με αυτόν που δίνεται από το [91]. Η βιβλιοθήκη HRVanalysis είναι ένα χρήσιμο εργαλείο (module) σε γλώσσα προγραμματισμού Python το οποίο έχει ως στόχο την ανάλυση και επεξεργασία σημάτων καρδιακών παλμών (Heart Rate). Περιλαμβάνει συναρτήσεις για καθαρισμό των καρδιακών παλμών (data cleaning) και εξαγωγή χαρακτηριστικών (Heart Rate Variability features) από το πεδίο του χρόνου, συχνότητας και διάφορα άλλα, όπως γεωμετρικά και μη γραμμικά χαρακτηριστικά (geometrical and non-linear).

Όσον αφορά τα δεδομένα ακτιγραφίας, αφαιρούνται οι εποχές που υπάρχουν λάθη καταγραφής ή δεν έχει γίνει κατηγοριοποίηση από τους ιατρούς. Αυτή η πληροφορία δίνεται από την μεταβλητή "EXCLUDED" που υπάρχει στο σύνολο δεδομένων ακτιγραφίας και περιέχει 2 διαφορετικές τιμές ("ACTIVE" και "EXCLUDED"). Όπως είναι προφανές, για τη μοντελοποίηση κρατάμε τις εποχές που έχουν ετικέτες ACTIVE. Τέλος, *συνενώνουμε τα δεδομένα ακτιγραφίας με τα δεδομένα RR-Intervals και σαν ετικέτα (label) κάθε ακολουθίας 30 δευτερολέπτων χρησιμοποιούμε εκείνη των δεδομένων ακτιγραφίας.* Στον Πίνακα 2.1 παρατίθενται τα στατιστικά των δεδομένων Ακτιγραφίας για το σύνολο δεδομένων και στο Σχήμα 2.2.2 και 2.2.3 το αντίστοιχο ιστόγραμμα και διάγραμμα.

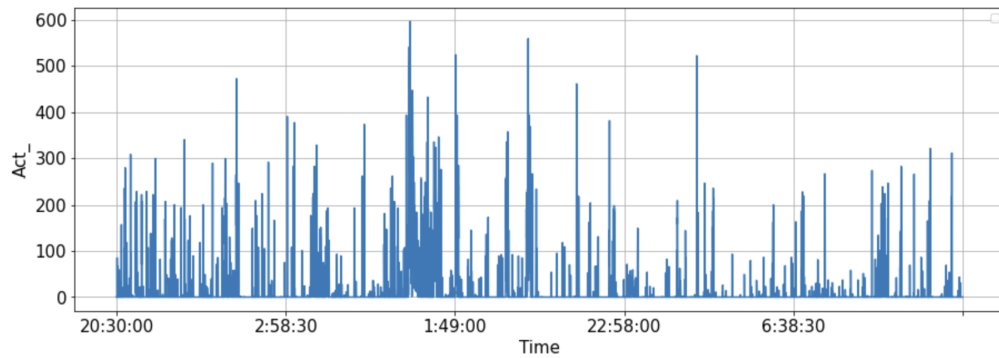
<i>Statistics</i>	<i>Values</i>
count	1903481
mean	14.437
std	47.81
min	0.0
25%	0.0
50%	0.0
75%	0.0
max	1700

Πίνακας 2.1: Στατιστικά των δεδομένων Ακτιγραφίας.

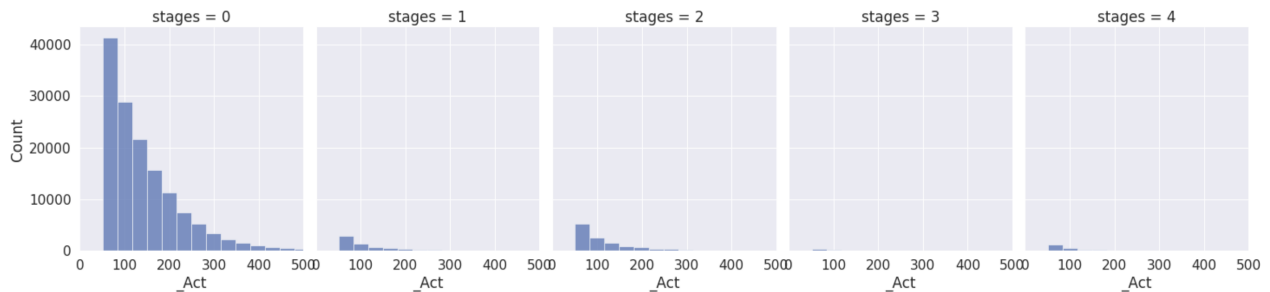
Από τα Σχήματα (2.2.2., 2.2.3) παρατηρούμε ότι οι περισσότερες τιμές "δραστηριότητας" (**activity counts**) συσσωρεύονται κοντά στο 15 με τυπική απόκλιση κοντά στο 50 και μέγιστη τιμή 1700. Αυτό δικαιολογείται επειδή ο ύπνος χαρακτηρίζεται από χαμηλές τιμές δραστηριότητας σε σχέση με την δραστηριότητα εκτός ύπνου, δηλαδή κατά τη διάρκεια της ημέρας (activity during day), αλλά υπάρχουν σημεία κατά τη διάρκεια του ύπνου τα οποία χαρακτηρίζονται από έκρηξη κίνησης (έντονη κινητικότητα) ή στάδια με σταθερή κινητικότητα, όπως κατά τα στάδια Light όπως φαίνεται στο Σχήμα 2.2.4.



Σχήμα 2.2.2: Ιστόγραμμα δεδομένων Ακτιγραφίας.



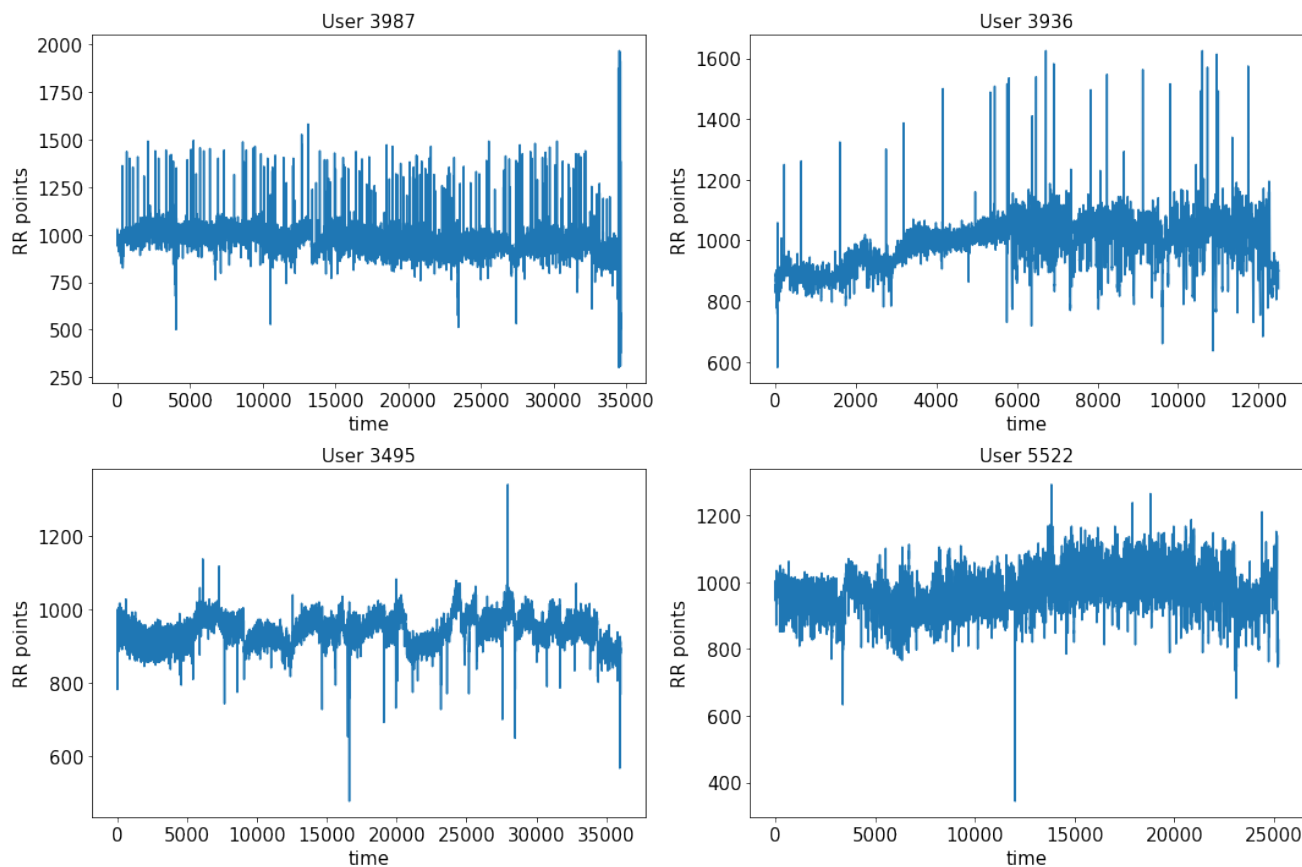
Σχήμα 2.2.3: Διάγραμμα δεδομένων Ακτιγραφίας για 1000 σημεία.



Σχήμα 2.2.4: Ιστόγραμμα δεδομένων Ακτιγραφίας ανά κλάση. **Wake:0, NREM<sub>1</sub>:1, NREM<sub>2</sub>:2, NREM<sub>3</sub>:3, REM:4**

Τα **δεδομένα RR-Intervals** χωρίζονται ανά 30 δευτερόλεπτα και αντιστοιχούν σε μια τιμή activity counts όπως εξηγήσαμε παραπάνω. Για 4 τυχαίους χρήστες από το σύνολο δεδομένων MESA παρουσιάζουμε τα παρακάτω Σχήματα 2.2.5 και 2.2.6.

Παρατηρούμε ότι η κατανομή των δεδομένων είναι κανονική κατανομή (Σχήμα 2.2.6) με μέση τιμή κοντά στο 1 δευτερόλεπτο (1000msec) και τυπική απόκλιση κοντά στα 0.1 δευτερόλεπτα (100msec). Επίσης όπως αναφέραμε και παραπάνω, οι τιμές κυμαίνονται μεταξύ 0.3 δευτερόλεπτα (300msec) και 2 δευτερόλεπτα (2000msec). Από τα διαγράμματα στο Σχήμα 2.2.5 διαπιστώνουμε ότι τα RR-Intervals έχουν σταθερές περιοχές, όπως πχ. στην αρχή του ύπνου, αλλά και περιοχές με έντονη διακύμανση (User3936, πάνω δεξιά στο Σχήμα 2.2.5). Το γεγονός αυτό αποδεικνύει ότι ο κάθε χρήστης περνά

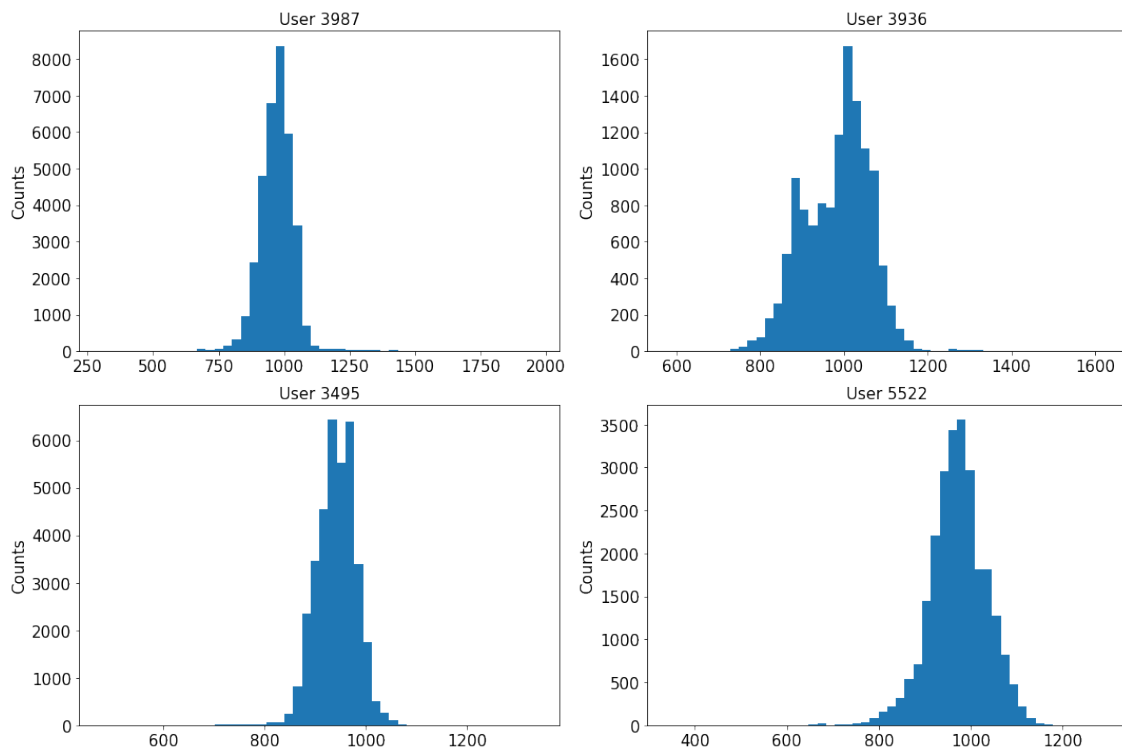


Σχήμα 2.2.5: Διαγράμματα RR-Intervals για 4 χρήστες.

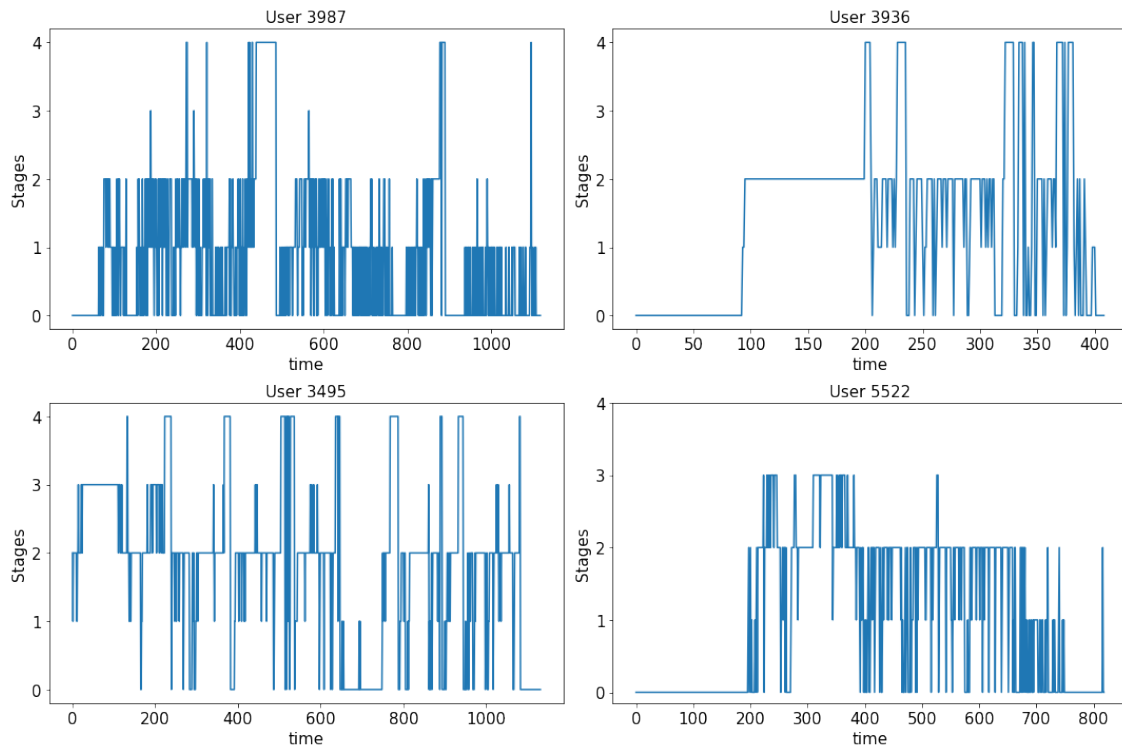
από τα διαφορετικά στάδια του ύπνου όπου η καρδιακή δραστηριότητα είναι διαφορετική ή παρόμοια. Επιπλέον, δεν παρατηρούμε κυκλικότητα με βάση τα διαγράμματα RR-Intervals. Στο Σχήμα 2.2.7 παρατίθεται τα αντίστοιχα υπονογράμματα των 4 τυχαίων χρηστών. Όπως είναι φυσιολογικό οι χρήστες δεν έχουν τον ίδιο ύπνο άρα δεν ακολουθούν το ίδιο μοτίβο οι εναλλαγές ανάμεσα στις 4 κλάσεις. Επίσης, κάποιος χρήστης μπορεί να μην περνούν καθόλου από κάποια στάδια ύπνου είτε διότι δεν κοιμήθηκαν καλά, είτε έχουν πρόβλημα με τον ύπνο, είτε ο ύπνος τους ήταν πολύ μικρός σε διάρκεια. Όμως, η πληροφορία αυτή δεν μας απασχολεί διότι αντιμετωπίζουμε το πρόβλημα ταξινόμησης σε 4 στάδια **χωρίς να λαμβάνουμε υπόψιν τα χαρακτηριστικά του κάθε χρήστη**. Τέλος, από το Σχήμα 2.2.7 παρατηρούμε ότι τα στάδια ύπνου δεν έχουν την ίδια χρονική διάρκεια και ότι η εναλλαγή μεταξύ τους δεν ακολουθεί απαραίτητα κάποια κυκλικότητα (sleep cycle). Στην επόμενη ενότητα θα χρησιμοποιήσουμε τα δεδομένα RR-Intervals για την εξαγωγή χαρακτηριστικών που θα βοηθήσουν την ταξινόμηση του ύπνου και στη περαιτέρω κατανόηση του προβλήματος που έχουμε να αντιμετωπίσουμε.

### 2.2.3 Εξαγωγή χαρακτηριστικών (Feature Extraction)

Σε αυτήν την υποενότητα θα εξάγουμε χαρακτηριστικά από τις ακολουθίες RR-Intervals μήκους 30 δευτερολέπτων που θα βοηθήσουν στην ταξινόμηση του ύπνου σε 4 στάδια. Αυτή η διαδικασία περιλαμβάνει **εξαγόμενα χαρακτηριστικά** από το πεδίο του χρόνου (**Time Domain**), συχνότητας (**Frequency Domain**), γεωμετρικά (**Geometrical Domain**) και μη-γραμμικά χαρακτηριστικά (**Non-Linear Domain**). Τα χαρακτηριστικά αυτά είναι Heart Rate Variability (HRV) χαρακτηριστικά που ερμηνεύουν, από διαφορετική οπτική το καθένα, την **χρονική διακύμανση** μεταξύ των σημείων



Σχήμα 2.2.6: Ιστογράμματα RR-Intervals για 4 χρήστες.



Σχήμα 2.2.7: Υπνογράμματα από 4 χρήστες.

Rpoints και μέσω αυτών θα προσπαθήσουμε να εξάγουμε σχετική πληροφορία με τον ύπνο, που θα ερμηνεύουν και θα διαχωρίζουν τα στάδια μεταξύ τους στον βέλτιστο βαθμό.

Οπότε, χρησιμοποιώντας τα δεδομένα RR-Intervals κάθε χρήστη, εξάγουμε χαρακτηριστικά HRV με την βοήθεια της βιβλιοθήκης `hrv-analysis` που προσφέρεται σε γλώσσα προγραμματισμού Python και σύμφωνα με το [87]. **Για κάθε εποχή 30 δευτερολέπτων**, εξάγουμε αρχικά χαρακτηριστικά από το πεδίο του χρόνου. Δηλαδή, χαρακτηριστικά όπως παρατίθενται παρακάτω. Για επιπλέον πληροφορίες παρατίθεται η έρευνα [92].

#### ***Time domain features:***

- *MeanNNI*: Μέση τιμή των RR-Intervals.
- *SDNN*: Τυπική απόκλιση των RR-Intervals.
- *SDSD*: Τυπική απόκλιση της διαφοράς των RR-Intervals.
- *NN50*: Αριθμός RR-Intervals με τιμή >50msec.
- *pNN50*: Λόγος μεταξύ του NN50 με τον συνολικό αριθμό RR-Intervals. (NN50/ # σύνολο RR)
- *NN20*: Πλήθος RR-Intervals με τιμή >20msec.
- *pNN20*: Λόγος μεταξύ του NN20 με τον συνολικό αριθμό RR-Intervals. (NN20/ # σύνολο RR)
- *RMSSD*: Ρίζα του μέσου όρου των RR-Intervals.
- *MedianNN*: Διάμεση τιμή των RR-Intervals.
- *RangeNN*: Απόσταση (σε msec) μεταξύ της μέγιστης και της ελάχιστης τιμής των RR-Intervals.
- *CVSD*: Συντελεστής διακύμανσης των διαδοχικών διαφορών (το RMSSD διαιρούμενο με το μέσο όρο των RR-Intervals).
- *CVNNI*: Ο συντελεστής διακύμανσης των RR-Intervals, δηλαδή ο λόγος του SDNN διαιρούμενος με το μέσο όρο των RR-Intervals.
- *MeanHR*: Μέση τιμή του Heart Rate (HR = 60000 / RR-Intervals).
- *MaxHR*: Μέγιστη τιμή του Heart Rate (HR = 60000 / RR-Intervals).
- *MinHR*: Ελάχιστη τιμή του Heart Rate (HR = 60000 / RR-Intervals).
- *STDHR*: Τυπική απόκλιση του Heart Rate (HR = 60000 / RR-Intervals).

Παρακάτω, βλέπουμε τα χαρακτηριστικά που έχουμε εξάγει από το πεδίο συχνοτήτων:

#### ***Frequency domain features:***

- *Low Frequency (LF)*: Διακύμανση (δηλαδή, η ισχύς) στο HRV στη χαμηλή συχνότητα (0.04 έως 0.15 Hz). Αντανακλά ένα μείγμα της συμπαθητικής και της παρασυμπαθητικής δραστηριότητας.
- *High Frequency (HF)*: Διακύμανση (δηλαδή, η ισχύς) στο HRV στην υψηλή συχνότητα (0.15 έως 0.40 Hz). Αντικατοπτρίζει τις γρήγορες αλλαγές μεταβλητότητας (Beat-to-Beat) λόγω παρασυμπαθητικής (vagal) δραστηριότητας.



- *Variance in Low Frequency (VLF)*: Διακύμανση (δηλαδή, η ισχύς) στο HRV στην Πολύ χαμηλή συχνότητα (0.003 έως 0.04 Hz). Αντικατοπτρίζει έναν εγγενή ρυθμό που παράγεται από την καρδιά, η οποία ρυθμίζεται κυρίως από τη συμπαθητική δραστηριότητα.
- *Low/High Frequency Ratio*: Η αναλογία LF προς HF.
- *Normalised Low Frequency Ratio*: Κανονικοποιημένη αναλογία της χαμηλής συχνότητας που υπολογίζεται από τις ακατέργαστες (raw) τιμές της ζώνης χαμηλής συχνότητας διαιρεμένη με τη συνολική φασματική ισχύ.
- *Normalised High Frequency Ratio*: Κανονικοποιημένη αναλογία της υψηλής συχνότητας που υπολογίζεται από τις ακατέργαστες (raw) τιμές της ζώνης υψηλής συχνότητας διαιρεμένη με τη συνολική φασματική ισχύ.
- *Mean NNi*: Μέση τιμή των RR-Intervals.
- *Total Power*: Ολική ισχύς του φάσματος πυκνότητας (spectral density).

Ακολουθούν τα γεωμετρικά χαρακτηριστικά:

#### ***Geometrical domain features:***

- *Triangular Index*: Η μέτρηση του τριγωνικού δείκτη HRV είναι το ολοκλήρωμα της κατανομής πυκνότητας (δηλαδή, ο αριθμός όλων των RR-Intervals) διαιρούμενο με το μέγιστο της κατανομής πυκνότητας (πλάτος κατηγορίας 8 ms)

Και τέλος, τα μη-γραμμικά χαρακτηριστικά:

#### ***Non-linear domain features:***

- *Cardiac Sympathetic IdNx (csi)*: Καρδιακός Δείκτης του συμπαθητικού.
- *Modified Cardiac Sympathetic IdNx (mcsi)*: Τροποποιημένος Καρδιακός Δείκτης του συμπαθητικού. ( $\frac{SD1^2}{SD2}$ )
- *Cardiac Vagal IndeNx (cvi)*: Καρδιακός Πνευμονογαστρικός Δείκτης.
- *SD1*: Τυπική απόκλιση της γραφικής παράστασης Πουανκαρέ (Poincare Plot) κάθετη στη γραμμή ταυτότητας.
- *SD2*: Τυπική απόκλιση της γραφικής παράστασης Πουανκαρέ (Poincare Plot) κατα μήκος της γραμμής ταυτότητας.
- *SD1/SD2 Ratio*: Λόγος SD1 προς SD2.

Από το σύνολο των παραπάνω χαρακτηριστικών (features), επιλέγουμε, σύμφωνα με το [1], **8 HRV χαρακτηριστικά** συν την μεταβλητή "**δραστηριότητας**" (activity). Οπότε, συνολικά επιλέχθηκαν 9 χαρακτηριστικά σε πρώτο στάδιο, ώστε να μπορούμε να συγκρίνουμε τα αποτελέσματά μας με το [1], και να προσπαθήσουμε να τα βελτιώσουμε. Συγκεκριμένα, αυτά τα χαρακτηριστικά επιλέχθηκαν από τους ερευνητές διότι επιδίωξαν η μοντελοποίηση τους να μην λαμβάνει υπόψιν την πολυπλοκότητα της συσκευής καταγραφής δεδομένων. Δηλαδή, αποφάσισαν να μην χρησιμοποιήσουν τα δεδομένα ηλεκτροκαρδιογραφίας (ECG), επειδή η καταγραφή δεδομένων ECG είναι δαπανηρή διαδικασία και δεν είναι προς το παρόν διαθέσιμα στους φορητούς αισθητήρες. Έτσι, χρησιμοποίησαν ένα σύνολο 8 HRV χαρακτηριστικών που μπορούν να καταγραφούν από πολλούς φορητούς καρδιακούς αισθητήρες, όπως το Eatica ή το ActiHeart. Για τα δεδομένα κίνησης, απλά χρησιμοποίησαν τις μετρήσεις "δραστηριότητας" από τη συσκευή ακτιγραφίας που φοριέται στον καρπό.

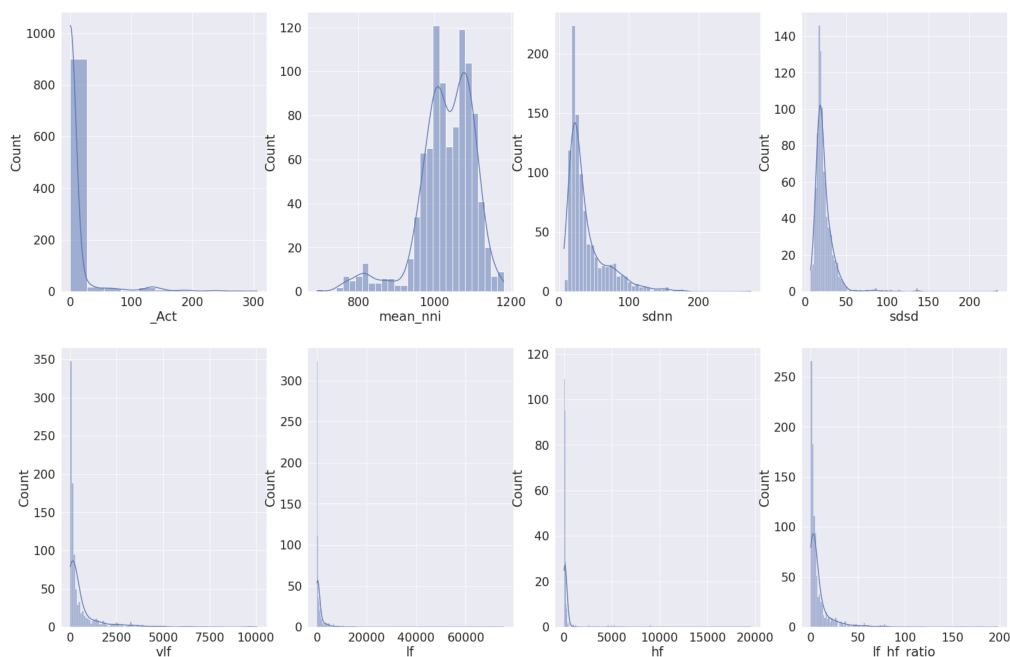


## 2.2.4 Ανάλυση χαρακτηριστικών Activity Counts και HRV (Explanatory Data Analysis)

Εν τέλει, από το σύνολο των εξαγόμενων χαρακτηριστικών, καταλήξαμε στα εξής **10** χαρακτηριστικά:

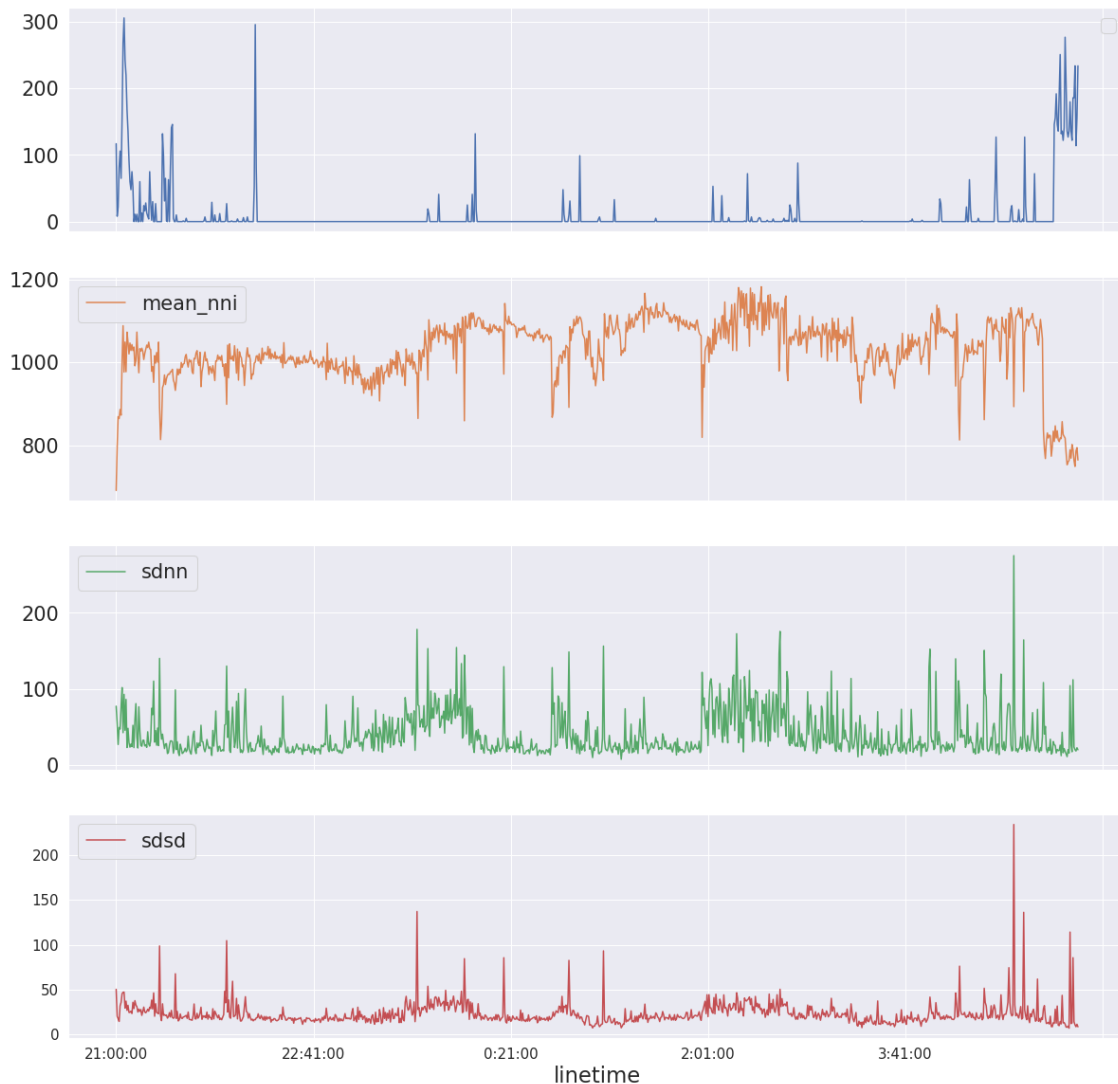
- Activity Counts
- Mean NNI,
- Standard Derivation of RR interval (SDNN),
- RR interval differences (SDSD),
- Very Low Frequency (VLF),
- Low Frequency (LF),
- High Frequency Bands (HF),
- Low Frequency to High Frequency Ratio (LF/HF),
- Total Power

Οπότε για τα παραπάνω χαρακτηριστικά παραθέτουμε τα Σχήματα 2.2.8, Σχήμα 2.2.9, Σχήμα 2.2.10. Από τα Σχήματα θα μπορέσουμε διαισθητικά να κάνουμε κάποιες παρατηρήσεις.



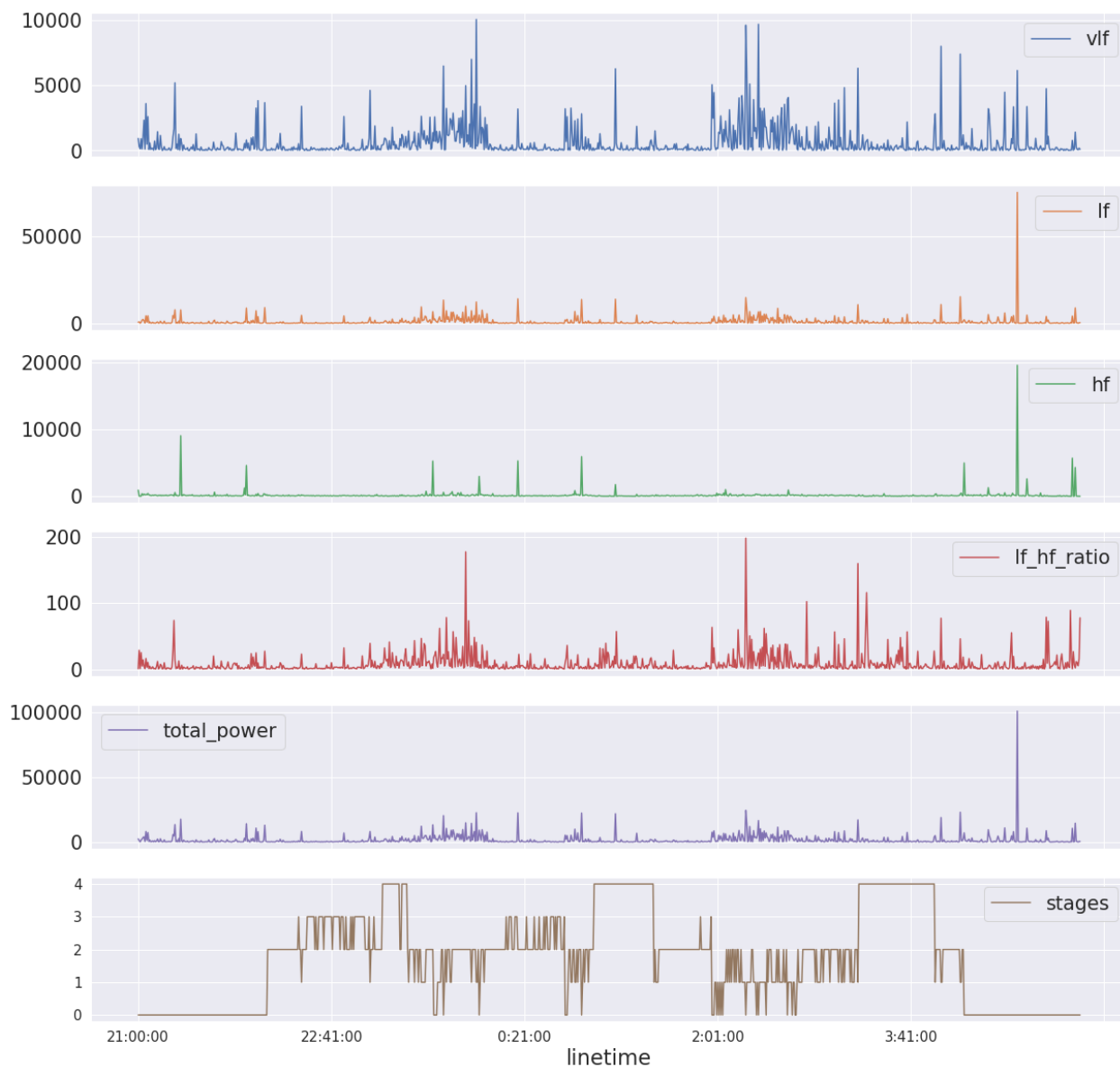
Σχήμα 2.2.8: Ιστογράμματα συνόλου χαρακτηριστικών ενός χρήστη (UserId=2982).

Τα παραπάνω Σχήματα (2.2.8, 2.2.9 και 2.2.10) που παραθέσαμε αντιστοιχούν σε διαγράμματα που περιγράφουν τον ύπνο και αφορούν έναν χρήστη. Αρχικά, από το ιστόγραμμα των activity counts (Σχήμα 2.2.8) παρατηρούμε ότι οι περισσότερες τιμές δραστηριότητας είναι κάτω από 50 και κάποιες πολύ μικρές σε αριθμό τιμές είναι άνω του 50. Από αυτό συμπεραίνουμε ότι ο χρήστης δεν κινείται πολύ στον ύπνο του που διαρκεί από τις 21:00 έως τις 4:00 (σύνολο 7 ώρες). Επίσης, παρατηρούμε ότι στα πρώτα στάδια του ύπνου του, για περίπου 40 λεπτά, η κινητικότητα είναι υψηλή με έντονες διακυμάνσεις



Σχήμα 2.2.9: Διαγράμματα χαρακτηριστικών ενός χρήστη (UserId=2982).

και με τη πάροδο του χρόνου μειώνεται και παρατηρείται μικρότερη διακύμανση [93]. Η παρατήρηση αυτή είναι λογική καθώς ο χρήστης βρίσκεται στην αρχή του ύπνου, οπότε κινείται πιο έντονα (στάδιο Wake). Επακόλουθα, το ίδιο συμπέρασμα προκύπτει και από τη μεταβλητή Mean NNi (Σχήμα 2.2.9), όπου για το ίδιο χρονικό διάστημα, κατά μέσο όρο τα RR-Intervals (Mean NNi) έχουν τιμές μικρότερες του 1 δευτερολέπτου, δηλαδή οι διαδοχικοί καρδιακοί παλμοί είναι πιο σύντομοι μεταξύ τους επειδή ο χρήστης κινείται (Σχήμα 2.2.9 το δεύτερο Διάγραμμα). Επίσης, παρατηρούμε ότι το χρονικό διάστημα 2:00 με 3:00, ο χρήστης εναλλάσσει στάδια μεταξύ των N1 και N2 (Light Sleep). Σε αυτό το διάστημα, οι τιμές όλων των μεταβλητών είναι υψηλές και συγκεκριμένα οι μεταβλητές Mean NNi, Total Power και LF/HF. Η παρατήρηση αυτή επιβεβαιώνει ότι κατά τη διάρκεια του σταδίου Light, ο άνθρωπος τείνει να χαλαρώσει και αυτό δικαιολογεί τις υψηλές τιμές. Επίσης, για το ίδιο χρονικό διάστημα, παρατηρούμε συγκεκριμένα ότι η μεταβλητή LF/HF (Σχήμα 2.2.10), δηλαδή ο λόγος της ισχύος του HRV στο εύρος χαμηλών συχνοτήτων προς την ισχύ του HRV στο εύρος των υψηλών συχνοτήτων, έχει υψηλές τιμές. Η παρατήρηση επιβεβαιώνει την υψηλή καρδιακή δραστηριότητα κατά τα στάδια αυτά όπως φαίνεται στο Σχήμα 2.2.10 (*lf hf ratio*). Αντίθετα, στο χρονικό διάστημα 3:00 έως 3:50, ο χρήστης είναι στο στάδιο REM και οι τιμές των Mean NNi, VLF, activity counts και Total Power έχουν χαμηλές τιμές.



Σχήμα 2.2.10: Διαγράμματα χαρακτηριστικών ενός χρήστη (UserId=2982). Για το τελευταίο διάγραμμα που παρουσιάζει το υπνόγραμμα του χρήστη UserId=2982, παραθέτουμε την αντιστοίχιση των σταδίων ύπνου: Wake:0, N1:1, N2:2, Deep:3, REM:4

Το γεγονός αυτό δικαιολογείται επειδή κατά το στάδιο REM, ο άνθρωπος ονειρεύεται, δηλαδή έχει πολύ έντονη εγκεφαλική κινητικότητα, αλλά αντίθετα οι μύες έχουν παραλύσει, οπότε οι τιμές των μεταβλητών activity counts και Total Power είναι πολύ χαμηλές [2]. Τέλος, για το χρονικό διάστημα 22:00 έως 23:00, παρατηρούμε ότι ο χρήστης εναλλάσσει στάδια μεταξύ Light και Deep Sleep. Κατά το στάδιο Deep, όλες τις μεταβλητές έχουν χαμηλές τιμές αλλά το ενδιαφέρον εστιάζεται στις μεταβλητές Mean NNi, SDNN και Total Power. Συγκεκριμένα, η μεταβλητή Mean NNi είναι σταθερή με μικρή διακύμανση και γι αυτό η μεταβλητή SDNN έχει χαμηλές τιμές. Επίσης, η ολική ισχύς του φάσματος πυκνότητας (Total Power) είναι σε χαμηλά επίπεδα λόγω της χαμηλής κινητικότητας του χρήστη. Συμπερασματικά, οι χαμηλές τιμές και η σταθερότητα των παραπάνω μεταβλητών δικαιολογείται επειδή κατά το στάδιο Deep ο άνθρωπος ξεκουράζεται, οι καρδιακοί παλμοί είναι χαμηλοί και η εγκεφαλική

δραστηριότητα είναι επίσης χαμηλή [2].

Στη συνέχεια της ανάλυσης, θα εξετάσουμε τη γραμμική συσχέτιση (με τη μέθοδο Pearson) μεταξύ των μεταβλητών/χαρακτηριστικών και θα εξάγουμε συμπεράσματα που έχουν ενδιαφέρον για το πρόβλημα μας. Ο συντελεστής συσχέτισης Pearson είναι ένα μέτρο **γραμμικής συσχέτισης** μεταξύ δύο συνόλων δεδομένων. Είναι ο λόγος μεταξύ της συνδιακύμανσης (covariance) δύο μεταβλητών και του γινόμενου των τυπικών αποκλίσεων τους ( $\rho = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y}$  και  $\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$ ). Επομένως, είναι ουσιαστικά μια κανονικοποιημένη μέτρηση της συνδιακύμανσης, έτσι ώστε το αποτέλεσμα να έχει πάντα μια τιμή μεταξύ -1 και 1. Όπως και με την ίδια τη συνδιακύμανση, το μέτρο μπορεί να αντικατοπτρίζει μόνο μια γραμμική συσχέτιση μεταβλητών και αγνοεί πολλούς άλλους τύπους σχέσης ή συσχέτισης<sup>2</sup>. Συγκεκριμένα, το Σχήμα 2.2.11 απεικονίζει τον Πίνακα Συσχετίσεων μεταξύ των 10 χαρακτηριστικών για όλο το σύνολο δεδομένων, δηλαδή για όλους τους χρήστες.

	_Act	mean_nni	sdnn	sdsd	vlf	lf	hf	lf_hf_ratio	total_power
_Act	1.00	-0.17	0.13	0.12	0.08	0.12	0.16	0.02	0.14
mean_nni	-0.17	1.00	0.16	0.21	0.08	0.05	-0.00	-0.04	0.05
sdnn	0.13	0.16	1.00	0.90	0.69	0.73	0.64	0.12	0.79
sdsd	0.12	0.21	0.90	1.00	0.53	0.67	0.72	-0.06	0.73
vlf	0.08	0.08	0.69	0.53	1.00	0.76	0.36	0.18	0.82
lf	0.12	0.05	0.73	0.67	0.76	1.00	0.60	0.08	0.98
hf	0.16	-0.00	0.64	0.72	0.36	0.60	1.00	-0.06	0.71
lf_hf_ratio	0.02	-0.04	0.12	-0.06	0.18	0.08	-0.06	1.00	0.08
total_power	0.14	0.05	0.79	0.73	0.82	0.98	0.71	0.08	1.00

Σχήμα 2.2.11: Πίνακας Γραμμικής Συσχέτισης των 9 χαρακτηριστικών για όλους τους χρήστες.

Παρατηρούμε αρχικά ότι, τα δεδομένα ακτιγραφίας (activity counts, Σχήμα 2.2.11 ως μεταβλητή **Act**) δεν έχουν μεγάλη γραμμική συσχέτιση με κανένα από τα υπόλοιπα χαρακτηριστικά. Η παρατήρηση αυτή είναι λογική επειδή τα δεδομένα ακτιγραφίας είναι δεδομένα από τον αισθητήρα κίνησης, ενώ τα HRV χαρακτηριστικά αφορούν τους καρδιακούς παλμούς. Επίσης, παρατηρούμε πολλή υψηλή γραμμική συσχέτιση μεταξύ της ισχύος του HRV στην χαμηλή συχνότητα (LF) και της ισχύς (Total Power) των σημάτων RR-Intervals (98%). Η παρατήρηση αυτή επιβεβαιώνει ότι τα RR-Intervals έχουν συνολική ισχύ (Total Power) στο εύρος των χαμηλών συχνοτήτων (LF και VLF (συσχέτιση 82%) και λιγότερο στο εύρος των υψηλών συχνοτήτων (συσχέτιση 71%). Γνωρίζουμε ότι η συνολική ισχύς των RR-Intervals (Total power) είναι το άθροισμα των ισχύων των επιμέρους συχνοτήτων, δηλαδή των VLF, LF, and HF για μικρού μήκους καταγραφές [92]. Αυτό απεικονίζεται επίσης και από τα διαγράμματα ενός χρήστη στο Σχήμα 2.2.10. Παρατηρούμε, δηλαδή ότι η ισχύς των υψηλών συχνοτήτων (HF) έχει χαμηλότερες τιμές (με πολύ λίγα peaks) από το διάγραμμα χαμηλών συχνοτήτων (LF), ενώ το διάγραμμα της ισχύος στο εύρος των χαμηλών συχνοτήτων έχει μεγαλύτερες τιμές και ομοιάζει πολύ με το διάγραμμα της μεταβλητής Total Power. Άρα κατά τη διάρκεια του ύπνου, τα σήματα των RR-Intervals αποτελούνται κυρίως από χαμηλές και λιγότερο από τις υπόλοιπες συχνότητες. Τέλος, όσον αφορά τις υπόλοιπες μεταβλητές, οι τιμές των συσχετίσεων είναι λογικές όπως για παράδειγμα η συσχέτιση των μεταβλητών SDNN και SDDSD (90%) αφού ερμηνεύουν παραπλήσια χαρακτηριστικά των σημάτων RR-Intervals.

<sup>2</sup>Pearson Correlation: [https://en.wikipedia.org/wiki/Pearson\\_correlation\\_coefficient](https://en.wikipedia.org/wiki/Pearson_correlation_coefficient)

Επακόλουθα στο Σχήμα 2.2.12, παρατίθενται τα στατιστικά δεδομένα των 10 μεταβλητών.

	_Act	mean_nni	sdnn	sdsd	vlf	lf	hf	lf_hf_ratio	total_power
count	217319.000000	217319.000000	217319.000000	217319.000000	217319.000000	217319.000000	217319.000000	217319.000000	217319.000000
mean	14.809027	948.043189	51.239508	39.598084	982.392281	2707.101772	936.560190	6.108309	4626.060176
std	48.992521	162.775582	54.968888	42.145161	3666.492060	10973.532424	3611.426216	14.562753	16290.609115
min	0.000000	382.812500	0.000000	0.000000	0.000000	0.000000	0.000000	0.003509	0.000000
25%	0.000000	836.272321	19.365169	14.863197	35.853480	77.406471	42.961335	0.828109	222.912729
50%	0.000000	943.725586	32.991822	25.189438	137.518102	269.400986	126.521201	2.153309	657.522143
75%	0.000000	1047.230440	59.850564	46.808393	543.734666	1076.304547	414.552455	5.650590	2248.119337
max	1415.000000	1869.140625	659.962092	617.031416	158800.412415	314335.229947	145377.965388	1282.377476	440351.421129

Σχήμα 2.2.12: Πίνακας στατιστικών των 9 χαρακτηριστικών για όλους τους χρήστες.

Με βάση τα **στατιστικά δεδομένα** του Σχήματος 2.2.12, για κάθε μεταβλητή, έχουμε τις ακόλουθες παρατηρήσεις:

- Η μεταβλητή Activity έχει μεγάλο αριθμό μηδενικών (76%). Αυτό δικαιολογείται επειδή ο άνθρωπος δεν κινείται στο μεγαλύτερο μέρος του ύπνου του.
- Η μέση τιμή των RR-Intervals (Mean NNI), έχει το 75% των τιμών κάτω του 1.47 δευτερολέπτων. Δηλαδή, για όλους τους χρήστες, η μέση τιμή μεταξύ δύο διαδοχικών καρδιακών παλμών, στο 75% ύπνου, είναι μικρότερη από 1.47 δευτερόλεπτα. Άρα, έχουμε διαισθητικά μια τιμή για τη χρονική απόσταση μεταξύ διαδοχικών καρδιακών παλμών.
- Από τις ισχύς των διαφορετικών συχνοτήτων VLF, LF και HF, παρατηρούμε ότι η VLF έχει 75% = 543.7, η LF έχει 75% = 1076.3 και η HF έχει 75% = 414.5. Άρα, και σε αυτό το σημείο αποδεικνύεται ότι τα RR-Intervals αποτελούνται κυρίως από χαμηλές συχνοτήτες.
- Τέλος, από τον λόγο LF/HF παρατηρούμε ότι η μέση τιμή είναι μεγαλύτερη του 1 και το 50% των τιμών είναι και αυτό μεγαλύτερο του 1. Αυτό σημαίνει ότι γενικά κατά μέσο όρο, η ισχύς των χαμηλών συχνοτήτων είναι μεγαλύτερη των υψηλών συχνοτήτων.

Στο τελικό στάδιο της ανάλυσης, θα παρατηρήσουμε την **συμπεριφορά των μεταβλητών/χαρακτηριστικών για κάθε στάδιο ύπνου (Wake, Light, Deep, REM)** και θα αναλύσουμε τα συμπεράσματα. Στα επόμενα Σχήματα (2.2.13 - 2.2.16) παραθέτουμε τους στατιστικούς πίνακες για κάθε στάδιο ύπνου.

	_Act	mean_nni	sdnn	sdsd	vlf	lf	hf	lf_hf_ratio	total_power
count	73237.000000	73237.000000	73237.000000	73237.000000	73237.000000	73237.000000	73237.000000	73237.000000	73237.000000
mean	37.217254	891.823349	57.121127	41.618374	1265.535618	3771.663747	1323.471648	8.092336	6360.688617
std	74.311246	161.956278	65.859513	50.671549	4562.695733	14628.786813	5170.108682	18.610451	21741.837345
min	0.000000	382.812500	0.000000	0.000000	0.000000	0.000000	0.000000	0.007913	0.000000
25%	0.000000	775.699013	19.176650	13.308736	37.334390	81.383783	30.963627	1.072039	207.672909
50%	0.000000	882.697610	34.476766	23.500526	155.757480	314.797604	96.086420	2.774244	697.336657
75%	39.000000	994.010417	65.755470	47.212867	660.701786	1295.640791	394.527410	7.587076	2658.608122
max	1415.000000	1869.140625	659.962092	617.031416	158800.412415	314335.229947	145377.965388	1282.377476	440351.421129

Σχήμα 2.2.13: Πίνακας στατιστικών των 9 χαρακτηριστικών για όλους τους χρήστες για το στάδιο **Wake**.

	_Act	mean_nni	sdnn	sdsd	vlf	lf	hf	lf_hf_ratio	total_power
count	103497.000000	103497.000000	103497.000000	103497.000000	103497.000000	103497.000000	103497.000000	103497.000000	103497.000000
mean	3.808603	982.905624	49.907954	39.776525	900.573120	2315.041423	790.150983	5.130789	4005.765526
std	21.451450	158.149255	49.502816	37.742556	3229.091295	8772.354854	2546.464834	12.022329	12981.583708
min	0.000000	420.623779	0.000000	0.000000	0.000000	0.000000	0.000000	0.004093	0.000000
25%	0.000000	879.250919	20.038910	16.277196	37.819790	81.723874	56.022313	0.733930	247.861007
50%	0.000000	976.562500	33.426363	26.853312	140.385523	273.370798	152.404490	1.909970	697.375142
75%	0.000000	1074.218750	59.220079	47.966372	536.866403	1074.714125	451.113259	4.854650	2259.688595
max	699.000000	1834.601151	496.112195	435.465723	129351.185837	187953.401210	68990.479225	695.410795	272754.074002

Σχήμα 2.2.14: Πίνακας στατιστικών των 9 χαρακτηριστικών για όλους τους χρήστες για το στάδιο **Light**.

	_Act	mean_nni	sdnn	sdsd	vlf	lf	hf	lf_hf_ratio	total_power
count	14641.000000	14641.000000	14641.000000	1.464100e+04	14641.000000	14641.000000	14641.000000	14641.000000	14641.000000
mean	2.212212	957.917406	36.318314	3.396524e+01	391.499955	1158.022107	513.655339	2.827213	2063.177401
std	18.607808	144.439972	36.133753	3.048905e+01	1878.298075	5697.739383	1723.712147	6.083085	8450.960230
min	0.000000	479.361979	1.029387	9.845569e-14	0.002853	0.067947	0.071564	0.008031	0.142363
25%	0.000000	856.026786	16.419937	1.531208e+01	18.593920	47.118807	53.518168	0.455417	172.903852
50%	0.000000	952.758789	25.733858	2.445875e+01	60.085501	136.036414	146.037216	1.134530	419.977212
75%	0.000000	1041.487069	41.493937	4.183854e+01	198.124038	452.727034	363.019438	2.847658	1103.038789
max	551.000000	1639.648438	437.617321	3.742048e+02	83855.297487	156189.263869	48245.475440	176.235706	231797.727839

Σχήμα 2.2.15: Πίνακας στατιστικών των 9 χαρακτηριστικών για όλους τους χρήστες για το στάδιο **Deep**.

	_Act	mean_nni	sdnn	sdsd	vlf	lf	hf	lf_hf_ratio	total_power
count	25944.000000	25944.000000	25944.000000	25944.000000	25944.000000	25944.000000	25944.000000	25944.000000	25944.000000
mean	2.545290	962.098331	48.368747	36.361980	842.965967	2140.181484	667.073188	6.258816	3650.220639
std	18.607269	149.425524	48.114801	37.191501	3127.862022	8590.199521	2359.632356	13.425654	12701.855069
min	0.000000	457.877604	0.000000	0.000000	0.000000	0.000000	0.000000	0.003509	0.000000
25%	0.000000	863.058036	19.260667	13.930912	40.812934	77.538852	36.162176	1.004976	209.967799
50%	0.000000	955.267137	32.591857	23.077496	149.586469	257.272243	101.317460	2.477791	605.251665
75%	0.000000	1049.973060	58.245201	43.901454	541.385521	956.414551	331.819267	6.112930	2021.650091
max	507.000000	1776.110197	626.408325	386.974058	86337.654831	196542.380273	73763.690894	396.703497	284314.730984

Σχήμα 2.2.16: Πίνακας στατιστικών των 9 χαρακτηριστικών για όλους τους χρήστες για το στάδιο **REM**.

Αρχικά, για τη μεταβλητή **Activity Counts** παρατηρούμε ότι μόνο το στάδιο Wake έχει υψηλές τιμές, δηλαδή μέση τιμή = 27.3, τυπική απόκλιση = 74.3 και μέγιστη τιμή 1415, ενώ στα υπόλοιπα στάδια, τα αντίστοιχα στατιστικά είναι πολύ χαμηλά και έχουν παραπλήσιες τιμές (Light, Deep, REM). Η παρατήρηση αυτή δικαιολογείται επειδή κατά το στάδιο Wake ο χρήστης προφανώς κινείται σε σχέση με τα υπόλοιπα στάδια όπου η κινητικότητα είναι εμφανώς μειωμένη. Άρα η μεταβλητή activity counts μπορεί να βοηθήσει στον διαχωρισμό του σταδίου Wake από τα υπόλοιπα.



Για τη μεταβλητή **Mean NNi** παρατηρούμε ότι η διαφορά μεταξύ των σταδίων ύπνου εμφανίζεται πιο έντονα στην τυπική απόκλιση όπου για τα στάδια Deep και REM είναι 144msec ενώ στα στάδια Wake και Light είναι 160msec. Δηλαδή κατά μέσο όρο τα RR-Intervals είναι πιο σταθερά γύρω από τη μέση τιμή στα στάδια Deep και Light σε σχέση με τα στάδια Wake και Light. Αλλά και πάλι δεν μπορούμε εποπτικά να βγάλουμε κάποιο ασφαλές συμπέρασμα διαχωρισμού των σταδίων ύπνου μέσω της μεταβλητής Mean NNi. Αντίστοιχο συμπέρασμα προκύπτει και για τη στατιστική συμπεριφορά των μεταβλητών **SDNN** και **SDSD**.

Για τη μεταβλητή **VLF** παρατηρούμε ότι η διαφορά μεταξύ των σταδίων ύπνου εμφανίζεται έντονα στην μέση τιμή και τυπική απόκλιση όπου για το στάδιο Deep είναι (391.5, 1878)  $ms^2$  ενώ στα στάδια Wake και Light και REM είναι (1265.5, 4562.7), (900.6, 3229) και (842.9, 3127.8)  $ms^2$ . Δηλαδή κατά το στάδιο Deep η ισχύς της πολύ χαμηλής συχνότητας είναι μικρότερη από την αντίστοιχη ισχύ της πολύ χαμηλής συχνότητας στα υπόλοιπα στάδια. Από αυτό συμπεραίνουμε ότι το στάδιο Deep δεν αποτελείται σε μεγάλο βαθμό από πολύ χαμηλές συχνότητες σε σχέση με τα υπόλοιπα στάδια. Αντίστοιχο συμπέρασμα προκύπτει και για τις μεταβλητές **LF**, **HF** και **LF/HF Ratio**. Όμως, μόνο από αυτές τις παρατηρήσεις δεν μπορούμε να βγάλουμε κάποιο ασφαλές συμπέρασμα εφόσον το σύνολο δεδομένων είναι μη ισορροπημένο οπότε δεν έχουμε μεγάλο εύρος διαφορετικών δεδομένων για την κλάση Deep αφού είναι η κλάση με το χαμηλότερο αριθμό δεδομένων που έχει η βάση MESA.

Για τη μεταβλητή **Total Power** παρατηρούμε ότι η διαφορά μεταξύ των σταδίων ύπνου εμφανίζεται σε όλα τα στατιστικά. Δηλαδή, για τα στάδια Wake και Light η συνολική ισχύς των RR-Intervals έχει υψηλότερες τιμές σε σχέση με την ισχύ των RR-Intervals των σταδίων Deep και REM. Δηλαδή κατά στάδια Wake και Light επειδή ο άνθρωπος έχει υψηλότερη κινητικότητα σε σχέση με τα στάδια Deep και Light, η αντίστοιχη ισχύς των σημάτων RR-Intervals είναι υψηλότερη.

Συνολικά από όλη την ανάλυση που προηγήθηκε, συμπεραίνουμε ότι οι παραπάνω μεταβλητές μπορούν να αποτελέσουν ικανοποιητικό σύνολο χαρακτηριστικών για την ταξινόμηση των δεδομένων στα 4 στάδια ύπνου διότι, έστω και σε διαισθητικό επίπεδο μπορούν να διαχωριστούν οι 4 κλάσεις μεταξύ τους σε έναν βαθμό.

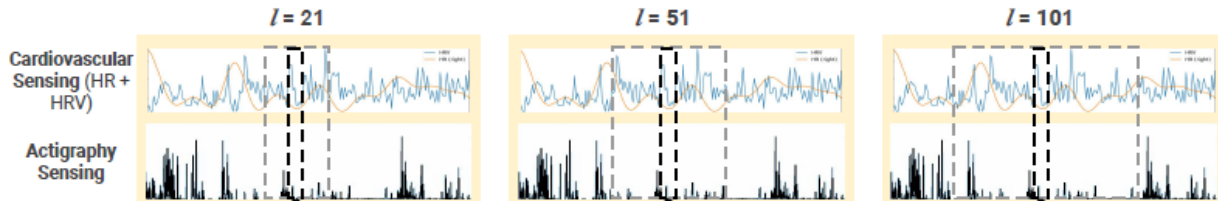
## 2.3 Ταξινόμηση δεδομένων με αλγόριθμους Βαθιάς Μάθησης (Deep Learning)

Στο κεφάλαιο αυτό θα αναπτύξουμε τη διαδικασία που ακολουθήθηκε για την ταξινόμηση των δεδομένων MESA στα 4 στάδια ύπνου (Wake/Light/Deep/REM) με την χρήση αλγορίθμων Deep Learning. Η μεθοδολογία έχει δύο στάδια: Σε πρώτο στάδιο στοχεύουμε στην επαλήθευση των αποτελεσμάτων που αναφέρονται στην έρευνα [1], και σε δεύτερο στάδιο θα προσπαθήσουμε να βελτιώσουμε τα αποτελέσματα των αλγορίθμων Deep Learning. Επειδή η έρευνα [1] αποτελεί πολύ πρόσφατη έρευνα στον τομέα του ύπνου και τη βάση δεδομένων MESA, η οποία χρησιμοποιείται ευρέως για την ταξινόμηση ύπνου, αυτός είναι και ο λόγος που θα προσπαθήσουμε να βελτιώσουμε τα αποτελέσματα της. Επίσης, η ταξινόμηση του ύπνου σε 4 στάδια (Wake/Light/Deep/REM), αποτελεί το σύνολο σταδίων ύπνου που χρησιμοποιείται ευρέως σε φορητές συσκευές όπως Apple Watch και Samsung Smartwatch [94].

Ξεκινώντας, όπως αναφέραμε στην προηγούμενη Ενότητα 2.2, για την ταξινόμηση, θα χρησιμοποιήσουμε δεδομένα ακτιγραφίας και HRV χαρακτηριστικά. Συνοπτικά, τα δεδομένα ακτιγραφίας και HRV χαρακτηριστικών που διαθέτει η βάση MESA είναι ανα 30 δευτερόλεπτα και κάθε ακολουθία 30 δευτερολέπτων αντιστοιχεί σε μια από τις 4 κλάσεις. Για την ταξινόμηση σε 4 στάδια, παραθέτουμε τον εξής **σχεδιασμό**. Αρχικά, χωρίζουμε το σύνολο δεδομένων ακτιγραφίας και HRV χαρακτηριστικών (1+8 μεταβλητές) σε σύνολο δεδομένων εκπαίδευσης και σε σύνολο δεδομένων αξιολόγησης (train και test) με αναλογία 80% (1395 χρήστες) και 20% (348 χρήστες) με βάση τον αριθμό Id των χρηστών. Η επι-



λογή των χρηστών για το σύνολο δεδομένων train και test γίνεται με τυχαίο τρόπο. Στη συνέχεια, προκειμένου οι αλγόριθμοι Deep Learning να ανιχνεύουν χαρακτηριστικά σε μεγαλύτερο διάστημα από 30 δευτερόλεπτα (30 δευτερόλεπτα = Διάσταση (1,9) = (1,1 ακτιγραφία + 8 HRV)) και να ανιχνεύουν χαρακτηριστικά για κάθε κλάση τα οποία εκτείνονται πέραν των 30 δευτερολέπτων, χωρίζουμε τα δεδομένα με βάση ένα "κινούμενο παράθυρο" (sliding window) (παραθύρωση δεδομένων). Αυτό είναι λογικό επειδή κάθε στάδιο έχει μέσο χρόνο πολύ περισσότερο από 30 δευτερόλεπτα. Τυπικές τιμές που χρησιμοποιούνται από το [1] για την παραθύρωση των δεδομένων είναι: 21,51,101 και αντιστοιχούν περίπου σε 10, 25 και 50 λεπτά. Άρα, προκύπτουν μεγάλα παράθυρα δεδομένων με διαστάσεις (21,9), (51,9) και (101,9), τα οποία περιλαμβάνουν χαρακτηριστικά για τα αντίστοιχα χρονικά διαστήματα που προκύπτουν από την παραθύρωση των δεδομένων, όπως φαίνεται στο Σχήμα 2.3.1. Τελικά, η τροφοδότηση των αλγόριθμων Deep Learning γίνεται με αυτή την μορφή των δεδομένων.



Σχήμα 2.3.1: Διαισθητική απεικόνιση της παραθύρωσης δεδομένων με μήκη 21, 51 και 101 (γκρι πλαίσιο) [1]. Η επικάλυψη (overlapping) μεταξύ διαδοχικών παραθύρων είναι 30 δευτερόλεπτα. Το μαύρο πλαίσιο μπορεί να αγνοηθεί. Τα μοντέλα δέχονται σαν είσοδο το γκρι πλαίσιο.

Ο σχεδιασμός για την ταξινόμηση εκτείνεται και στους αλγόριθμους Deep Learning πέραν της παραθύρωσης των δεδομένων, όπου χρησιμοποιούμε ένα **Συνελικτικό Νευρωνικό δίκτυο (CNN)**, ένα **Νευρωνικό Δίκτυο Μακροπρόθεσμης Μνήμης (LSTM)** και ένα συνδυαστικό μοντέλο των δύο παραπάνω (**Ensemble model**). Το τελευταίο μοντέλο (Ensemble model), έχει ως στόχο να πετύχει καλύτερα αποτελέσματα από τα CNN και LSTM, επειδή το καθένα από τα υπομοντέλα (CNN, LSTM) προσπαθεί να εξάγει διαφορετικά χαρακτηριστικά από τα δεδομένα για κάθε κλάση λόγω της διαφορετικής δομής τους. Επίσης, το μοντέλο Ensemble, εκπαιδεύεται με βάση δύο διαφορετικές τεχνικές αξιολόγησης (scoring): 1) Ταξινόμηση με βάση τη μέση τιμή της ύστερης (mean posterior) πιθανότητας πρόβλεψης των μοντέλων που αποτελείται (model-averaging), 2) Ταξινόμηση με βάση την μεγαλύτερη ύστερη (max posterior) πιθανότητα πρόβλεψης από όλα τα μοντέλα (maximum posterior selection). Πιο συγκεκριμένα, υποθέτοντας ότι έχουμε  $M$  αριθμό ταξινομητών/μοντέλων όπου ο κάθε ένας προβλέπει ένα διάνυσμα ύστερης πιθανότητας  $\mathbf{p}$  στον χώρο των  $K$  διαστάσεων ( $\mathbf{p} \in \mathbb{R}^K$ ), όπου  $K$  είναι ο αριθμός των κλάσεων ( $K=4$ ), τότε για το Ensemble μοντέλο και την **model-averaging μέθοδο** αξιολόγησης προκύπτει ότι:

$$\mathbf{p}^{mean}_t = \frac{1}{M} \sum_{m=1}^M \mathbf{p}^m_t, \mu \in \mathbf{p}^m_t$$

η ύστερη (posterior) πιθανότητα του μοντέλου  $m$ , και η προβλεπόμενη κλάση  $\hat{\mathbf{k}}$  του μοντέλου Ensemble, είναι αυτή με την μεγαλύτερη ύστερη πιθανότητα, δηλαδή:

$$\hat{\mathbf{k}} = \arg \max_k \mathbf{p}^{fusion}_t$$

Αντίστοιχα, για την μέθοδο αξιολόγησης **maximum posterior selection** προκύπτει ότι:

$$\hat{k} = \arg \max_k P^{fusion}_t, P^{fusion}_t = [p_t^1, p_t^2, p_t^3, p_t^4], \mu \in p_t^k$$

η ύστερη (posterior) πιθανότητα της κλάσης k.

Για την αξιολόγηση όλων των μοντέλων θα χρησιμοποιήσουμε τις παρακάτω δημοφιλείς μετρικές ταξινόμησης, με βάση τα δεδομένα που ταξινομήθηκαν στη σωστή κλάση (True Positives (TP) και True Negatives (TN)) και τα δεδομένα που ταξινομήθηκαν λάθος (False Positives (FP) και False Negatives (FN)), όπου Positive και Negative ορίζονται ως δύο κλάσεις και οι παρακάτω μετρικές μπορούν να επεκταθούν και σε περισσότερες. Οπότε, έχουμε τα εξής:

- **Accuracy:** Ο αριθμός των σωστά ταξινομημένων κλάσεων, προς τον συνολικό αριθμό δεδομένων

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

- **Recall:** Το ποσοστό των σωστά ταξινομημένων δεδομένων, δηλαδή των δεδομένων που ανιχνεύθηκαν σωστά από το μοντέλο.

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

- **Specificity:** Ποσοστό True Negative Rate. Δηλαδή το ποσοστό των σωστά ταξινομημένων δεδομένων της κλάσης Negative.

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

- **Precision:** Ο λόγος των σωστά ταξινομημένων δεδομένων της κλάσης Positive, προς το σύνολο των δεδομένων που ταξινομήθηκαν είτε σωστά είτε λάθος στη κλάση Positive.

$$Precision = \frac{TP}{TP + FP}$$

- **F1-Score:** Αρμονικός μέσος της μετρικής Precision και Recall. Πολύ χρήσιμος για μη ισορροπημένο σύνολο δεδομένων (imbalance dataset).

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

- **Cohen-Kappa:** Υπολογίζει την "συμφωνία" (agreement) πρόβλεψης/ταξινόμησης μεταξύ της προβλεπόμενης συμφωνίας (predicted agreement) και της πραγματικής/παρατηρούμενης συμφωνίας (observed agreement). Δηλαδή:

$$\kappa = \frac{P_0 - P_e}{1 - P_e}$$

όπου  $P_0$  το πραγματικό/παρατηρούμενο ποσοστό συμφωνίας (observed agreement) και  $P_e$  το αναμενόμενο ποσοστό συμφωνίας (predicted agreement). Πρακτικά, η Cohen-Kappa ( $\kappa$ ) μετρική, υπολογίζει το ποσοστό τυχαίας ταξινόμησης δεδομένων που αναδύεται από το μη ισορροπημένο σύνολο δεδομένων (imbalance dataset). Όσο μεγαλύτερο το ποσοστό %, τόσο μικρότερη είναι η πιθανότητα ένα μοντέλο να έχει ταξινομήσει σωστά τις κλάσεις κατά τύχη. Οπότε επιθυμούμε ένα μοντέλο να έχει ταξινομήσει τις κλάσεις συνειδητά σωστά και όχι κατά τύχη. (Εύρος [0,100]%).

Επίσης, για την αξιολόγηση των μοντέλων ταξινόμησης χρησιμοποιούνται ο Πίνακας Σύγχυσης (Confusion Matrix) και η χρονική απόκλιση ενός σταδίου ύπνου σε λεπτά (Time Deviation Metric). Ένας πίνακας σύγχυσης είναι μια σύνοψη των αποτελεσμάτων πρόβλεψης σε ένα πρόβλημα ταξινόμησης. Ο αριθμός των σωστών και εσφαλμένων προβλέψεων συνοψίζεται με τιμές των μετρικών (Recall κλπ.) και αναλύεται ανά κλάση. Η Time Deviation μετρική, ορίζεται ως εξής: Έχοντας σύνολο  $N$  δεδομένων (30 δευτερολέπτων το καθένα) της κλάσης  $k$ , ορίζεται η χρονική απόκλιση (σε λεπτά) της κλάσης  $k$  ως,

$$TD_k = \frac{1}{N} \sum_{i=1}^N (Pred_k^i - True_k^i)$$

όπου  $Pred_k$  είναι η πρόβλεψη του μοντέλου για την κλάση  $k$  και  $True_k$  είναι η πραγματική ετικέτα (label) της κλάσης  $k$ . Οπότε μέσω της μετρικής  $TD_k$  μπορούμε να αξιολογήσουμε για πόσο χρόνο (λεπτά) ένα μοντέλο υπερεκτιμά ή υποεκτιμά (overestimate, underestimate) τη κλάση  $k$ . **Η μετρική αυτή είναι πολύ σημαντική για τους γιατρούς οι οποίοι έχουν την εξειδίκευση για την αξιολόγηση του χρόνου παραμονής του ανθρώπου σε κάθε στάδιο ύπνου.**

Πολλές μελέτες ταξινόμησης ύπνου χρησιμοποίησαν την ακρίβεια (accuracy) ή το F1-Score για να μετρήσουν την απόδοση του μοντέλου τους, ωστόσο αυτές είναι μετρήσεις υψηλού επιπέδου (high-level) που δεν λαμβάνουν υπόψη την απόδοση ανά κλάση. Οι πίνακες σύγχυσης, από την άλλη πλευρά, παρέχουν προβλέψεις ανά κλάση και αντίστοιχους τύπους μετρικών (Recall, Precision κλπ.). Ωστόσο, για τους κλινικούς ιατρούς και άλλους επαγγελματίες υγείας, αυτοί οι πίνακες δεν είναι ο πιο προφανής τρόπος για να αναπαραστήσουν τη χρονική απόκλιση των σταδίων ύπνου, επειδή περιλαμβάνουν πολλές λεπτομέρειες χαμηλού επιπέδου (low-level). Η προτεινόμενη χρονική απόκλιση είναι μια μέτρηση μεσαίου επιπέδου (mid-level), η οποία συνοψίζει την απόδοση ταξινόμησης με διαισθητικό τρόπο (intuitive). Ως εκ τούτου, μπορεί να χρησιμοποιηθεί ως συμπληρωματική μετρική σε αυτό που προσφέρεται από παραδοσιακές μεθόδους αξιολόγησης (πίνακας σύγχυσης, accuracy και F1-Score), που επιτρέπουν στους επαγγελματίες υγείας να αποκτήσουν μια διαισθητική κατανόηση της αξιοπιστίας ενός μοντέλου.

Στη επόμενη ενότητα θα αναλυθούν οι πειραματισμοί που πραγματοποιήθηκαν για το πρόβλημα ταξινόμησης των 4 σταδίων ύπνου με τους αλγορίθμους Deep Learning.

### 2.3.1 Προετοιμασία Δεδομένων

Στο κεφάλαιο αυτό θα παραθέσουμε τη διαδικασία τροφοδότησης των μοντέλων Deep Learning που αναφέρθηκε στην εισαγωγή της Ενότητας 2.3. Συγκεκριμένα, ο **πυρήνας** της διαδικασίας τροφοδότησης των μοντέλων Deep Learning είναι η παραθύρωση των δεδομένων (sliding window), καθώς κάθε στάδιο ύπνου έχει διάρκεια μεγαλύτερη των 30 δευτερολέπτων<sup>3</sup>, οπότε με αυτόν τον τρόπο θα μπορούν τα μοντέλα Βαθιάς Μάθησης (Deep Learning) να ανιχνεύσουν και να μάθουν χαρακτηριστικά (features) μακράς εμβέλειας (long-term). Με τον όρο μακράς εμβέλειας εννοούμε ότι αφού τα στάδια ύπνου διαρκούν από ένα λεπτό μέχρι και μέχρι και μία ώρα, τότε τα μοντέλα Βαθιάς Μάθησης θα είναι σε θέση να "μάθουν" την φυσιολογία των σταδίων ύπνου με την εξαγωγή χαρακτηριστικών (features). Οπότε αναμένουμε η παραθύρωση να παίζει σημαντικό ρόλο στην καλή απόδοση των μοντέλων Βαθιάς Μάθησης.

Επομένως, για την προετοιμασία των δεδομένων όπως αναφέραμε και στην αρχή της Ενότητας 2.3, χωρίζουμε το σύνολο δεδομένων ακτιγραφίας και HRV χαρακτηριστικών (1+8 μεταβλητές) σε σύνολο δεδομένων εκπαίδευσης και σε σύνολο δεδομένων αξιολόγησης (train και test) με **αναλογία 80% (1395 χρήστες) και 20% (348 χρήστες)** με βάση τον αριθμό Id των χρηστών. Η επιλογή των χρηστών για το σύνολο δεδομένων train και test γίνεται με τυχαίο τρόπο. Για την αξιολόγηση των

---

<sup>3</sup>Physiology, Sleep Stages: <https://www.ncbi.nlm.nih.gov/books/NBK526132/>

μοντέλων βαθιάς Μάθησης κατά τη διάρκεια της εκπαίδευσής τους θα χρειαστούμε και ένα σύνολο επιβεβαίωσης (validation). Οπότε, χρησιμοποιούμε το 20% του συνόλου δεδομένων εκπαίδευσης ως σύνολο επιβεβαίωσης (validation). Στη συνέχεια, **κανονικοποιούμε** τα δεδομένα με μέση τιμή (mean) μηδέν 0 και τυπική απόκλιση (std) 1. Για να αποφύγουμε την **διαρροή πληροφορίας (data leakage)** από το σύνολο δεδομένων εκπαίδευσης στο σύνολο δεδομένων αξιολόγησης, κανονικοποιούμε μόνο το σύνολο δεδομένων εκπαίδευσης (train) και απλά μετά εφαρμόζουμε την τεχνική κανονικοποίησης (μέση τιμή και τυπική απόκλιση) στο σύνολο δεδομένων αξιολόγησης. Με αυτόν τον τρόπο δεν διαρρέεται πληροφορία για την αξιολόγηση του συνόλου test και η αξιολόγηση των μοντέλων Βαθιάς Μάθησης δεν θα είναι "προκατειλημμένη" (Biased).

Επακόλουθα, μετατρέπουμε τις ετικέτες (labels) των δεδομένων σε μορφή OneHotEncoding<sup>4</sup>. Δηλαδή κάθε κλάση με αριθμό  $i$ ,  $i \in [1, 4]$  μετατρέπεται σε δυαδική μορφή  $[0,1]$ . Για παράδειγμα, η κλάση Wake αντιστοιχεί σε ετικέτα (label) 0, τότε η δυαδική μορφή της κλάσης είναι  $[1, 0, 0, 0]$ <sup>5</sup>. Επομένως, αφού έχουμε χωρίσει το σύνολο δεδομένων σε δεδομένα εκπαίδευσης (train), δεδομένα αξιολόγησης (test) και επιβεβαίωσης (validation) καταλήγουμε στα εξής:

- Διαστάσεις συνόλου  $X_{train}$  : **(1215509, k, 9)**
- Διαστάσεις συνόλου  $X_{val}$  : **(384513, k, 9)**
- Διαστάσεις συνόλου  $X_{test}$  : **(303878, k, 9)**
- Διαστάσεις συνόλου (labels) κλάσεων  $y_{train}$  : **(1215509, 4)**
- Διαστάσεις συνόλου (labels) κλάσεων  $y_{val}$  : **(384513, 4)**
- Διαστάσεις συνόλου (labels) κλάσεων  $y_{test}$  : **(303878, 4)**

όπου  $k$ , το μήκος παραθύρωσης **(21, 51, 101)**.

Στα παρακάτω Κεφάλαια θα αναλυθούν τα πειράματα ταξινόμησης των δεδομένων στα 4 στάδια ύπνου (Wake, Light, Deep, REM). Υπενθυμίζουμε τα μοντέλα Βαθιάς Μάθησης που χρησιμοποιήθηκαν:

- Συνελικτικό Νευρωνικό Δίκτυο (CNN)
- Νευρωνικό Δίκτυο Μακροπρόθεσμης Μνήμης (LSTM)
- Μοντέλο Ensemble (Ensemble Model)

Σε πρώτο στάδιο, η όλη διαδικασία που ακολουθήθηκε βασίζεται στην έρευνα [1]. Δηλαδή, η προετοιμασία των δεδομένων για μοντελοποίηση και τα καλύτερα μοντέλα Βαθιάς Μάθησης προέρχονται από την προαναφερθείσα έρευνα για την επαλήθευση των αποτελεσμάτων.

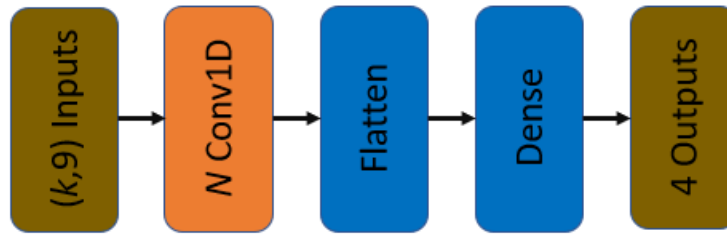
### 2.3.2 Ταξινόμηση με Συνελικτικό Νευρωνικό Δίκτυο (CNN)

Το πρώτο μέρος του πειραματικού σταδίου αφορά την ταξινόμηση των δεδομένων στα 4 στάδια ύπνου με τη χρήση ενός Συνελικτικού Νευρωνικού Δικτύου. Συγκεκριμένα, η αρχιτεκτονική του δικτύου παρουσιάζεται στο Σχήμα 2.3.2.

Από την αρχιτεκτονική του μοντέλου έχουμε ότι, το CNN μοντέλο δέχεται ως είσοδο δεδομένα με διαστάσεις **(k, 9)**. Η πρώτη διάσταση (αριθμός k) απεικονίζει το μήκος παραθύρωσης (sliding window) των δεδομένων και η δεύτερη διάσταση τα  $8+1 = 9$  (ακτιγραφία, HRV) χαρακτηριστικά. Το μήκος παραθύρωσης όπως αναφέραμε στο προηγούμενο Κεφάλαιο 3.3, παίρνει τιμές (21, 51, 101). Η έξοδος

<sup>4</sup>OneHotEncoder: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html>

<sup>5</sup>Wake:0, Light:1, Deep:2, REM:4



Σχήμα 2.3.2: Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου. Μεταβλητή των μοντέλων αποτελεί το μήκος παραθύρωσης (sliding window) και αναφέρεται ως  $k$ ,  $k \in (21, 51, 101)$ .

του μοντέλου έχει διάσταση (4,1) και απεικονίζει τα 4 στάδια του ύπνου. Το διάνυσμα αυτό της εξόδου είναι σε δυαδική μορφή (One Hot Encoding).

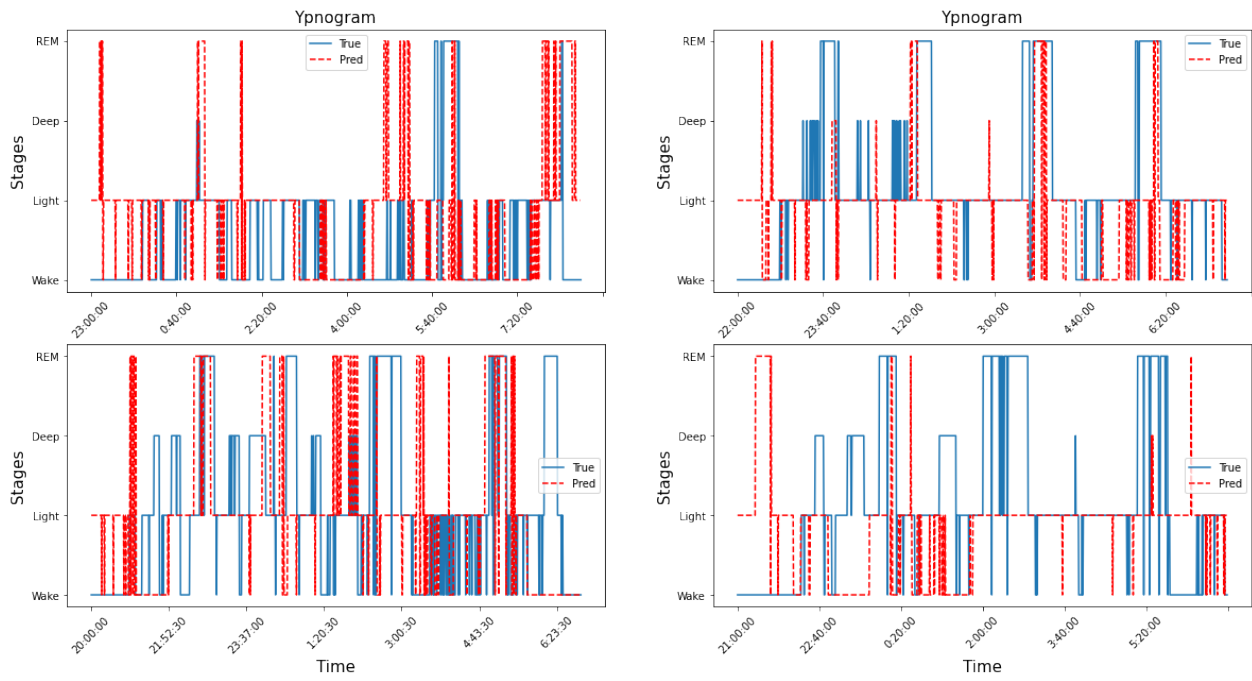
Λεπτομερέστερα, τα φίλτρα (filters/kernels) του επιπέδου συνέλιξης (1D Convolutional Layer) είναι  $N = 64$  σε πλήθος και έχουν μέγεθος 3 (kernel size) το κάθε ένα και stride ίσο με 1. Μετέπειτα, εφόσον το συνελικτικό επίπεδο έχει εξάγει χαρακτηριστικά από τα δεδομένα, τα τελευταία "περνούν" από την συνάρτηση ReLU ( $Relu(x) = \max(\mathbf{0}, x)$ ), ακολουθεί η επιπέδωση των δεδομένων (Flatten) και η έξοδός τους με 4 νευρώνες, που αποτελούν τα 4 στάδια ύπνου. Πρέπει να αναφερθεί ότι κατά την έξοδο (επίπεδο Dense = 4 νευρώνες), το επίπεδο Dense επιστρέφει πιθανότητες σύμφωνα με την συνάρτηση Softmax.  $\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^4 e^{z_j}}$   $i = 1, 2, 3, 4$ . Δηλαδή, στην έξοδο του Dense επιπέδου εφαρμόζεται η συνάρτηση Softmax, η οποία μετατρέπει την είσοδο που δέχεται σε κανονικοποιημένη πιθανότητα.

Μετά τον πειραματισμό (και σύμφωνα με το [1]), προκύπτουν 3 μοντέλα, ένα για κάθε μήκος παραθύρωσης. Δηλαδή,  $CNN_{21}$ ,  $CNN_{51}$ ,  $CNN_{101}$ . Η εκπαίδευση των μοντέλων γίνεται για 100 εποχές, χρησιμοποιώντας batch size ίσο με 512. Επίσης, όπως αναφέραμε και στο προηγούμενο Κεφάλαιο (3.3), το 20% των δεδομένων εκπαίδευσης χρησιμοποιείται για validation ώστε να αξιολογηθεί η απόδοση του μοντέλου κατά την εκπαίδευση. Τέλος, για την εκπαίδευση και βελτιστοποίηση των βαρών του δικτύου (weights), χρησιμοποιείται ο αλγόριθμος RMSProp<sup>6</sup>. Σύμφωνα με τα παραπάνω γίνεται η εκπαίδευση των CNN μοντέλων στο σύνολο δεδομένων εκπαίδευσης (train), και μετά γίνεται η πρόβλεψη/ταξινόμηση στο σύνολο δεδομένων αξιολόγησης (test). Στο Σχήμα 2.3.3 - 2.3.5 παρατίθενται τα υπνογράμματα τεσσάρων χρηστών του συνόλου αξιολόγησης (test), δηλαδή διαγράμματα πρόβλεψης/ταξινόμησης των σταδίων του ύπνου τεσσάρων χρηστών, στα τέσσερα στάδια Wake, Light, Deep και REM. Επιπλέον, τα αποτελέσματα της ταξινόμησης για τα διαφορετικά μήκη παραθύρωσης παρατίθενται στον Πίνακα 2.2 με τις κλασικές δημοφιλείς μετρικές. Τέλος, στους Πίνακες 2.3 - 2.5 παρουσιάζονται τα αποτελέσματα ταξινόμησης για κάθε στάδιο ύπνου (Wake, Light, Deep και REM) για τα τρία διαφορετικά μήκη παραθύρωσης και στα Σχήματα 2.3.6 - 2.3.8 παρουσιάζονται οι Πίνακες Σύγχυσης για τα διαφορετικά μήκη παραθύρωσης.

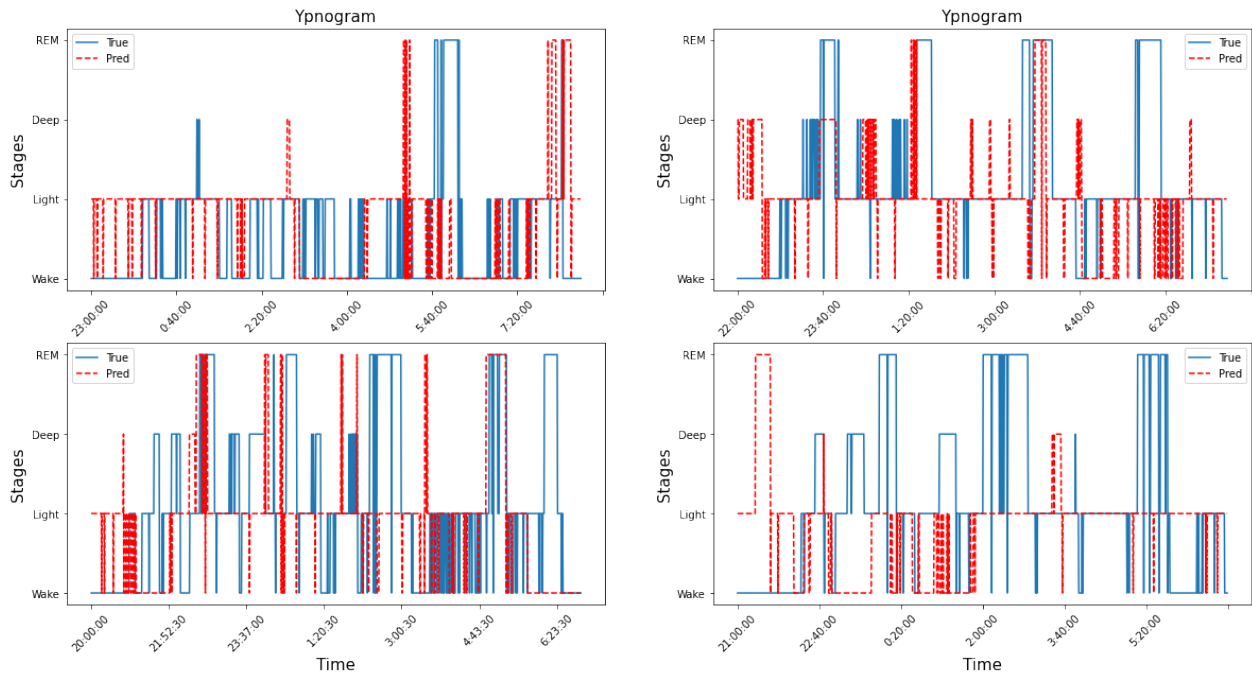
Εποπτικά, από τα υπνογράμματα του Σχήματος 2.3.3 - 2.3.5, παρατηρούμε ότι η κλάση Deep δεν μπορεί να αναγνωριστεί καλά από το συνελικτικό δίκτυο. Συγκεκριμένα, επειδή η κλάση Light αναγνωρίζεται πολύ καλά από το μοντέλο, φαίνεται ότι η κλάση Deep αναγνωρίζεται ως κλάση Light λανθασμένα (misclassification). Οι υπόλοιπες κλάσεις, κυρίως οι κλάσεις Wake και Light, μπορούν να αναγνωριστούν καλύτερα (αφού επικαλύπτονται οι δύο γραμμές True και Predictions). Το συμπέρασμα αυτό μπορεί να διαγνωσθεί και από τους Πίνακες μετρικών του συνόλου αξιολόγησης (Πίνακες 2.3 - 2.5) και από τους Πίνακες σύγχυσης (Confusion Matrix - Σχήματα 2.3.6 - 2.3.8) όπου η κλάση Deep έχει τις χειρότερες αποδόσεις σε όλες τις μετρικές, ακολουθεί με λίγο καλύτερες αποδόσεις η κλάση REM, μετά η κλάση

<sup>6</sup>RMSProp Optimizer: <https://keras.io/api/optimizers/rmsprop/>



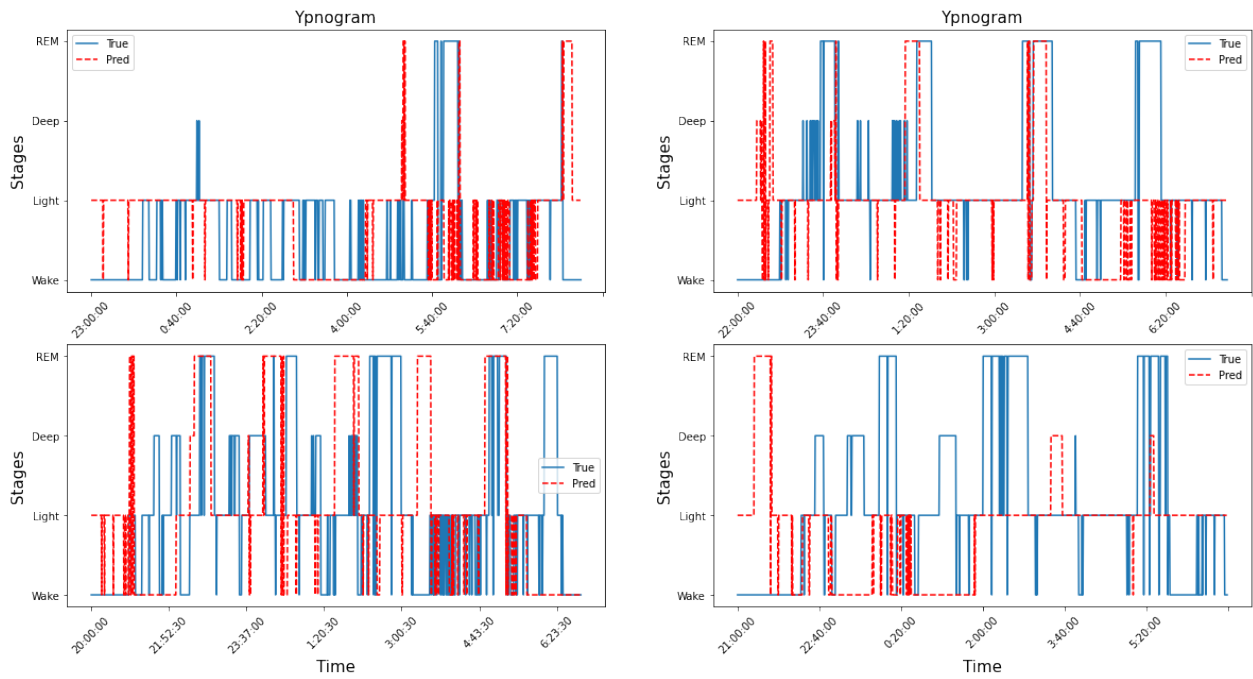


Σχήμα 2.3.3: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 21.



Σχήμα 2.3.4: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 51.

Wake, και τέλος τις υψηλότερες αποδόσεις τις έχει η κλάση Light.



Σχήμα 2.3.5: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 101.

<i>Metrics%</i>	<i>SeqLen21</i>	<i>SeqLen51</i>	<i>SeqLen101</i>
Accuracy	68.9 ± 1.0	70.1 ± 1.0	70.4 ± 1.1
Precision	54.8 ± 1.2	61.4 ± 1.5	61.4 ± 1.5
Recall	51.7 ± 0.9	51.4 ± 1.0	51.4 ± 1.0
F1-Score	49.5 ± 0.9	50.5 ± 1.1	50.5 ± 1.1
Cohen-Kappa	50.3 ± 1.7	50.3 ± 1.8	53.4 ± 1.9
Specificity	86.6 ± 0.4	86.7 ± 0.4	86.7 ± 0.4

Πίνακας 2.2: Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης.

Λεπτομερέστερα, από τον Πίνακα μετρικών 2.2, συμπεραίνουμε ότι τα μοντέλα  $CNN_{51}$  και  $CNN_{101}$  έχουν τις υψηλότερες αποδόσεις σε όλες τις μετρικές. Η μεγάλη διαφορά αντικατοπτρίζεται στη μετρική Precision όπου τα μοντέλα  $CNN_{51}$  και  $CNN_{101}$  έχουν καλύτερα ποσοστά σε σχέση με το μοντέλο  $CNN_{21}$  (απόλυτη διαφορά 7%). Δηλαδή, τα μοντέλα  $CNN_{51}$  και  $CNN_{101}$  ταξινομούν περισσότερα δεδομένα σωστά. Λόγω του μη ισορροπημένου συνόλου δεδομένων, η μετρική Accuracy δεν είναι αξιόπιστη για την αξιολόγηση των μοντέλων. Επίσης, παρατηρούμε ότι η μετρική Specificity είναι υψηλή σε όλα τα μοντέλα CNN. Το γεγονός αυτό καταδεικνύει ότι είναι μικρό το ποσοστό των λάθος ταξινομημένων δεδομένων, γεγονός που είναι σημαντικό. Αντίθετα, η μετρική F1-Score και Cohen-Kappa είναι σχετικά χαμηλή αλλά αποκαλύπτει την πραγματική απόδοση των μοντέλων εξαιτίας του μη ισορροπημένου συνόλου δεδομένων (imbalance dataset). Οπότε, για αυτό το λόγο η μετρική Specificity είναι υψηλή, επειδή η κλάση Light έχει το μεγαλύτερο ποσοστό δεδομένων συνολικά και αναγνωρίζεται καλά από τα μοντέλα. οπότε δεν φαίνεται η δυσκολία των μοντέλων να αναγνωρίσουν καλά τις κλάσεις με το χαμηλότερο πλήθος δεδομένων, δηλαδή Deep και REM.

Εξαιτίας αυτού του φαινομένου, παρατηρούμε τους Πίνακες Σύγκρισης 2.3 - 2.5 όπου παρουσιάζονται

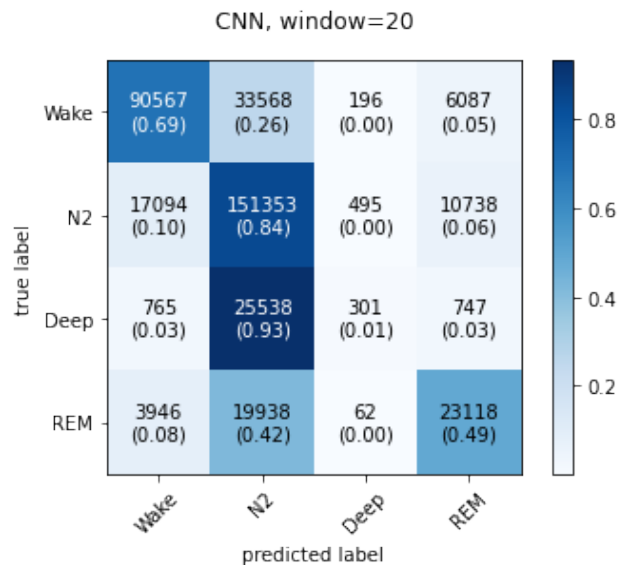


<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	81.0	69.0	75.0
Light	66.0	84.0	74.0
Deep	29.0	1.0	2.0
REM	57.0	49.0	53.0

Πίνακας 2.3: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 21.

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	84.0	69.0	76.0
Light	65.0	89.0	75.0
Deep	47.0	9.0	15.0
REM	69.0	38.0	49.0

Πίνακας 2.4: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 51.

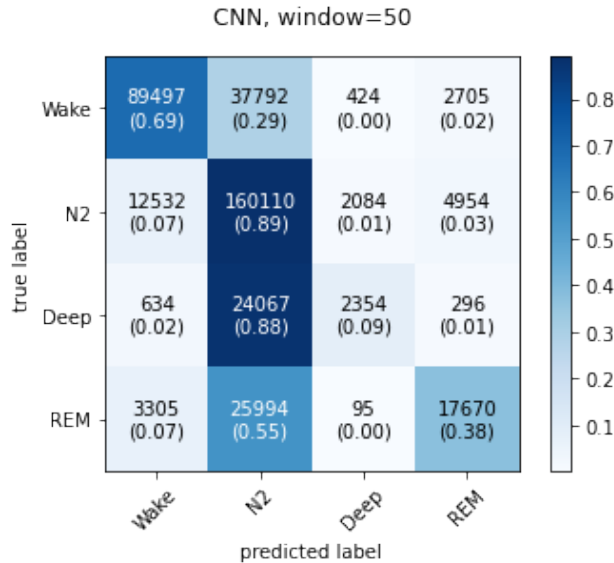


Σχήμα 2.3.6: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 21.

οι αποδόσεις των μοντέλων ανά κλάση. Είναι φανερό ότι τα μοντέλα, λόγω του μη ισορροπημένου συνόλου δεδομένων (imbalance dataset), δεν μπορούν να αναγνωρίσουν σχεδόν καθόλου την κλάση Deep (F1-Score < 20 %) και ακολουθεί η κλάση REM. Επίσης, παρατηρούμε ότι τα μοντέλα  $CNN_{51}$  και  $CNN_{101}$  είναι καλύτερα σε όλες τις μετρικές ανά κλάση, σε σχέση με το μοντέλο  $CNN_{21}$ . Αν εστιάσουμε στην κλάση REM, Deep και Wake, παρατηρούμε ότι το μοντέλο  $CNN_{101}$  έχει τη καλύτερη απόδοση F1-Score και Recall από όλα τα μοντέλα. Δηλαδή, η μεγαλύτερη παραθύρωση βοηθά περισσότερο τις κλάσεις να αναγνωριστούν. Αυτό αποκαλύπτει ότι για να αναγνωριστούν καλά τα στάδια του ύπνου, πρέπει τα μοντέλα να επεξεργάζονται μεγαλύτερα παράθυρα δεδομένων (sliding windows) ώστε να αναγνωρίσουν τα long-term χαρακτηριστικά που διαχωρίζουν τα στάδια του ύπνου.

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	76.0	78.0	77.0
Light	69.0	81.0	74.0
Deep	45.0	5.0	9.0
REM	65.0	51.0	57.0

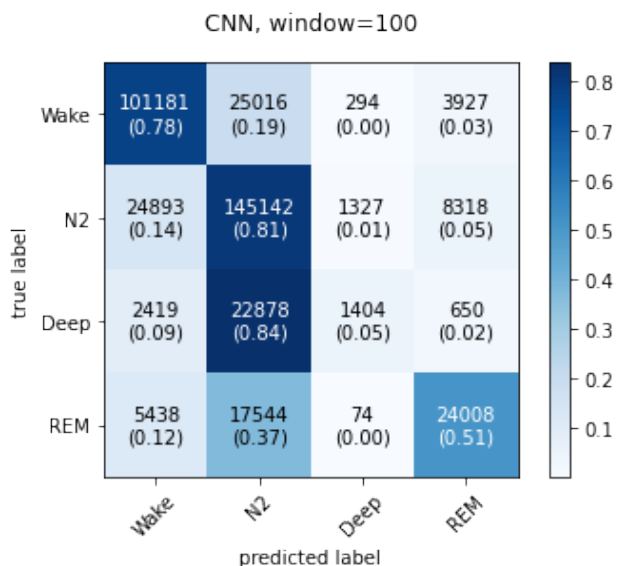
Πίνακας 2.5: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 101.



Σχήμα 2.3.7: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 51.

Τα παραπάνω συμπεράσματα συνοψίζονται στους Πίνακες Σύγχυσης 2.3.6 - 2.3.8. Δηλαδή, όλα τα μοντέλα δυσκολεύονται να αναγνωρίσουν τις κλάσεις Deep και REM και αναγνωρίζουν πολύ καλά τις κλάσεις Wake και Light, παρατηρώντας τα ποσοστά Recall της διαγωνίου. Το γεγονός αυτό είναι λογικό εφόσον το σύνολο δεδομένων είναι μη ισορροπημένο. Οπότε, περισσότερα δεδομένα από μια κλάση, διαφορετικά μεταξύ τους, τόσο καλύτερα τα μοντέλα θα την αναγνωρίσουν.

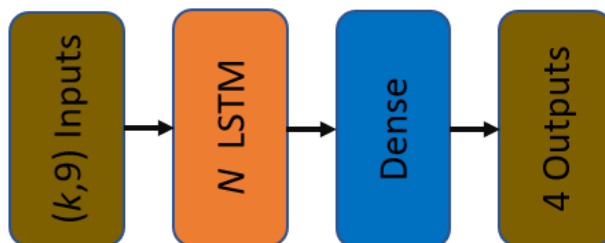
Στην επόμενη ενότητα, θα παρουσιάσουμε τα μοντέλα LSTM και θα παραθέσουμε τα αντίστοιχα αποτελέσματα.



Σχήμα 2.3.8: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 101.

### 2.3.3 Ταξινόμηση με Νευρωνικό Δίκτυο Μακροπρόθεσμης Μνήμης (LSTM)

Στη παρούσα ενότητα θα παρουσιάσουμε τα πειράματα που πραγματοποιήθηκαν για την ταξινόμηση των δεδομένων στα 4 στάδια ύπνου με τη χρήση ενός Νευρωνικού Δικτύου Μακροπρόθεσμης Μνήμης (LSTM). Συγκεκριμένα, η αρχιτεκτονική του δικτύου παρουσιάζεται στο Σχήμα 2.3.9.



Σχήμα 2.3.9: Αρχιτεκτονική Νευρωνικού Δικτύου Μακροπρόθεσμης Μνήμης. Μεταβλητή των μοντέλων αποτελεί το μήκος παραθύρωσης (sliding window) και αναφέρεται ως  $k$ ,  $k \in (21, 51, 101)$ .

Από την αρχιτεκτονική του μοντέλου έχουμε ότι, το LSTM μοντέλο δέχεται ως είσοδο δεδομένα με διαστάσεις  $(k, 9)$  όπως και στην περίπτωση του Συνελικτικού Νευρωνικού Δικτύου. Η πρώτη διάσταση (αριθμός  $k$ ) απεικονίζει το μήκος παραθύρωσης (sliding window) των δεδομένων και η δεύτερη διάσταση τα  $8+1 = 9$  (ακτιγραφία, HRV) χαρακτηριστικά. Το μήκος παραθύρωσης όπως αναφέραμε στο προηγούμενο Κεφάλαιο 2.3, παίρνει τιμές  $(21, 51, 101)$ . Η έξοδος του μοντέλου έχει διάσταση  $(4,1)$  και απεικονίζει τα 4 στάδια του ύπνου. Το διάνυσμα αυτό της εξόδου είναι σε δυαδική μορφή (One Hot Encoding).

Λεπτομερέστερα, το μέγεθος των κρυμμένων μονάδων (hidden units  $h_t$ ) του επιπέδου μακροπρόθεσμης μνήμης (LSTM Layer) είναι  $N = 32$ . Μετέπειτα, εφόσον το επίπεδο μακροπρόθεσμης μνήμης έχει

εξαγάγει χαρακτηριστικά από τα δεδομένα εισόδου, η έξοδος μόνο της τελευταίας κρυμμένης μονάδας  $h_t$  "περνά" από την συνάρτηση υπερβολικής εφαπτομένης  $\tanh(x)$ , και ακολουθεί η έξοδος με 4 νευρώνες, που αποτελούν τα 4 στάδια ύπνου. Πρέπει να αναφερθεί ότι κατά την έξοδο (επίπεδο Dense = 4 νευρώνες), το επίπεδο Dense επιστρέφει πιθανότητες σύμφωνα με την συνάρτηση Softmax  $\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^4 e^{z_j}}$   $i = 1, 2, 3, 4$ . Δηλαδή, στην έξοδο του Dense επιπέδου εφαρμόζεται η συνάρτηση Softmax, η οποία μετατρέπει την είσοδο που δέχεται σε κανονικοποιημένη πιθανότητα. Επίσης, σημαντικό να αναφερθεί είναι ότι ο αριθμός των επιπέδων μακροπρόθεσμης μνήμης (LSTM Layer) είναι 1 σύμφωνα με το [1]. Δηλαδή, κάθε επίπεδο μακροπρόθεσμης μνήμης επιστρέφει τις εξόδους από όλες τις κρυμμένες μονάδες του (hidden units  $h_t$ ) στο επόμενο επίπεδο μακροπρόθεσμης μνήμης (LSTM Layer), εκτός από το τελευταίο επίπεδο μακροπρόθεσμης μνήμης που επιστρέφει **μόνο την έξοδο της τελευταίας κρυμμένης μονάδας του  $h_t$**  και περνά στην έξοδο Dense.

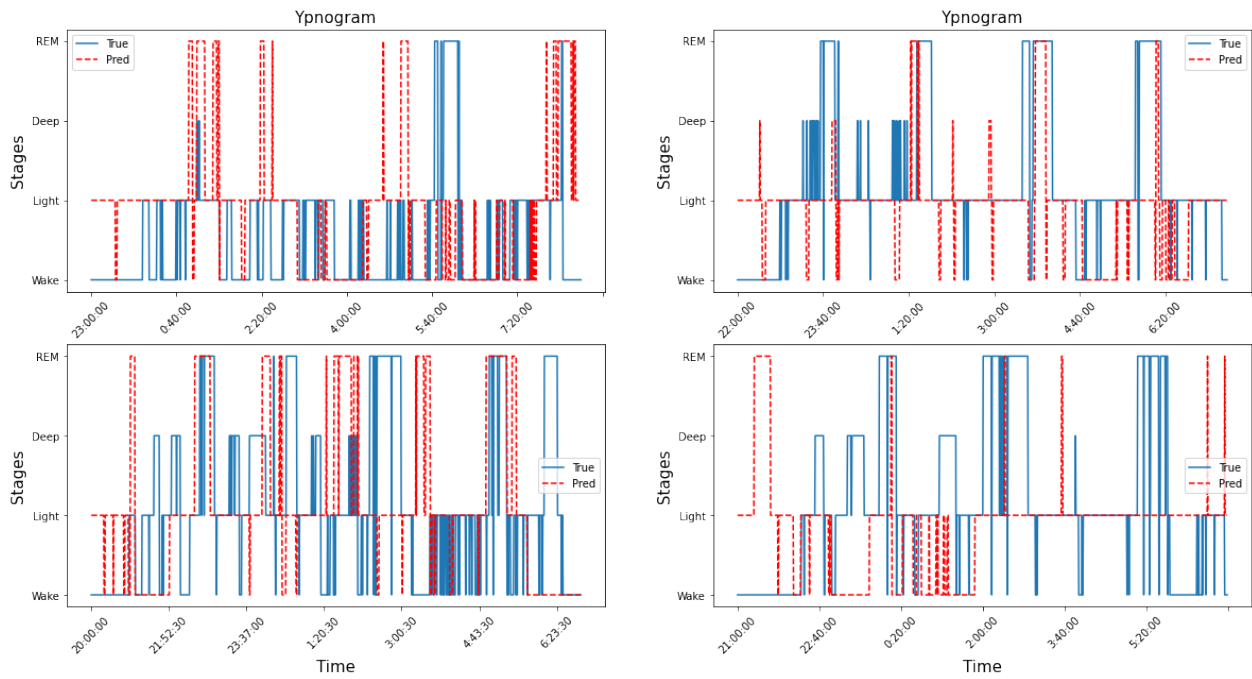
Μετά τον πειραματισμό (και σύμφωνα με το [1]), προκύπτουν 3 μοντέλα, ένα για κάθε μήκος παραθύρωσης. Δηλαδή,  $LSTM_{21}$ ,  $LSTM_{51}$ ,  $LSTM_{101}$ . Η εκπαίδευση των μοντέλων γίνεται για 100 εποχές, χρησιμοποιώντας batch size ίσο με 512. Επίσης, όπως αναφέραμε και στο προηγούμενο Κεφάλαιο (2.3), το 20% των δεδομένων εκπαίδευσης χρησιμοποιείται για validation ώστε να αξιολογηθεί η απόδοση του μοντέλου κατά την εκπαίδευση. Τέλος, για την εκπαίδευση και βελτιστοποίηση των βαρών του δικτύου (weights), χρησιμοποιείται ο αλγόριθμος RMSProp<sup>7</sup>. Σύμφωνα με τα παραπάνω γίνεται η εκπαίδευση των LSTM μοντέλων στο σύνολο δεδομένων εκπαίδευσης (train), και μετά γίνεται η πρόβλεψη/ταξινόμηση στο σύνολο δεδομένων αξιολόγησης (test). Στα Σχήματα 2.3.10 - 2.3.12 παρατίθενται τα υπογράμματα τεσσάρων χρηστών του συνόλου αξιολόγησης (test), δηλαδή διαγράμματα πρόβλεψης/ταξινόμησης των σταδίων του ύπνου τεσσάρων χρηστών, στα τέσσερα στάδια Wake, Light, Deep και REM. Επιπλέον, τα αποτελέσματα της ταξινόμησης για τα διαφορετικά μήκη παραθύρωσης παρατίθενται στον Πίνακα 2.6 με τις κλασικές δημοφιλείς μετρικές. Τέλος, στους Πίνακες 2.7 - 2.9 παρουσιάζονται τα αποτελέσματα ταξινόμησης για κάθε στάδιο ύπνου (Wake, Light, Deep και REM) για τα τρία διαφορετικά μήκη παραθύρωσης και στα Σχήματα 2.3.13 - 2.3.15 παρουσιάζονται οι Πίνακες Σύγχυσης για τα διαφορετικά μήκη παραθύρωσης.

Εποπτικά, από τα υπογράμματα του Σχήματος 2.3.10 - 2.3.12, παρατηρούμε ότι η κλάση Deep δεν μπορεί να αναγνωριστεί καλά από το νευρωνικό δίκτυο μακροπρόθεσμης μνήμης, η κλάση Light και Wake αναγνωρίζονται πολύ καλά από το μοντέλο όπως και στη περίπτωση του συνελκτικού δικτύου CNN. Δηλαδή παρατηρείται το ίδιο όπως και στη μοντελοποίηση με CNN με μικρές διαφορές, αλλά λίγο καλύτερα φαίνονται τα υπογράμματα που προκύπτουν από το μοντέλο LSTM. Η κλάση REM αναγνωρίζεται σε μέτριο βαθμό, καλύτερα βέβαια από την κλάση Deep. Το συμπέρασμα αυτό μπορεί να διαγνωσθεί και από τους Πίνακες μετρικών του συνόλου αξιολόγησης (Πίνακες 2.7 - 2.9) και από τους Πίνακες σύγχυσης (Confusion Matrix - Σχήματα 2.3.13 - 2.3.15) όπου η κλάση Deep έχει τις χειρότερες αποδόσεις σε όλες τις μετρικές, ακολουθεί με λίγο καλύτερες αποδόσεις η κλάση REM, μετά η κλάση Wake, και τέλος τις υψηλότερες αποδόσεις τις έχει η κλάση Light.

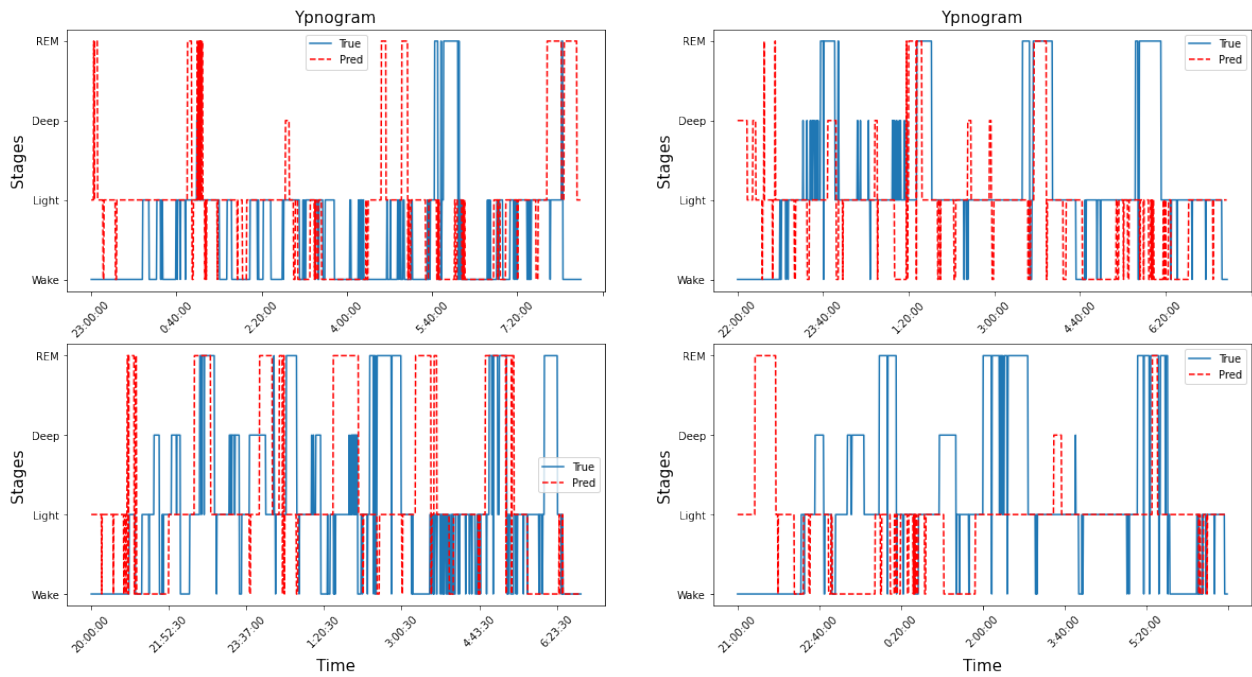
Λεπτομερέστερα, από τον Πίνακα μετρικών 2.6, συμπεραίνουμε ότι τα μοντέλα  $LSTM_{51}$  και  $LSTM_{101}$  έχουν τις υψηλότερες αποδόσεις σε όλες τις μετρικές, αντίστοιχα με τα μοντέλα CNN. Παρατηρούμε ότι όλες οι μετρικές έχουν παραπλήσια ποσοστά για όλα τα μήκη παραθύρωσης των δεδομένων, με καλύτερο το μοντέλο  $LSTM_{51}$ . Οι μετρικές Precision και Recall είναι της τάξης του 51 % και 58 % αντίστοιχα και γι αυτό και η μετρική F1-Score κυμαίνεται στα ίδια επίπεδα. Όπως αναφέρθηκε και στη μοντελοποίηση με CNN, λόγω του μη ισορροπημένου συνόλου δεδομένων, η μετρική Accuracy δεν είναι αξιόπιστη για την αξιολόγηση των μοντέλων. Επίσης, παρατηρούμε ότι η μετρική Specificity είναι υψηλή σε όλα τα μοντέλα LSTM. Το γεγονός αυτό καταδεικνύει ότι είναι μικρό το ποσοστό των λάθος ταξινομημένων δεδομένων, γεγονός που είναι σημαντικό. Αντίθετα, η μετρική F1-Score και Cohen-Kappa είναι σχετικά χαμηλή αλλά αποκαλύπτει την πραγματική απόδοση των μοντέλων εξαιτίας του

---

<sup>7</sup>RMSProp Optimizer: <https://keras.io/api/optimizers/rmsprop/>

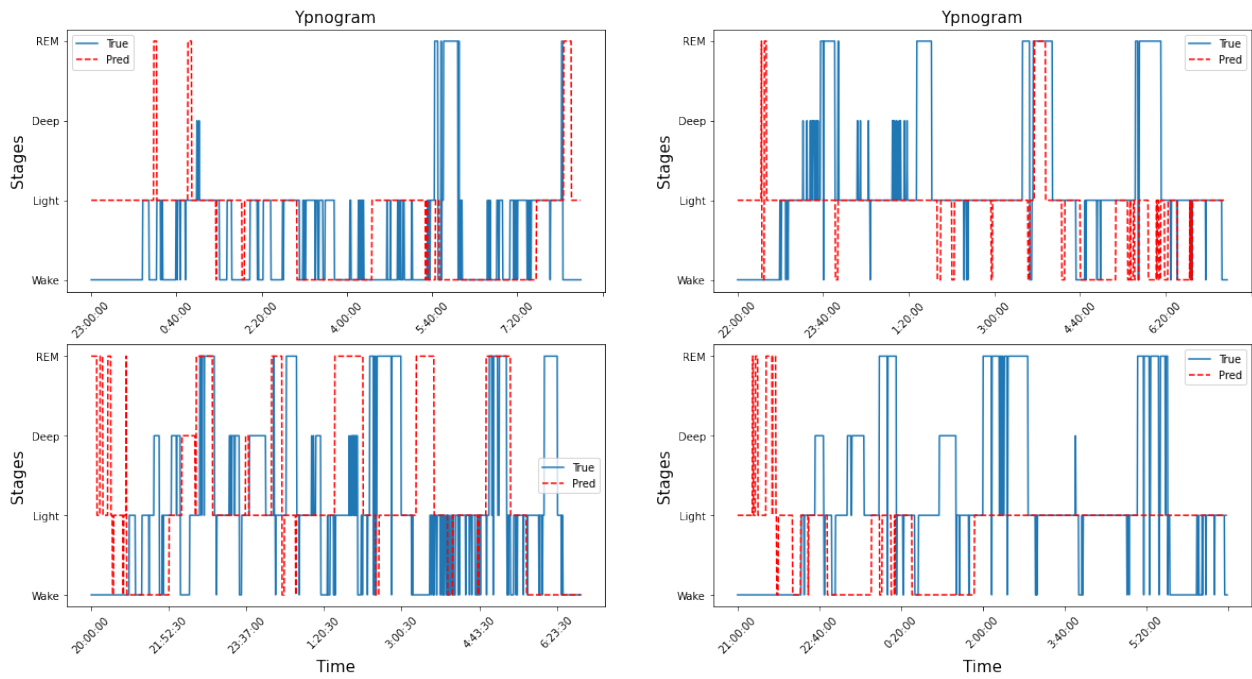


Σχήμα 2.3.10: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 21.



Σχήμα 2.3.11: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 51.

μη ισορροπημένου συνόλου δεδομένων (imbalanced dataset) όπως και στην μοντελοποίηση που πραγματοποιήθηκε με τα μοντέλα CNN. Οπότε, γι αυτό τον λόγο η μετρική Specificity είναι υψηλή, επειδή η



Σχήμα 2.3.12: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 101.

<i>Metrics%</i>	<i>SeqLen21</i>	<i>SeqLen51</i>	<i>SeqLen101</i>
Accuracy	68.7 ± 1.0	70.2 ± 1.0	69.5 ± 1.0
Precision	56.4 ± 1.4	57.2 ± 1.4	58.5 ± 1.4
Recall	51.8 ± 0.9	53.1 ± 1.0	52.9 ± 1.1
F1-Score	49.7 ± 0.9	51.1 ± 1.0	51.4 ± 1.1
Cohen-Kappa	50.4 ± 1.8	52.8 ± 1.9	51.7 ± 1.9
Specificity	86.6 ± 0.4	87.2 ± 0.4	86.8 ± 0.4

Πίνακας 2.6: Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης.

κλάση Light έχει το μεγαλύτερο ποσοστό δεδομένων συνολικά και αναγνωρίζεται καλά από τα μοντέλα. Αυτό έχει ως αποτέλεσμα, να μην φαίνεται η δυσκολία των μοντέλων να αναγνωρίσουν καλά τις κλάσεις με το χαμηλότερο πλήθος δεδομένων, δηλαδή Deep και REM.

Εξαιτίας αυτού του φαινομένου, παρατηρούμε τους Πίνακες Σύγκρισης 2.7 - 2.9 όπου παρουσιάζονται οι αποδόσεις των μοντέλων ανα κλάση. Είναι φανερό ότι τα μοντέλα, λόγω του μη ισορροπημένου συνόλου δεδομένων (imbalance dataset), δεν μπορούν να αναγνωρίσουν σχεδόν καθόλου την κλάση Deep (F1-Score  $\leq 20\%$ ) και ακολουθεί η κλάση REM. Επίσης, παρατηρούμε ότι τα μοντέλα  $LSTM_{51}$  και  $LSTM_{101}$  είναι καλύτερο σε όλες τις μετρικές ανα κλάση, σε σχέση με το μοντέλο  $LSTM_{21}$ . Αν εστιάσουμε στην κλάση REM, Deep και Wake, παρατηρούμε ότι το μοντέλο  $LSTM_{50}$  έχει τη καλύτερη απόδοση F1-Score και Recall από όλα τα μοντέλα. Δηλαδή, η βέλτιστη παραθύρωση είναι αυτή με μήκος 51 και βοηθά περισσότερο τις κλάσεις να αναγνωριστούν από ότι τα μήκη παραθύρωσης 21 και 101, παρατηρώντας ειδικά για τις μετρικές F1-Score και Recall. Αυτό έρχεται σε αντίθεση με τα αποτελέσματα των CNN μοντέλων, δηλαδή για να αναγνωριστούν καλά τα στάδια του ύπνου, πρέπει τα μοντέλα να επεξεργάζονται

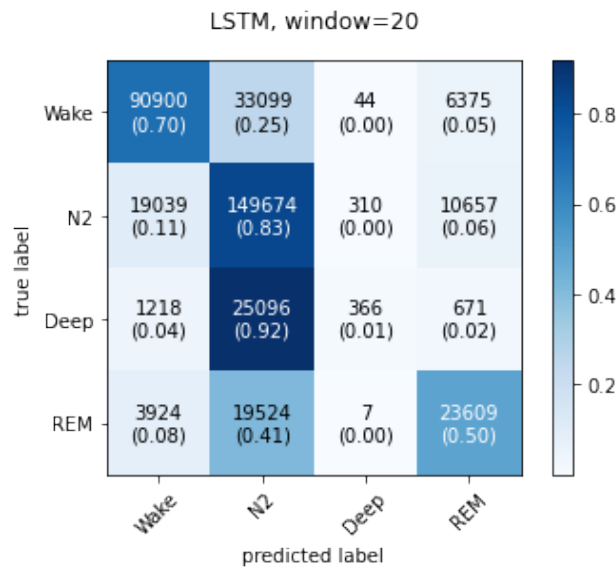


<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	79.0	70.0	74.0
Light	66.0	83.0	74.0
Deep	50.0	1.0	3.0
REM	57.0	50.0	53.0

Πίνακας 2.7: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 21.

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	80.0	73.0	76.0
Light	68.0	82.0	75.0
Deep	43.0	13.0	20.0
REM	63.0	56.0	59.0

Πίνακας 2.8: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 51.



Σχήμα 2.3.13: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 21.

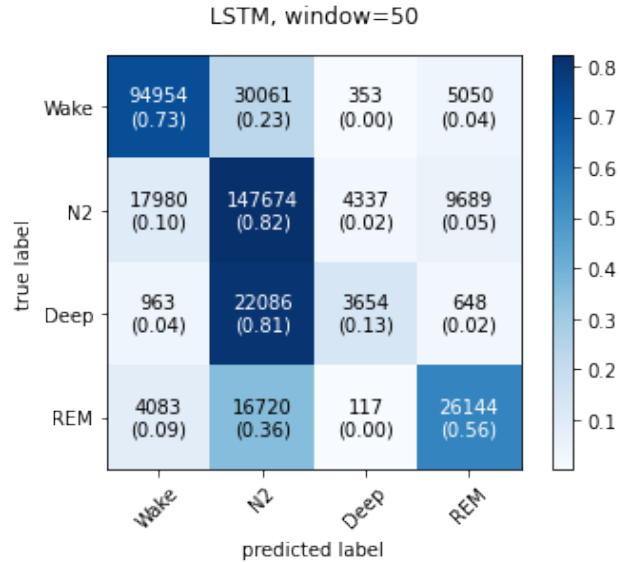
μεγαλύτερα παράθυρα δεδομένων (sliding windows) ώστε να αναγνωρίσουν τα long-term χαρακτηριστικά που διαχωρίζουν τα στάδια του ύπνου.

Τα παραπάνω συμπεράσματα συνοψίζονται στους Πίνακες Σύγχυσης 2.3.13 - 2.3.15. Δηλαδή, όλα τα μοντέλα δυσκολεύονται να αναγνωρίσουν τις κλάσεις Deep και REM και αναγνωρίζουν πολύ καλά τις κλάσεις Wake και Light, παρατηρώντας τα ποσοστά Recall της διαγωνίου. Το γεγονός αυτό είναι λογικό εφόσον το σύνολο δεδομένων είναι μη ισορροπημένο. Οπότε, περισσότερα δεδομένα από μια κλάση, διαφορετικά μεταξύ τους, τόσο καλύτερα τα μοντέλα θα την αναγνωρίσουν. **Σημαντική παρατήρηση αποτελεί ότι το μοντέλο  $LSTM_{51}$ , ενώ έχει παραπλήσιες αποδόσεις με το μοντέλο  $LSTM_{101}$  στις γενικές δημοφιλείς μετρικές (Πίνακας 2.6), υπερτερεί στις μετρικές ανα κλάση.**



<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	81.0	71.0	76.0
Light	66.0	84.0	74.0
Deep	53.0	8.0	13.0
REM	58.0	48.0	53.0

Πίνακας 2.9: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 101.



Σχήμα 2.3.14: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 51.

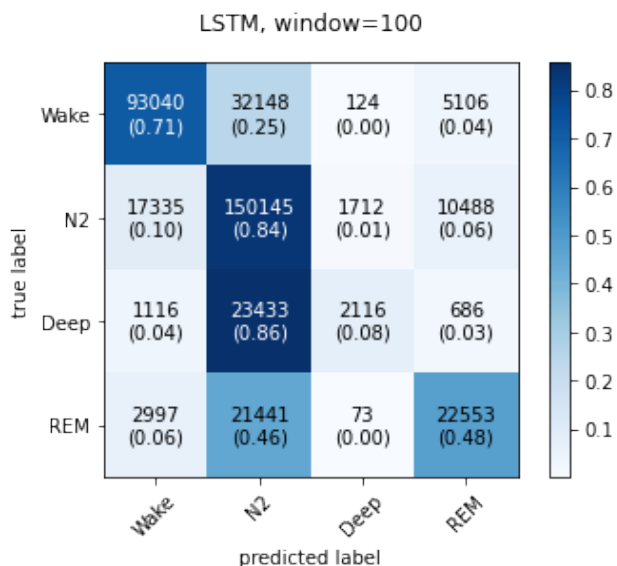
Στην επόμενη ενότητα, θα παρουσιάσουμε τα μοντέλα Ensemble και θα παραθέσουμε τα αντίστοιχα αποτελέσματα.

### 2.3.4 Ταξινόμηση με τη Μέθοδο Ensemble (Ensemble Method)

Στο παρούσα ενότητα, προκειμένου:

- Να εκμεταλλευτούμε τις **δυνατότητες των μοντέλων Βαθιάς Μάθησης** που χρησιμοποιήσαμε, δηλαδή των CNN και LSTM μοντέλων, τα οποία εξάγουν διαφορετικά χαρακτηριστικά από τα δεδομένα ύπνου, και
- Να εκμεταλλευτούμε τη μοντελοποίηση με **διαφορετική παραθύρωση (sliding window)** δεδομένων, που οδηγεί σε διαφορετικά εξαγόμενα χαρακτηριστικά από τα μοντέλα Βαθιάς Μάθησης (Deep Learning)

εφαρμόζουμε την μέθοδο Ensemble που περιγράφηκε στο Κεφάλαιο 3.3. Συνοπτικά, με βάση τα αποτελέσματα ταξινόμησης που δίνουν τα μοντέλα Βαθιάς Μάθησης που πειραματιστήκαμε, εφαρμόζονται δύο μέθοδοι τελικής ταξινόμησης των δεδομένων ή αλλιώς **μεικτής αξιολόγησης (Score Fusion)** των αποτελεσμάτων: 1) **Μέθοδος Max Ensemble** και 2) **Μέθοδος Mean Ensemble**. Και στις δύο μεθόδους μεικτής αξιολόγησης, χρησιμοποιούνται οι ύστερες πιθανότητες (posterior probabilities) που προκύπτουν από τα μοντέλα Βαθιάς Μάθησης (CNN και LSTM) μέσω της συνάρτησης



Σχήμα 2.3.15: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 101.

Softmax:  $\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^4 e^{z_j}}$   $i = 1, 2, 3, 4$ . Δηλαδή, για κάθε σύνολο 8+1 χαρακτηριστικών χρονικής διάρκειας 30 δευτερολέπτων του συνόλου αξιολόγησης (test dataset), τα μοντέλα προβλέπουν τις 4 πιθανότητες (μια για κάθε στάδιο ύπνου Wake, Light, Deep και REM).

Πιο συγκεκριμένα, έχουμε 2 μοντέλα (CNN και LSTM) επί 3 διαφορετικά μήκη παραθύρωσης (**21, 51, 101**), άρα 6 στον αριθμό διαφορετικών ταξινομητών/μοντέλων όπου το κάθε ένα προβλέπει ένα διάνυσμα ύστερης πιθανότητας  $\mathbf{p}$  στον χώρο των 4 διαστάσεων ( $\mathbf{p} \in \mathbb{R}^K$ ), (όπου  $K$  είναι ο αριθμός των κλάσεων ( $K=4$ )), για κάθε σύνολο 8+1 χαρακτηριστικών χρονικής διάρκειας 30 δευτερολέπτων του συνόλου αξιολόγησης (test dataset). Τότε για το Ensemble μοντέλο και την **Mean Ensemble** μέθοδο αξιολόγησης προκύπτει ότι:  $\mathbf{p}^{mean}_t = \frac{1}{6} \sum_{m=1}^6 \mathbf{p}^m_t$ , με  $\mathbf{p}^m_t$  η ύστερη (posterior) πιθανότητα του μοντέλου  $\mathbf{m} \in [1, 6]$ . Προκύπτει ότι  $\mathbf{p}^{mean}_t$  είναι ένα διάνυσμα ύστερης πιθανότητας στον χώρο  $\mathbb{R}^4$ . Η προβλεπόμενη κλάση  $\hat{\mathbf{k}}$  του μοντέλου Mean Ensemble, είναι αυτή με την μεγαλύτερη ύστερη πιθανότητα, δηλαδή  $\hat{\mathbf{k}} = \arg \max_k \mathbf{p}^{mean}_t$  και αντιστοιχεί σε μία από τις 4 κλάσεις με την μεγαλύτερη ύστερη πιθανότητα. Αντίστοιχα, για την μέθοδο αξιολόγησης **Max Ensemble** προκύπτει ότι η τελική ταξινόμηση στην κλάση  $\hat{\mathbf{k}}$  είναι:  $\hat{\mathbf{k}} = \arg \max_k \mathbf{P}^{max}_t$ , όπου  $\mathbf{P}^{max}_t = [p_t^1, p_t^2, p_t^3, p_t^4, p_t^5, p_t^6]$ , με  $\mathbf{p}^m_t$  η ύστερη (posterior) πιθανότητα του μοντέλου  $\mathbf{m} \in [1, 6]$  και  $\mathbf{p}^m_t \in \mathbb{R}^4$  και  $\mathbf{P}^{max}_t \in \mathbb{R}^{6 \times 4}$ .

Σύμφωνα με τα παραπάνω προκύπτει ο Πίνακας αποτελεσμάτων 2.10.

Από τον Πίνακα 2.10 παρατηρούμε ότι πετυχαίνουμε καλά αποτελέσματα σε όλες τις μετρικές, τα οποία πρέπει να συγκριθούν με τις μετρικές των επιμέρους μοντέλων για να μπορούμε να αξιολογήσουμε την αποδοτικότητα του μοντέλου Ensemble στην ταξινόμηση των σταδίων ύπνου. Όμως αυτό που παρατηρούμε είναι ότι και οι δύο μέθοδοι (Mean και Max Ensemble) μεικτής αξιολόγησης (Score Fusion) είναι αποδοτικοί, με **λίγο καλύτερα αποτελέσματα να δίνει η μέθοδος Max Ensemble στις μετρικές Precision και Recall**.

Στην επόμενη ενότητα θα συνοψίσουμε τα αποτελέσματα που προέκυψαν από όλη την ανάλυση που πραγματοποιήθηκε με τις διαφορετικές προσεγγίσεις που ακολουθήσαμε για την ταξινόμηση των δεδομένων actigraphy + HRV σε 4 στάδια ύπνου.

<i>Metrics%</i>	<i>Mean Ensemble</i>	<i>Max Ensemble</i>
Accuracy	72.1 ± 1.0	71.9 ± 1.0
Precision	58.8 ± 1.3	59.5 ± 1.4
Recall	53.8 ± 1.0	54.2 ± 1.0
F1-Score	52.1 ± 1.0	52.5 ± 1.0
Cohen-Kappa	55.7 ± 1.8	55.7 ± 1.8
Specificity	87.6 ± 0.4	87.7 ± 0.4

Πίνακας 2.10: Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης.

### 2.3.5 Σύνοψη Αποτελεσμάτων

Στην παρούσα ενότητα θα παραθέσουμε και θα συνοψίσουμε τα αποτελέσματα που προέκυψαν από την μοντελοποίηση των δεδομένων ακτιγραφίας και HRV χαρακτηριστικών μέσω των μοντέλων CNN, LSTM και τη μέθοδο Ensemble, και θα συγκρίνουμε τα πειράματά μας με τα αντίστοιχα της έρευνας [1] για να δούμε αν επαληθεύονται τα αποτελέσματα που δίνουν οι ερευνητές, οπότε θα έχουμε μια βάση για να εξελίξουμε και να βελτιώσουμε αυτά τα αποτελέσματα της έρευνας.

Συγκεκριμένα, για κάθε μοντέλο και μέθοδο ταξινόμησης, παραθέτουμε τον Πίνακα 2.11 ο οποίος περιλαμβάνει τα αποτελέσματα της μετρικής "χρονικής απόκλισης" (Time Deviation) που παραθέσαμε στην εισαγωγή του Κεφαλαίου 3.3. Συνοπτικά, για μια κλάση  $k$ , και  $N$  αριθμό χρηστών (users), η μετρική "χρονικής απόκλισης" ορίζεται ως  $TD_k = \frac{1}{N} \sum_{i=1}^N (Pred_k^i - True_k^i)$ , όπου  $Pred_k$  είναι ο συνολικός χρόνος σε λεπτά για την προβλεπόμενη κλάση  $k$  (predicted label) του μοντέλου και  $True_k$  είναι ο αντίστοιχος συνολικός χρόνος σε λεπτά της πραγματικής κλάσης  $k$  (true label). Οπότε σύμφωνα με τα παραπάνω έχουμε τον Πίνακα 2.11 της "χρονικής απόκλισης" για κάθε μοντέλο ανά κλάση.

<i>Model%</i>	<i>Deep</i>	<i>Light</i>	<i>REM</i>	<i>Wake</i>
<i>CNN<sub>21</sub></i>	-41.4	72.9 ± 7.6	-13.2	-32.6
<i>CNN<sub>51</sub></i>	-35.8	98.1 ± 7.5	-34.4	-41.8
<i>CNN<sub>101</sub></i>	-38.4	44.4 ± 8.5	-18.2	5.0 ± 8.6
<i>LSTM<sub>21</sub></i>	-41.9	68.6 ± 7.7	-12.3	-28.9
<i>LSTM<sub>51</sub></i>	-40.8	58.9 ± 7.6	-15.4	-16.6
<i>LSTM<sub>101</sub></i>	-37	68.2 ± 7.3	-15.5	-29.4
<i>MaxEnsemble</i>	-40.8	64.8 ± 7.4	-20.4	-17.6
<i>MeanEnsemble</i>	-41.9	78.7 ± 7.5	-23.8	-26.8

Πίνακας 2.11: Μετρική Χρονικής Απόκλισης (Time Deviation) ανα κλάση σε λεπτά (minutes).

Από τον Πίνακα 2.11 παρατηρούμε σε πρώτο βαθμό ότι για τις κλάσεις Deep, REM και Wake, **υποεκτιμάται** ο χρόνος παραμονής των δεδομένων κλάσεων (λόγω του αρνητικού πρόσημου στα αποτελέσματα), ενώ αντίθετα ο χρόνος παραμονής στην κλάση Light **υπερεκτιμάται** (λόγω του θετικού πρόσημου στα αποτελέσματα). Λεπτομερέστερα, η κλάση Deep υποεκτιμάται από όλα τα μοντέλα, όμως το *CNN<sub>51</sub>* μοντέλο δίνει τη μικρότερη τιμή απόκλισης (-35.8 λεπτά). Αντίστοιχα, η κλάση REM υποεκτιμάται από όλα τα μοντέλα, όμως το *LSTM<sub>21</sub>* μοντέλο έχει τη μικρότερη τιμή απόκλισης (-12.3 λεπτά). Επίσης, η κλάση Wake υποεκτιμάται από όλα τα μοντέλα, όμως το *LSTM<sub>51</sub>* μοντέλο δίνει τη μικρότερη τιμή απόκλισης (-16.6 λεπτά). Αντίθετα, η κλάση **Light υπερεκτιμάται** από όλα τα μοντέλα, όμως το *CNN<sub>101</sub>* μοντέλο δίνει το μικρότερο χρόνο απόκλισης (44.4 λεπτά). Με βάση τα

παραπάνω αποτελέσματα συμπεραίνουμε ότι είναι λογικό η κλάση Light να υπερεκτιμάται εξαιτίας του μη ισορροπημένου συνόλου δεδομένων (imbalance dataset), οπότε τα μοντέλα τείνουν να ταξινομούν τα δεδομένα ύπνου στη κλάση με το μεγαλύτερο πλήθος δεδομένων, δηλαδή στη κλάση Light. Αυτό το συμπέρασμα είναι σημαντικό γιατί μας υποδεικνύει ότι πρέπει να αντιμετωπιστεί το πρόβλημα του μη ισορροπημένου συνόλου δεδομένων.

Επιπρόσθετα, από όλα τα μοντέλα με τα οποία πειραματιστήκαμε, συμπεραίνουμε ότι ο συνδυασμός των Συνελικτικών Νευρωνικών Δικτύων (CNN) με τα Νευρωνικά Δίκτυα Μακροπρόθεσμης Μνήμης (LSTM), με τη μέθοδο Ensemble, οδηγεί σε καλύτερα αποτελέσματα αφού το κάθε ένα μοντέλο ξεχωριστά εστιάζει στην εξαγωγή διαφορετικού είδους χαρακτηριστικών. Δηλαδή, τα CNN μοντέλα εξάγουν χωρικά χαρακτηριστικά (spatial features), δηλαδή επεξεργάζονται τα δεδομένα στατικά, ενώ τα μοντέλα LSTM εξάγουν χρονικά χαρακτηριστικά (temporal features), δηλαδή αντιμετωπίζουν τα δεδομένα βάσει της χρονικής τους διάστασης. Οπότε, ο συνδυασμός των αποτελεσμάτων που προκύπτουν από τα CNN και τα LSTM μοντέλα μέσω της μεθόδου Ensemble, οδηγεί σε καλύτερα αποτελέσματα ταξινόμησης σε όλες τις δημοφιλείς μετρικές όπως φαίνεται στον Πίνακα 2.10.

Σε επόμενο στάδιο παραθέτουμε τα αποτελέσματα, που δίνουν οι ερευνητές του [1], των μοντέλων Βαθιάς Μάθησης (Deep Learning) για το πρόβλημα ταξινόμησης των δεδομένων ύπνου σε 4 στάδια. Ακολουθεί ο Πίνακας 2.13 με τις αντίστοιχες δημοφιλείς μετρικές και ο Πίνακας 2.14 με τις αντίστοιχες δημοφιλείς μετρικές για τις μεθόδους Ensemble. Επίσης, παρουσιάζονται στον Πίνακα 2.12 τα αποτελέσματα που προέκυψαν από την υλοποίηση μας με βάση την έρευνα [1].

<i>Metrics%</i>	<i>CNN<sub>21</sub></i>	<i>CNN<sub>51</sub></i>	<i>CNN<sub>101</sub></i>	<i>LSTM<sub>21</sub></i>	<i>LSTM<sub>51</sub></i>	<i>LSTM<sub>101</sub></i>
Accuracy	68.9 ± 1.0	70.1 ± 1.0	70.4 ± 1.1	68.7 ± 1.0	70.2 ± 1.0	69.5 ± 1.0
Precision	54.8 ± 1.2	61.4 ± 1.5	61.4 ± 1.5	56.4 ± 1.4	57.2 ± 1.4	58.5 ± 1.4
Recall	51.7 ± 0.9	51.4 ± 1.0	51.4 ± 1.0	51.8 ± 0.9	53.1 ± 1.0	52.9 ± 1.1
F1-Score	49.5 ± 0.9	50.5 ± 1.1	50.5 ± 1.1	49.7 ± 0.9	51.1 ± 1.0	51.4 ± 1.1
Cohen-Kappa	50.3 ± 1.7	50.3 ± 1.8	53.4 ± 1.9	50.4 ± 1.8	52.8 ± 1.9	51.7 ± 1.9
Specificity	86.6 ± 0.4	86.7 ± 0.4	86.7 ± 0.4	86.6 ± 0.4	87.2 ± 0.4	86.8 ± 0.4

Πίνακας 2.12: Πίνακας αποτελεσμάτων που προκύπτουν από τις υλοποιήσεις μας με βάση την έρευνα [1].

<i>Metrics%</i>	<i>CNN<sub>21</sub></i>	<i>CNN<sub>51</sub></i>	<i>CNN<sub>101</sub></i>	<i>LSTM<sub>21</sub></i>	<i>LSTM<sub>51</sub></i>	<i>LSTM<sub>101</sub></i>
Accuracy	67.0 ± 1.0	68.0 ± 1.1	69.5 ± 1.0	68.3 ± 1.0	70.3 ± 1.0	70.2 ± 1.0
Precision	52.2 ± 1.1	54.8 ± 1.3	59.5 ± 1.4	55.0 ± 1.2	57.9 ± 1.3	59.9 ± 1.5
Recall	50.5 ± 0.9	53.5 ± 1.0	54.4 ± 1.0	51.6 ± 0.9	54.0 ± 1.0	52.4 ± 1.0
F1-Score	47.8 ± 0.9	49.9 ± 1.0	52.1 ± 1.1	49.5 ± 0.9	51.9 ± 1.0	51.3 ± 1.1
Cohen-Kappa	48.3 ± 1.7	51.6 ± 1.9	52.6 ± 1.8	50.0 ± 1.7	53.8 ± 1.9	51.7 ± 1.8
Specificity	85.9 ± 0.4	86.9 ± 0.4	87.2 ± 0.4	86.5 ± 0.4	87.4 ± 0.4	86.9 ± 0.4

Πίνακας 2.13: Πίνακας αποτελεσμάτων που δίνουν οι ερευνητές στο [1].

Από τους δύο Πίνακες (2.12, 2.13) παρατηρούμε ότι τα μοντέλα που αναπτύξαμε πέτυχαν σχεδόν τα ίδια αποτελέσματα (με μικρές αποκλίσεις) με τους ερευνητές σε όλες τις δημοφιλείς μετρικές, ακολουθώντας την ερευνά τους. Επίσης, από τους Πίνακες 2.10 και 2.14 παρατηρούμε ότι και στις δύο υλοποιήσεις, η μέθοδος Max Ensemble δίνει τα καλύτερα αποτελέσματα και επίσης είναι σημαντικό που επιβεβαιώσαμε τα αποτελέσματα της έρευνας [1] διότι έτσι έχουμε μια βάση, ώστε σε επόμενο στάδιο να προσπαθήσουμε να λύσουμε τα προβλήματα που αναδύθηκαν στις προηγούμενες ενότητες του Κεφαλαίου 2.3 όπου αναπτύξαμε τα μοντέλα Βαθιάς Μάθησης σύμφωνα με το [1].

<i>Metrics%</i>	<i>Mean Ensemble</i>	<i>Max Ensemble</i>
Accuracy	71.6 ± 1.0	71.1 ± 1.0
Precision	57.1 ± 1.2	58.3 ± 1.3
Recall	54.3 ± 1.0	54.8 ± 1.0
F1-Score	52.1 ± 1.0	52.4 ± 1.0
Cohen-Kappa	56.1 ± 1.8	55.7 ± 1.8
Specificity	87.6 ± 0.4	87.7 ± 0.4

Πίνακας 2.14: Πίνακας μετρικών για τις μεθόδους Ensemble που δίνουν οι ερευνητές στο [1].

Με βάση τα πειράματα που αναπτύχθηκαν στο Κεφάλαιο 2.3, διαπιστώνουμε τελικά ότι:

- Το **μη ισορροπημένο σύνολο δεδομένων** επηρεάζει σε μεγάλο βαθμό την απόδοση των μοντέλων Βαθιάς Μάθησης (Deep Learning) και πρέπει να αντιμετωπιστεί με τεχνικές/μεθόδους που θα παρουσιασθούν στο επόμενο Κεφάλαιο 3.4. Συγκεκριμένα, το πρόβλημα εστιάζεται στο γεγονός ότι οι κλάσεις Deep και REM δεν έχουν μεγάλο πλήθος δεδομένων και αντίθετα ότι η κλάση Light έχει πολύ μεγάλο πλήθος δεδομένων σε σχέση με τις κλάσεις Deep και REM.
- Ο πειραματισμός με **πιο "βαθιά" μοντέλα**, δηλαδή προσθήκη περισσότερων επιπέδων εξαγωγής χαρακτηριστικών (Convolutional Layers η LSTM Layers) ή και με **άλλου είδους μοντέλων** (πχ. Μοντέλο Μετασχηματιστή (Transformer model)) τα οποία θα διαχειριστούν τα δεδομένα με διαφορετικό τρόπο, ευελπιστούμε ότι θα βοηθήσουν στην περαιτέρω καλύτερη απόδοση των μοντέλων βαθιάς Μάθησης.
- Τέλος, ο πειραματισμός με **διαφορετικό μήκος παραθύρωσης (sliding window)** επηρεάζει σημαντικά τα αποτελέσματα ταξινόμησης σε όλα τα μοντέλα, οπότε μπορεί να οδηγήσει σε περαιτέρω βελτίωση των αποτελεσμάτων, επειδή στις προηγούμενες ενότητες του Κεφαλαίου 3.3, συμπεράναμε ότι όσο μεγαλώνει το μήκος παραθύρωσης, τόσο βελτιώνεται η απόδοσης των μοντέλων Βαθιάς Μάθησης.

## 2.4 Βελτίωση Αλγόριθμων Βαθιάς Μάθησης (Deep Learning)

Στο κεφάλαιο αυτό θα προσπαθήσουμε να βελτιώσουμε τα μοντέλα Βαθιάς Μάθησης που αναπτύχθηκαν στο Κεφάλαιο 2.3 για την ταξινόμηση των δεδομένων MESA στα 4 στάδια του ύπνου, με διάφορες **τεχνικές αντιμετώπισης του προβλήματος του μη ισορροπημένου συνόλου δεδομένων (imbalance dataset)**, με το **πειραματισμό μέσω νέων μοντέλων Βαθιάς Μάθησης (Deep Learning)** αλλά και με τον **πειραματισμό με διαφορετικά μήκη παραθύρωσης (sliding window)**. Αρχικά, θα παραθέσουμε τον σχεδιασμό των πειραμάτων που υλοποιήσαμε, μετέπειτα θα παραθέσουμε την κάθε μέθοδο/μοντέλο ξεχωριστά με τα αντίστοιχα αποτελέσματα τους και στο τέλος του Κεφαλαίου θα παρουσιάσουμε τα τελικά αποτελέσματα όλων των πειραμάτων συγκεντρωτικά.

### 2.4.1 Περιγραφή διαδικασίας - Προεπεξεργασία Δεδομένων

Στο προηγούμενο Κεφάλαιο (2.3), για την ταξινόμηση των δεδομένων MESA με τη χρήση δεδομένων ακτιγραφίας και HRV χαρακτηριστικών, χρησιμοποιήσαμε μοντέλα Βαθιάς Μάθησης και συγκεκριμένα ένα Συνελικτικό Νευρωνικό Δίκτυο (CNN) και ένα Νευρωνικό Δίκτυο Μακροπρόθεσμης Μνήμης (LSTM). Σε αυτό το Κεφάλαιο καλούμαστε να βασιστούμε στα παραπάνω μοντέλα του προηγούμενου Κεφαλαίου και να τα εξελίξουμε με την προσθήκη περαιτέρω επιπέδων εξαγωγής χαρακτηριστικών (Fea-



ture Extraction Layers) αλλά και ότι είναι απαραίτητο για την σωστή εκπαίδευση των μοντέλων Βαθιάς Μάθησης, όπως για παράδειγμα η εφαρμογή τεχνικών Regularization. Επιπλέον, δοκιμάζουμε το μοντέλο Μετασχηματιστή (Transformer model) [84] και παρατηρούμε τα αποτελέσματα που προκύπτουν. Σε όλους τους επιπρόσθετους πειραματισμούς που υλοποιήθηκαν, προσπαθούμε να αντιμετωπίσουμε το πρόβλημα του μη ισορροπημένου συνόλου δεδομένων (imbalance dataset). Δηλαδή, εφαρμόζουμε τεχνικές Data Augmentation ώστε να "ισορροπήσουμε" το σύνολο δεδομένων και να παρατηρήσουμε τις επιδράσεις τους στο τελικό αποτέλεσμα ταξινόμησης. Τα προβλήματα των μοντέλων του Κεφαλαίου 3.3 οδήγησαν την έρευνα μας να εστιάσει στις τεχνικές Data Augmentation. Τέλος, πειραματιζόμαστε με διαφορετικό μήκος παραθύρωσης (sliding window) ώστε να εκτιμήσουμε την επίδραση του μήκους παραθύρωσης των δεδομένων στο τελικό αποτέλεσμα της ταξινόμησης.

Με βάσει όλα τα παραπάνω θα αναλύσουμε τι βελτιώσεις πετύχαμε από τους πειραματισμούς και να προτείνουμε τα επόμενα βήματα που θα μπορούσαν να γίνουν για την περαιτέρω βελτίωση των αποτελεσμάτων για ένα δύσκολο πρόβλημα όπως το πρόβλημα ταξινόμησης των δεδομένων ύπνου σε 4 στάδια ύπνου.

### Κανονικοποίηση Δεδομένων:

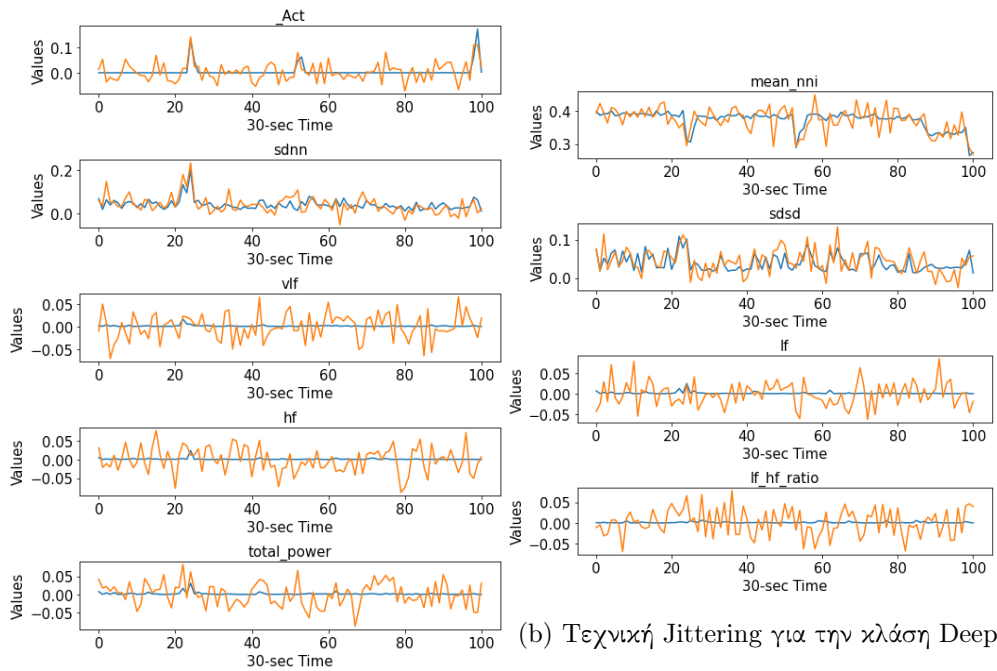
Λεπτομερέστερα, η βασική αλλαγή κατά της προ-επεξεργασία των δεδομένων ακτιγραφίας και HRV χαρακτηριστικών είναι ότι προκειμένου τα τελευταία να τροφοδοτηθούν στα μοντέλα Βαθιάς Μάθησης πρέπει να είναι κανονικοποιημένα, δηλαδή στο ίδιο εύρος τιμών. Στο Κεφάλαιο 3.3 κατά τη μοντελοποίηση, χρησιμοποιήθηκε η μέθοδος κανονικοποίησης με μέση τιμή 0 και τυπική απόκλιση 1 (Standard Scaling). Στο παρόν Κεφάλαιο 3.4, χρησιμοποιείται η **MinMax κανονικοποίηση** στα δεδομένα προς μοντελοποίηση. Δηλαδή, για κάθε χαρακτηριστικό HRV (8 στο σύνολο) και στα δεδομένα ακτιγραφίας, εφαρμόζεται η συνάρτηση MinMax<sup>8</sup>:  $\mathbf{X}_{scaled} = \mathbf{X}_{std} * (\mathbf{a} - \mathbf{b}) + \mathbf{a}$  όπου  $\mathbf{X}_{std} : \mathbf{X}_{std} = \frac{\mathbf{X} - \mathbf{X}_{min}}{\mathbf{X}_{max} - \mathbf{X}_{min}}$ ,  $\mathbf{a} = \mathbf{max}, \mathbf{b} = \mathbf{min}$  για ένα τυχαίο εύρος κανονικοποίησης  $[\mathbf{a}, \mathbf{b}]$ . Στο πρόβλημα μας εφαρμόζουμε την συνάρτηση MinMax στο εύρος:  $[\mathbf{a}, \mathbf{b}] = [0, 1]$ . Ευελπιστούμε ότι η κανονικοποίηση MinMax θα βοηθήσει τα μοντέλα Βαθιάς Μάθησης περισσότερο από την κανονικοποίηση με μέση τιμή 0 και τυπική απόκλιση 1 (Standard Scaling), δηλαδή ότι η αναπαράσταση των δεδομένων με MinMax θα είναι καλύτερη από την αναπαράσταση με Standard Scaling. Ο λόγος είναι επειδή τα δεδομένα ακτιγραφίας και HRV χαρακτηριστικών **δεν ακολουθούν ή ομοιάζουν με Κανονική Κατανομή** ( $\sim \mathcal{N}(0, 1)$ ) όπως φαίνεται στα Σχήματα 2.2.4 και 2.2.8. Οπότε είναι λογικό να εφαρμόσουμε την μέθοδο MinMax Scaling επειδή περιλαμβάνει την ευρωστία σε πολύ μικρές τυπικές αποκλίσεις των χαρακτηριστικών και τη διατήρηση μηδενικών σε αραιά δεδομένα<sup>9</sup>. Η τελευταία παρατήρηση αποκαλύπτεται στο Σχήμα 2.2.12 με την στατιστική περιγραφή του κάθε χαρακτηριστικού (ακτιγραφίας + 8 HRV). Συγκεκριμένα, στο Σχήμα 2.2.12 παρατηρούμε ότι τα δεδομένα έχουν μικρές τυπικές αποκλίσεις σε σχέση με το εύρος τιμών τους και έχουν κατανομή "δεξιά λοξή" (Right Skewed), οπότε δικαιολογείται σε ένα βαθμό η εφαρμογή της συνάρτησης MinMax για κανονικοποίηση των δεδομένων, παρά τη παρουσία ακραίων τιμών (outliers). Τέλος, στη πράξη είδαμε ότι τα μοντέλα Βαθιάς Μάθησης απέδωσαν καλύτερα σε σχέση με τη κανονικοποίηση των χαρακτηριστικών που εφαρμόζαμε στα προηγούμενα Κεφάλαια (Standard Scaling).

### Τεχνικές Τεχνητής Αύξησης του συνόλου δεδομένων (Data Augmentation Techniques):

Σε επόμενο στάδιο, προσπαθήσαμε να αντιμετωπίσουμε το πρόβλημα της ανισοκατανομής του συνόλου δεδομένων (imbalance dataset). Συγκεκριμένα, σε όλους του πειραματισμούς που ακολουθούν στα παρακάτω Κεφάλαια, εφαρμόσαμε τεχνικές Αύξησης του συνόλου δεδομένων (Data Augmentation

<sup>8</sup>MinMax Scaling: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>

<sup>9</sup>Προεπεξεργασία Δεδομένων: <https://scikit-learn.org/stable/modules/preprocessing.html>



(a) Τεχνική Jittering για την κλάση Deep

(b) Τεχνική Jittering για την κλάση Deep

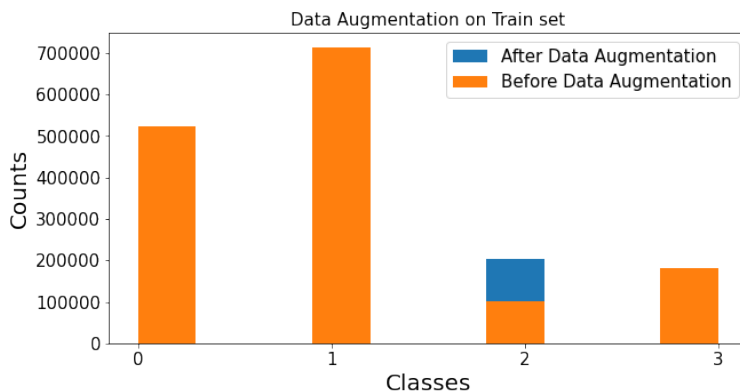
Σχήμα 2.4.1: Τεχνική Jittering για την κλάση Deep για κάθε χαρακτηριστικό ακτιγραφίας και HRV.

Techniques) στα δεδομένα ύπνου, ώστε να αντιμετωπίσουμε το πρόβλημα αυτό. Η αύξηση δεδομένων μπορεί να θεωρηθεί ως **προσθήκη προηγούμενης γνώσης σχετικά με τις αμετάβλητες ιδιότητες των δεδομένων έναντι ορισμένων μετασχηματισμών**. Τα επαυξημένα δεδομένα (augmented dataset) μπορούν να καλύψουν ανεξερεύνητες πτυχές του συνόλου δεδομένων εισόδου (prior knowledge), να αποτρέψουν την υπερβολική προσαρμογή των μοντέλων στο σύνολο δεδομένων εκπαίδευσης (overfitting) και να βελτιώσουν την ικανότητα γενίκευσης των μοντέλων Βαθιάς Μάθησης (generalization performance) [95]. Συνηθισμένες τεχνικές που εφαρμόζονται σε δεδομένα αισθητήρων (και όχι μόνο) είναι: **τρεμούλιαση/προσθήκη θορύβου (Jittering)**, **κλιμάκωση (Scaling)**, **περικοπή (Cropping)**, **παραμόρφωση (Warping)** και **περιστροφή (Rotating)**. Στο πρόβλημα ταξινόμησης δεδομένων ακτιγραφίας και HRV χαρακτηριστικών έχουμε παρατηρήσει ότι η κλάση Deep έχει το μικρότερο αριθμό δεδομένων. Οπότε σε αυτήν την κλάση θα προσθέσουμε τεχνητά δεδομένα, βασισμένα στα πραγματικά, ώστε να εξισορροπήσουμε το σύνολο δεδομένων εκπαίδευσης (balance dataset). Η πιο απλή τεχνική είναι να προσθέσουμε θόρυβο (Jittering) στην κλάση Deep, δηλαδή σε κάθε χαρακτηριστικό (feature), του συνόλου δεδομένων εκπαίδευσης, προσθέτουμε λευκό θόρυβο (White noise) με μέση τιμή 0 και τυπική απόκλιση 0.01:  $\mathbf{X}_{aug} = \mathbf{X} + \mathbf{Random}(\mathcal{N}(0, 0.01))$ , όπου  $\mathbf{X}_{aug}$  είναι τα επαυξημένα δεδομένα και  $\mathbf{Random}(\mathcal{N}(0, 0.01))$  είναι τυχαίοι αριθμοί της Κανονικής Κατανομής  $\mathcal{N}(0, 0.01)$ . Το αποτέλεσμα της τεχνικής Data Augmentation για κάθε χαρακτηριστικό (ακτιγραφία + 8 HRV) της κλάσης Deep φαίνεται στο Σχήμα 2.4.1.

Από το Σχήμα 2.4.1 παρατηρούμε ότι για ένα τυχαίο δείγμα του συνόλου δεδομένων (με μήκος παραθύρωσης 101) δεν αλλάζουν πολύ οι τιμές των πραγματικών δεδομένων με την προσθήκη θορύβου. Αυτό αιτιολογείται από την τιμή της τυπικής απόκλισης 0.01, η οποία επιλέχθηκε επίτηδες πολύ μικρή ώστε να μην δημιουργηθούν επαυξημένα δεδομένα τα οποία να μην ανταποκρίνονται στη πραγματικότητα (artificial data). Σε μια τέτοια περίπτωση, με μεγαλύτερες τιμές τιμής απόκλισης ( $> 0.01$ ), τα μοντέλα θα μπερδεύονται και δεν θα μαθαίνουν σωστά την κλάση Deep επειδή η μεγάλη προσθήκη θορύβου θα αλλάζουν την φυσιολογία της κλάσης Deep. Οπότε, από τη διαδικασία τεχνητής αύξησης δεδομένων,



διπλασιάζεται ο αριθμός των δεδομένων της κλάσης Deep εφόσον στα ήδη υπάρχοντα δεδομένα προσθέτουμε αυτά που δημιουργήσαμε με την προσθήκη λευκού θορύβου. Το αποτέλεσμα της αύξησης του συνόλου δεδομένων παρουσιάζεται στο Σχήμα 2.4.2.



Σχήμα 2.4.2: Ιστόγραμμα συνόλου δεδομένων εκπαίδευσης (train set). Αύξηση συνόλου δεδομένων της κλάσης Deep.

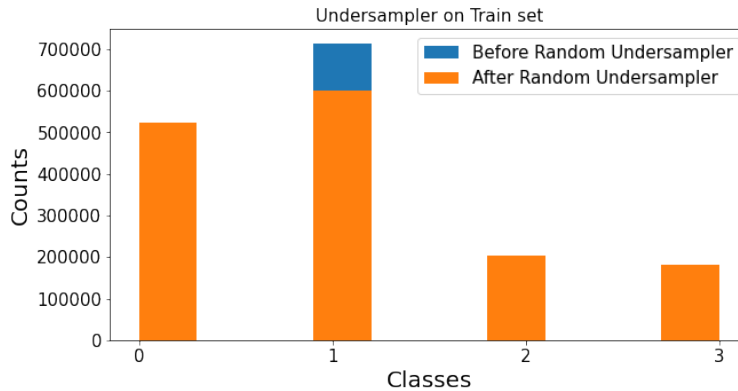
Στο Σχήμα 2.4.2 φαίνεται ξεκάθαρα ότι διπλασιάστηκε το σύνολο δεδομένων της κλάσης Deep, οπότε μπορούμε να υποθέσουμε ότι η κίνηση αυτή θα βοηθήσει τα μοντέλα Βαθιάς Μάθησης να διαχωρίζουν καλύτερα τις κλάσεις μεταξύ τους και να αναγνωρίζουν την κλάση Deep καλύτερα, εφόσον είναι η κλάση με το μικρότερο πλήθος δεδομένων. Επίσης, πρέπει να τονιστεί ότι η διαδικασία Augmentation, εφαρμόζεται μόνο στα δεδομένα εκπαίδευσης των μοντέλων Βαθιάς Μάθησης ώστε να μην γίνει αξιολόγηση των μοντέλων σε τεχνητά δεδομένα (artificial data) που δεν είναι πραγματικά. Τέλος, ο λόγος που εφαρμόστηκε η τεχνική αύξησης μόνο στη κλάση Deep, είναι επειδή η κλάση Deep έχει το μικρότερο πλήθος δεδομένων και είδαμε ότι κυρίως σε αυτήν τα μοντέλα Βαθιάς Μάθησης δυσκολεύονται να αναγνωρίσουν, είτε λόγω μικρού πλήθους είτε λόγω φυσιολογίας της κλάσης (δηλαδή να ομοιάζει με άλλες κλάσεις ως στάδιο ύπνου N3)<sup>10</sup>. Οπότε, είναι επιτακτική ανάγκη σε πρώτο στάδιο να εφαρμόσουμε την τεχνική αύξησης δεδομένων στη κλάση αυτή ώστε να αντιμετωπιστεί το πρόβλημα του μη ισορροπημένου συνόλου δεδομένων. Αν εφαρμόζαμε ταυτόχρονα και στη κλάση REM, η οποία είναι η δεύτερη σε μικρότερο αριθμό δεδομένων, τότε θα διατηρούνταν το πρόβλημα της ανισοκατανομής των δεδομένων, οπότε επιλέξαμε να κινηθούμε με αυτό τον σχεδιασμό.

### Πρόβλημα όγκου δεδομένων (Memory Leakage)

Επιπρόσθετα, επειδή προκύπτει πρόβλημα με τον όγκο δεδομένων για μήκος παραθύρωσης 101, δηλαδή δεν μπορούν όλα τα δεδομένα να φορτωθούν στην μνήμη RAM, εφαρμόζεται μόνο για τους πειραματισμούς με μήκος παραθύρωσης 101, τυχαία υποδειγματοληψία (Random Undersampling) της κλάσης Light. Ενώ θα μπορούσαμε να μειώσουμε το σύνολο δεδομένων εκπαίδευσης αφαιρώντας δεδομένα από όλες τις κλάσεις, επιλέξαμε να αφαιρέσουμε μόνο από την κλάση Light ώστε να παρατηρήσουμε αν έχει θετική επίδραση στο πρόβλημα του μη ισορροπημένου συνόλου δεδομένων. Το αποτέλεσμα της τυχαίας υποδειγματοληψίας δεδομένων οδηγεί στην μείωση τους συνόλου δεδομένων της κλάσης Light κατά:  $loss\% = \frac{(1620910 - 1507118) * 100}{1620910} = 7.02\%$ . Με αυτόν τον τρόπο πετυχαίνουμε δύο πράγματα: 1) Ελευθερώνεται η μνήμη RAM ώστε να μπορούν τα εκπαιδευτούν τα μοντέλα Βαθιάς Μάθησης, 2) εξισορροπείται περισσότερο το σύνολο δεδομένων εκπαίδευσης (balance dataset), οπότε μπορεί να οδηγήσει σε καλύτερη απόδοση των

<sup>10</sup><https://www.sleepfoundation.org/stages-of-sleep>

**μοντέλων Βαθιάς Μάθησης.** Με την περαιτέρω εξισορρόπηση του συνόλου δεδομένων, περιμένουμε καλά αποτελέσματα. Στο Σχήμα 2.4.3 παρατίθεται το αποτέλεσμα της τυχαίας υποδειγματοληψίας της κλάσης Light.



Σχήμα 2.4.3: Ιστόγραμμα συνόλου δεδομένων εκπαίδευσης (train set) για μήκος παραθύρωσης 101, μετά από υποδειγματοληψία της κλάσης Light.

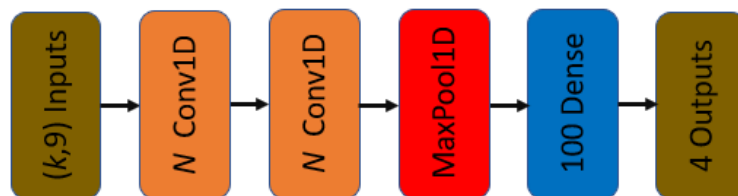
Οπότε, έχοντας ολοκληρώσει την κανονικοποίηση των δεδομένα εκπαίδευσης (train και validation) και τις τεχνικές Data Augmentation, μετατρέπουμε τις ετικέτες (labels) των δεδομένων σε δυαδική μορφή με τη μέθοδο One Hot Encoding επειδή οι αλγόριθμοι Deep Learning που έχουν υλοποιηθεί με την βιβλιοθήκη Keras <sup>11</sup>, χρειάζονται ως έξοδο τους οι ετικέτες (labels) να έχουν την δυαδική μορφή One Hot Encoding. Μετά, χωρίζουμε τα δεδομένα σε δεδομένα εκπαίδευσης και αξιολόγησης (train, validation), οπότε τα δεδομένα έχουν τα παρακάτω στοιχεία: Διαστάσεις συνόλου εκπαίδευσης  $\mathbf{X}_{train}$ : (1.134.637, k, 9), Διαστάσεις συνόλου αξιολόγησης  $\mathbf{X}_{val}$ : (486.273, k, 9), Διαστάσεις συνόλου εκπαίδευσης(labels) κλάσεων  $\mathbf{y}_{train}$ : (1.134.637, 4), Διαστάσεις συνόλου αξιολόγησης (labels) κλάσεων  $\mathbf{y}_{val}$ : (486.273, 4), όπου  $k$  το μήκος παραθύρωσης (21,51,101). Στα δεδομένα test όπως αναφέραμε προηγουμένως δεν εφαρμόζονται οι τεχνικές Data Augmentation.

Εφόσον έχουν ολοκληρωθεί οι αλλαγές που υλοποιήσαμε για το κομμάτι της προ-επεξεργασίας δεδομένων με την συνάρτηση κανονικοποίησης MinMax και τις τεχνικές Data Augmentation, ακολουθεί το κομμάτι του πειραματισμού με τα μοντέλα Βαθιάς Μάθησης (Deep Learning).

<sup>11</sup>Βιβλιοθήκη Keras: <https://keras.io>

## 2.4.2 Ταξινόμηση με Συνελικτικό Νευρωνικό Δίκτυο (CNN)

Το πρώτο μέρος του πειραματικού σταδίου αφορά την ταξινόμηση των δεδομένων στα 4 στάδια ύπνου με τη χρήση ενός Συνελικτικού Νευρωνικού Δικτύου. Συγκεκριμένα, η αρχιτεκτονική του δικτύου παρουσιάζεται στο Σχήμα 2.4.4.

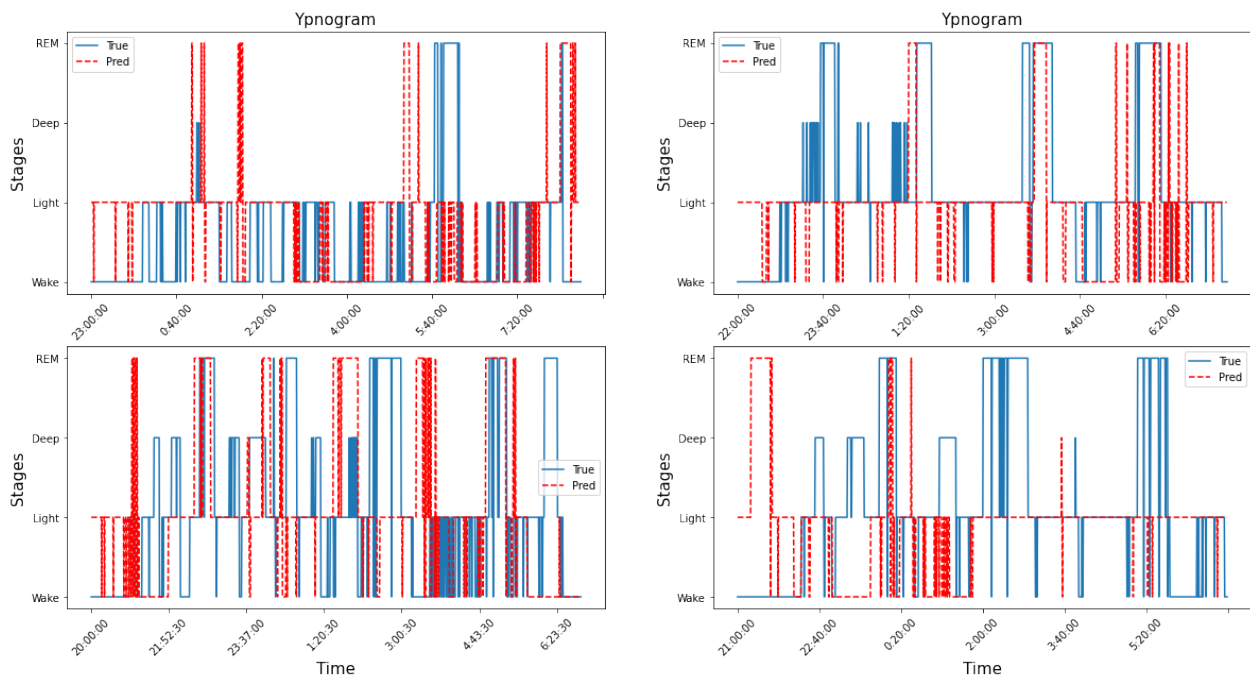


Σχήμα 2.4.4: Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου. Μεταβλητή των μοντέλων αποτελεί το μήκος παραθύρωσης (sliding window) και αναφέρεται ως  $k$ ,  $k \in (21, 51, 101)$ .

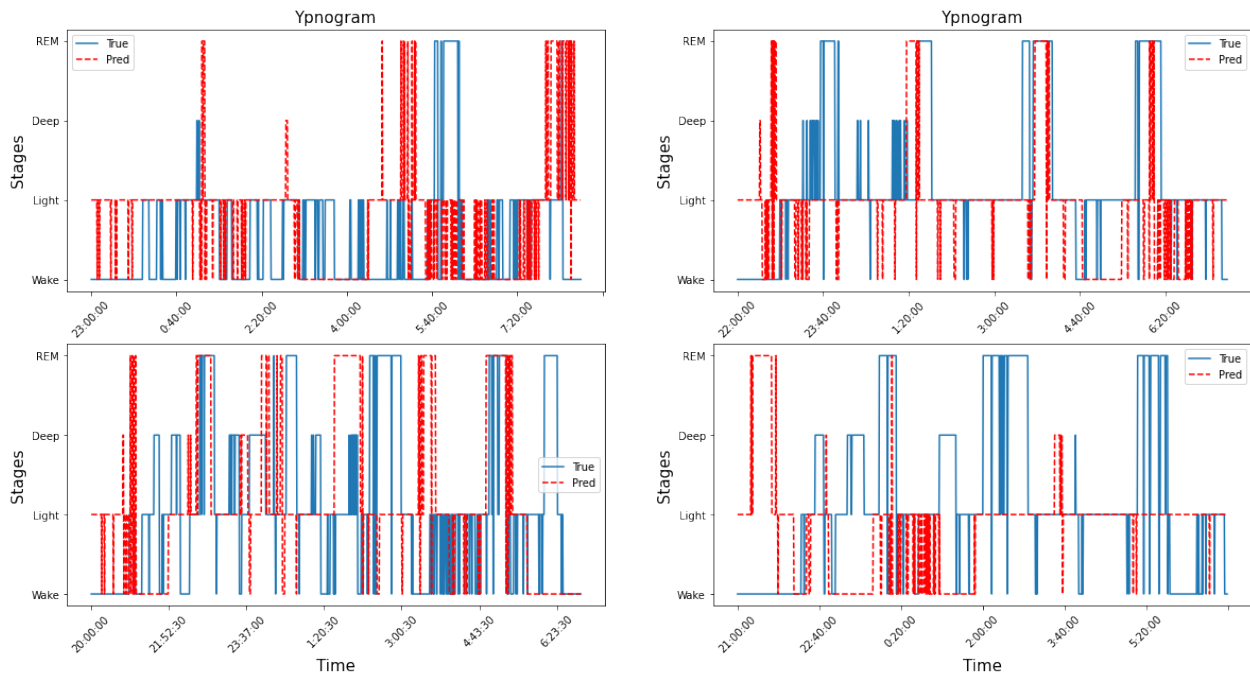
Από την αρχιτεκτονική του μοντέλου έχουμε ότι, το CNN μοντέλο δέχεται ως είσοδο δεδομένα με διαστάσεις  $(k, 9)$ . Η πρώτη διάσταση (αριθμός  $k$ ) απεικονίζει το μήκος παραθύρωσης (sliding window) των δεδομένων και η δεύτερη διάσταση τα  $8+1 = 9$  (ακτιγραφία, HRV) χαρακτηριστικά. Το μήκος παραθύρωσης όπως αναφέραμε στο προηγούμενο Κεφάλαιο 3.3, παίρνει τιμές  $(21, 51, 101)$ . Η έξοδος του μοντέλου έχει διάσταση  $(4,1)$  και απεικονίζει τα 4 στάδια του ύπνου. Το διάνυσμα αυτό της εξόδου είναι σε δυαδική μορφή (One Hot Encoding).

Λεπτομερέστερα, τα φίλτρα (filters/kernels) του επιπέδου συνέλιξης (1D Convolutional Layer) είναι  $N = 64$  σε πλήθος και έχουν μέγεθος 3 (kernel size) το κάθε ένα και stride ίσο με 1. Μετέπειτα, εφόσον το συνελικτικό επίπεδο έχει εξάγει χαρακτηριστικά από τα δεδομένα, τα τελευταία "περνούν" από την συνάρτηση ReLU ( $Relu(x) = \max(0, x)$ ), ακολουθεί η τεχνική Dropout με τιμή 50%, η οποία είναι τεχνική Regularization ώστε να γίνει σωστά η εκπαίδευση του μοντέλου και να αποφευχθεί το overfitting, δηλαδή το μοντέλο να "μάθει" πάρα πολύ καλά τα δεδομένα εκπαίδευσης και έτσι δεν θα μπορεί να προβλέπει σωστά καινούργια δεδομένα στα οποία δεν έχει εκπαιδευτεί. Έπειτα, ακολουθεί ένα επιπλέον επίπεδο συνέλιξης (1D Convolutional Layer) με  $N = 64$  αριθμό φίλτρων (filters/kernels) που έχουν μέγεθος 3 (kernel size) το κάθε ένα και stride ίσο με 1. Μετά από την εξαγωγή χαρακτηριστικών από τα δύο επίπεδα συνέλιξης, ακολουθεί η επιπέδωση των δεδομένων (Flatten) και η έξοδός τους με 4 νευρώνες, που αποτελούν τα 4 στάδια ύπνου.

Μετά τον πειραματισμό, προκύπτουν 3 μοντέλα, ένα για κάθε μήκος παραθύρωσης. Δηλαδή,  $CNN_{21}, CNN_{51}, CNN_{101}$ . Η εκπαίδευση των μοντέλων γίνεται για 100 εποχές, χρησιμοποιώντας batch size ίσο με 512. Επίσης, όπως αναφέραμε και στο Κεφάλαιο (3.3), το 20% των δεδομένων εκπαίδευσης χρησιμοποιείται για validation ώστε να αξιολογηθεί η απόδοση του μοντέλου κατά την εκπαίδευση. Τέλος, για την εκπαίδευση και βελτιστοποίηση των βαρών του δικτύου (weights), χρησιμοποιείται ο αλγόριθμος Adam με τις τυπικές τιμές που δίνεται από την βιβλιογραφία της βιβλιοθήκης Keras. Σύμφωνα με τα παραπάνω γίνεται η εκπαίδευση των CNN μοντέλων στο σύνολο δεδομένων εκπαίδευσης (train), και μετά γίνεται η πρόβλεψη/ταξινόμηση στο σύνολο δεδομένων αξιολόγησης (test). Στο Σχήμα 2.4.5 - 2.4.7 παρατίθεται τα υπνογράμματα τεσσάρων χρηστών του συνόλου αξιολόγησης (test), δηλαδή διαγράμματα πρόβλεψης/ταξινόμησης των σταδίων του ύπνου τεσσάρων χρηστών, στα τέσσερα στάδια Wake, Light, Deep και REM. Επιπλέον, τα αποτελέσματα της ταξινόμησης για τα διαφορετικά μήκη παραθύρωσης παρατίθενται στον Πίνακα 2.15 με τις κλασικές δημοφιλείς μετρικές. Τέλος, στους Πίνακες 2.16 - 2.18 παρουσιάζονται τα αποτελέσματα ταξινόμησης για κάθε στάδιο ύπνου (Wake, Light, Deep και REM) για τα τρία διαφορετικά μήκη παραθύρωσης και στα Σχήματα 2.4.8 - 2.4.10 παρουσιάζονται οι Πίνακες Σύγκρισης για τα διαφορετικά μήκη παραθύρωσης.

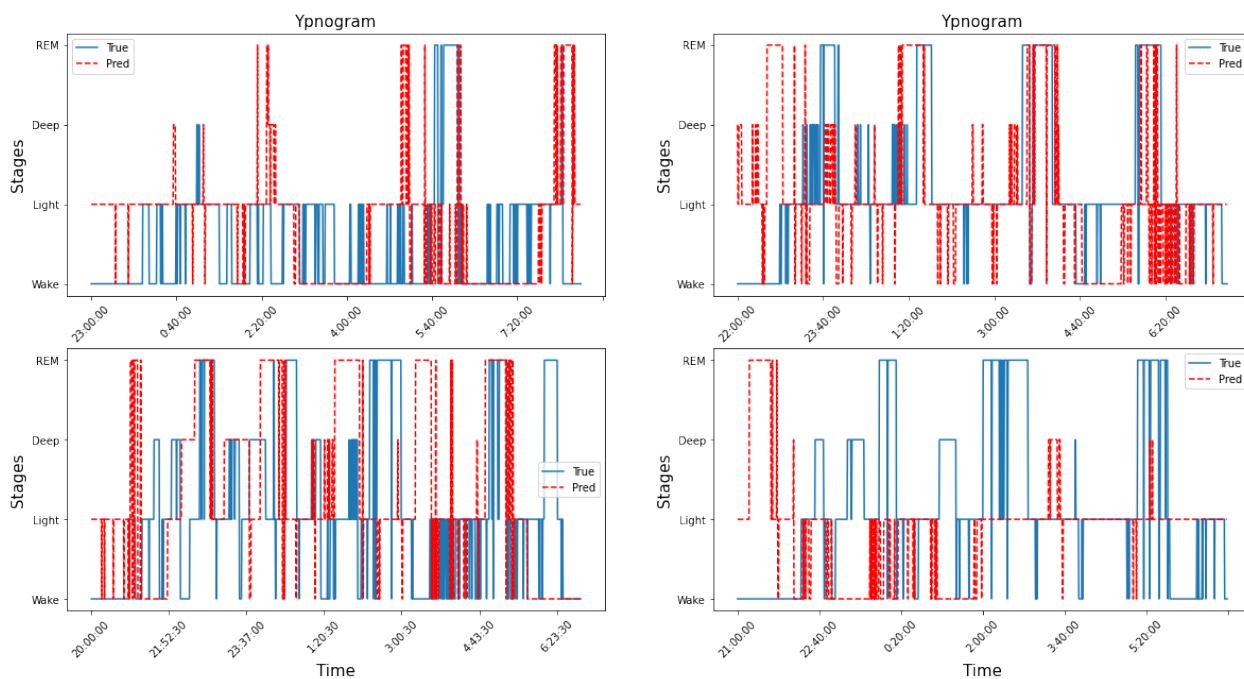


Σχήμα 2.4.5: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 21.



Σχήμα 2.4.6: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 51.

Εποπτικά, από τα υπνογράμματα του Σχήματος 2.4.5 - 2.4.7, παρατηρούμε ότι η κλάση Deep δεν μπορεί να αναγνωριστεί καλά από το συνελικτικό δίκτυο ενώ οι υπόλοιπες κλάσεις, κυρίως οι κλάσεις Wake



Σχήμα 2.4.7: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 101.

και Light, μπορούν να αναγνωριστούν καλύτερα αντίστοιχα με τις μοντελοποιήσεις του προηγούμενου Κεφαλαίου 2.3. Επίσης, σε σχέση με την προηγούμενη μοντελοποίηση με CNN που υλοποιήσαμε (Κεφάλαιο 2.3), παρατηρούμε ότι έχουμε περισσότερο επικαλυπτόμενα (True και Predictions) υπνογράμματα για όλες τις κλάσεις. Το συμπέρασμα αυτό μπορεί να διαγνωσθεί και από τους Πίνακες μετρικών του συνόλου αξιολόγησης (Πίνακας 2.16 - 2.18) και από τους Πίνακες σύγχυσης (Confusion Matrix - Σχήματα 2.4.8 - 2.4.10) όπου η κλάση Deep έχει τις χειρότερες αποδόσεις σε όλες τις μετρικές, ακολουθεί με καλύτερες αποδόσεις η κλάση REM, μετά η κλάση Wake, και τέλος τις υψηλότερες αποδόσεις τις έχει η κλάση Light.

<i>Metrics%</i>	<i>SeqLen21</i>	<i>SeqLen51</i>	<i>SeqLen101</i>
Accuracy	70.1 ± 0.9	71.5 ± 1.0	71.0 ± 1.0
Precision	60.4 ± 1.3	62.7 ± 1.4	62.0 ± 1.3
Recall	53.3 ± 0.9	55.3 ± 1.0	58.9 ± 1.1
F1-Score	51.7 ± 1.0	54.0 ± 1.0	56.9 ± 1.1
Cohen-Kappa	52.6 ± 1.7	55.6 ± 1.7	57.3 ± 1.8
Specificity	87.2 ± 0.4	87.9 ± 0.4	88.3 ± 0.4

Πίνακας 2.15: Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης.

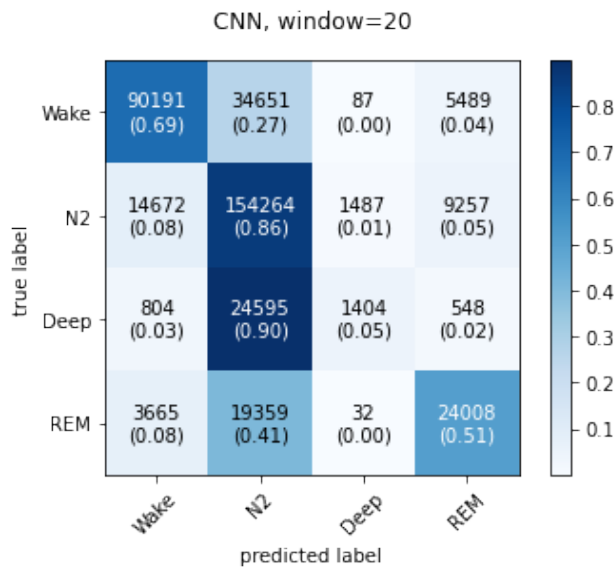
Λεπτομερέστερα, από τον Πίνακα μετρικών 2.15, συμπεραίνουμε ότι το μοντέλο  $CNN_{101}$  έχει τις υψηλότερες αποδόσεις σε όλες τις μετρικές. Η μεγάλη διαφορά αντικατοπτρίζεται στη μετρική F1-Score όπου τα μοντέλα  $CNN_{51}$  και  $CNN_{101}$  έχουν καλύτερα ποσοστά σε σχέση με το μοντέλο  $CNN_{21}$  (απόλυτη διαφορά **5.2%** μεταξύ  $CNN_{101}$  και  $CNN_{21}$ , ενώ στη προηγούμενη μοντελοποίηση με CNN πριν τις αλλαγές μας, η απόλυτη διαφορά ήταν **1%** μεταξύ  $CNN_{101}$  και  $CNN_{21}$ ). Δηλαδή,

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	82.0	69.0	75.0
Light	66.0	86.0	75.0
Deep	47.0	5.0	9.0
REM	61.0	51.0	56.0

Πίνακας 2.16: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 21.

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	82.0	74.0	78.0
Light	68.0	84.0	75.0
Deep	45.0	9.0	15.0
REM	64.0	53.0	58.0

Πίνακας 2.17: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 51.



Σχήμα 2.4.8: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 21.

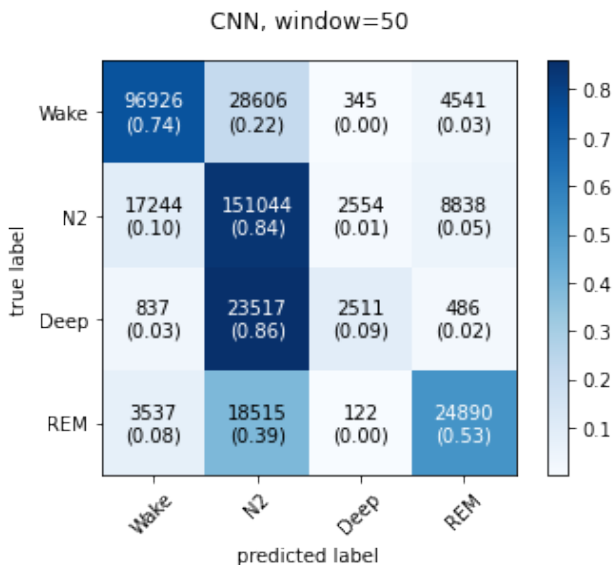
το μοντέλο  $CNN_{101}$  ταξινομεί καλύτερα τα δεδομένα διότι έχουν τα υψηλότερα ποσοστά Precision και Recall. Αντίστοιχα με την μετρική F1-Score πηγαίνει και η μετρική Cohen-Kappa και αποκαλύπτει την πραγματική απόδοση των μοντέλων εξαιτίας του μη ισορροπημένου συνόλου δεδομένων (imbalance dataset) όπως και η μετρική F1-Score.

Εξαιτίας αυτού του φαινομένου, παρατηρούμε τους Πίνακες Σύγχυσης 2.16 - 2.18 όπου παρουσιάζονται οι αποδόσεις των μοντέλων ανα κλάση. Είναι φανερό ότι τα μοντέλα, λόγω του μη ισορροπημένου συνόλου δεδομένων (imbalance dataset), δεν μπορούν να αναγνωρίσουν καλά την κλάση Deep (F1-Score  $\leq 30\%$ ) και ακολουθεί η κλάση REM (F1-Score  $\leq 60\%$ ). Επίσης, παρατηρούμε ότι το μοντέλο  $CNN_{101}$  είναι καλύτερο σε όλες τις μετρικές ανα κλάση, σε σχέση με τα μοντέλα  $CNN_{51}$  και  $CNN_{21}$ . Αν εστιάσουμε στην κλάση REM, Deep, παρατηρούμε ότι το μοντέλο  $CNN_{101}$  έχει τη



<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	82.0	74.0	78.0
Light	70.0	79.0	74.0
Deep	38.0	23.0	29.0
REM	61.0	60.0	61.0

Πίνακας 2.18: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 101.

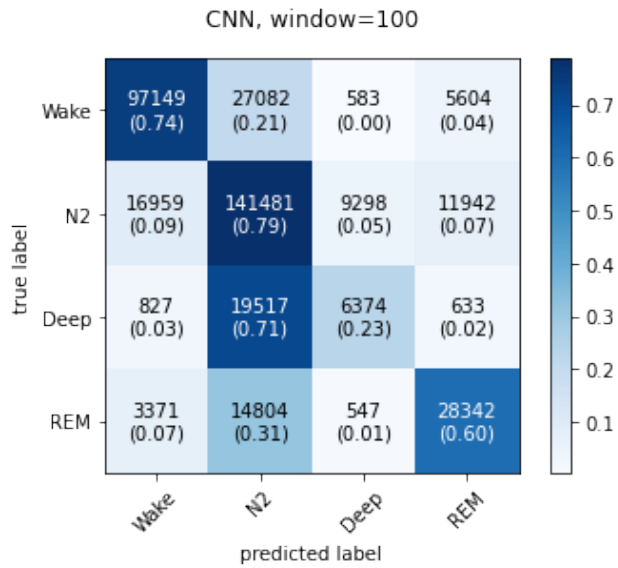


Σχήμα 2.4.9: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 51.

καλύτερη απόδοση F1-Score και Recall από όλα τα μοντέλα. Δηλαδή, 1) η μεγαλύτερη παραθύρωση και 2) η υποδειγματοληψία της κλάσης **Light** βοηθούν τις κλάσεις να αναγνωριστούν καλύτερα. Αυτό αποκαλύπτει ότι για αναγνωριστούν καλά τα στάδια του ύπνου, πρέπει τα μοντέλα να επεξεργάζονται μεγαλύτερα παράθυρα δεδομένων (sliding windows) ώστε να αναγνωρίσουν τα long-term χαρακτηριστικά που διαχωρίζουν τα στάδια του ύπνου αλλά και η υποδειγματοληψία της κλάσης **Light** δίνει την δυνατότητα στο μοντέλο να μην μαθαίνει τόσο καλά τη κλάση αυτή και να μαθαίνει καλύτερα τις υπόλοιπες κλάσεις οι οποίες έχουν μικρότερο πλήθος δεδομένων.

Τα παραπάνω συμπεράσματα συνοψίζονται στους Πίνακες Σύγχυσης 2.4.8 - 2.4.10. Δηλαδή, όλα τα μοντέλα δυσκολεύονται να αναγνωρίσουν τη κλάση **Deep** και αναγνωρίζουν πολύ καλά τις κλάσεις **Wake** και **Light** και **REM**, παρατηρώντας τα ποσοστά Recall της διαγωνίου. Το γεγονός αυτό είναι λογικό εφόσον το σύνολο δεδομένων είναι μη ισορροπημένο. Οπότε, περισσότερα δεδομένα από μια κλάση, διαφορετικά μεταξύ τους, τόσο καλύτερα τα μοντέλα θα την αναγνωρίσουν. Στο τέλος του κεφαλαίου θα συγκρίνουμε τα βελτιωμένα μοντέλα με τα μοντέλα του Κεφαλαίου 2.3.

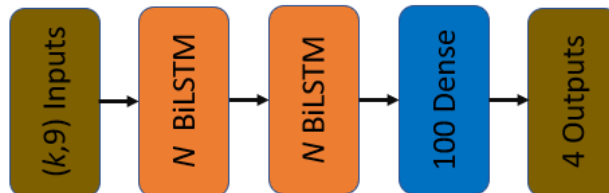
Στην επόμενη ενότητα, θα παρουσιάσουμε τα μοντέλα LSTM και θα παραθέσουμε τα αντίστοιχα αποτελέσματα.



Σχήμα 2.4.10: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 101.

### 2.4.3 Ταξινόμηση με Νευρωνικό Δίκτυο Μακροπρόθεσμης Μνήμης Δύο Κατευθύνσεων (Bidirectional LSTM)

Στη παρούσα ενότητα θα παρουσιάσουμε τα πειράματα που πραγματοποιήθηκαν για την ταξινόμηση των δεδομένων στα 4 στάδια ύπνου με τη χρήση ενός Νευρωνικού Δικτύου Μακροπρόθεσμης Μνήμης Δύο Κατευθύνσεων (Bidirectional LSTM - BiLSTM). Συγκεκριμένα, η αρχιτεκτονική του δικτύου παρουσιάζεται στο Σχήμα 2.4.11.



Σχήμα 2.4.11: Αρχιτεκτονική Νευρωνικού Δικτύου Μακροπρόθεσμης Μνήμης Δύο Κατευθύνσεων. Μεταβλητή των μοντέλων αποτελεί το μήκος παραθύρωσης (sliding window) και αναφέρεται ως  $k$ ,  $k \in (21, 51, 101)$ .

Από την αρχιτεκτονική του μοντέλου έχουμε ότι, το BiLSTM μοντέλο δέχεται ως είσοδο δεδομένα με διαστάσεις  $(k, 9)$  όπως και στην περίπτωση του Συνελικτικού Νευρωνικού Δικτύου. Η πρώτη διάσταση (αριθμός  $k$ ) απεικονίζει το μήκος παραθύρωσης (sliding window) των δεδομένων και η δεύτερη διάσταση τα  $8+1 = 9$  (ακτιγραφία, HRV) χαρακτηριστικά. Το μήκος παραθύρωσης όπως αναφέραμε στο προηγούμενο Κεφάλαιο 3.4, παίρνει τιμές  $(21, 51, 101)$ . Η έξοδος του μοντέλου έχει διάσταση  $(4,1)$  και απεικονίζει τα 4 στάδια του ύπνου. Το διάνυσμα αυτό της εξόδου είναι σε δυαδική μορφή (One Hot Encoding).

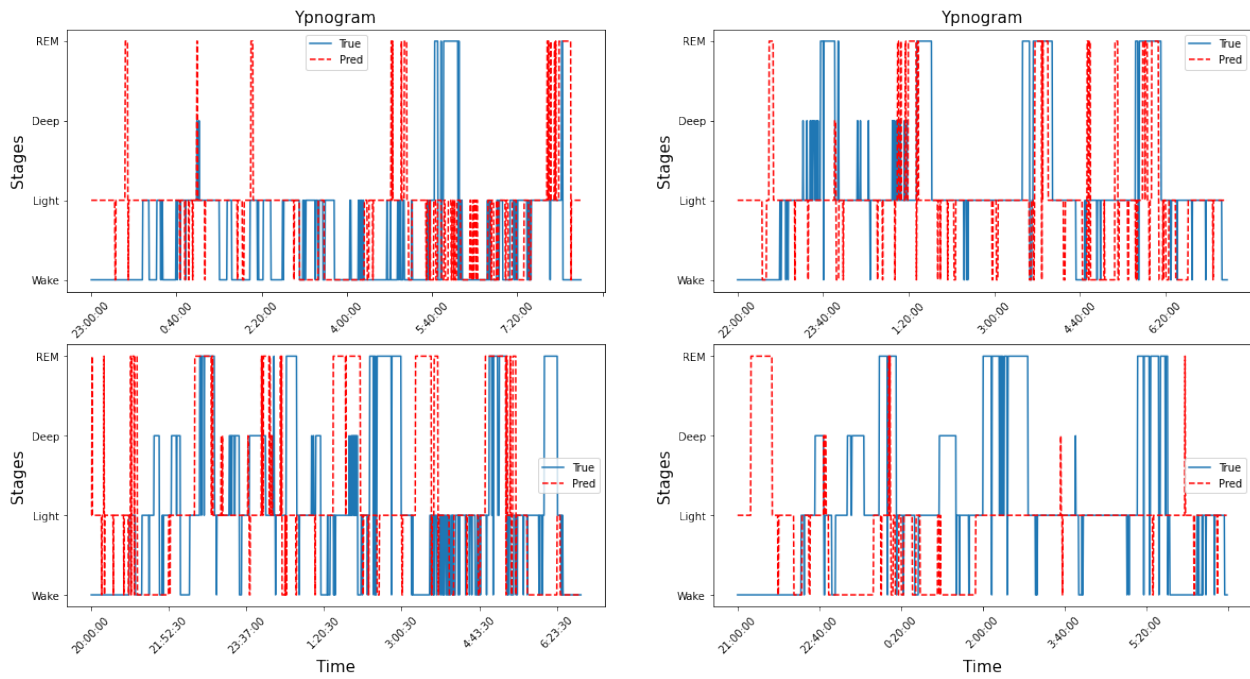
Λεπτομερέστερα, το μέγεθος των κρυμμένων μονάδων (hidden units  $h_t$ ) του επιπέδου μακροπρόθεσμης μνήμης (BiLSTM Layer) είναι  $N = 64$ . Μετέπειτα, εφόσον το επίπεδο μακροπρόθεσμης μνήμης δύο

κατευθύνσεων έχει εξαγάγει χαρακτηριστικά από τα δεδομένα εισόδου, η έξοδος **όλων** των κρυμμένων μονάδων  $\mathbf{h}_t, \forall t$  "περνούν" από την συνάρτηση υπερβολικής εφαπτομένης  $\tanh(\mathbf{x})$ . Ακολουθεί η τεχνική Dropout, με τιμή 20%, η οποία είναι τεχνική Regularization ώστε να γίνει σωστά η εκπαίδευση του μοντέλου και να αποφευχθεί το overfitting, δηλαδή το μοντέλο να "μάθει" πάρα πολύ καλά τα δεδομένα εκπαίδευσης και έτσι δεν θα μπορεί να προβλέπει σωστά καινούργια δεδομένα στα οποία δεν έχει εκπαιδευτεί. Στη συνέχεια, τα δεδομένα "περνούν" και από δεύτερο επίπεδο μακροπρόθεσμης μνήμης δύο κατευθύνσεων (BiLSTM Layer) με  $N = 64$  αριθμό κρυμμένων μονάδων, αλλά αυτό το δεύτερο επίπεδο μακροπρόθεσμης μνήμης δύο κατευθύνσεων επιστρέφει ως έξοδο μόνο την έξοδο της τελευταίας κρυμμένης μονάδας  $\mathbf{h}_{final}, \mathbf{t}_{final}$  και ακολουθεί όπως και στο προηγούμενο επίπεδο η συνάρτηση υπερβολικής εφαπτομένης  $\tanh(\mathbf{x})$  και εφαρμόζεται και εδώ η τεχνική Regularization, δηλαδή η τεχνική Dropout, με τιμή 20%. Εφόσον έχει γίνει η εξαγωγή των χαρακτηριστικών από τα επίπεδα μακροπρόθεσμης μνήμης δύο κατευθύνσεων (BiLSTM Layer), ακολουθεί ένα feed-forward νευρωνικό δίκτυο με 100 νευρώνες (Dense Layer) και μετά η έξοδος με 4 νευρώνες, που αποτελούν τα 4 στάδια ύπνου.

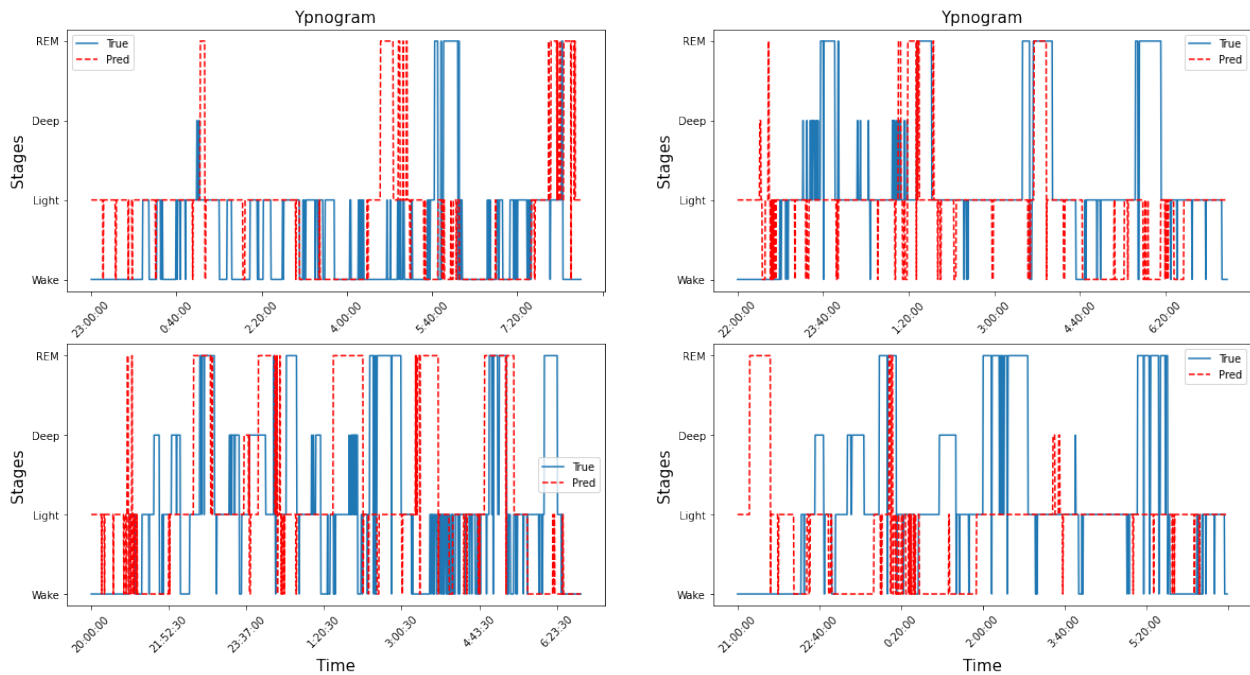
Μετά τον πειραματισμό, προκύπτουν 3 μοντέλα, ένα για κάθε μήκος παραθύρωσης. Δηλαδή, **BiLSTM<sub>21</sub>**, **BiLSTM<sub>51</sub>**, **BiLSTM<sub>101</sub>**. Η εκπαίδευση των μοντέλων γίνεται για 100 εποχές, χρησιμοποιώντας batch size ίσο με 512. Επίσης, όπως αναφέραμε και στο Κεφάλαιο 2.3, το 20% των δεδομένων εκπαίδευσης χρησιμοποιείται για validation ώστε να αξιολογηθεί η απόδοση του μοντέλου κατά την εκπαίδευση. Τέλος, για την εκπαίδευση και βελτιστοποίηση των βαρών του δικτύου (weights), χρησιμοποιείται ο αλγόριθμος Adam με τις τυπικές τιμές που δίνεται από την βιβλιογραφία της βιβλιοθήκης Keras. Σύμφωνα με τα παραπάνω γίνεται η εκπαίδευση των BiLSTM μοντέλων στο σύνολο δεδομένων εκπαίδευσης (train), και μετά γίνεται η πρόβλεψη/ταξινόμηση στο σύνολο δεδομένων αξιολόγησης (test). Στα Σχήματα 2.4.12 - 2.4.14 παρατίθενται τα υπνογράμματα τεσσάρων χρηστών του συνόλου αξιολόγησης (test), δηλαδή διαγράμματα πρόβλεψης/ταξινόμησης των σταδίων του ύπνου τεσσάρων χρηστών, στα τέσσερα στάδια Wake, Light, Deep και REM. Επιπλέον, τα αποτελέσματα της ταξινόμησης για τα διαφορετικά μήκη παραθύρωσης παρατίθενται στον Πίνακα 3.19 με τις κλασικές δημοφιλείς μετρικές. Τέλος, στους Πίνακες 3.20 - 3.22 παρουσιάζονται τα αποτελέσματα ταξινόμησης για κάθε στάδιο ύπνου (Wake, Light, Deep και REM) για τα τρία διαφορετικά μήκη παραθύρωσης και στα Σχήματα 2.4.15 - 2.4.17 παρουσιάζονται οι Πίνακες Σύγχυσης για τα διαφορετικά μήκη παραθύρωσης.

Εποπτικά, από τα υπνογράμματα του Σχήματος 2.4.12 - 2.4.14, παρατηρούμε τα ίδια αποτελέσματα με το Συνελικτικό Δίκτυο CNN, που αναπτύξαμε προηγουμένως. Δηλαδή, ότι η κλάση Deep δεν μπορεί να αναγνωριστεί καλά από το νευρωνικό δίκτυο μακροπρόθεσμης μνήμης ενώ οι υπόλοιπες κλάσεις, κυρίως οι κλάσεις Wake και Light, μπορούν να αναγνωριστούν καλύτερα και ότι η κλάση Light αναγνωρίζεται πολύ καλά από το μοντέλο. Άρα υπάρχει μια συνέπεια στα αποτελέσματά μας, γεγονός σημαντικό. Επίσης, σε σχέση με την προηγούμενη μοντελοποίηση με LSTM που υλοποιήσαμε, παρατηρούμε ότι έχουμε περισσότερο επικαλυπτόμενα (True και Predictions) υπνογράμματα για όλες τις κλάσεις. Τα συμπεράσματα αυτά μπορούν να διαγνωσθούν και από τους Πίνακες μετρικών του συνόλου αξιολόγησης (Πίνακας 2.20 - 2.22) και από τους Πίνακες σύγχυσης (Confusion Matrix - Σχήματα 2.4.15 - 2.4.17) όπου η κλάση Deep έχει τις χειρότερες αποδόσεις σε όλες τις μετρικές, ακολουθεί με καλύτερες αποδόσεις η κλάση REM, μετά η κλάση Wake, και τέλος τις υψηλότερες αποδόσεις τις έχει η κλάση Light.

Λεπτομερέστερα, από τον Πίνακα μετρικών 2.19, συμπεραίνουμε ότι το μοντέλο **BiLSTM<sub>101</sub>** έχει τις υψηλότερες αποδόσεις σε όλες τις μετρικές, αντίστοιχα με το μοντέλο **CNN<sub>101</sub>**. Παρατηρούμε ότι όλες οι μετρικές έχουν παραπλήσια ποσοστά για όλα τα μήκη παραθύρωσης των δεδομένων, με καλύτερο το μοντέλο **LSTM<sub>101</sub>**. Οι μετρικές Precision και Recall είναι της τάξης του 63 % και 61 % αντίστοιχα και γι αυτό και η μετρική F1-Score κυμαίνεται στα ίδια επίπεδα. Λόγω του μη ισορροπημένου συνόλου δεδομένων, η μετρική Accuracy δεν είναι αξιόπιστη για την αξιολόγηση των μοντέλων. Επίσης,

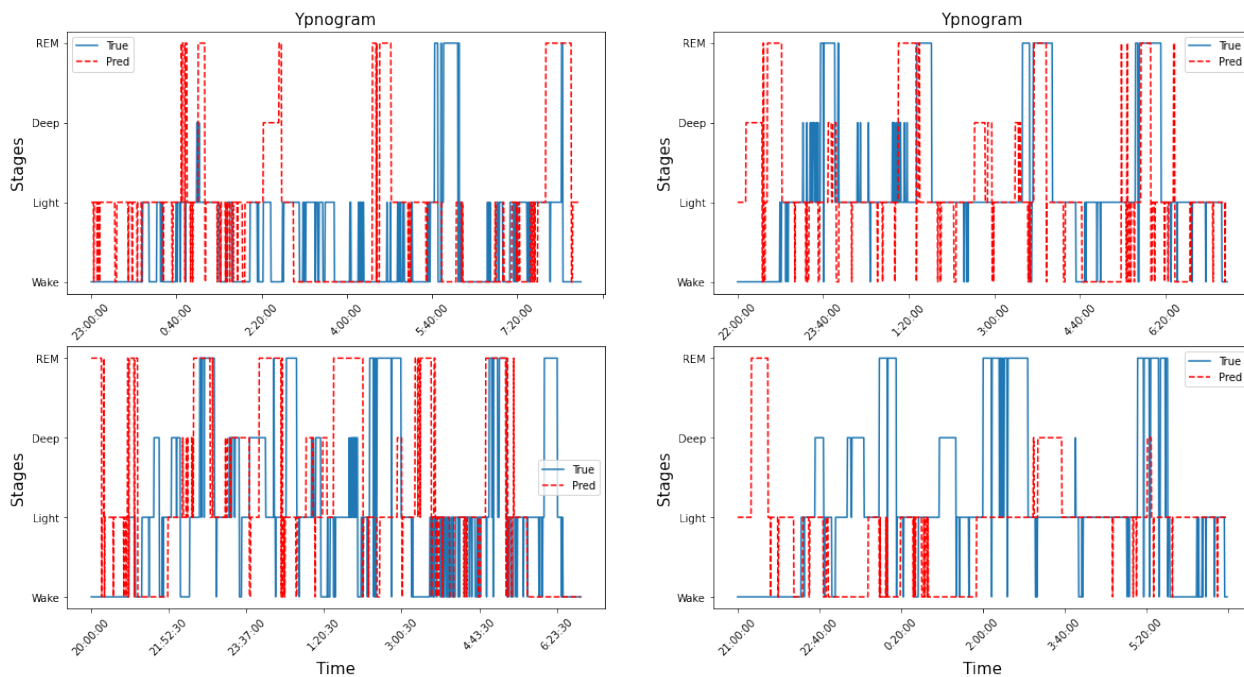


Σχήμα 2.4.12: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 21.

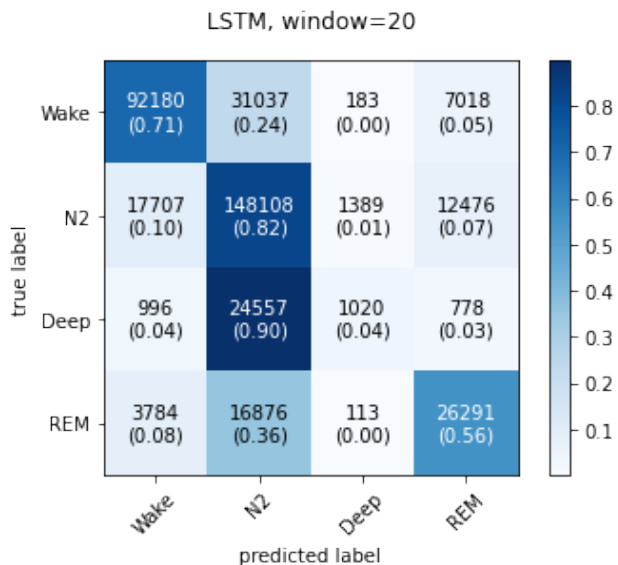


Σχήμα 2.4.13: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 51.

παρατηρούμε ότι η μετρική Specificity είναι υψηλή σε όλα τα μοντέλα BiLSTM. Το γεγονός αυτό καταδεικνύει ότι είναι μικρό το ποσοστό των λάθος ταξινομημένων δεδομένων, γεγονός που είναι



Σχήμα 2.4.14: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 101.



Σχήμα 2.4.15: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 21.

σημαντικό.

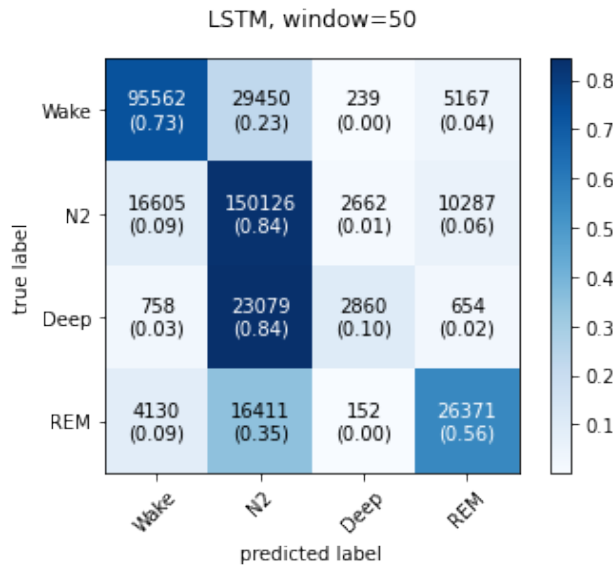
Εξαιτίας αυτού του φαινομένου, παρατηρούμε τους Πίνακες Σύγχυσης 2.20 - 2.22 όπου παρουσιάζονται οι αποδόσεις των μοντέλων ανα κλάση. Είναι φανερό ότι τα μοντέλα, λόγω του μη ισορροπημένου συνόλου δεδομένων, δεν μπορούν να αναγνωρίσουν σχεδόν καθόλου την κλάση Deep (F1-Score <=

<i>Metrics%</i>	<i>SeqLen21</i>	<i>SeqLen51</i>	<i>SeqLen101</i>
Accuracy	69.5 ± 1.0	71.4 ± 1.0	71.7 ± 1.0
Precision	58.0 ± 1.4	62.1 ± 1.4	63.3 ± 1.3
Recall	53.8 ± 0.9	56.0 ± 1.1	61.2 ± 1.2
F1-Score	51.7 ± 0.9	54.5 ± 1.1	58.7 ± 1.2
Cohen-Kappa	52.5 ± 1.7	55.4 ± 1.8	58.9 ± 1.8
Specificity	87.2 ± 0.4	87.9 ± 0.4	88.7 ± 0.4

Πίνακας 2.19: Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης.

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	80.0	71.0	75.0
Light	67.0	82.0	74.0
Deep	38.0	4.0	7.0
REM	56.0	56.0	56.0

Πίνακας 2.20: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 21.



Σχήμα 2.4.16: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 51.

35 %) και ακολουθεί η κλάση REM (F1-Score ≤ 61 %). Επίσης, παρατηρούμε ότι τα μοντέλα *BiLSTM*<sub>51</sub> και *BiLSTM*<sub>101</sub> είναι καλύτερα σε όλες τις μετρικές ανα κλάση, σε σχέση με το μοντέλο *BiLSTM*<sub>21</sub>, ενώ στη προηγούμενη μοντελοποίηση με απλό LSTM μοντέλο το καλύτερο μοντέλο ήταν το *LSTM*<sub>51</sub>, γεγονός που έρχεται σε αντίθεση με τα μοντέλα *BiLSTM*. Σε γενικά όμως πλαίσια, ακολουθείται η λογική ότι όσο μεγαλύτερο το μήκος παραθύρωσης, βελτιώνονται οι αποδόσεις των μοντέλων τόσο στα μοντέλα LSTM όσο και στα CNN. Επιπλέον, αν εστιάσουμε στην κλάση REM, Deep και Wake, παρατηρούμε ότι το μοντέλο *BiLSTM*<sub>101</sub> έχει τη καλύτερη απόδοση F1-Score και

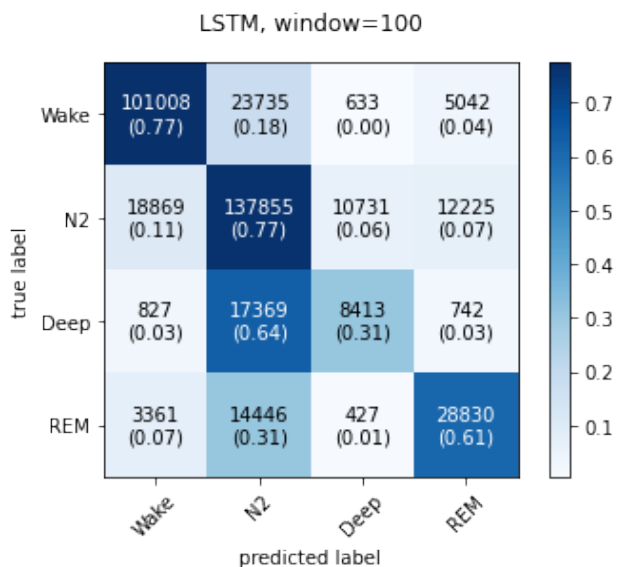


<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	82.0	73.0	77.0
Light	69.0	84.0	75.0
Deep	48.0	10.0	17.0
REM	62.0	56.0	59.0

Πίνακας 2.21: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 51.

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	81.0	77.0	79.0
Light	71.0	77.0	74.0
Deep	42.0	31.0	35.0
REM	62.0	61.0	61.0

Πίνακας 2.22: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 101.



Σχήμα 2.4.17: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 101.

Recall από όλα τα μοντέλα. Δηλαδή, η βέλτιστη παραθύρωση είναι αυτή με μήκος 101 και βοηθά περισσότερο τις κλάσεις να αναγνωριστούν από ότι τα μήκη παραθύρωσης 21 και 51, παρατηρώντας ειδικά για τις μετρικές F1-Score και Recall. Ακόμα, όπως και την προηγούμενη μοντελοποίηση με CNN, 1) η μεγαλύτερη παραθύρωση και 2) η υποδειγματοληψία της κλάσης Light βοηθούν τις κλάσεις να αναγνωριστούν καλύτερα.

Τα παραπάνω συμπεράσματα συνοψίζονται στους Πίνακες Σύγχυσης 2.4.15 - 2.4.17. Δηλαδή, όλα τα μοντέλα δυσκολεύονται να αναγνωρίσουν τις κλάσεις Deep και αναγνωρίζουν καλύτερα τη κλάση REM και πολύ καλά τις κλάσεις Wake και Light, παρατηρώντας τα ποσοστά Recall της διαγωνίου. Το γεγονός αυτό είναι λογικό εφόσον το σύνολο δεδομένων είναι μη ισορροπημένο. Οπότε, περισσότερα

δεδομένα από μια κλάση, διαφορετικά μεταξύ τους, τόσο καλύτερα τα μοντέλα θα την αναγνωρίσουν.

Στην επόμενη ενότητα, θα παρουσιάσουμε τα μοντέλα με διαφορετικό μήκος παραθύρωσης και θα παραθέσουμε τα αντίστοιχα αποτελέσματα.

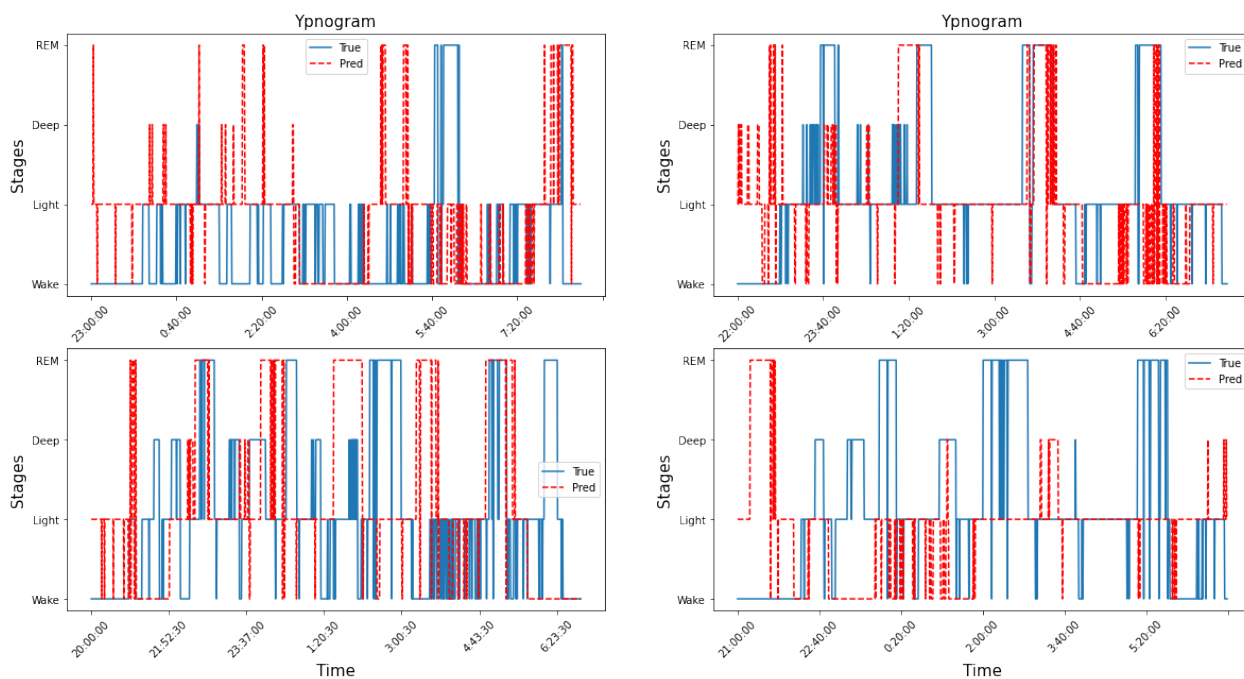
#### 2.4.4 Ταξινόμηση με διαφορετικό μήκος παραθύρωσης (Sliding Window)

Στη παρούσα ενότητα θα παρουσιάσουμε τα πειράματα που πραγματοποιήθηκαν για την ταξινόμηση των δεδομένων στα 4 στάδια ύπνου με μήκος παραθύρωσης 71. Δηλαδή, εκτός των μηκών παραθύρωσης 21, 51 και 101, δοκιμάζουμε και το μήκος παραθύρωσης 71 βασιζόμενοι στο συμπέρασμα ότι όσο αυξάνουμε το μήκος παραθύρωσης, υπάρχουν ενδείξεις από τους πειραματισμούς των Εννοτήτων 3.4.4 και 3.4.5 ότι βελτιώνεται η απόδοση των μοντέλων Βαθιάς Μάθησης σε όλες τις δημοφιλείς μετρικές που παρουσιάζουμε. Συνοπτικά, στη παρούσα ενότητα παρουσιάζονται οι πειραματισμοί μας για την ταξινόμηση των δεδομένων ακτιγραφίας και HRV χαρακτηριστικών με τη χρήση ενός Νευρωνικού Δικτύου Μακροπρόθεσμης Μνήμης Δύο Κατευθύνσεων (Bidirectional LSTM - BiLSTM) και ενός Συνελικτικού Νευρωνικού Δικτύου (CNN). Τα μοντέλα αυτά έχουν την αρχιτεκτονική που παρουσιάστηκαν στις Ενότητες 2.4.5 και 2.4.4 για τα μοντέλα BiLSTM και CNN αντίστοιχα. Συγκεκριμένα, η αρχιτεκτονική του Συνελικτικού Νευρωνικού δικτύου,  $CNN_{71}$ , παρουσιάζεται στο Σχήμα 2.4.4 και του Νευρωνικού Δικτύου Μακροπρόθεσμης Μνήμης Δύο Κατευθύνσεων,  $BiLSTM_{71}$ , στο Σχήμα 2.4.11.

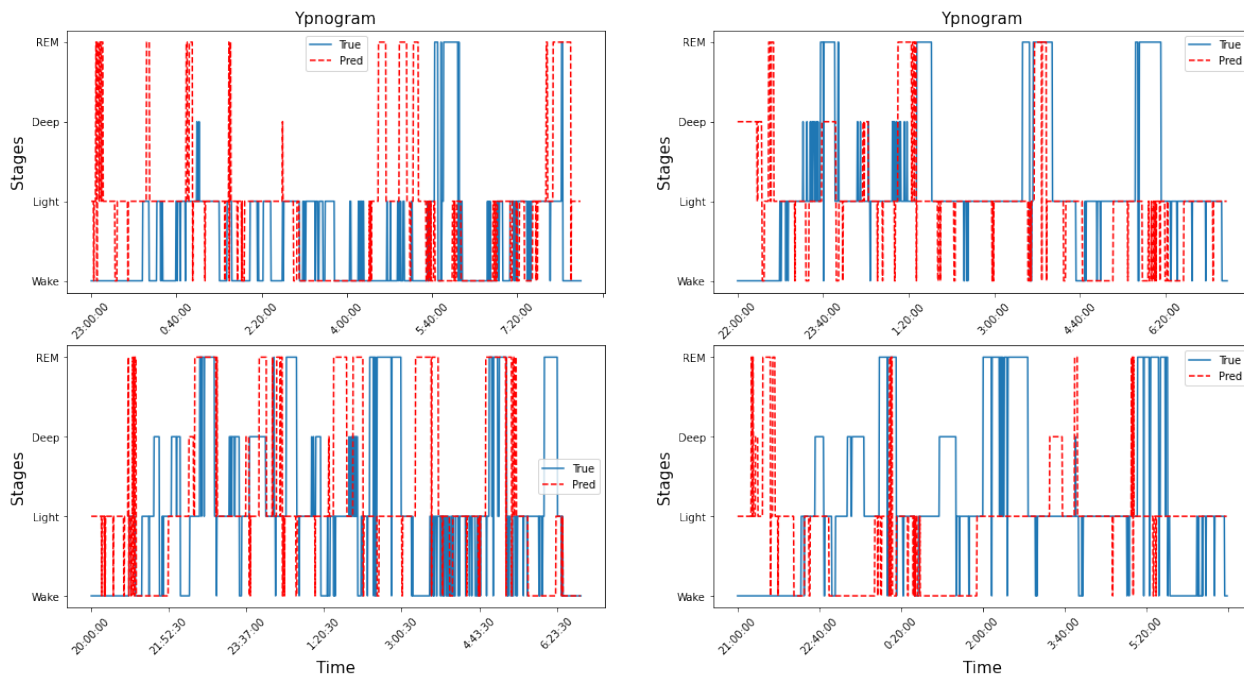
Συνοπτικά, από την αρχιτεκτονική των μοντέλων έχουμε ότι, τα CNN και BiLSTM μοντέλα δέχονται ως είσοδο δεδομένα με διαστάσεις (71, 9). Η πρώτη διάσταση απεικονίζει το μήκος παραθύρωσης (sliding window) των δεδομένων και η δεύτερη διάσταση τα  $8+1 = 9$  (ακτιγραφία, HRV) χαρακτηριστικά. Η έξοδος του μοντέλου έχει διάσταση (4,1) και απεικονίζει τα 4 στάδια του ύπνου. Το διάλυσμα αυτό της εξόδου είναι σε δυαδική μορφή (One Hot Encoding). Η λεπτομερής ανάλυση της αρχιτεκτονικής των μοντέλων CNN και BiLSTM παρουσιάζεται στις Ενότητες 3.4.4 και 3.4.5 αντίστοιχα.

Μετά τον πειραματισμό, προκύπτουν 2 μοντέλα, δηλαδή,  $CNN_{71}$ ,  $BiLSTM_{71}$ . Η εκπαίδευση των μοντέλων γίνεται για 100 εποχές, χρησιμοποιώντας batch size ίσο με 512. Επίσης, όπως αναφέραμε και στο Κεφάλαιο 3.3, το 20% των δεδομένων εκπαίδευσης χρησιμοποιείται για validation ώστε να αξιολογηθεί η απόδοση του μοντέλου κατά την εκπαίδευση. Τέλος, για την εκπαίδευση και βελτιστοποίηση των βαρών του δικτύου (weights), χρησιμοποιείται ο αλγόριθμος Adam με τις τυπικές τιμές που δίνονται από την βιβλιογραφία της βιβλιοθήκης Keras. Σύμφωνα με τα παραπάνω γίνεται η εκπαίδευση των μοντέλων στο σύνολο δεδομένων εκπαίδευσης (train), και μετά γίνεται η πρόβλεψη/ταξινόμηση στο σύνολο δεδομένων αξιολόγησης (test). Στα Σχήματα 2.4.18 και 2.4.19 παρατίθενται τα υπνογράμματα τεσσάρων χρηστών του συνόλου αξιολόγησης (test), δηλαδή διαγράμματα πρόβλεψης/ταξινόμησης των σταδίων του ύπνου τεσσάρων χρηστών, στα τέσσερα στάδια Wake, Light, Deep και REM για τα δύο μοντέλα CNN και LSTM. Επιπλέον, τα αποτελέσματα της ταξινόμησης για μήκος παραθύρωσης 71, παρατίθενται στον Πίνακα 2.23 με τις κλασικές δημοφιλείς μετρικές. Τέλος, στους Πίνακες 2.24 και 2.25 παρουσιάζονται τα αποτελέσματα ταξινόμησης για κάθε στάδιο ύπνου (Wake, Light, Deep και REM) για τα δύο μοντέλα και στα Σχήματα 2.4.20 - 2.4.21 παρουσιάζονται οι Πίνακες Σύγχυσης για τα διαφορετικά μήκη παραθύρωσης.

Εποπτικά, από τα υπνογράμματα του Σχήματος 2.4.18 - 2.4.19, παρατηρούμε ότι η κλάση Deep δεν μπορεί να αναγνωρισθεί καλά από το νευρωνικό δίκτυο μακροπρόθεσμης μνήμης ενώ οι υπόλοιπες κλάσεις, κυρίως οι κλάσεις Wake και Light, μπορούν να αναγνωριστούν καλύτερα (αφού επικαλύπτονται οι δύο γραμμές True και Predictions). Η κλάση REM αναγνωρίζεται σε πολύ καλό βαθμό, καλύτερα προφανώς από την κλάση Deep. Το συμπέρασμα αυτό μπορεί να διαγνωσθεί και από τους Πίνακες μετρικών του συνόλου αξιολόγησης (Πίνακες 2.24 - 2.25) και από τους Πίνακες σύγχυσης (Confusion Matrix - Σχήματα 2.4.20 - 2.4.21) όπου η κλάση Deep έχει τις χειρότερες αποδόσεις σε όλες



Σχήμα 2.4.18: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 71. Μοντέλο CNN.



Σχήμα 2.4.19: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 71. Μοντέλο LSTM.

τις μετρικές, ακολουθεί με καλύτερες αποδόσεις η κλάση REM, μετά η κλάση Wake, και τέλος τις υψηλότερες αποδόσεις τις έχει η κλάση Light.

<i>Metrics%</i>	<i>CNN<sub>71</sub></i>	<i>LSTM<sub>71</sub></i>
Accuracy	71.9 ± 1.0	71.8 ± 1.0
Precision	63.1 ± 1.4	63.2 ± 1.3
Recall	57.5 ± 1.1	58.1 ± 1.1
F1-Score	55.8 ± 1.1	56.6 ± 1.1
Cohen-Kappa	57.3 ± 1.8	57.8 ± 1.8
Specificity	88.2 ± 0.4	88.3 ± 0.4

Πίνακας 2.23: Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά τα μοντέλα CNN και LSTM με μήκος παραθύρωσης 71.

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	81.0	76.0	78.0
Light	70.0	82.0	75.0
Deep	45.0	14.0	22.0
REM	63.0	58.0	61.0

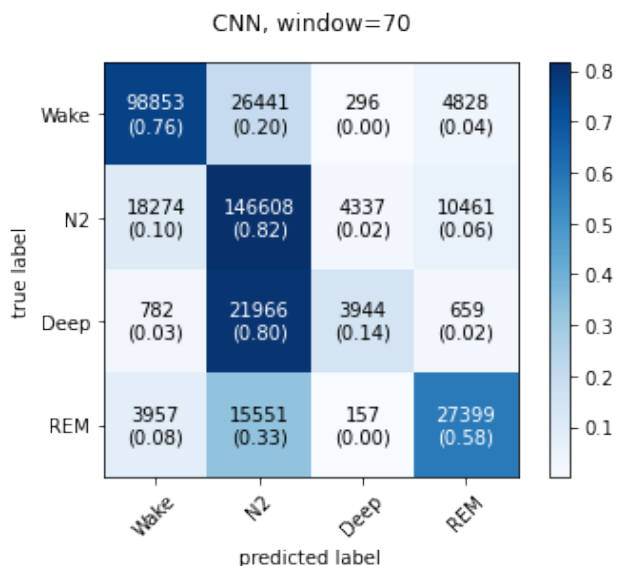
Πίνακας 2.24: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση για το μοντέλο CNN. Μήκος παραθύρωσης 71.

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	83.0	73.0	78.0
Light	69.0	83.0	75.0
Deep	42.0	16.0	24.0
REM	63.0	60.0	62.0

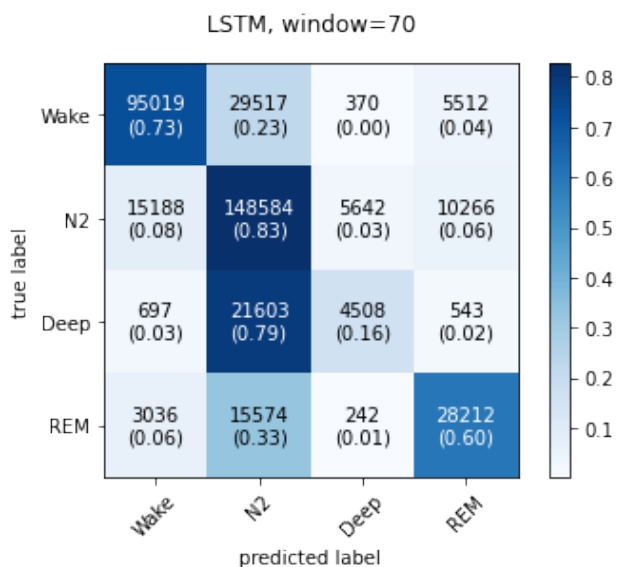
Πίνακας 2.25: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 51.

Λεπτομερέστερα, από τον Πίνακα μετρικών 2.23, συμπεραίνουμε ότι το μοντέλο *BiLSTM<sub>71</sub>* έχει τις υψηλότερες αποδόσεις σε όλες τις μετρικές, με μικρές διαφορές από τις αποδόσεις του μοντέλου *CNN<sub>71</sub>*. Οι μετρικές Precision και Recall είναι της τάξης του 63 % και 58 % αντίστοιχα και γι αυτό και η μετρική F1-Score κυμαίνεται στα ίδια επίπεδα 56.6%. Επιπλέον, η μετρική F1-Score και Cohen-Kappa είναι υψηλή σε σχέση με όλα τα προηγούμενα μοντέλα που πειραματιστήκαμε και αποκαλύπτουν την πραγματική απόδοση των μοντέλων εξαιτίας του μη ισορροπημένου συνόλου δεδομένων (imbalance dataset).

Εξαιτίας αυτού του φαινομένου, παρατηρούμε τους Πίνακες Σύγκρισης 2.24 - 2.25 όπου παρουσιάζονται οι αποδόσεις των μοντέλων ανα κλάση. Είναι φανερό ότι τα μοντέλα, λόγω του μη ισορροπημένου συνόλου δεδομένων (imbalance dataset), δεν μπορούν να αναγνωρίσουν σχεδόν καθόλου την κλάση Deep (F1-Score ≤ 24 %) και ακολουθεί η κλάση REM. Αν εστιάσουμε στην κλάση REM και Deep, παρατηρούμε ότι το μοντέλο *BiLSTM<sub>71</sub>* έχει τη καλύτερη απόδοση F1-Score και Recall από το μοντέλο *CNN<sub>71</sub>*, οπότε είναι λίγο καλύτερο το μοντέλο *BiLSTM<sub>71</sub>*. Επίσης, παρατηρούμε ότι η παραθύρωση των δεδομένων εκπαίδευσης με μήκος 71, βοηθά περισσότερο τις κλάσεις να αναγνωριστούν από ότι τα μήκη παραθύρωσης 21 και 51, παρατηρών-



Σχήμα 2.4.20: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης του μοντέλου CNN. Μήκος παραθύρωσης 71.



Σχήμα 2.4.21: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης του μοντέλου LSTM. Μήκος παραθύρωσης 71.

τας ειδικά για τις μετρικές **F1-Score** και **Recall** στις Ενότητες 3.4.4 και 3.4.5. Άρα, υπάρχει συνέπεια του συμπεράσματος ότι όσο αυξάνουμε το μήκος παραθύρωσης τόσο βελτιώνεται η αποδόσεις των μοντέλων σε όλες τις μετρικές. Αυτό αποκαλύπτει ότι για αναγνωριστούν καλά τα στάδια του ύπνου, πρέπει τα μοντέλα να επεξεργάζονται μεγαλύτερα παράθυρα δεδομένων (sliding windows) ώστε να αναγνωρίσουν τα long-term χαρακτηριστικά που διαχωρίζουν τα στάδια του ύπνου.

Τα παραπάνω συμπεράσματα συνοψίζονται στους Πίνακες Σύγχυσης 2.4.20 - 2.4.21. Δηλαδή, όλα

τα μοντέλα δυσκολεύονται να αναγνωρίσουν τις κλάσεις Deep και αναγνωρίζουν καλύτερα τη κλάση REM και πολύ καλά τις κλάσεις Wake και Light, παρατηρώντας τα ποσοστά Recall της διαγωνίου. Το γεγονός αυτό είναι λογικό εφόσον το σύνολο δεδομένων είναι μη ισορροπημένο. Οπότε, περισσότερα δεδομένα από μια κλάση, διαφορετικά μεταξύ τους, τόσο καλύτερα τα μοντέλα θα την αναγνωρίσουν.

Στην επόμενη ενότητα, θα παρουσιάσουμε τη μέθοδο Ensemble Model για όλα τα μήκη παραθύρωσης, για όλα τα μοντέλα, και θα παραθέσουμε τα αντίστοιχα αποτελέσματα.

### 2.4.5 Ταξινόμηση με τη Μέθοδο Ensemble (Ensemble Method)

Στο παρούσα ενότητα, προκειμένου:

- Να εκμεταλλευτούμε τις **δυνατότητες των μοντέλων Βαθιάς Μάθησης** που χρησιμοποιήσαμε, δηλαδή των CNN και LSTM μοντέλων, τα οποία εξάγουν διαφορετικά χαρακτηριστικά από τα δεδομένα ύπνου, και
- Να εκμεταλλευτούμε τη μοντελοποίηση με **διαφορετική παραθύρωση (sliding window)** δεδομένων, που οδηγεί σε διαφορετικά εξαγόμενα χαρακτηριστικά από τα μοντέλα Βαθιάς Μάθησης (Deep Learning) και όπως συμπεράναμε στην προηγούμενη Ενότητα, **όσο αυξάνουμε το μήκος παραθύρωσης βελτιώνονται οι αποδόσεις των μοντέλων,**

εφαρμόζουμε την μέθοδο Ensemble που περιγράφηκε στο Κεφάλαιο 2.3. Συνοπτικά, με βάση τα αποτελέσματα ταξινόμησης που δίνουν τα μοντέλα Βαθιάς Μάθησης (Deep Learning) που πειραματιστήκαμε, εφαρμόζονται δύο μέθοδοι τελικής ταξινόμησης των δεδομένων ή αλλιώς **μεικτής αξιολόγησης (Score Fusion)** των αποτελεσμάτων: 1) **Μέθοδος Max Ensemble** και 2) **Μέθοδος Mean Ensemble**. Και στις δύο μεθόδους μεικτής αξιολόγησης, χρησιμοποιούνται οι ύστερες πιθανότητες (posterior probabilities) που προκύπτουν από τα μοντέλα Βαθιάς Μάθησης (CNN και BiLSTM) μέσω της συνάρτησης Softmax:  $\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^4 e^{z_j}}$   $i = 1, 2, 3, 4$ . Δηλαδή, για κάθε σύνολο 8+1 χαρακτηριστικών χρονικής διάρκειας 30 δευτερολέπτων του συνόλου αξιολόγησης (test dataset), τα μοντέλα προβλέπουν τις 4 πιθανότητες (μια για κάθε στάδιο ύπνου Wake, Light, Deep και REM).

Πιο συγκεκριμένα, έχουμε 2 μοντέλα (CNN και LSTM) επί 4 διαφορετικά μήκη παραθύρωσης (**21, 51, 71, 101**), άρα 8 στον αριθμό διαφορετικών ταξινομητών/μοντέλων όπου το κάθε ένα προβλέπει ένα διάνυσμα ύστερης πιθανότητας  $\mathbf{p}$  στον χώρο των 4 διαστάσεων ( $\mathbf{p} \in \mathbb{R}^K$ ), (όπου  $K$  είναι ο αριθμός των κλάσεων ( $K=4$ )), για κάθε σύνολο 8+1 χαρακτηριστικών χρονικής διάρκειας 30 δευτερολέπτων του συνόλου αξιολόγησης (test dataset). Τότε για το Ensemble μοντέλο: την **Mean Ensemble** μέθοδο αξιολόγησης και την **Max Ensemble** μέθοδο προκύπτει ο Πίνακας αποτελεσμάτων 2.26

<i>Metrics%</i>	<i>Mean Ensemble</i>	<i>Max Ensemble</i>
Accuracy	74.6 ± 1.0	73.8 ± 1.0
Precision	64.6 ± 1.5	65.5 ± 1.4
Recall	59.0 ± 1.1	60.5 ± 1.1
F1-Score	57.3 ± 1.1	58.5 ± 1.1
Cohen-Kappa	61.9 ± 1.8	61.5 ± 1.7
Specificity	89.0 ± 0.4	89.1 ± 0.4

Πίνακας 2.26: Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης.

Από τον Πίνακα 2.26 παρατηρούμε ότι πετυχαίνουμε καλά αποτελέσματα σε όλες τις μετρικές, τα οποία



πρέπει να συγκριθούν με τις μετρικές των επιμέρους μοντέλων για να μπορούμε να αξιολογήσουμε την αποδοτικότητα του μοντέλου Ensemble στην ταξινόμηση των σταδίων ύπνου. Όμως αυτό που παρατηρούμε είναι ότι και οι δύο μέθοδοι (Mean και Max Ensemble) μεικτής αξιολόγησης (Score Fusion) είναι αποδοτικοί, με λίγο καλύτερα αποτελέσματα να δίνει η μέθοδος Max Ensemble στις μετρικές Precision, Recall και F1-Score. Επίσης, παρατηρούμε ότι αν συγκρίνουμε τα αποτελέσματα μας με το προηγούμενο Ensemble μοντέλο (Πίνακας 2.10), παρατηρούμε ότι πετυχαίνουμε πολύ καλύτερα αποτελέσματα σε όλες τις δημοφιλείς μετρικές, τόσο για τη μέθοδο Mean όσο και για τη Max.

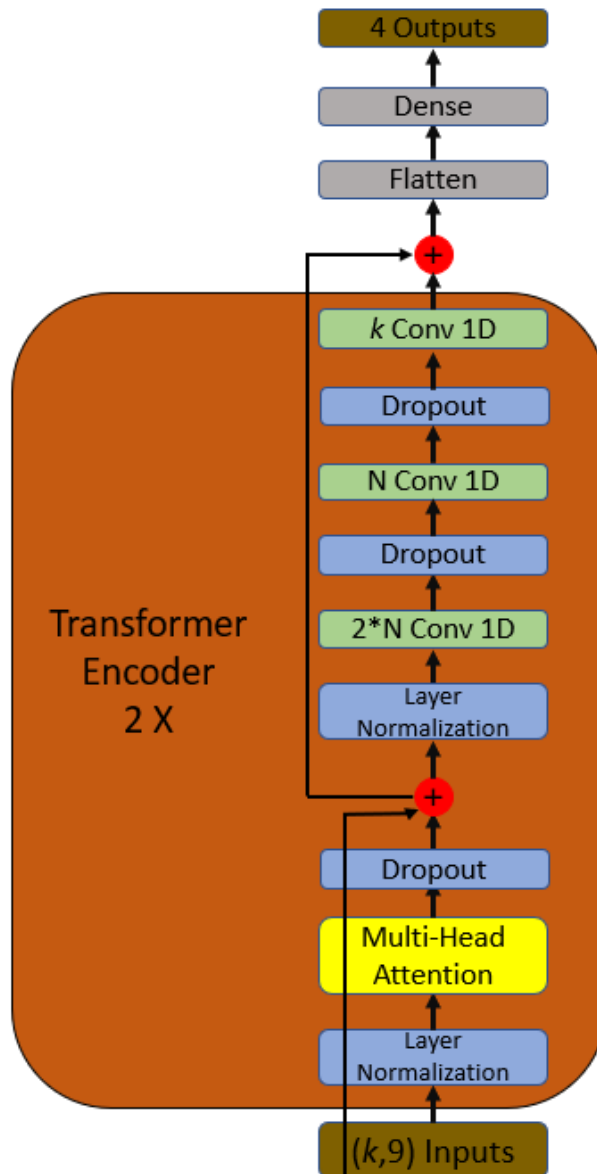
Στην επόμενη ενότητα θα πειραματιστούμε με το μοντέλο Μετασχηματιστή (Transformer Model) και θα παρουσιάσουμε τα αντίστοιχα αποτελέσματα.

#### 2.4.6 Ταξινόμηση με Μοντέλο Μετασχηματιστή (Transformer Model)

Στη παρούσα ενότητα θα παρουσιάσουμε τα πειράματα που πραγματοποιήθηκαν για την ταξινόμηση των δεδομένων στα 4 στάδια ύπνου με τη χρήση ενός μοντέλου Μετασχηματιστή (Transformer model). Συγκεκριμένα, η αρχιτεκτονική του δικτύου παρουσιάζεται στο Σχήμα 3.4.22.

Από την αρχιτεκτονική του μοντέλου έχουμε ότι, ο κύριος κορμός του μοντέλου αποτελείται από **2 επαναλαμβανόμενα διαδοχικά επίπεδα Κωδικοποίησης Μετασχηματιστή (Transformer Encoder)** και ακολουθεί ένα δίκτυο νευρώνων (Dense Layer) και η έξοδος του μοντέλου. Λεπτομερέστερα, η είσοδος του μοντέλου είναι τα δεδομένα ακτιγραφίας και τα 8 HRV χαρακτηριστικά,  $(k, 9)$ , με μήκος παραθύρωσης  $k \in (21, 51, 101)$ . Στη συνέχεια, τα δεδομένα περνούν από κανονικοποίηση επιπέδου (Layer Normalization) και μετέπειτα από έναν **πολλαπλό μηχανισμό προσοχής (Multi-Head Attention)** σύμφωνα με το [84]. Κατά τον πολλαπλό μηχανισμό προσοχής το μοντέλο εφαρμόζει τον μηχανισμό προσοχής (Attention Mechanism) πολλές φορές (number of attention heads). Στη περίπτωση τους δικού μας πειραματισμού ο αριθμός αυτός είναι 4 (number of attention heads = 4). Κατά συνέπεια το μοντέλο μαθαίνει πού να δίνει προσοχή σε πολλαπλά επίπεδα, δηλαδή θέλουμε το μοντέλο μετασχηματιστή να συνδυάζει γνώσεις από διαφορετικές συμπεριφορές του ίδιου μηχανισμού προσοχής (attention mechanism), όπως η καταγραφή εξαρτήσεων διαφόρων περιόδων (π.χ. μικρότερου εύρους και μεγαλύτερου εύρους) - (long and short term dependencies) μέσα σε μια ακολουθία δεδομένων ακτιγραφίας και HRV χαρακτηριστικών. Αφού τα δεδομένα έχουν περάσει από τον πολλαπλό μηχανισμό προσοχής ακολουθεί η **τεχνική regularization Dropout** και συνενώνονται τα δεδομένα εισόδου με τα δεδομένα που έχουν περάσει από τον πολλαπλό μηχανισμό προσοχής ( $x = inputs + MultiHead_{output}$ ). Στη συνέχεια ακολουθεί η εξαγωγή χαρακτηριστικών, όπου τα δεδομένα περνούν από κανονικοποίηση επιπέδου (Layer Normalization) και μετά από 3 συνελικτικά επίπεδα (Convolutional Layers). Ο αριθμός των φίλτρων των συνελικτικών επιπέδων είναι  $N = 64$  με stride ίσο με 3. Στο τέλος η έξοδος της εξαγωγής χαρακτηριστικών του τελευταίου 3ου επιπέδου, συνενώνεται με τα δεδομένα πριν περάσουν από τα συνελικτικά επίπεδα ( $x = MultiHead_{output} + CNN_{s_{output}}$ ). Η διαδικασία αυτή επαναλαμβάνεται **2 φορές και αποτελεί διαδοχικά επίπεδα Κωδικοποίησης Μετασχηματιστή (Transformer Encoder)**. Τέλος, τα εξαγόμενα χαρακτηριστικά οριζοντιώνονται (Flatten), ακολουθεί ένα feed-forward νευρωνικό δίκτυο με 100 νευρώνες (Dense Layer) και μετά η έξοδος με 4 νευρώνες, που αποτελούν τα 4 στάδια ύπνου. Το διάλυμα αυτό της εξόδου είναι σε δυαδική μορφή (One Hot Encoding). Πρέπει να αναφερθεί ότι κατά την έξοδο (επίπεδο Dense = 4 νευρώνες), το επίπεδο Dense επιστρέφει πιθανότητες σύμφωνα με την συνάρτηση Softmax.  $\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^4 e^{z_j}}$   $i = 1, 2, 3, 4$ . Δηλαδή, στην έξοδο του Dense επιπέδου εφαρμόζεται η συνάρτηση Softmax, η οποία μετατρέπει την είσοδο που δέχεται σε κανονικοποιημένη πιθανότητα.

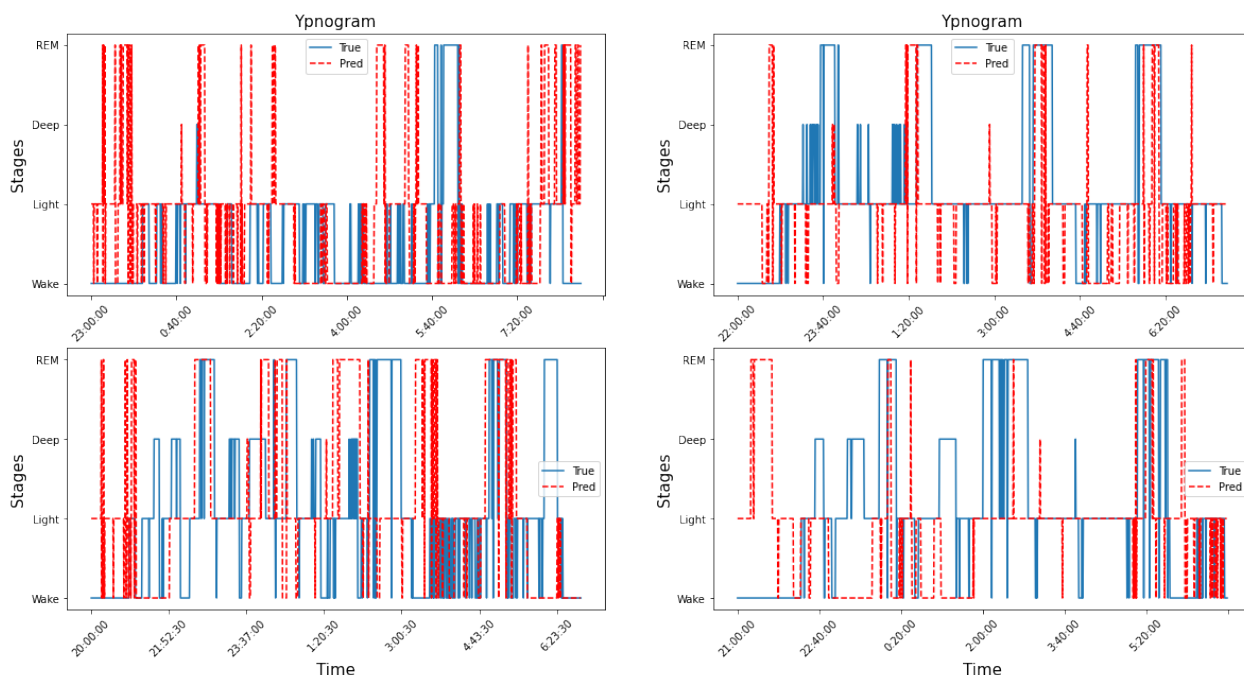
Μετά τον πειραματισμό, προκύπτουν 3 μοντέλα, ένα για κάθε μήκος παραθύρωσης. Δηλαδή,



Σχήμα 2.4.22: Αρχιτεκτονική μοντέλου Μετασχηματιστή (Transformer). Μεταβλητή των μοντέλων αποτελεί το μήκος παραθύρωσης (sliding window) και αναφέρεται ως  $k$ ,  $k \in (21, 51, 101)$ .

***Transformer<sub>21</sub>, Transformer<sub>51</sub>, Transformer<sub>101</sub>***. Η εκπαίδευση των μοντέλων γίνεται για 100 εποχές, χρησιμοποιώντας batch size ίσο με 512. Επίσης, όπως αναφέραμε και στο Κεφάλαιο 3.3, το 20% των δεδομένων εκπαίδευσης χρησιμοποιείται για validation ώστε να αξιολογηθεί η απόδοση του μοντέλου κατά την εκπαίδευση. Τέλος, για την εκπαίδευση και βελτιστοποίηση των βαρών του δικτύου (weights), χρησιμοποιείται ο αλγόριθμος Adam με τις τυπικές τιμές που δίνεται από την βιβλιογραφία της βιβλιοθήκης Keras. Σύμφωνα με τα παραπάνω γίνεται η εκπαίδευση των μοντέλων στο σύνολο δεδομένων εκπαίδευσης (train), και μετά γίνεται η πρόβλεψη/ταξινόμηση στο σύνολο δεδομένων αξιολόγησης (test). Στα Σχήματα 2.4.23 - 2.4.25 παρατίθενται τα υπογράμματα τεσσάρων χρηστών του συνόλου αξιολόγησης (test), δηλαδή διαγράμματα πρόβλεψης/ταξινόμησης των σταδίων του ύπνου τεσσάρων χρηστών, στα τέσσερα στάδια Wake, Light, Deep και REM. Επιπλέον, τα αποτελέσματα της ταξινόμησης για τα διαφορετικά μήκη παραθύρωσης παρατίθενται στον Πίνακα 2.27 με τις κλασικές

δημοφιλείς μετρικές. Τέλος, στους Πίνακες 2.28 - 2.30 παρουσιάζονται τα αποτελέσματα ταξινόμησης για κάθε στάδιο ύπνου (Wake, Light, Deep και REM) για τα τρία διαφορετικά μήκη παραθύρωσης και στα Σχήματα 2.4.26 - 2.4.28 παρουσιάζονται οι Πίνακες Σύγχυσης για τα διαφορετικά μήκη παραθύρωσης.

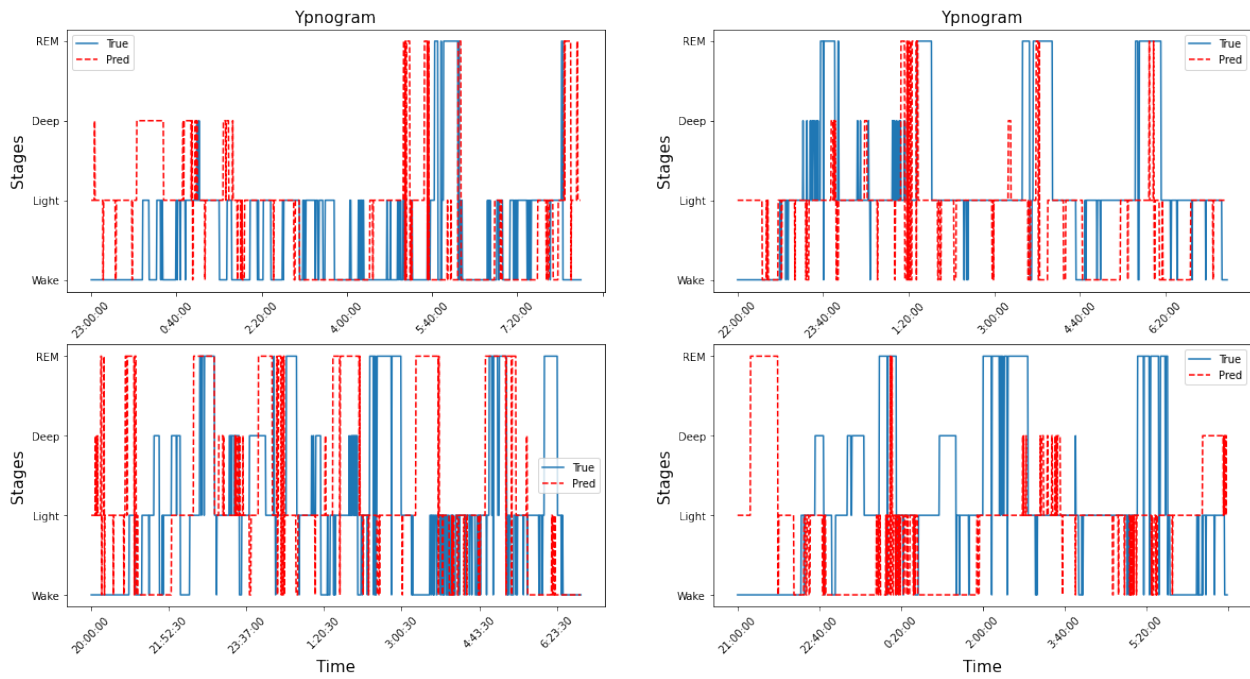


Σχήμα 2.4.23: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 21.

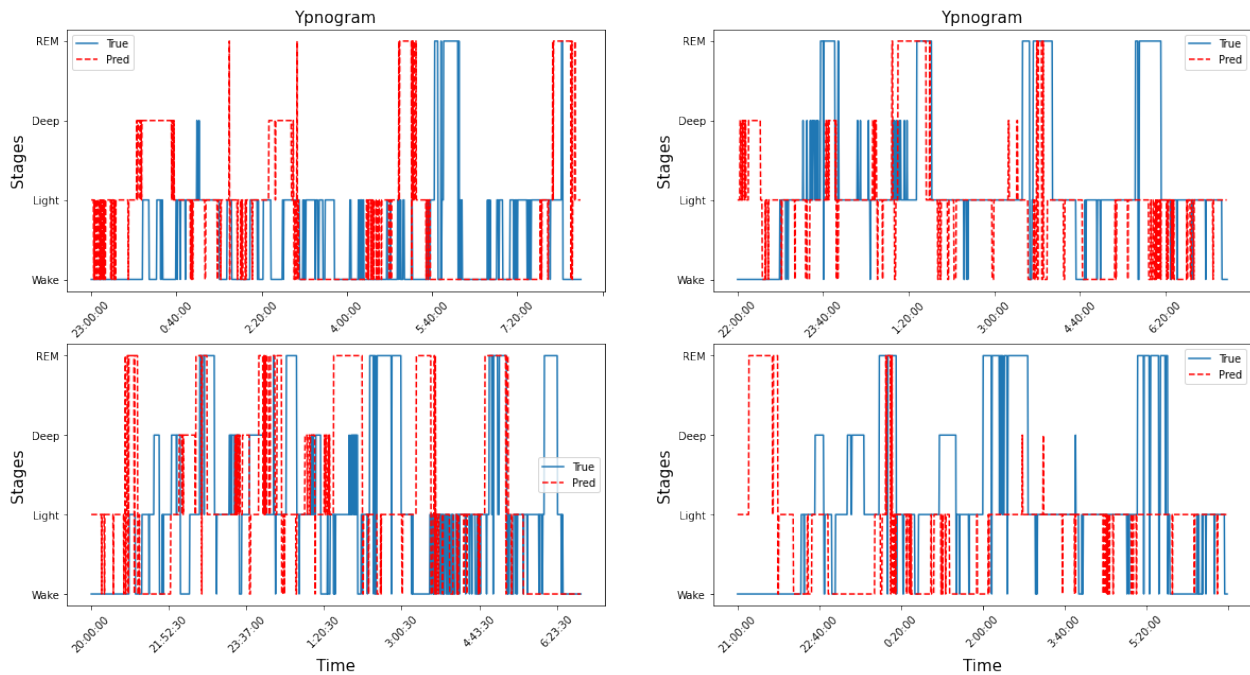
Εποπτικά, από τα υπνογράμματα του Σχήματος 2.4.23 - 2.4.25, παρατηρούμε ότι η κλάση Deep δεν μπορεί να αναγνωριστεί καλά από το μοντέλο Μετασχηματιστή ενώ οι υπόλοιπες κλάσεις, κυρίως οι κλάσεις Wake και Light, μπορούν να αναγνωριστούν καλύτερα (αφού επικαλύπτονται οι δύο γραμμές True και Predictions). Η κλάση REM αναγνωρίζεται σε πολύ καλό βαθμό, καλύτερα προφανώς από την κλάση Deep. Το συμπέρασμα αυτό μπορεί να διαγνωσθεί και από τους Πίνακες μετρικών του συνόλου αξιολόγησης (Πίνακες 2.28 - 2.30) και από τους Πίνακες σύγχυσης (Confusion Matrix - Σχήματα 2.4.26 - 2.4.28) όπου η κλάση Deep έχει τις χειρότερες αποδόσεις σε όλες τις μετρικές, ακολουθεί με καλύτερες αποδόσεις η κλάση REM, μετά η κλάση Wake, και τέλος τις υψηλότερες αποδόσεις τις έχει η κλάση Light.

<i>Metrics%</i>	<i>SeqLen21</i>	<i>SeqLen51</i>	<i>SeqLen101</i>
Accuracy	69.9 ± 1.0	71.8 ± 1.0	71.4 ± 1.0
Precision	51.4 ± 1.4	53.8 ± 1.4	56.3 ± 1.3
Recall	65.0 ± 0.9	67.9 ± 1.1	64.5 ± 1.2
F1-Score	51.4 ± 0.9	54.4 ± 1.1	57.9 ± 1.2
Cohen-Kappa	53.4 ± 1.7	57.6 ± 1.8	57.5 ± 1.8
Specificity	88.6 ± 0.4	89.4 ± 0.4	89.0 ± 0.4

Πίνακας 2.27: Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για διαφορετικά μήκη παραθύρωσης.



Σχήμα 2.4.24: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 51.



Σχήμα 2.4.25: Υπνόγραμμα 4 χρηστών από το σύνολο αξιολόγησης (test). Απεικονίζονται οι πραγματικές τιμές των σταδίων ύπνου (True) και οι προβλεπόμενες (Pred). Μήκος παραθύρωσης 101.

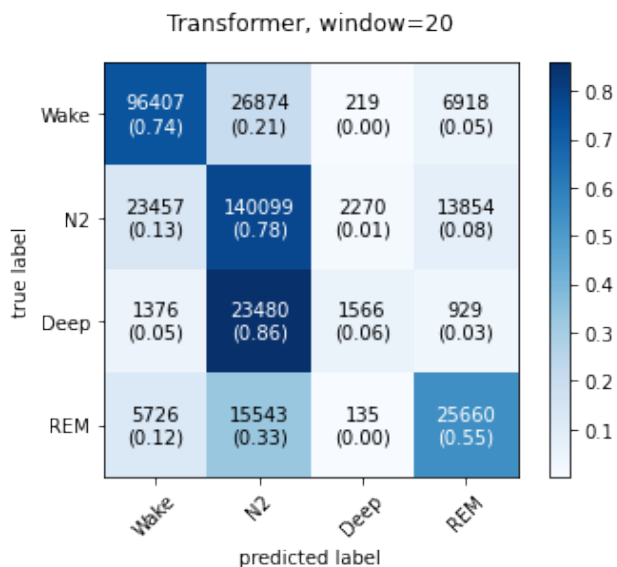
Λεπτομερέστερα, από τον Πίνακα μετρικών 2.27, συμπεραίνουμε ότι το μοντέλο **Transformer**<sub>101</sub> έχει τις υψηλότερες αποδόσεις στις περισσότερες τις μετρικές, και ακολουθεί το μοντέλο

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	76.0	74.0	75.0
Light	68.0	78.0	73.0
Deep	37.0	6.0	10.0
REM	54.0	55.0	54.0

Πίνακας 2.28: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 21.

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	75.0	77.0	76.0
Light	70.0	75.0	72.0
Deep	39.0	17.0	24.0
REM	56.0	56.0	56.0

Πίνακας 2.29: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 51.



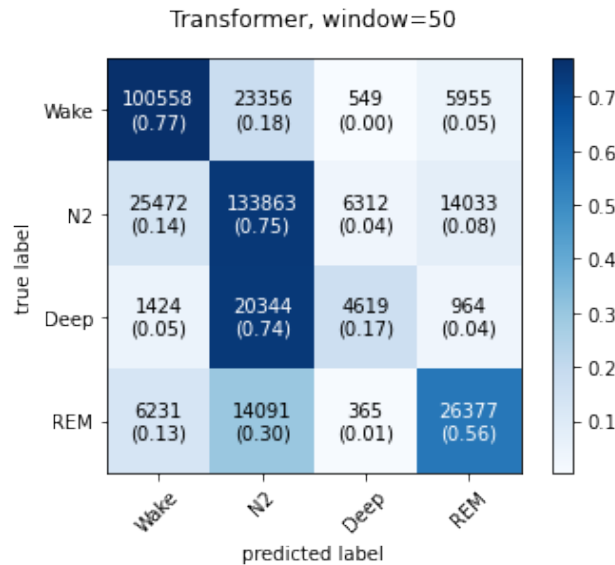
Σχήμα 2.4.26: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 21.

**Transformer<sub>51</sub>**. Παρατηρούμε ότι όλες οι μετρικές έχουν παραπλήσια ποσοστά για όλα τα μήκη παραθύρωσης των δεδομένων, με καλύτερο το μοντέλο **Transformer<sub>101</sub>**. Οι μετρικές Precision και Recall είναι της τάξης του 56 % και 67 % αντίστοιχα και γι αυτό και η μετρική F1-Score κυμαίνεται στα ίδια επίπεδα. Επίσης, η μετρική F1-Score και Cohen-Kappa είναι υψηλή.

Εξαιτίας αυτού του φαινομένου, παρατηρούμε τους Πίνακες Σύγχυσης 2.28 - 2.30 όπου παρουσιάζονται οι αποδόσεις των μοντέλων ανα κλάση. Είναι φανερό ότι τα μοντέλα, λόγω του μη ισορροπημένου συνόλου δεδομένων (imbalance dataset), δεν μπορούν να αναγνωρίσουν σχεδόν καθόλου την κλάση Deep (F1-Score  $\leq 35\%$ ) και ακολουθεί η κλάση REM. Επίσης, παρατηρούμε ότι τα μοντέλα **Transformer<sub>51</sub>** και **Transformer<sub>101</sub>** είναι καλύτερα σε όλες τις μετρικές ανα κλάση, σε σχέση με το μοντέλο **Transformer<sub>21</sub>**. Αν εστιάσουμε στην κλάση REM, Deep και Wake, παρατηρούμε ότι

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	81.0	77.0	79.0
Light	71.0	77.0	74.0
Deep	42.0	31.0	35.0
REM	62.0	61.0	61.0

Πίνακας 2.30: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση. Μήκος παραθύρωσης 101.



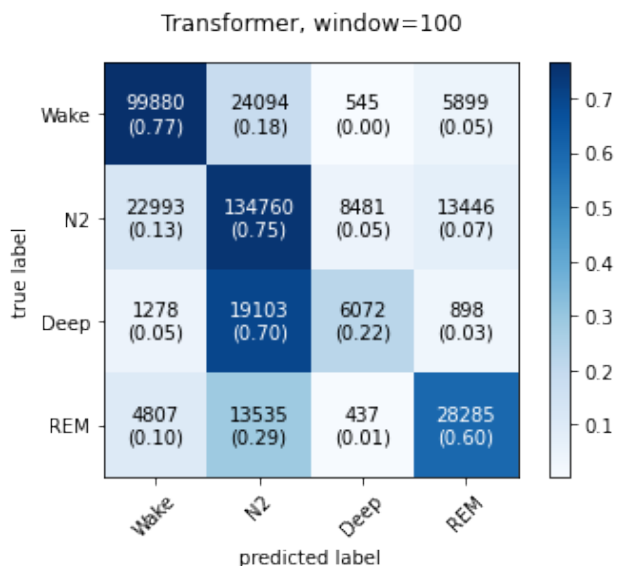
Σχήμα 2.4.27: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 51.

το μοντέλο **Transformer<sub>101</sub>** έχει τη καλύτερη απόδοση F1-Score και Recall από όλα τα μοντέλα. όπως και στις προηγούμενες μοντελοποιήσεις η βέλτιστη παραθύρωση είναι αυτή με μήκος 101 και βοηθά περισσότερο τις κλάσεις να αναγνωριστούν από ότι τα μήχη παραθύρωσης 21 και 51, παρατηρώντας ειδικά για τις μετρικές F1-Score και Recall.

Τα παραπάνω συμπεράσματα συνοψίζονται στους Πίνακες Σύγχυσης 2.4.26 - 2.4.28. Δηλαδή, όλα τα μοντέλα δυσκολεύονται να αναγνωρίσουν τις κλάσεις Deep και αναγνωρίζουν καλύτερα τη κλάση REM και πολύ καλά τις κλάσεις Wake και Light, παρατηρώντας τα ποσοστά Recall της διαγωνίου.

Στην επόμενη ενότητα θα συνοψίσουμε τα αποτελέσματα που προέκυψαν από όλη την ανάλυση που πραγματοποιήθηκε με τις διαφορετικές προσεγγίσεις που ακολουθήσαμε για την ταξινόμηση των δεδομένων actigraphy + HRV σε 4 στάδια ύπνου.





Σχήμα 2.4.28: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης. Μήκος παραθύρωσης 101.

## 2.5 Σύνοψη Αποτελεσμάτων - Συμπεράσματα

Στο παρόν Κεφάλαιο μοντελοποιήσαμε τα δεδομένα ακτιγραφίας και HRV χαρακτηριστικών της βάσης δεδομένων MESA. Αρχικά, προσπαθήσαμε να αναπαράξουμε τα αποτελέσματα της έρευνας [1] και είδαμε ότι χρησιμοποιώντας τα ίδια μοντέλα προκύπτουν τα ίδια αποτελέσματα. Στη συνέχεια, η προαναφερθείσα έρευνα αποτέλεσε βάση της δικής μας έρευνας ώστε να μπορέσουμε να διερευνήσουμε αν μπορούμε να βελτιώσουμε τα αποτελέσματα. Εστίασαμε στα **μειονεκτήματα της έρευνας [1] και της βάσης δεδομένων MESA** και δομήσαμε την έρευνας μας πάνω σε αυτά. Τα κύρια προβλήματα της βάσης MESA και στο πρόβλημα της ταξινόμησης ύπνου είναι, όπως έχουμε τονίσει σε όλες τις ενότητες, **το μη ισορροπημένο σύνολο δεδομένων (imbalance dataset)**. Το πρόβλημα αυτό αντιμετωπίστηκε με τη χρήση τεχνικών Αύξησης Συνόλου Δεδομένων (**Data Augmentation**) ώστε να δημιουργηθούν δεδομένα τεχνητά και όσο γίνεται πιο κοντά στη πραγματικότητα. Με τη τεχνική αυτή καταφέραμε **να βελτιώσουμε τα αποτελέσματα όλων των μοντέλων όσον αφορά τις δημοφιλείς μετρικές και τις μετρικές ανά κλάση**. Συγκρίνοντας τα αποτελέσματα της έρευνας [1] και της δικής μας, **για το ίδιο σύνολο δεδομένων αξιολόγησης (test dataset)**, προκύπτουν τα εξής αποτελέσματα που παρουσιάζονται στον Πίνακα 2.31, με τα μοντέλα που δίνουν τα καλύτερα αποτελέσματα. Είναι σημαντικό να τονιστεί ότι σε όλους τους πειραματισμούς μας κρατήσαμε το σύνολο δεδομένων αξιολόγησης σταθερό με τους ίδιους χρήστες (MESA id), ώστε να μπορούμε να συγκρίνουμε τα μοντέλα μεταξύ τους. **Το καλύτερο μοντέλο της έρευνας [1] με βάση τις δημοφιλείς μετρικές είναι η μέθοδος Max Ensemble ενώ από τη δική μας έρευνα προέκυψε ότι η μέθοδος Mean Ensemble δίνει τα καλύτερα αποτελέσματα.**

Παρατηρούμε ότι σε όλες τις μετρικές υπάρχει βελτίωση αλλά τη μεγαλύτερη από όλες τις μετρικές την έχει η μετρική Cohen-Kappa και Precision με **απόλυτο ποσοστό αύξησης απόδοσης ίσο με 6.3 μονάδες ( $Absolute Diff = |metric_{model_1} - metric_{model_2}|$ )**. Με βάση αυτές τις δυο μετρικές συμπεραίνουμε ότι η τεχνική Data Augmentation και η αλλαγή των μοντέλων, οδηγεί όντως στην βελτίωση των αποτελεσμάτων ταξινόμησης. Η βοήθεια της τεχνικής Data Augmentation για την κλάση Deep βοηθά τον καλύτερο διαχωρισμό των κλάσεων. Συγκεκριμένα, από όλους τους Πίνακες

<i>Metrics%</i>	<i>Max Ensemble</i>	<i>Mean Ensemble</i>	<i>Absolute Diff</i>
Accuracy	71.1 ± 1.0	74.6 ± 1.0	+3.5
Precision	58.3 ± 1.3	64.6 ± 1.5	+6.3
Recall	54.8 ± 1.0	59.0 ± 1.1	+4.2
F1-Score	52.4 ± 1.0	57.3 ± 1.1	+4.9
Cohen-Kappa	55.7 ± 1.8	61.9 ± 1.8	+6.2
Specificity	87.7 ± 0.4	89.0 ± 0.4	+1.3

Πίνακας 2.31: Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για τα καλύτερα μοντέλα της έρευνας [1] (Max Ensemble) και της δικής μας (Mean Ensemble).

του Κεφαλαίου 2.4 παρατηρούμε ότι με την **προσθήκη θορύβου (Jittering)**, αυξάνονται τα ποσοστά των κλάσεων **Deep** και **REM** και μειώνονται ελάχιστα τα ποσοστά των κλάσεων **Wake** και **Light**. Το συμπέρασμα αυτό αντικατοπτρίζεται στους Πίνακες Σύγκρισης όλων των μοντέλων που πειραματιστήκαμε, αφού **αυξάνεται το ποσοστό της μετρικής Recall για τις κλάσεις Deep, REM και μειώνεται των Wake και REM σε όλα τα μοντέλα**. Άρα, τα μοντέλα Βαθιάς Μάθησης αναγνώρισαν καλύτερα τις κλάσεις REM και Deep μέσω της προσθήκης θορύβου στα δεδομένα Deep, οπότε εξισορροπήθηκε η απόδοση της αναγνώρισης των κλάσεων. Με περαιτέρω διερεύνηση των τεχνικών **Data Augmentation** πολύ πιθανό να βελτιωθούν περισσότερο οι αποδόσεις των μοντέλων σύμφωνα με τα [96] και [95]. Παραδείγματα άλλων τεχνικών Data Augmentation πέραν της τεχνικής Jittering, περιγράφηκαν στο Κεφάλαιο 2.4.1 (Περιγραφή Διαδικασίας - Προεπεξεργασία Δεδομένων).

Για κάθε μοντέλο και μέθοδο ταξινόμησης, παραθέτουμε τους Πίνακες 2.32 και 2.33 οι οποίοι περιλαμβάνουν τα **αποτελέσματα της μετρικής "χρονικής απόκλισης" (Time Deviation)** που παραθέσαμε στην εισαγωγή του Κεφαλαίου 2.3. Συνοπτικά, για μια κλάση  $k$ , και  $N$  αριθμό χρηστών (users), η μετρική "χρονικής απόκλισης" ορίζεται ως  $TD_k = \frac{1}{N} \sum_{i=1}^N (Pred_k^i - True_k^i)$ , όπου  $Pred_k$  είναι ο συνολικός χρόνος σε λεπτά για την προβλεπόμενη κλάση  $k$  (predicted label) του μοντέλου και  $True_k$  είναι ο αντίστοιχος συνολικός χρόνος σε λεπτά της πραγματικής κλάσης  $k$  (true label). Οπότε σύμφωνα με τα παραπάνω έχουμε τους Πίνακες 2.32 και 2.32 της "χρονικής απόκλισης" για κάθε μοντέλο ανά κλάση.

<i>Model%</i>	<i>Deep</i>	<i>Light</i>	<i>REM</i>	<i>Wake</i>
<i>CNN<sub>21</sub></i>	-42.5	69.9 ± 8.5	11.1 ± 5.4	-39
<i>CNN<sub>51</sub></i>	-41.4	32.5 ± 8.8	18.5 ± 6.2	-20.3
<i>CNN<sub>101</sub></i>	-36.9	55.9 ± 7.9	6.9 ± 4.7	-36.4
<i>LSTM<sub>21</sub></i>	-41.3	55.7 ± 7.7	-10.7	-18.2
<i>LSTM<sub>51</sub></i>	-39.7	42.8 ± 7.4	-9.6	-7.9
<i>LSTM<sub>101</sub></i>	-35.9	76.0 ± 7.3	-28.4	-25.5
<i>MaxEnsemble</i>	-40.9	52.0 ± 7.8	-8.1	-17.8
<i>MeanEnsemble</i>	-42	65.2 ± 7.8	-13.3	-24.5

Πίνακας 2.32: Μετρική Χρονικής Απόκλισης (Time Deviation) ανα κλάση σε λεπτά (minutes) της έρευνας [1].

Από τους Πίνακα 2.32 και 2.33 παρατηρούμε σε πρώτο βαθμό ότι για τις κλάσεις Deep, REM και Wake, **υποεκτιμάται** ο χρόνος παραμονής των δεδομένων κλάσεων (λόγω του αρνητικού πρόσημου στα αποτελέσματα), ενώ αντίθετα ο χρόνος παραμονής στην κλάση Light υπερεκτιμάται (λόγω του

<i>Model%</i>	<i>Deep</i>	<i>Light</i>	<i>REM</i>	<i>Wake</i>
<i>CNN</i> <sub>21</sub>	-35	62.2 ± 7.0	-10.2	-30.8
<i>CNN</i> <sub>51</sub>	-28.6	46.8 ± 6.9	-10.9	-21
<i>CNN</i> <sub>101</sub>	-24.4	10.3 ± 6.6	-6.6	13.3 ± 6.6
<i>BiLSTM</i> <sub>21</sub>	-35.2	53.0 ± 7.3	-11.3	-20.6
<i>BiLSTM</i> <sub>51</sub>	-30.8	53.0 ± 7.3	-11.6	-24.6
<i>BiLSTM</i> <sub>101</sub>	-30.3	48.6 ± 6.9	-10.4	-22.1
<i>MaxEnsemble</i>	-35.1	39.6 ± 6.8	-8.3	-9.8
<i>MeanEnsemble</i>	-38	62.2 ± 6.9	-17.1	-20.4

Πίνακας 2.33: Μετρική Χρονικής Απόκλισης (Time Deviation) ανα κλάση σε λεπτά (minutes) των βελτιωμένων μοντέλων (CNN και BiLSTM).

θετικού πρόσημου στα αποτελέσματα). Λεπτομερέστερα, η κλάση **Deep** υποεκτιμάται από όλα τα μοντέλα, όμως το **LSTM**<sub>101</sub> μοντέλο δίνει τη μικρότερη τιμή απόκλισης (**-35.9 λεπτά**) με βάση τα αποτελέσματα της έρευνας [1]. Όμως με βάση τη δική μας μοντελοποίηση, ο χρόνος αυτός μειώνεται (**-24.4 λεπτά**, μοντέλο **CNN**<sub>101</sub>). Αντίστοιχα, η κλάση **REM** υποεκτιμάται από όλα τα μοντέλα, όμως η μέθοδος **MaxEnsemble** έχει τη μικρότερη τιμή απόκλισης (**-8.1 λεπτά**). Με βάση τη δικιά μας μοντελοποίηση ο χρόνος αυτός μειώνεται κι άλλο από το μοντέλο **CNN**<sub>101</sub> με τιμή (**-6.6 λεπτά**). Επίσης, η κλάση **Wake** υποεκτιμάται από όλα τα μοντέλα εκτός, όμως το **LSTM**<sub>51</sub> μοντέλο δίνει τη μικρότερη τιμή απόκλισης (**-7.9 λεπτά**). Με βάση τη δική μας μοντελοποίηση ο χρόνος αυτός είναι λίγο μεγαλύτερος με τιμή (**-9.8 λεπτά**) που τον δίνει η μέθοδος **MaxEnsemble**. Αντίθετα, η κλάση **Light** **υπερεκτιμάται** από όλα τα μοντέλα, όμως το **CNN**<sub>52</sub> μοντέλο δίνει το μικρότερο χρόνο απόκλισης (**+32.5 λεπτά**). Με βάση τη δική μας μοντελοποίηση, ο χρόνος αυτός μειώνεται περαιτέρω από το μοντέλο **CNN**<sub>101</sub> με τιμή (**+10.3 λεπτά**). Με βάση τα παραπάνω αποτελέσματα συμπεραίνουμε ότι είναι λογικό η κλάση **Light** να **υπερεκτιμάται** εξαιτίας του μη ισορροπημένου συνόλου δεδομένων (imbalance dataset), οπότε τα μοντέλα τείνουν να ταξινομούν τα δεδομένα ύπνου στην κλάση με το μεγαλύτερο πλήθος δεδομένων, δηλαδή στην κλάση **Light**. Αυτό το συμπέρασμα είναι σημαντικό γιατί μας υποδεικνύει ότι πρέπει να αντιμετωπιστεί το πρόβλημα του μη ισορροπημένου συνόλου δεδομένων.

Ακόμα, από τους πειραματισμούς μας συμπεράναμε ότι η **αύξηση του μήκους παραθύρωσης** των δεδομένων βοηθά στην καλύτερη ταξινόμηση των σταδίων του ύπνου σε όλες τις δημοφιλείς μετρικές και στις μετρικές ανά κλάση, δηλαδή ο πειραματισμός μας με μήκος παραθύρωσης 71, έδωσε καλύτερα αποτελέσματα από ότι οι πειραματισμοί με μήκη 21 και 51. Αντίστοιχα, τα μοντέλα με μήκη παραθύρωσης 101 έδωσαν καλύτερα αποτελέσματα από τα μοντέλα με μήκος παραθύρωσης 71. Τα παραπάνω αποτελέσματα αποδείχθηκαν για όλα τα μοντέλα Βαθιάς Μάθησης (Deep Learning models) σε όλες τις δημοφιλείς μετρικές. Τα αποτελέσματα φαίνονται στους Πίνακες 2.34 και 2.35.

Από τους Πίνακες 2.34 και 2.35 παρατηρούμε αυτό το φαινόμενο. Ότι η **αύξηση του μήκους παραθύρωσης** αυξάνει και τις αποδόσεις των μοντέλων σε όλες τις μετρικές. Άρα μπορούμε να υποθέσουμε ότι αν αυξήσουμε περαιτέρω το μήκος παραθύρωσης, μεγαλύτερο του 101, πολύ πιθανόν τα αποτελέσματα να βελτιωθούν ακόμη περισσότερο. Επιπλέον, η παρατήρηση αυτή επιβεβαιώνει ότι τα δεδομένα ύπνου (actigraphy + 8 HRV features), προκειμένου να ταξινομηθούν σωστά, τα μοντέλα πρέπει να επεξεργάζονται και να εξάγουν χαρακτηριστικά μακράς εμβέλειας (long-term features) επειδή τα στάδια του ύπνου διαρκούν περισσότερο από  $30 * 100 = 300$  δευτερόλεπτα =  $300/60 = 5$  λεπτά που είναι τα δεδομένα με μήκος παραθύρωσης 100 [2].

Επιπρόσθετα, ο πειραματισμός μας με το μοντέλο **Μετασχηματιστή** επιβεβαιώνει το

<i>Metrics%</i>	<i>CNN<sub>21</sub></i>	<i>CNN<sub>51</sub></i>	<i>CNN<sub>71</sub></i>	<i>CNN<sub>101</sub></i>
Accuracy	70.1 ± 0.9	71.5 ± 1.0	71.9 ± 1.0	71.0 ± 1.0
Precision	60.4 ± 1.3	62.7 ± 1.4	63.1 ± 1.4	62.0 ± 1.3
Recall	53.3 ± 0.9	55.3 ± 1.0	57.5 ± 1.1	58.9 ± 1.1
F1-Score	51.7 ± 1.0	54.0 ± 1.0	55.8 ± 1.1	56.9 ± 1.1
Cohen-Kappa	52.6 ± 1.7	55.6 ± 1.7	57.3 ± 1.8	57.3 ± 1.8
Specificity	87.2 ± 0.4	87.9 ± 0.4	88.2 ± 0.4	88.3 ± 0.4

Πίνακας 2.34: Πίνακας αποτελεσμάτων για όλα τα μήκη παραθύρωσης (CNN).

<i>Metrics%</i>	<i>BiLSTM<sub>21</sub></i>	<i>BiLSTM<sub>51</sub></i>	<i>BiLSTM<sub>71</sub></i>	<i>BiLSTM<sub>101</sub></i>
Accuracy	69.5 ± 1.0	71.4 ± 1.0	71.8 ± 1.0	71.7 ± 1.0
Precision	58.0 ± 1.4	62.1 ± 1.4	63.2 ± 1.3	63.3 ± 1.3
Recall	53.8 ± 0.9	56.0 ± 1.1	58.1 ± 1.1	61.2 ± 1.2
F1-Score	51.7 ± 0.9	54.5 ± 1.1	56.6 ± 1.1	58.7 ± 1.2
Cohen-Kappa	52.5 ± 1.7	55.4 ± 1.8	57.8 ± 1.8	58.9 ± 1.8
Specificity	87.2 ± 0.4	87.9 ± 0.4	88.3 ± 0.4	88.7 ± 0.4

Πίνακας 2.35: Πίνακας αποτελεσμάτων για όλα τα μήκη παραθύρωσης (BiLSTM).

γεγονός ότι το κυριότερο πρόβλημα του προβλήματος ταξινόμησης είναι το μη ισορροπημένο σύνολο δεδομένων και όχι η επιλογή μοντέλου, καθώς το μοντέλο Μετασχηματιστή έδωσε τα ίδια, με μικρή διαφορά χειρότερα, αποτελέσματα από το **καλύτερο μοντέλο που είναι το *BiLSTM<sub>101</sub>*** όπως φαίνεται στον Πίνακα 2.36. Δηλαδή, στον Πίνακα 2.36 παρατηρούμε ότι το μοντέλο ***BiLSTM<sub>101</sub>*** είναι καλύτερο σχεδόν σε όλες τις μετρικές κυρίως στη μετρική Precision (+7 μονάδες).

<i>Metrics%</i>	<i>BiLSTM<sub>101</sub></i>	<i>Transformer<sub>101</sub></i>	<i>Diff</i>
Accuracy	71.7 ± 1.0	71.4 ± 1.0	+0.3
Precision	63.3 ± 1.3	56.3 ± 1.5	+7.0
Recall	61.2 ± 1.2	64.5 ± 1.1	-3.3
F1-Score	58.7 ± 1.2	57.9 ± 1.1	+0.8
Cohen-Kappa	58.9 ± 1.8	57.5 ± 1.8	+1.4
Specificity	88.7 ± 0.4	89.0 ± 0.4	+0.3

Πίνακας 2.36: Μετρικές στο σύνολο δεδομένων αξιολόγησης (test) για τα καλύτερα μοντέλα της έρευνας μας εκτός των μεθόδων Ensemble.

Επίσης, πειραματισμός με επιπλέον εξαγόμενα χαρακτηριστικά (**features engineering**) μπορεί να οδηγήσει σε καλύτερα αποτελέσματα. Τέτοια, χαρακτηριστικά μπορούν να κατασκευαστούν με τη βοήθεια του τομέα της ιατρικής (sleep medicine) και να εξαχθούν από τα δεδομένα ακτιγραφίας και των καρδιακών παλμών (Heart Rate). Επιπλέον, η **προσθήκη και άλλου είδους δεδομένων** μπορεί να βοηθήσει την ταξινόμηση, όπως για παράδειγμα ο κιρκάδιος κύκλος (circadian cycle) [94], δεδομένα στάσης χεριών και σώματος (hand and body posture) [97], δεδομένα αναπνοής (Respiration Signal) και δεδομένα επιτάχυνσης χεριών τριών αξόνων (3-axis accelerometer) [98] τα οποία μπορούν να εμπλουτίσουν περαιτέρω την πληροφορία που επεξεργάζονται τα μοντέλα, οπότε τα τελευταία θα μπορούν να διακρίνουν καλύτερα τις διαφορές ανάμεσα στα στάδια του ύπνου.

Τέλος, γνωρίζουμε ότι η απόδοση των μοντέλων καθορίζεται από την επιλογή του συνόλου δεδομένων εκπαίδευσης και αξιολόγησης (train και test). Δηλαδή, αν επιθυμούμε να κάνουμε ταξινόμηση των σταδίων του ύπνου σε υγιείς χρήστες τότε θα πρέπει να γίνει μια ανάλυση ώστε να αφαιρεθούν οι χρήστες που δεν έχουν υγιή ύπνο (patients). Κανονικά η πληροφορία αυτή θα πρέπει να μας δίνεται αλλά για θέματα προστασίας δεδομένων δεν δίνεται από τη βάση MESA.

Συνολικά, από όλους τους πειραματισμούς μας καταφέραμε την αύξηση της απόδοσης των δημοφιλών μετρικών ταξινόμησης (και ανά κλάση) για όλα τα μοντέλα και επιπλέον βγάλαμε πολύ χρήσιμα συμπεράσματα για την επίδραση των τεχνικών Data Augmentation, Undersampling και για τον μήκος παραθύρωσης των δεδομένων. Επίσης, προτείνουμε διάφορες μεθόδους που θα μπορούσαν να βελτιώσουν ακόμα περισσότερο τις αποδόσεις των μοντέλων Βαθιάς Μάθησης (Deep Learning).

Στο επόμενο Κεφάλαιο θα αντιμετωπίσουμε με διαφορετικό τρόπο τα δεδομένα της βάσης MESA χρησιμοποιώντας μόνο τα δεδομένα καρδιακών παλμών (Heart Rate) χωρίς την εξαγωγή χαρακτηριστικών (raw data).





## Κεφάλαιο 3

# Μέθοδος Ταξινόμησης Σταδίων Ύπνου με χρήση ακατέργαστων (raw) δεδομένων RR-Intervals

---

<b>3.1 Βιβλιογραφική Έρευνα</b> . . . . .	<b>98</b>
<b>3.2 Δεδομένα RR- Intervals - Προεπεξεργασία</b> . . . . .	<b>99</b>
3.2.1 Περιγραφή διαδικασίας - Προετοιμασία δεδομένων για μοντελοποίηση . . . . .	101
3.2.2 Εφαρμογή τεχνικών αύξησης συνόλου δεδομένων (Data Augmentation) . . . . .	110
<b>3.3 Μοντελοποίηση Δεδομένων RR-Intervals με αλγορίθμους Deep Learning</b> . . . . .	<b>112</b>
3.3.1 Μοντελοποίηση με Συνελικτικό Νευρωνικό Δίκτυο - CNN . . . . .	114
3.3.2 Μοντελοποίηση με LSTM . . . . .	117
3.3.3 Μοντελοποίηση με LSTM και μηχανισμό Προσοχής (Attention) . . . . .	119
<b>3.4 Ανάλυση Αποτελεσμάτων και Συμπεράσματα</b> . . . . .	<b>122</b>

---

### 3.1 Βιβλιογραφική Έρευνα

Όπως αναφέραμε και στην εισαγωγή, λόγω του υψηλού κόστους και της διεισδυτικής φύσης των κλινικών μελετών ύπνου, όπως για παράδειγμα η χρήση εξοπλισμού PSG σε ειδικές κλινικές για καταγραφή δεδομένων ύπνου, σήμερα είναι αναγκαίο να αναπτυχθούν μέθοδοι/τεχνικές, όπου με μη διεισδυτικό τρόπο να μπορούν να καταγραφούν δεδομένα και μέσω της επεξεργασίας τους, να επιτυγχάνεται η ταξινόμηση των σταδίων του ύπνου του ανθρώπου στο βέλτιστο βαθμό. Οι συσκευές που χρησιμοποιούνται ευρέως σήμερα για την καταγραφή της ανθρώπινης δραστηριότητας είναι τα επονομαζόμενα wearables τα οποία συνήθως καταγράφουν δεδομένα κίνησης (accelerometer, actigraphy) και καρδιακούς παλμούς (heart rate) [32][99]. Το κέρδος που έχουμε από την χρήση των φορητών συσκευών (wearables, smartwatch) είναι ο μη διεισδυτικός τρόπος καταγραφής δεδομένων χωρίς μεγάλο κόστος, αφού οι χρήστες απλά φορούν τα "έξυπνα ρολόγια" (smartwatch) και αυτά καταγράφουν δεδομένα. Στο [94], αναφέρεται ότι είναι πλέον συνηθισμένες οι φορητές καταναλωτικές συσκευές για την καταγραφή βιοσημάτων και για τα πειράματά τους χρησιμοποίησαν μια δική τους εφαρμογή (mobile application) για τη συλλογή ακατέργαστων (raw) δεδομένων επιτάχυνσης και καρδιακών παλμών από το Apple Watch που φορούσαν οι συμμετέχοντες. Χρησιμοποιώντας αυτά τα δεδομένα, οι ερευνητές συνέχριναν τις συνεισφορές πολλαπλών χαρακτηριστικών όπως κίνηση (movement), τοπική τυπική απόκλιση του καρδιακού ρυθμού και clock proxy. (Το clock proxy χαρακτηριστικό ορίζεται ως η αλλαγή κίνησης που κινάδιου ρολογιού (change of the circadian clock)). Πέραν της δικής τους βάσης δεδομένων, χρησιμοποίησαν δεδομένα 188 εθελοντών από τη βάση δεδομένων MESA [100][101][102] για γενίκευση των αποτελεσμάτων χωρίς να γίνει εκπαίδευση σε αυτή τη βάση. Τελικά, το καλύτερο μοντέλο, ένα Νευρωνικό Δίκτυο (multilayer perceptron (MLP)), πέτυχε **accuracy 69%** και **Cohen Kappa 40%** για ταξινόμηση 3 σταδίων ύπνου (**Wake/NREM/REM**). Σε αντίστοιχη έρευνα [68], οι ερευνητές χρησιμοποίησαν μόνο τους ακατέργαστους καρδιακούς παλμούς (raw heart rate) για την ταξινόμηση του ύπνου σε 4 στάδια (**Wake/Light/Deep/REM**) με δεδομένα από τη βάση MESA και SHHS για την εκπαίδευση του μοντέλου και τη βάση δεδομένων CinC για γενίκευση αποτελεσμάτων. Συγκεκριμένα, εξήγαγαν τους καρδιακούς παλμούς από ηλεκτροκαρδιογράφημα ECG και για την μοντελοποίηση τροφοδότησαν ένα συνελικτικό δίκτυο (CNN) με τα ακατέργαστα δεδομένα καρδιακών παλμών. Τελικά, στη γενίκευση αποτελεσμάτων, το μοντέλο πέτυχε **accuracy 72%** και **Cohen Kappa 55%** στη βάση CinC και στη βάση δεδομένων MESA, σε 194 συμμετέχοντες, το μοντέλο πέτυχε **accuracy 80%** και **Cohen Kappa 69%**. Επιπρόσθετα, στο [103], οι ερευνητές χρησιμοποίησαν δεδομένα καρδιακών παλμών (heart rate), που είχαν εξαχθεί από ηλεκτροκαρδιογράφημα (ECG) από 50 εθελοντές κάτω από συγκεκριμένες/ελεγχόμενες συνθήκες θερμοκρασίας και υγρασίας (25deg, 55% υγρασία). Για την ταξινόμηση του ύπνου σε 5 στάδια (5-Stages), εξήγαγαν χαρακτηριστικά από RR-Intervals από το πεδίο συχνότητας (frequency domain features) και χρησιμοποίησαν ένα μοντέλο επαναλαμβανόμενο νευρωνικό δίκτυο (Recurrent Neural Network - RNN) και ένα μοντέλο Markov (Hidden Markov Model - HMM). Τα καλύτερα αποτελέσματα τα έδωσε το επαναλαμβανόμενο νευρωνικό δίκτυο (RNN) με **accuracy 66%**. Σκοπός τους όμως είναι στο επόμενο στάδιο να ταξινομήσουν τα δεδομένα ύπνου χωρίς δεδομένα καρδιακών παλμών (heart rate) που προέρχονται από ηλεκτροκαρδιογράφημα (ECG), που αποτελεί διεισδυτικό τρόπο καταγραφής δεδομένων (μέσω PSG), αλλά από φορητές συσκευές (wearables/smartwatches). Τέλος, στο [104], οι ερευνητές χρησιμοποίησαν δεδομένα ακτιγραφίας και RR-Intervals για την ταξινόμηση του ύπνου σε 5 στάδια εκτελώντας διάφορα πειράματα με συνδυασμό δεδομένων (ακτιγραφία + RR-Intervals) αλλά και με κάθε είδος δεδομένων ξεχωριστά (πχ μόνο RR-Intervals). Εξήγαγαν χαρακτηριστικά από τομέα του χρόνου και συχνότητας (Time and Frequency Domain) από τα δεδομένα ακτιγραφίας και RR-Intervals. Το καλύτερο μοντέλο που δέχεται ως είσοδο μόνο εξαγόμενα χαρακτηριστικά από τα RR-Intervals, ήταν ένα Bidirectional LSTM (δηλαδή LSTM δύο κατευθύνσεων) και απέδωσε **accuracy 56%** και **F1-Score 53%** για την ταξινόμηση του ύπνου σε 5 στάδια. Οπότε συμπεραίνουμε ότι η εξαγωγή χαρακτηριστικών από δεδομένα RR-Intervals είναι πολύ συνηθισμένη και δίνει ικανοποιητικά αποτελέσματα στην ταξινόμηση των δεδομένων ύπνου

και σε 5 στάδια ύπνου.

Συμπερασματικά, με βάση τις πρόσφατες έρευνες, η ταξινόμηση του ύπνου σε 4 στάδια μόνο με την χρήση δεδομένων καρδιακών παλμών από φορητές συσκευές (wearables) αποτελεί ένα δύσκολο πρόβλημα στο τομέα του ύπνου είτε εξάγουμε χαρακτηριστικά από αυτά είτε όχι.

Λαμβάνοντας υπόψιν τις παραπάνω πρόσφατες έρευνες στο τομέα της ταξινόμησης των σταδίων του ύπνου στη βάση δεδομένων MESA με χρήση ακατέργαστων δεδομένων, αναπτύξαμε παρόμοια μεθοδολογία χρησιμοποιώντας τα ακατέργαστα RR-Intervals της βάσης δεδομένων MESA. Συγκεκριμένα, στόχος είναι να συγκρίνουμε τις προηγούμενες μεθόδους, όπου εξηγάγαμε χαρακτηριστικά (actigraphy + HRV) από τα ακατέργαστα βιοσήματα, με τις μεθόδους χρήσης ακατέργαστων (raw) δεδομένων χωρίς εξαγωγή χαρακτηριστικών και να συγκρίνουμε ποιοτικά ποια μπορεί να αποτελέσει **καλύτερη προσέγγιση του προβλήματος ταξινόμησης** του ύπνου με βάση δύο άξονες: Πρώτον, λόγω της τεχνολογικής ανάπτυξης φορητών συσκευών (wearables) και της δυσκολίας απόκτησης επαρκούς όγκου δεδομένων και διαφορετικού είδους (π.χ ακτιγραφίας, καρδιακοί παλμοί, επίπεδα οξυγόνου), είναι αναγκαία η ταξινόμηση του ύπνου με όσο το δυνατόν λιγότερα διαφορετικού είδους δεδομένα, όπως για παράδειγμα μόνο με τη χρήση καρδιακών παλμών. Και δεύτερον, η απόδοση των μοντέλων Μηχανικής Μάθησης εξαρτάται άμεσα από τη διαδικασία επεξεργασίας δεδομένων, δηλαδή την εξαγωγή ή όχι χαρακτηριστικών από τα βιοσήματα ώστε να βγάλουμε συμπεράσματα. Επομένως, εφόσον τα μοντέλα δέχονται ως είσοδο διαφορετικού είδους δεδομένα τότε εξαρτάται σε μεγάλο βαθμό και η απόδοση τους στο πρόβλημα ταξινόμησης. Τέλος, ένας επιπλέον στόχος είναι να πετύχουμε τα καλύτερα αποτελέσματα, με βάση κάποιες μετρικές (π.χ. accuracy, F1-Score), με όσον το δυνατόν μικρότερη ποικιλομορφία δεδομένων, όπως για παράδειγμα με τη χρήση μόνο των καρδιακών παλμών. Με αυτόν τον τρόπο θα συμπεράνουμε **αν αρκεί η χρήση μόνο των καρδιακών παλμών για την σωστή και έγκυρη ταξινόμηση των δεδομένων ύπνου.**

### 3.2 Δεδομένα RR- Intervals - Προεπεξεργασία

Στη μεθοδολογία που ακολουθεί χρησιμοποιήσαμε μόνο τα RR-Intervals του κάθε χρήστη της βάσης δεδομένων MESA. Λεπτομερέστερα, για κάθε χρήστη ακολουθήθηκε η διαδικασία των προηγούμενων μεθόδων του Κεφαλαίου 2, που εξηγάγαμε χαρακτηριστικά από τα δεδομένα RR-Intervals, με τη διαφορά ότι τώρα χρησιμοποιούμε μόνο τα RR-Intervals για μοντελοποίηση και ταξινόμηση. Αρχικά, τα RR-Intervals, ορίζονται ως ο χρόνος μεταξύ δύο διαδοχικών R-Points, δηλαδή ο χρόνος μεταξύ δύο διαδοχικών R κορυφών του QRS Complex [105]. Οπότε, όσον αφορά την προεπεξεργασία των δεδομένων RR-Intervals, για κάθε χρήστη, αρχικά ευθυγραμμίζουμε τα δεδομένα ακτιγραφίας και R-points για κάθε εποχή 30sec όπου έχει γίνει η ταυτοποίηση της κλάσης (labeling) ώστε να αντιστοιχούν στο ίδιο χρονικό διάστημα 30sec. Η διαδικασία αυτή πραγματοποιείται με βάση το overlapping.csv αρχείο που προσφέρει η βάση MESA όπως περιγράψαμε και στο Κεφάλαιο 3. Δηλαδή, για κάθε χρήστη, το σήμα R-points κόβεται στο μέγεθος του αντίστοιχου αρχείου ακτιγραφίας επειδή ανά 30 δευτερόλεπτα έχουμε και μια τιμή "δραστηριότητας" ακτιγραφίας. Υπενθυμίζουμε, ότι η συχνότητα δειγματοληψίας των δεδομένων ακτιγραφίας είναι 30 δευτερόλεπτα:  $f_s = \frac{1}{30} \text{Hz}$ . Επίσης, αφαιρούνται τα R-points τα οποία επικαλύπτονται ή το ρολόι δεν φοριέται από τον χρήστη. Εν συνεχεία, το RR-Intervals σήμα υπολογίζεται με βάση τον ορισμό του ( $RR = (T_{R_{t+s}} - T_{R_t})$ , όπου  $T_R$  ο χρόνος που συμβαίνει ένα R-point και  $s$  ο χρόνος μεταξύ δύο διαδοχικών R-points). Επακόλουθα, τα RR-Intervals πολλαπλασιάζονται επί 1000 ώστε οι αλγόριθμοι ταξινόμησης να μην διαχειρίζονται δεδομένα πολύ μικρής τάξης μεγέθους, δηλαδή τα seconds να μετατραπούν σε milliseconds. Σημαντικό να αναφερθεί είναι ότι το σύνολο τιμών των δεδομένων RR-Intervals είναι από 0.1 δευτερόλεπτα έως κάποια δευτερόλεπτα ανάλογα τον χρήστη. Ακολουθεί, ο "καθαρισμός" (cleaning) των RR-Intervals με βάση μια τυπική δι-

αδικασία καθαρισμού των RR-Intervals <sup>1</sup>. Λεπτομερέστερα, αντικαθιστούμε με missing values (NaNs), τα ακραία σημεία (outliers) που βρίσκονται εκτός του χρονικού διαστήματος 300ms έως 2000ms, δηλαδή περιορίζουμε το σύνολο τιμών των RR-Intervals στο σύνολο [300, 2000]ms ώστε να γίνει σωστά η κανονικοποίηση των δεδομένων προς ταξινόμηση και να μην επηρεαστεί από ακραίες τιμές. Ακολούθως, γίνεται γραμμική παρεμβολή (interpolation) των σημείων ώστε να αντικατασταθούν τα missing values (NaNs) με πραγματικές τιμές RR-Intervals. Στη συνέχεια, αντικαθιστούμε με missing values τους εκτοπικούς παλμούς (ectopic beats) και με γραμμική παρεμβολή αντικαθιστούμε τα missing values (NaNs) με πραγματικές τιμές RR-Intervals. Οι έκτοποι παλμοί (ectopic beats) είναι πρώιμοι (πρώροι) ή επιπλέον καρδιακοί παλμοί, που μπορεί να προκαλέσουν επιπλέον αίσθημα παλμών. Πιο απλά, εκτοπικό σημαίνει εκτός τόπου, δηλαδή εκτός κανονικού χρόνου. Οπότε είναι λογικό να θέλουμε να καθαρίσουμε τα σήματα RR-Intervals από εκτοπικούς παλμούς.

Η διαδικασία που μόλις αναφέρθηκε αποτελεί συνολικά την προεπεξεργασία των δεδομένων RR-Intervals για τον καθαρισμό τους με βάση την τυπική διαδικασία καθαρισμού των RR-Intervals. Οπότε, με βάση τα "καθαρισμένα" RR-Intervals **αντί να εξάγουμε χαρακτηριστικά για κάθε 30sec RR-Intervals όπως κάναμε στο Κεφάλαιο 3, θα χρησιμοποιήσουμε το ακατέργαστο σήμα (raw) RR-Intervals για την μοντελοποίηση και ταξινόμηση των δεδομένων ύπνου σε 4 στάδια.**

Σημαντικό είναι να επισημανθεί, ότι μετά την προεπεξεργασία των δεδομένων RR-Intervals και την ευθυγράμμιση των δεδομένων ακτιγραφίας και RR-Intervals, προκύπτει ότι κάθε ακολουθία RR-Intervals 30sec έχει διαφορετικό πλήθος τιμών/σημείων. Δηλαδή, μια ακολουθία 30sec που αντιστοιχεί σε μία κλάση μπορεί να έχει μήκος από 2 έως 90 σημεία RR-Intervals. Αυτό συμβαίνει επειδή διαδοχικοί χτύποι δεν πραγματοποιούνται σε σταθερά διαστήματα, οπότε μέσα σε 30sec δεν περιλαμβάνεται ο ίδιος αριθμός σημείων, καθώς υπάρχει επιπλέον και το σφάλμα μέτρησης και ο "καθαρισμός"/προεπεξεργασία των δεδομένων. Σε αυτό το γεγονός θα βασιστεί η μοντελοποίηση των δεδομένων που θα αναλυθεί στην επόμενη υποενότητα.

Τελικά, μετά την επεξεργασία των δεδομένων καταλήγουμε σε 1.743 χρήστες με "καθαρά" δεδομένα. Το τελικό σύνολο δεδομένων περιέχει τα RR-Intervals για κάθε 30sec και την αντίστοιχη κλάση στην οποία ανήκει η ακολουθία RR-Intervals. Το σύνολο δεδομένων έχει διαστάσεις (1903481, 2), όπου περιέχει τις ακολουθίες RR-Intervals και την αντίστοιχη κλάση στην οποία ανήκει η κάθε 30sec-ακολουθία, όπως στο παρακάτω Σχήμα 3.2.1.

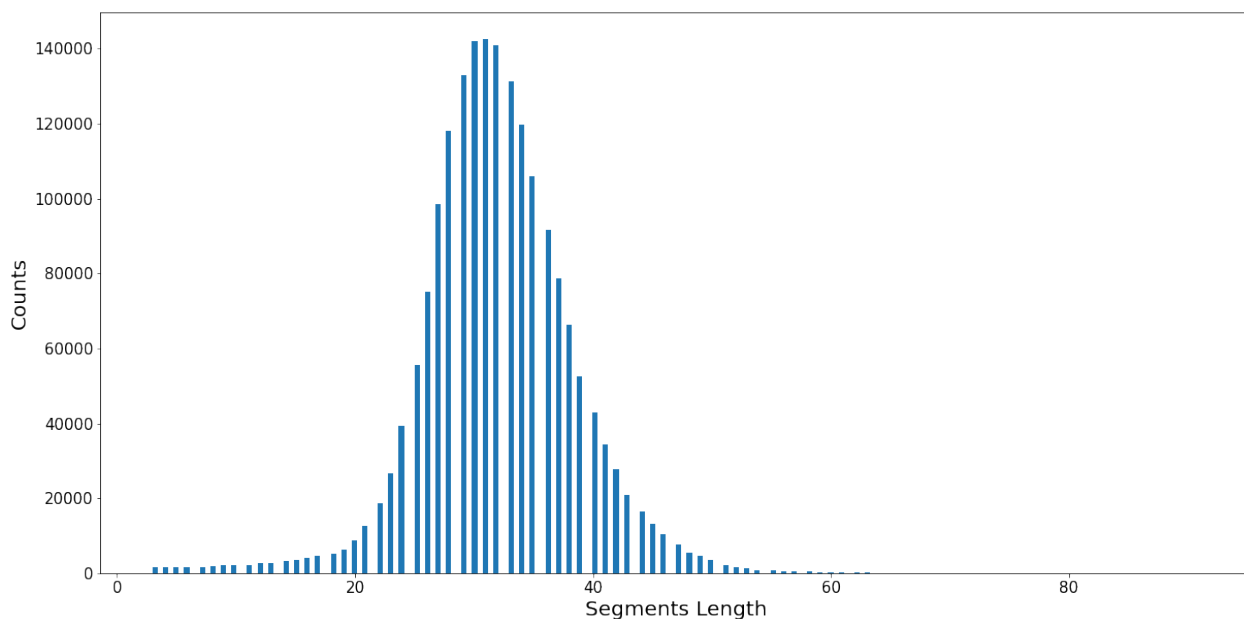
Task 3		
Sleep Stages	# Epochs	%
Wake	652,509(314,784)	34%(20%)
Light	893,472	47%(57%)
Deep	128,874	7%(8%)
REM	229,045	12%(15%)
<b>Total</b>	<b>1,903,900(1,566,175)</b>	<b>100%</b>

Σχήμα 3.2.1: Σύνολο δεδομένων ανά κλάση.

<sup>1</sup>HRV-Analysis: <https://github.com/Aura-healthcare/hrv-analysis>

### 3.2.1 Περιγραφή διαδικασίας - Προετοιμασία δεδομένων για μοντελοποίηση

Σε αυτήν την υποενότητα θα περιγράψουμε την διαδικασία που ακολουθήθηκε μέχρι πριν την τροφοδότηση των μοντέλων, δηλαδή την προετοιμασία των δεδομένων. Αρχικά, όπως αναφέρθηκε και προηγουμένως, οι ακολουθίες RR-Intervals 30 δευτερολέπτων δεν έχουν μεταξύ τους το ίδιο πλήθος σημείων επειδή οι καρδιακοί παλμοί ανά 30 δευτερόλεπτα δεν πραγματοποιούνται σε σταθερά χρονικά διαστήματα. Οπότε, έχοντας "καθαρίσει" τα δεδομένα RR-Intervals δημιουργούμε ένα ιστόγραμμα των μηκών της κάθε ακολουθίας για όλους τους χρήστες για να παρατηρήσουμε αυτό το φαινόμενο. Στο Σχήμα 3.2.2 παραθέτουμε το ιστόγραμμα.



Σχήμα 3.2.2: Ιστόγραμμα μηκών 30sec RR-Intervals για όλους τους χρήστες της βάσης δεδομένων MESA.

Παρατηρούμε ότι κάθε ακολουθία RR-Interval, δεν περιέχει τον ίδιο αριθμό σημείων. Δηλαδή το πλήθος των σημείων ακολουθεί γκαουσιανή και όχι ομοιόμορφη κατανομή. Αν ήταν ίδιος ο αριθμός σημείων της κάθε ακολουθίας RR-Intervals, τότε η κατανομή θα ήταν ομοιόμορφη. Η μέση τιμή είναι 31 σημεία και τυπική απόκλιση 6. Αυτό σημαίνει ότι οι καρδιακοί παλμοί κάθε χρήστη έχουν κυμαινόμενο μήκος RR-Intervals όπου οι περισσότερες 30sec ακολουθίες έχουν μήκος 31. Αντίστοιχα, αυτό μπορεί να φανεί στο Σχήμα 3.2.3 όπου παίρνουμε την **μέση κατανομή όλων των χρηστών**. Παρατηρούμε και σε αυτό το διάγραμμα ότι η μέση τιμή είναι 31 σημεία με τυπική απόκλιση 3 σημεία. Επίσης, από τη μέγιστη και ελάχιστη τιμή της γκαουσιανής κατανομής του Σχήματος 3.2.3, δηλαδή  $\min=10$  και  $\max=48$ , συμπεραίνουμε ότι το μέσο μήκος των RR-Intervals κάποιου χρήστη είναι 10 και ενός άλλου 48, που σημαίνει ότι ανα 30 δευτερόλεπτα, κάθε χρήστης έχει διαφορετικό μοτίβο καρδιακών παλμών.

Συγκεκριμένα, οι δύο κατανομές έχουν χαρακτηριστικά που απεικονίζονται στον Πίνακα 3.1 και 3.2.

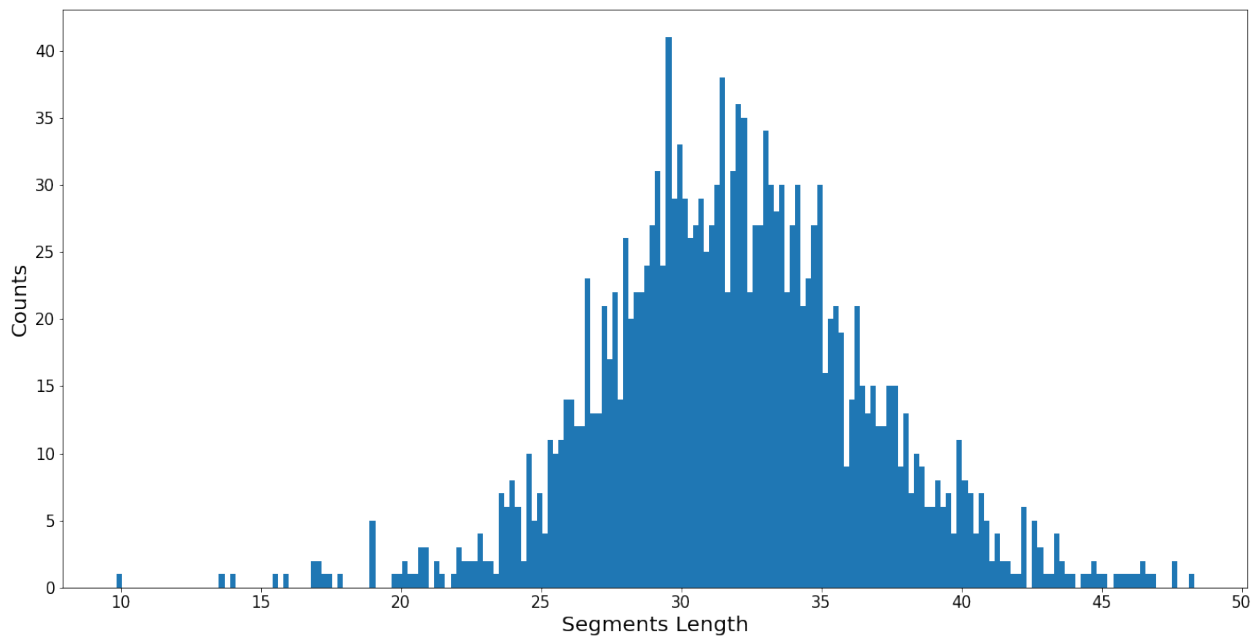
**Έχοντας υπόψιν τα παραπάνω ευρήματα καταλήγουμε ότι το μεγαλύτερο μέρος της πληροφορίας για κάθε χρήστη βρίσκεται στα 30sec RR-Intervals με μήκος 31.** Για μεγαλύτερο πλήθος πληροφορίας κρατάμε μόνο τις ακολουθίες οι οποίες έχουν μήκος 28 έως 35 με βάση τη τυπική απόκλιση του Πίνακα 3.2 και με βάση τα ιστογράμματα 3.2.2 και 3.2.3. Στο Σχήμα 3.2.4 παραθέτουμε το ιστόγραμμα των δεδομένων που κρατήσαμε για μοντελοποίηση. Ο λόγος που δεν κρατήσαμε όλες τις ακολουθίες ήταν επειδή κατά τη μοντελοποίηση, όλες οι ακολουθίες θα πρέπει

<i>Statistics</i>	<i>Values</i>
count	1.904.381
mean	31.1
std	6.3
min	3.0
25%	2.8
50%	3.2
75%	3.6
max	91.0

Πίνακας 3.1: Στατιστικά των μηκών 30sec RR-Intervals.

<i>Statistics</i>	<i>Values</i>
count	1743
mean	31.2
std	4.7
min	9.8
25%	28.95
50%	31.18
75%	34.75
max	48.2

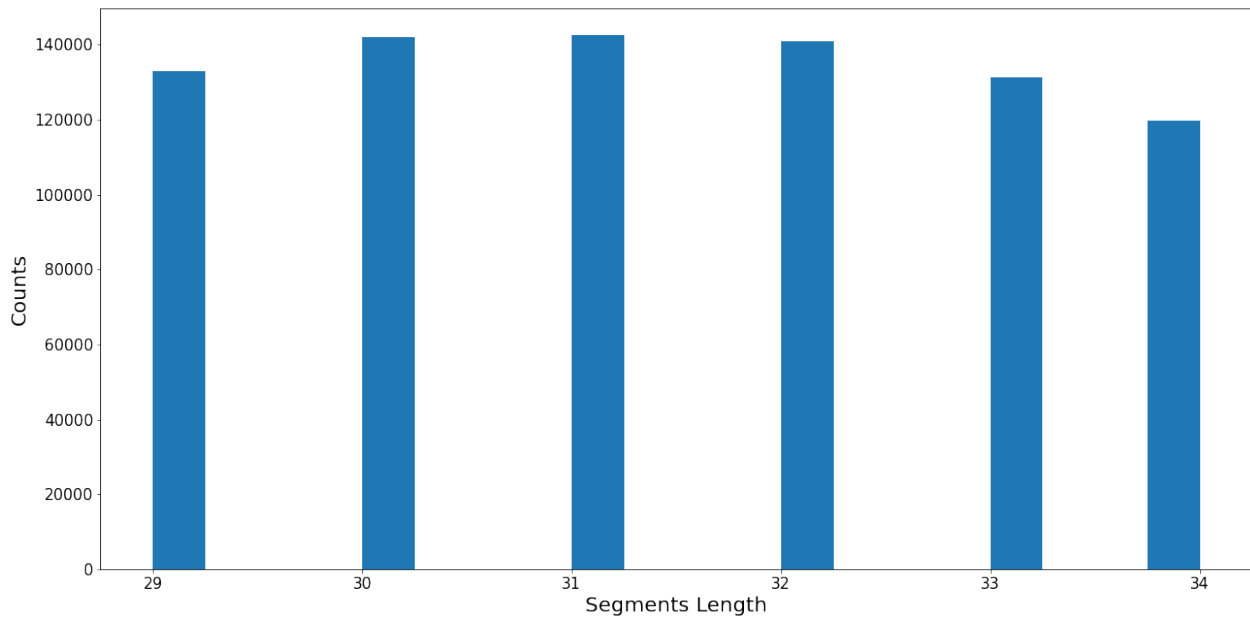
Πίνακας 3.2: Στατιστικά της μέσης κατανομής των μηκών 30sec RR-Intervals.



Σχήμα 3.2.3: Κατανομή μέσου μήκους 30sec RR-Intervals όλων των χρηστών.

να έχουν το ίδιο μήκος. Οπότε χρησιμοποιώντας μεθόδους όπως resampling ή padding να έχουμε το επιθυμητό αποτέλεσμα χωρίς να αλλοιώσουμε την ισχύουσα πληροφορία. Το τελικό μέγεθος του

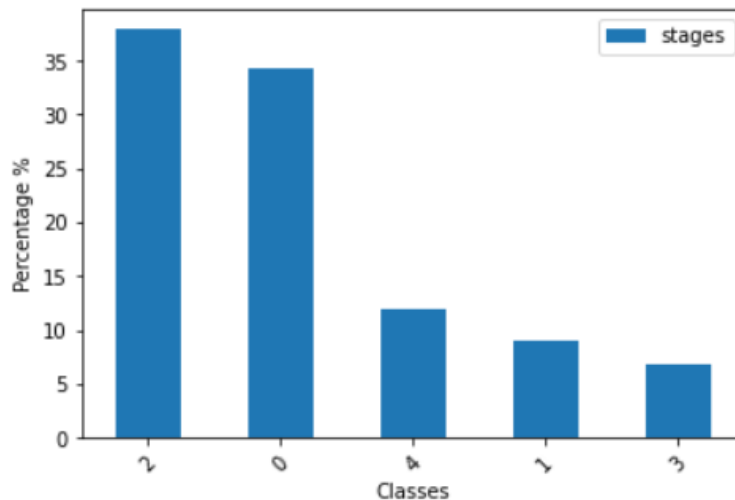




Σχήμα 3.2.4: Ιστόγραμμα 30sec RR-Intervals όλων των χρηστών με τη μεγαλύτερη πληροφορία.

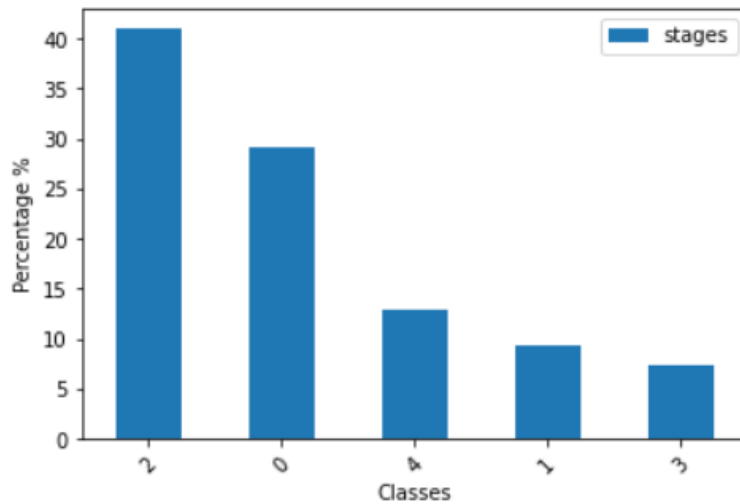
συνόλου δεδομένων είναι (809393, 2). Με βάση αυτές τις ακολουθίες θα γίνει η ταξινόμηση τους στις 4 κλάσεις (Wake/Light/Deep/REM).

Οπότε, έχουμε ακολουθίες RR-Intervals 30sec με μήκος 31 και σε κάθε ακολουθία αντιστοιχεί μια κλάση. Εύλογο είναι να αναρωτηθεί κανείς αν σε αυτές τις ακολουθίες που επιλέχθηκαν διατηρείται η κατανομή των σταδίων του ύπνου, δηλαδή η κατανομή των ετικετών (labels). Το αποτέλεσμα φαίνεται στα Σχήματα 3.2.5 και 3.2.6 <sup>2</sup>). Παρατηρούμε ότι αρχικά το 37% του συνόλου δεδομένων είναι κλάση N2 ενώ μετά τη μέθοδο είναι το 41%. Αντίστοιχα, η κλάση Wake 34% και 29%, η κλάση REM 12% και 13%, η κλάση N1 8.9% και 9.4%, και τέλος η κλάση N3 6.7% και 7.4%.



Σχήμα 3.2.5: Κατανομή κλάσεων πριν την μέθοδο επιλογής ακολουθιών.

<sup>2</sup>Wake: 0, N1:1, N2:2, Deep:3, REM:4

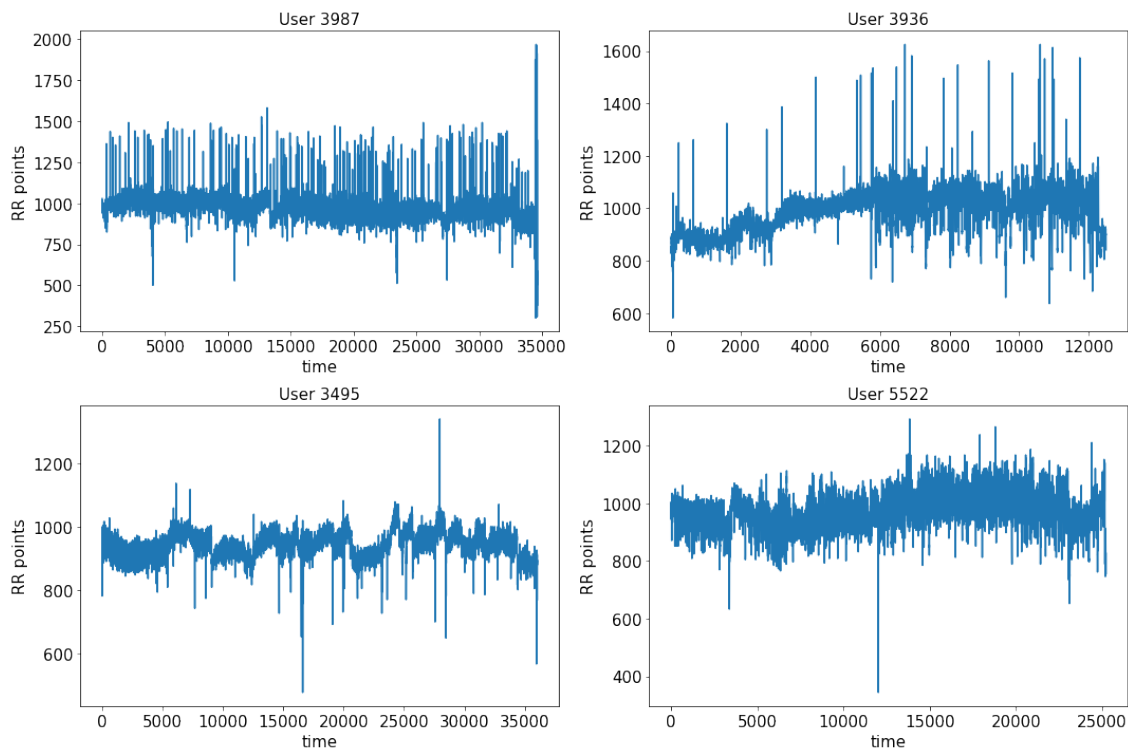


Σχήμα 3.2.6: Κατανομή κλάσεων μετά την μέθοδο επιλογής ακολουθιών.

Παρατηρούμε ότι μετά την επιλογή των ακολουθιών **δεν παρατηρείται να αλλάζει η κατανομή των κλάσεων** αλλά καταλαβαίνουμε ότι η κατανομή των labels στις κλάσεις/στάδια ύπνου είναι ισχυρά ανισόρροπη (imbalance), δηλαδή η κάθε κλάση δεν περιέχει τον ίδιο ή παραπλήσιο αριθμό δεδομένων. Οπότε, με την συγκεκριμένη μέθοδο επιλογής ακολουθιών RR-Intervals: 1) Μειώσαμε κατά πολύ το σύνολο δεδομένων αφού επιλέξαμε μόνο τις ακολουθίες με πλήθος σημείων 28 έως 35, αλλά 2) Διατηρήσαμε την κατανομή των κλάσεων όπως υποδεικνύουν τα Σχήματα 3.2.5 και 3.2.6. Άρα, συμπεραίνουμε ότι η προσέγγιση μπορεί να θεωρηθεί σωστή εφόσον το μόνο που άλλαξε δραματικά είναι το πλήθος των διαθέσιμων δεδομένων για ταξινόμηση.

Στη συνέχεια τα δεδομένα RR-Intervals "καθαρίστηκαν" από τις ελλιπείς τιμές (missing values) και η αντικατάστασή τους έγινε με γραμμική παρεμβολή (linear interpolation) όπως στο [106]. Στα παρακάτω σχήματα παρατίθενται διαγράμματα, ιστογράμματα σχετικά με τα δεδομένα και τους χρήστες.

Στο Σχήμα 3.2.7 παραθέτουμε τα διαγράμματα RR-Intervals για 4 χρήστες. Τα διαγράμματα αυτά αποτελούνται από 30sec RR-Intervals όπου για κάθε ένα γνωρίζουμε σε ποία κλάση ανήκει. Με βάση την προηγούμενη μέθοδο που αναπτύξαμε, κάθε 30sec RR-Intervals αποτελείται από 28 έως 35 σημεία το καθένα για όλους τους χρήστες. Για οπτικοποίηση, τα RR-Intervals του Σχήματος 3.2.7 αποτελούνται από ακολουθίες 30 δευτερολέπτων που περιέχουν από 28 έως 35 σημεία, τα οποία μετά από τη μέθοδο padding όλες οι ακολουθίες έχουν 31 σημεία. Η διαδικασία αυτή θα αναλυθεί παρακάτω. Επιπλέον, στο Σχήμα 3.2.8, παραθέτουμε τέσσερα διαδοχικά RR-Intervals για έναν χρήστη (User=3987) από την αρχή του ύπνου του. Παρατηρούμε ότι όλα τα RR-Intervals έχουν 31 πλήθος σημείων με τιμές κοντά στα 1000ms και ανήκουν στην κλάση 0 (Wake) αφού ο χρήστης είναι στην αρχή του ύπνου του. Εφόσον τα RR-Intervals δεν έχουν την ίδια χρονική απόσταση μεταξύ τους, ο άξονας X δεν απεικονίζει δευτερόλεπτα αλλά πλήθος σημείων, δηλαδή 31. Επίσης, πρέπει να αναφερθεί ότι ο ύπνος κάθε χρήστη δεν έχει την ίδια διάρκεια, γεγονός που απεικονίζεται από τον άξονα X (time) του Σχήματος 3.2.7 για 4 διαφορετικούς χρήστες. Ακόμα, στο Σχήμα 3.2.9 παρατηρούμε τα ιστογράμματα των ίδιων τεσσάρων διαφορετικών χρηστών και συμπεραίνουμε ότι το σύνολο τιμών είναι σε γενικές γραμμές το ίδιο μεταξύ [300, 2000]ms, ανεξαρτήτου χρήστη, προφανώς με διαφορές αφού: 1) οι καρδιακοί παλμοί του κάθε ανθρώπου είναι μοναδικοί και κάθε άνθρωπος κάνει διαφορετικό ύπνο, 2) ο συνολικός χρόνος ύπνου δεν είναι ίδιος για κάθε χρήστη αφού η καταγραφή δεδομένων δεν είναι ίσης χρονικής διάρκειας για κάθε χρήστη. **Επειδή δεν παρατηρείται κάποια ιδιαίτερη διαφορά στα ιστογράμματα μεταξύ των χρηστών, επιλέγουμε να αντιμετωπίσουμε το πρόβλημα ταξινόμησης,**



Σχήμα 3.2.7: Διαγράμματα RR-Intervals για 4 χρήστες. Ο άξονας time δεν απεικονίζει δευτερόλεπτα αλλά πλήθος σημείων όπου κάθε sample αποτελείται από 31 σημεία.

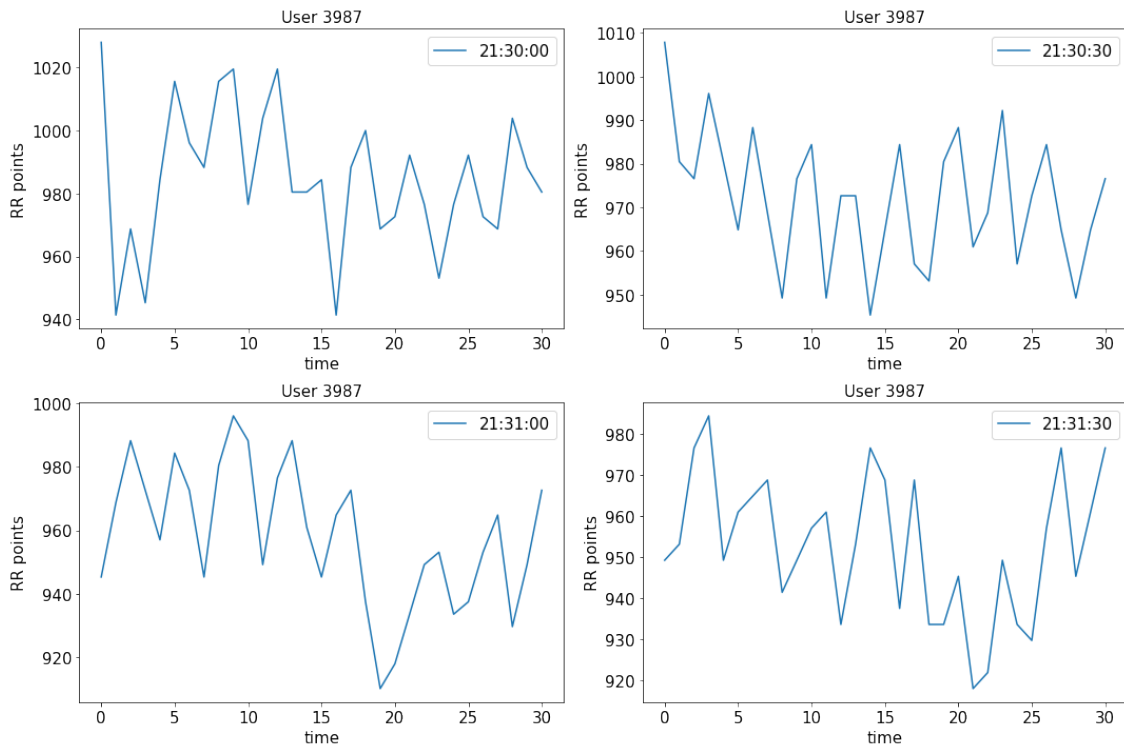
χωρίς να λαμβάνουμε υπόψιν το Id των χρηστών. Επιπρόσθετα, όπως αναφέραμε στην αρχή του παρόντος Κεφαλαίου 3.2, τα δεδομένα έχουν καθοριστεί στο εύρος 300 έως 2000 ms. Οι διαφορές στα ιστογράμματα των χρηστών, δηλαδή στο σύνολο τιμών των RR-Intervals μπορούν να αιτιολογηθούν από το Σχήμα 3.2.10 όπου παρατηρούμε τα υπογράμματα τεσσάρων χρηστών<sup>3</sup>. Παρατηρούμε ότι κάθε χρήστης δεν έχει την ίδια κατανομή κλάσεων, άρα έχει και διαφορετικό σήμα RR-Intervals, άρα και διαφορετική κατανομή των τιμών του σήματος. Επίσης, ο κάθε χρήστης δεν "περνάει" από όλα τα στάδια ύπνου, όπως φαίνεται στο Σχήμα 3.2.11 όπου ο χρήστης User 5522 δεν "περνά" καθόλου στάδιο 4 (REM). Τέλος, το ίδιο συμπέρασμα, που μας οδηγεί και στη ανισορροπία του συνόλου δεδομένων (imbalance dataset), απεικονίζεται στο Σχήμα 3.2.11 με την κατανομή των κλάσεων τεσσάρων διαφορετικών χρηστών. Άρα επιβεβαιώνεται ότι κάθε χρήστης κάνει ξεχωριστό/μοναδικό ύπνο.

#### Προετοιμασία συνόλου δεδομένων για μοντελοποίηση:

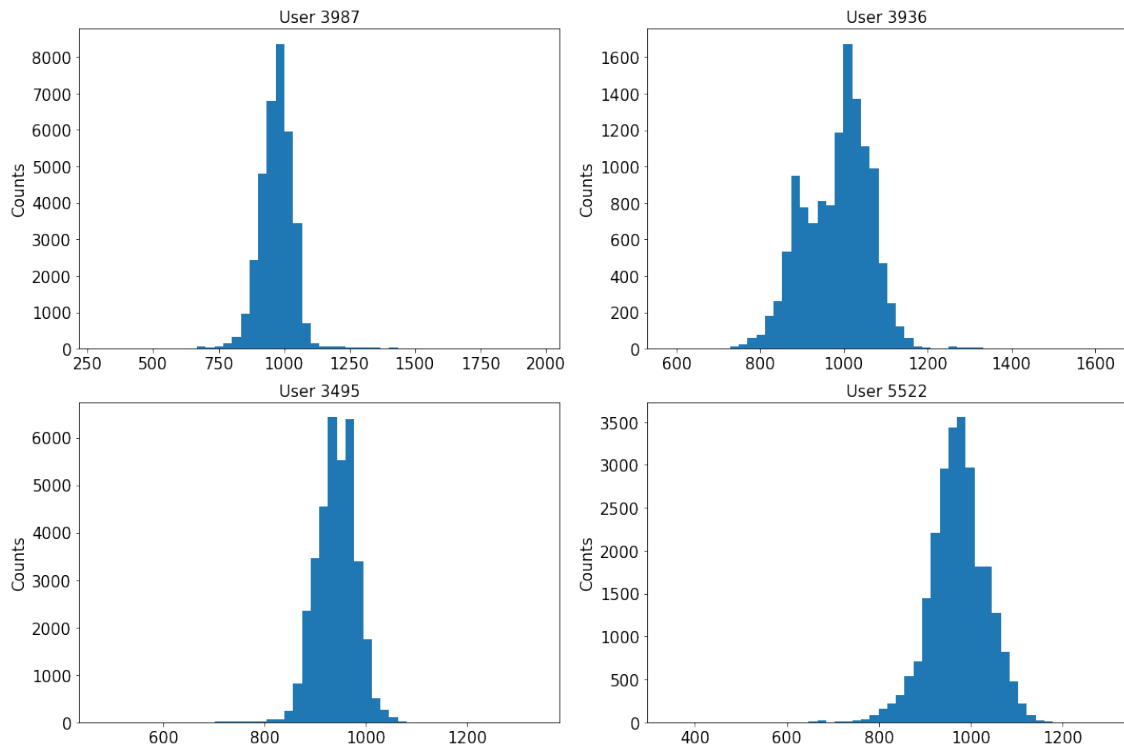
Στη συνέχεια, αφού έχουμε υλοποιήσει μια πρωταρχική ανάλυση δεδομένων ακολουθώντας την παραπάνω μεθοδολογία, σε αυτό το μέρος θα αναλύσουμε την προετοιμασία των δεδομένων για την τροφοδοσία των Deep Learning μοντέλων. Αναλυτικότερα, στη προηγούμενη ενότητα αναφέραμε ότι τα 30sec RR-Intervals διαφέρουν στο πλήθος των σημείων από 28 έως 35. Προκειμένου να έχουν τον ίδιο πλήθος σημείων, ώστε να είναι εφικτή η μοντελοποίησή τους, χωρίσαμε τις 30sec ακολουθίες ανα κλάση<sup>4</sup>) και με την τεχνική padding με μέγιστο μήκος 31, όλες οι ακολουθίες προκύπτουν να έχουν το ίδιο πλήθος 31 σημείων. Ο συγκεκριμένος αριθμός 31, επιλέχθηκε επειδή η μέση τιμή της κατανομής των 30sec RR-Intervals είναι 31. Αυτό απεικονίζεται στους στατιστικούς Πίνακες 4.1 και 4.2. Το padding των ακολουθιών έγινε ανα κλάση και

<sup>3</sup>Wake: 0, N1:1, N2:2, Deep:3, REM:4

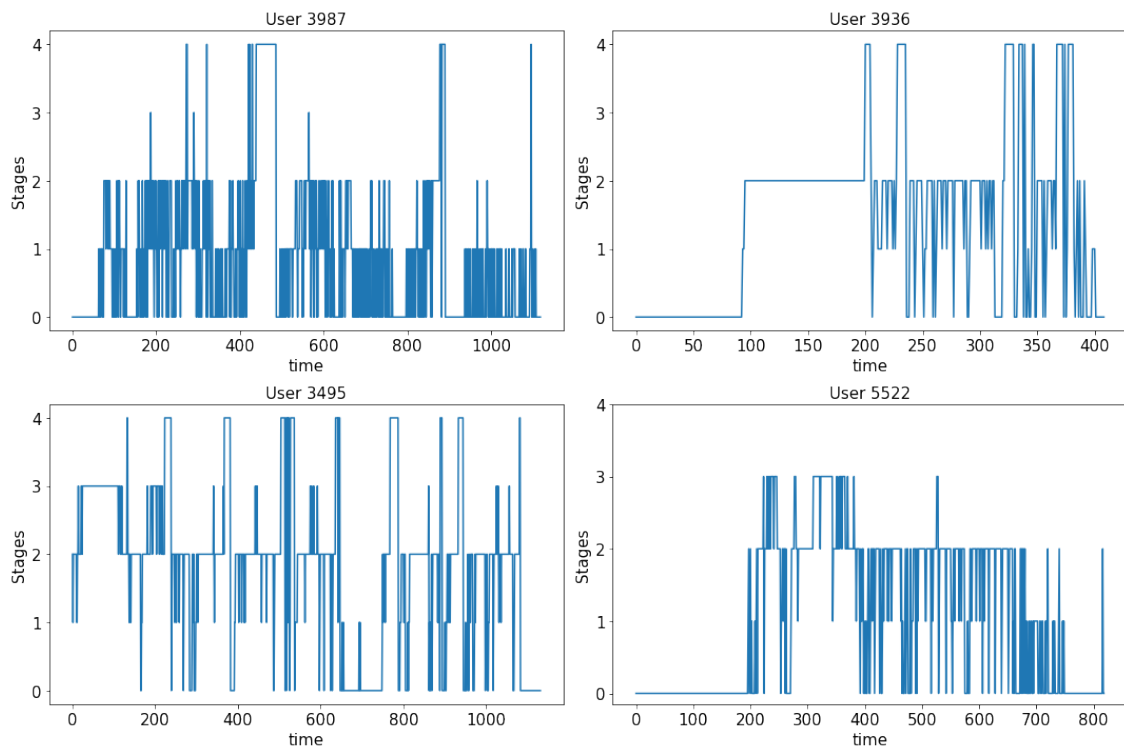
<sup>4</sup>Wake: 0, N1:1, N2:2, Deep:3, REM:4



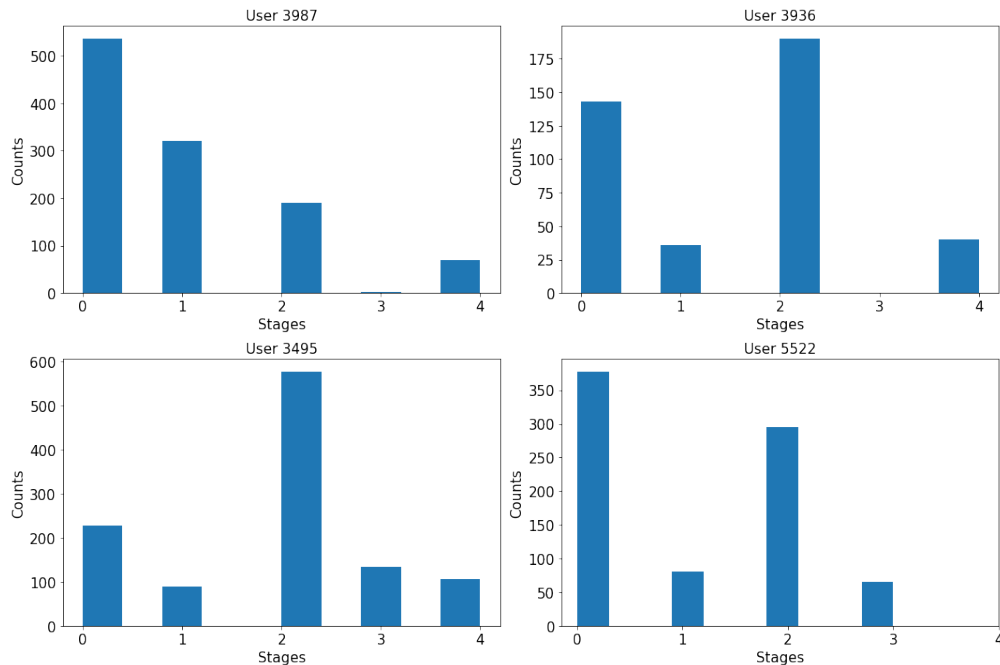
Σχήμα 3.2.8: Διαδοχικά Διαγράμματα RR-Intervals για 1 χρήστη κατά το στάδιο Wake.



Σχήμα 3.2.9: Ιστογράμματα RR-Intervals για 4 χρήστες.



Σχήμα 3.2.10: Υπογράμματα από 4 χρήστες. Ο άξονας time δεν απεικονίζει δευτερόλεπτα αλλά πλήθος σημείων/samples ανά 30 δευτερόλεπτα. Π.χ για τον User 3939 το σύνολο των σημείων RR-points είναι  $400 \cdot 31 = 12400$  όπως φαίνεται στο Σχήμα 3.2.7.



Σχήμα 3.2.11: Κατανομή κλάσεων από 4 χρήστες.

με βάση τη μέση τιμή της κάθε κλάσης. Δηλαδή, η μέση τιμή της κάθε κλάσης προστίθεται ως επιπλέον σημεία, στις ακολουθίες με μήκος μικρότερο του 31. Για παράδειγμα, αν μια ακολουθία, με πλήθος σημείων 28, ανήκει στην κλάση Deep, τότε στην ακολουθία αυτή θα προστεθούν 3 τιμές (στο τέλος της ακολουθίας) με την μέση τιμή της κλάσης Deep (951.5ms, βλέπε πίνακα 4.6), οπότε το συνολικό της μήκος θα είναι 31. Η διαδικασία του padding έγινε ανά κλάση επειδή οι ακολουθίες RR-Intervals έχουν διαφορετική συμπεριφορά μεταξύ τους ανάλογα το στάδιο ύπνου. Η διαφορετική στατιστική συμπεριφορά των σταδίων ύπνου αντικατοπτρίζεται στους Πίνακες 4.3 έως 4.7. Πρέπει να αναφερθεί ότι η τεχνική padding αναφέρθηκε και στην Εισαγωγή (Κεφάλαιο 1) με πιο λεπτομερείς επεξηγήσεις.

<i>Statistics</i>	<i>Values</i>
count	7.314.388
mean	932.65
std	100.45
min	300
25%	882
50%	933
75%	988
max	2000

Πίνακας 3.3: Στατιστικά των ακολουθιών 30sec RR-Intervals για την κλάση Wake.

<i>Statistics</i>	<i>Values</i>
count	2.368.431
mean	952.74
std	88.1
min	300
25%	898
50%	953
75%	1007
max	2000

Πίνακας 3.4: Στατιστικά των ακολουθιών 30sec RR-Intervals για την κλάση N1.

Από τους στατιστικούς πίνακες παρατηρούμε ότι όλες οι κλάσεις έχουν παρόμοια συμπεριφορά, δηλαδή παρόμοια στατιστικά μεγέθη, με τη διαφορά να είναι εμφανής στην τυπική απόκλιση (standard deviation) όπου η κλάση Wake έχει τη μεγαλύτερη (std=100ms) και η Deep Sleep κλάση έχει την μικρότερη (std=76ms). Το γεγονός αυτό αιτιολογείται διότι στην κλάση Deep η κατάσταση του ύπνου είναι σταθερή χωρίς πολλές διακυμάνσεις, ενώ αντίθετα στην κλάση Wake υπάρχουν πολύ περισσότερες διακυμάνσεις επειδή ο χρήστης κινείται περισσότερο από ό,τι στις άλλες κλάσεις<sup>5</sup>. Επίσης, η κλάση Wake έχει και τη χαμηλότερη μέση τιμή (mean=933ms) από όλες τις κλάσεις (mean 955ms) διότι στη κλάση Wake παρατηρείται η υψηλότερη κινητικότητα του ανθρώπου κατά τη διάρκεια του ύπνου οπότε τα R-Points έχουν μικρότερη χρονική απόσταση μεταξύ τους (υψηλότερη τιμή καρδιακών παλμών (Heart Rate)). Όπως είναι λογικό, αν είχαν επιλεγεί περισσότερες ακολουθίες, οι διαφορές ανάμεσα στις κλάσεις μπορεί ήταν πιο ξεκάθαρες. Ο λόγος που δεν επιλέχθηκαν περισσότερες ακολουθίες για την μοντελοποίηση είναι επειδή, όπως αναφέραμε και προηγουμένως στην προεπεξεργασία των RR-

<sup>5</sup>αναφορά στο Κεφάλαιο 3



<i>Statistics</i>	<i>Values</i>
count	1.028.571
mean	953
std	81
min	300
25%	902
50%	953
75%	1003
max	2000

Πίνακας 3.5: Στατιστικά των ακολουθιών 30sec RR-Intervals για την κλάση N2.

<i>Statistics</i>	<i>Values</i>
count	1.859.566
mean	954
std	76.7
min	304
25%	906
50%	953
75%	1000
max	2000

Πίνακας 3.6: Στατιστικά των ακολουθιών 30sec RR-Intervals για την κλάση Deep Sleep.

Intervals, οι ακολουθίες 30 δευτερολέπτων δεν έχουν το ίδιο πλήθος σημείων μεταξύ τους. Γι αυτό το λόγο επιλέξαμε τις ακολουθίες με τον μεγαλύτερο στατιστικά πλήθος σημείων, δηλαδή 28 έως 35 (Σχήματα 3.2.2 και 3.2.3). Επίσης, δεν είναι λογικό να εφαρμόσουμε την τεχνική padding σε ακολουθίες που περιέχουν πολύ λίγα σημεία μέσα σε 30 δευτερόλεπτα. Και αυτό διότι θα δημιουργούσαμε ακολουθίες 30 δευτερολέπτων που δεν είναι πραγματικές (artificial data).

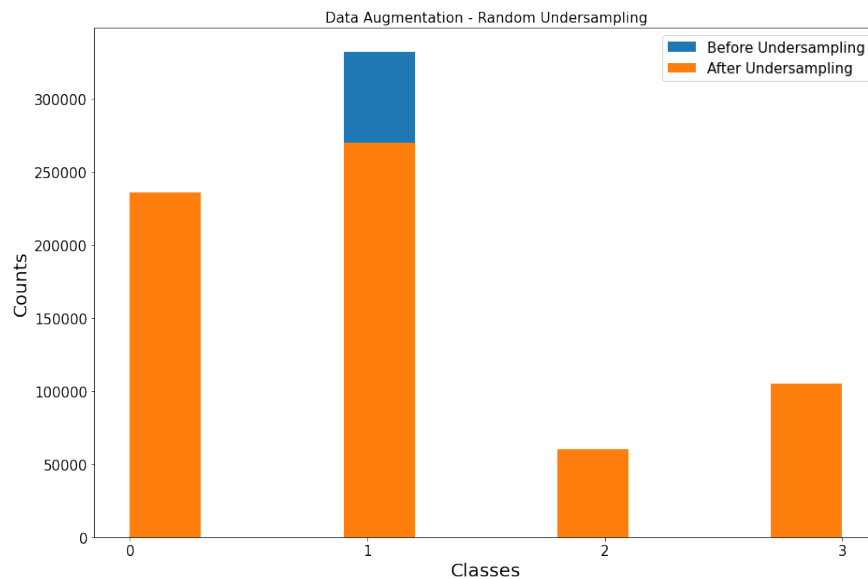
Στη συνέχεια, λόγω του γεγονότος ότι η κλάση N1 έχει πολύ λίγα δεδομένα και η κλάση N2 έχει τον υψηλότερο αριθμό δεδομένων που ανήκουν σε αυτήν, αποφασίσαμε να μην συμπεριλάβουμε την κλάση N1 στο σύνολο δεδομένων, καθώς οι κλάσεις N1 και N2 αφορούν την ίδια κλάση στο πρόβλημα των 4 κλάσεων (NREM stages) [1] (Ενότητα 3.3 Tasks). Αν την είχαμε συμπεριλάβει μαζί με την κλάση N2, τότε το σύνολο δεδομένων θα ήταν ακόμη πιο μη ισορροπημένο, καθώς η κλάση N2 είναι η κλάση με το μεγαλύτερο πλήθος δεδομένων και τότε όπως είναι λογικό τα μοντέλα θα "μάθαιναν" πολύ καλά αυτή την κλάση και λιγότερο τις υπόλοιπες (Wake, Deep, REM). Οπότε **ορίζουμε ως κλάση Light μόνο την κλάση N2 χωρίς την N1**. Τελικά, συγχωνεύοντας τις 5 κλάσεις σε 4 έχουμε: Wake, Light Sleep (N2), Deep Sleep(N3) και REM, δηλαδή η κλάση N2 απεικονίζει την κλάση Light στο πρόβλημα 4 σταδίων ύπνου.

<i>Statistics</i>	<i>Values</i>
count	3.263.091
mean	951.5
std	81.15
min	300
25%	902
50%	949
75%	1000
max	2000

Πίνακας 3.7: Στατιστικά των ακολουθιών 30sec RR-Intervals για την κλάση REM.

### 3.2.2 Εφαρμογή τεχνικών αύξησης συνόλου δεδομένων (Data Augmentation)

Παράλληλα, προκειμένου να περιορίσουμε το πρόβλημα της ανισορροπίας του συνόλου δεδομένων (imbalance dataset, Σχήμα 3.2.1) που αναφέραμε στην Ενότητα 3.2.1, εφαρμόζουμε μια σειρά από **τεχνικές τεχνητής (artificial) αύξησης του συνόλου δεδομένων (Data Augmentation)**. Πριν την διαδικασία Data Augmentation, χωρίζουμε τα δεδομένα, σε δεδομένα εκπαίδευσης (train 80%) και δεδομένα αξιολόγησης (test 20%), ώστε: 1) οι τεχνικές να εφαρμοστούν στο train σύνολο δεδομένων και 2) για να μην κάνουμε αξιολόγηση των αλγορίθμων σε τεχνητά δεδομένα (artificial data) που δεν είναι πραγματικά. Οπότε, αρχικά εφαρμόζουμε τυχαία υποδειγματοληψία (Random Undersampling) <sup>6</sup> στη κλάση Light Sleep επειδή είναι η κλάση με το μεγαλύτερο αριθμό δεδομένων. Αφαιρούμε τόσα δεδομένα ώστε να έχει λίγο περισσότερα δεδομένα από την δεύτερη σε μέγεθος κλάση (Wake). Στο Σχήμα 3.2.12 απεικονίζεται το αποτέλεσμα της υποδειγματοληψίας. <sup>7</sup>



Σχήμα 3.2.12: Υποδειγματοληψία της κλάσης Light Sleep.

Από το Σχήμα 3.2.12 παρατηρούμε ότι το αποτέλεσμα της τυχαίας υποδειγματοληψίας οδηγεί στην

<sup>6</sup>Random Undersampler: <https://imbalanced-learn.org/stable/references/index.html>

<sup>7</sup>Wake:0, Light:1, Deep:2, REM:3

μείωση τους συνόλου δεδομένων της κλάσης Light κατά:  $loss\% = \frac{(331797-270000)*100}{331797} = 8.4\%$ , οπότε εξισορροπείται το σύνολο δεδομένων σχετικά με τις κλάσεις Wake και Light. Η επιλογή του 18% έγινε ευριστικά και η διαφορά της κλάσης Light από τη κλάση Wake είναι 12%. Δηλαδή, περίπου το ποσοστό (18%) που αφαιρέσαμε από τη κλάση Light είναι και η διαφορά της, ποσοστιαία, με κλάση Wake (12%). Με αυτόν τον τρόπο: 1) εξισορροπείται το σύνολο δεδομένων ως προς τις κλάσεις Wake και Light, και 2) αποτρέπουμε τα μοντέλα να "μάθουν" περισσότερο την κλάση Light σε σχέση με τις υπόλοιπες κλάσεις.

Επιπλέον, από το Σχήμα 3.2.12, παρατηρούμε ότι η κλάση Deep Sleep και REM έχουν πολύ μικρό αριθμό δεδομένων σε σχέση με τις κλάσεις Wake και Light. Οπότε, προκειμένου να αντιμετωπίσουμε αυτό το πρόβλημα, εφαρμόζουμε επιπλέον τεχνικές Data Augmentation [95]. Μετά από πειραματισμό, για την κλάση Deep, αντιστρέψαμε τα δεδομένα ως προς τον χρόνο (Reverse) και επιπλέον τα αντιστρέψαμε και προσθέσαμε και θόρυβο 20msec (Jittering). Για την κλάση REM, μετά από πειραματισμό, απλά αντιστρέψαμε τα δεδομένα ως προς τον χρόνο (Reverse). Τα αποτελέσματα των τεχνικών Jittering, Reverse και ο συνδυασμός τους (Reverse+Jittering) απεικονίζονται στο Σχήμα 3.2.13. Ο πειραματισμός έγινε με βάση τα αποτελέσματα που δίνουν οι αλγόριθμοι Deep Learning που θα αναλυθούν στην επόμενη ενότητα. Καταλήξαμε ότι οι **τεχνικές Jittering και Reverse** δίνουν τα καλύτερα αποτελέσματα. Η κύρια ιδέα είναι ότι ανάλογα την φυσιολογία του κάθε σταδίου ύπνου, μας οδηγεί να εφαρμόσουμε την αντίστοιχη τεχνική Data Augmentation ώστε: **1) να διαχωριστούν οι κλάσεις μεταξύ τους, στον μέγιστο βαθμό για να αναγνωριστούν καλύτερα από τους αλγόριθμους Deep Learning και 2) Να εξισορροπηθεί το σύνολο των κλάσεων Deep και REM σε σχέση με τις κλάσεις Wake και Light όπου έχουν και το μεγαλύτερο πλήθος δεδομένων.** Στο Σχήμα 3.2.13 και 3.2.14 απεικονίζονται τα αποτελέσματα της διαδικασίας. Τέλος, συγκρίνοντας τα Σχήματα 3.2.12 και 3.2.14 παρατηρούμε τον διπλασιασμό του πλήθους των δεδομένων των κλάσεων Deep και REM. **Ο λόγος που δεν δημιουργήσαμε περισσότερα δεδομένα μέσω Data Augmentation για τις κλάσεις REM και Deep, είναι επειδή η αναλογία τεχνητών δεδομένων προς πραγματικών είναι 1:1 λόγω του διπλασιασμού των πραγματικών δεδομένων.** Άρα, εάν ήταν μεγαλύτερη του 1 η παραπάνω αναλογία, τότε τα μοντέλα θα μαθαίνουν τεχνητά μη πραγματικά δεδομένα, κάτι που δεν θέλουμε.

Συμπερασματικά, παρατηρούμε από το Σχήμα 3.2.14 ότι έχει αντιμετωπιστεί, σε μεγάλο ποσοστό, το πρόβλημα της ανισοκατανομής των δεδομένων (imbalance dataset) με τις τεχνικές Data Augmentation:

- Κλάση Light: Random Undersampling
- Κλάσεις Deep και REM: Jittering, Reverse

Στην επόμενη ενότητα θα παραθέσουμε αναλυτικά τον τρόπο με τον οποίο οι τεχνικές του Data Augmentation ενίσχυσαν τις αποδόσεις των αλγορίθμων Deep Learning για το πρόβλημα ταξινόμησης των 4 κλάσεων.

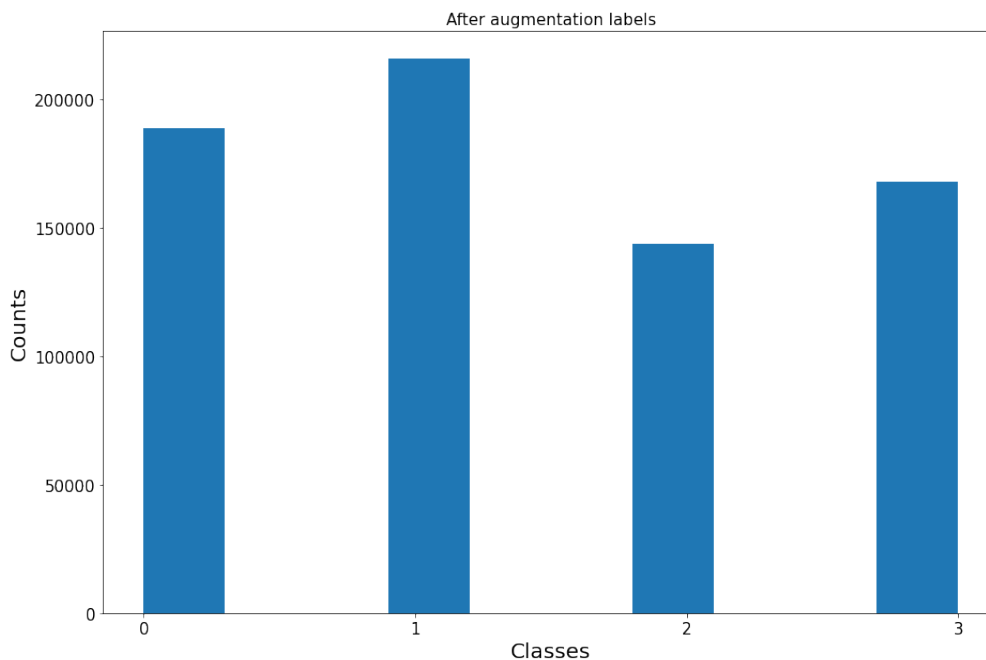


Σχήμα 3.2.13: Data Augmentation: (1) Reverse Deep Class, (2) Reverse + Jitter Deep Class και (3) Reverse REM Class.

### 3.3 Μοντελοποίηση Δεδομένων RR-Intervals με αλγορίθμους Deep Learning

Σε αυτή την ενότητα θα αναπτύξουμε την μεθοδολογία που ακολουθήθηκε για την μοντελοποίηση των δεδομένων RR-Intervals, με στόχο την ταξινόμηση τους σε 4 στάδια ύπνου. Αρχικά, έχοντας ολοκληρώσει τις τεχνικές data Augmentation στα δεδομένα εκπαίδευσης (train), κανονικοποιούμε το σύνολο των ακολουθιών με τέτοιο τρόπο ώστε να έχουν μέση τιμή 0 και τυπική απόκλιση 1 (Standard Scaling). Επιλέχθηκε αυτός ο τρόπος κανονικοποίησης δεδομένων επειδή έδωσε τα καλύτερα αποτελέσματα ταξινόμησης σε σχέση με τους Min-Max Scaler και Robust Scaler. Επίσης, το σύνολο δεδομένων των κλάσεων (labels) έχει μετατραπεί σε δυαδική μορφή με τη μέθοδο One Hot Encoding επειδή οι αλγόριθμοι Deep Learning που έχουν υλοποιηθεί με την βιβλιοθήκη Keras <sup>8</sup>, χρειάζονται ως έξοδο τους οι ετικέτες (labels) να έχουν την δυαδική μορφή One Hot Encoding. Τελικά, τα δεδομένα έχουν τα παρακάτω στοιχεία:

<sup>8</sup>Βιβλιοθήκη Keras: <https://keras.io>



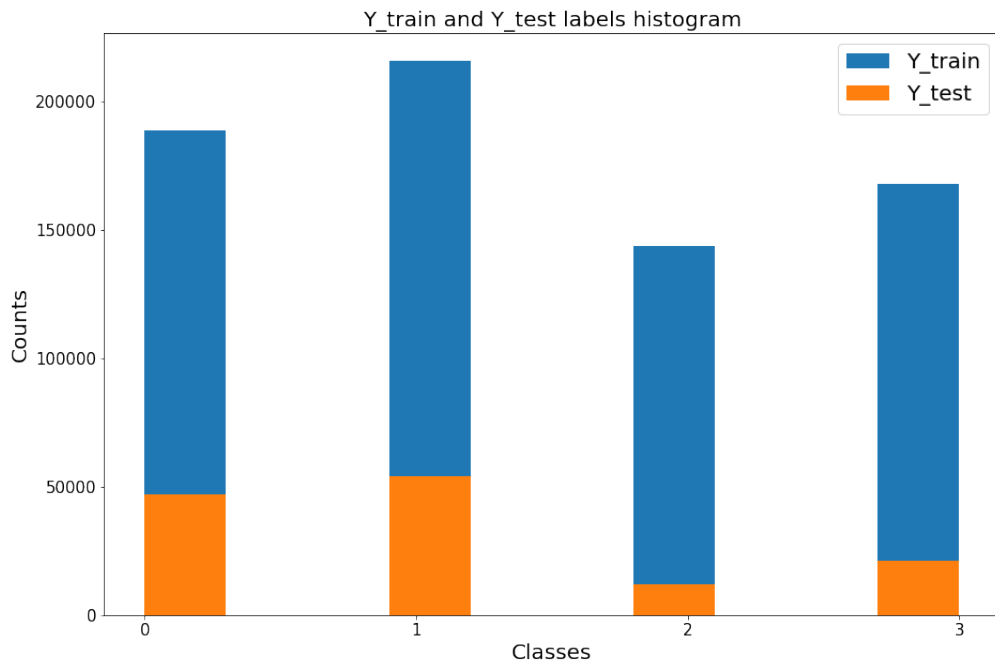
Σχήμα 3.2.14: Κατανομή κλάσεων μετά τις τεχνικές Data Augmentation.

- Διαστάσεις συνόλου  $\mathbf{X}_{train}$ : (717.093, 31, 1)
- Διαστάσεις συνόλου  $\mathbf{X}_{test}$ : (134.239, 31, 1)
- Διαστάσεις συνόλου (labels) κλάσεων  $\mathbf{y}_{train}$ : (717.093, 4)
- Διαστάσεις συνόλου (labels) κλάσεων  $\mathbf{y}_{test}$ : (134.239, 4)

Επίσης, πρέπει να επισημανθεί ότι κατά το διαχωρισμό των δεδομένων, σε train και test, η κατανομή των κλάσεων είναι η ίδια. Στο Σχήμα 3.3.1 φαίνεται το αποτέλεσμα, και είναι σημαντικό διότι **οι μετρικές που θα εξαχθούν για την αξιολόγηση των μοντέλων προϋποθέτουν την ίδια κατανομή**, τόσο κατά την μοντελοποίηση, όσο και κατά την αξιολόγηση στα αντίστοιχα σύνολα δεδομένων εξαιτίας της ανισορροπίας των κλάσεων (class imbalance).

Οπότε, για το πρόβλημα της ταξινόμησης των RR-Intervals σε 4 κλάσεις, υλοποιήσαμε τον παρακάτω σχεδιασμό ως προς τα επιλεγμένα μοντέλα προς εκπαίδευση:

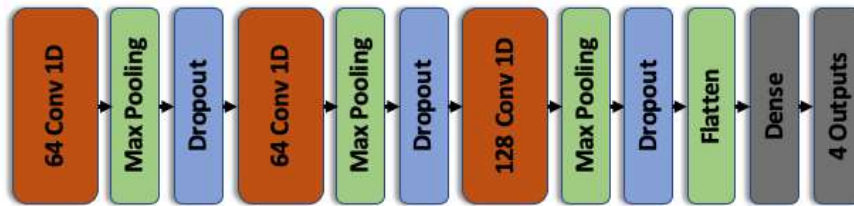
- Convolutional Neural Network (CNN)
- Long Short-Term Memory Network (LSTM)
- Long Short-Term Memory Network with Attention (LSTM + Attention)



Σχήμα 3.3.1: Κατανομή κλάσεων στο σύνολο δεδομένων train και test.

### 3.3.1 Μοντελοποίηση με Συνελικτικό Νευρωνικό Δίκτυο - CNN

Έχοντας προετοιμάσει τα δεδομένα, δημιουργήσαμε ένα συνελικτικό νευρωνικό δίκτυο που απεικονίζεται στο παρακάτω Σχήμα 3.3.2.



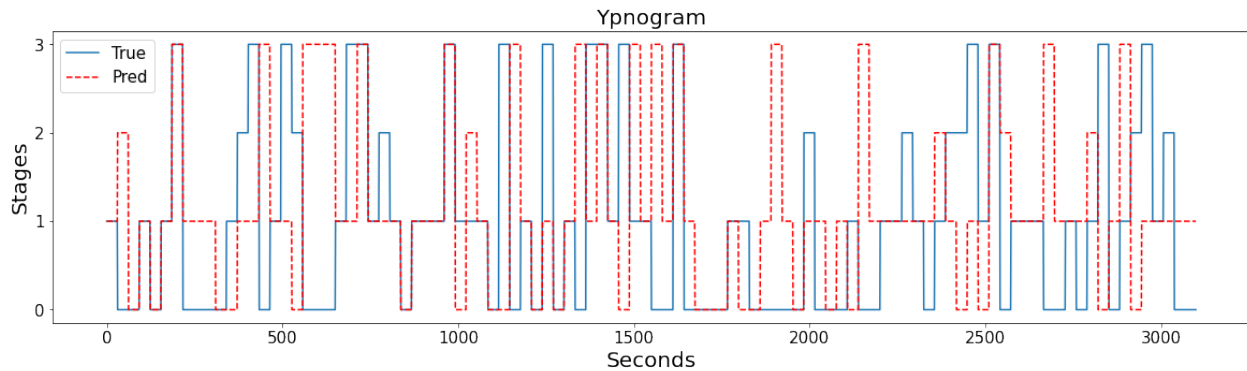
Σχήμα 3.3.2: Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου.

Το CNN μοντέλο δέχεται ως είσοδο τις 30sec RR-Intervals ακολουθίες με μέγεθος (31,1) και ως έξοδο την κλάση στην οποία ανήκει σε δυαδική μορφή (One Hot Encoding). Συνοπτικά, τα φίλτρα όλων των επιπέδων συνέλιξης (Convolutional Layers) έχουν μέγεθος 3 (kernel size) και εφαρμόζεται causal padding. Τα Max Pooling επίπεδα, έχουν μέγεθος pool ίσο με 2 και stride ίσο με 2, με same padding. Τα Dropout επίπεδα έχουν regularization rate ίσο με 20%. Τα εξαγόμενα χαρακτηριστικά από τα επίπεδα συνέλιξης οριζοντιώνονται (Flatten) και ακολουθεί ένα Feed-Forward νευρωνικό δίκτυο με 100 νευρώνες που καταλήγει στην έξοδο 4 νευρώνων, όσες δηλαδή και οι κλάσεις.

Η εκπαίδευση του μοντέλου γίνεται για 100 εποχές, χρησιμοποιώντας batch size ίσο με 1024 και την μέθοδο ReduceLRonPlateau με ποσοστό 0.1. Επίσης, το 20% των δεδομένων εκπαίδευσης χρησιμοποιείται για validation ώστε να αξιολογηθεί η απόδοση του μοντέλου.

Αφού έχει εκπαιδευτεί το CNN μοντέλο, γίνεται η πρόβλεψη στο σύνολο δεδομένων αξιολόγησης (test). Στο Σχήμα 3.3.3 παρατίθεται υπνόγραμμα για 100 διαδοχικές ακολουθίες του συνόλου αξιολόγησης

και τις αντίστοιχες πραγματικές κλάσεις στις οποίες ανήκουν οι ακολουθίες.



Σχήμα 3.3.3: Υπνόγραμμα 100 διαδοχικών ακολουθιών του συνόλου αξιολόγησης και των αντίστοιχων πραγματικών κλάσεων. Σημείωση - Wake:0, Light:1, Deep:2, REM:3.

Εποπτικά, στο Σχήμα 3.3.3, παρατηρούμε ότι η κλάση Deep δεν μπορεί να αναγνωριστεί καλά από το συνεκτικό δίκτυο ενώ οι υπόλοιπες κλάσεις μπορούν να αναγνωριστούν καλύτερα (αφού επικαλύπτονται οι δύο γραμμές True και Predictions). Το συμπέρασμα αυτό μπορεί να διαγνωσθεί και από τον πίνακα μετρικών του συνόλου αξιολόγησης και από τον πίνακα σύγχυσης (Confusion Matrix - Σχήμα 3.3.4). Δηλαδή, στο Πίνακα 3.8 απεικονίζονται τα αποτελέσματα στο σύνολο δεδομένων αξιολόγησης (test) και στον Πίνακα 3.9 απεικονίζονται οι μετρικές ανά κλάση για το σύνολο δεδομένων αξιολόγησης. Ο λόγος που τα ποσοστά των μετρικών απόδοσης της κλάσης Deep είναι υψηλά, είναι επειδή η κλάση Deep έχει πολύ λίγα δεδομένα σε σχέση με τις υπόλοιπες κλάσεις. Οπότε, αν το CNN δεν αναγνωρίζει καλά αυτά τα λίγα δεδομένα, τότε τα ποσοστά των μετρικών απόδοσης δεν θα είναι καλές.

<i>Metrics%</i>	<i>Values</i>
Accuracy	63.7
Precision	58.4
Recall	58.4
F1-Score	58.2
Cohen-Kappa	44.6
Specificity	86.8

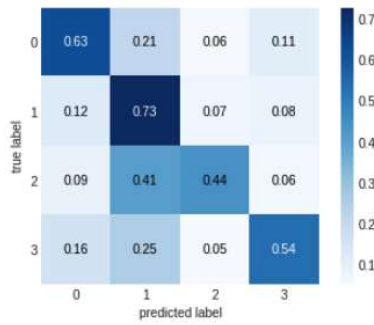
Πίνακας 3.8: Μετρικές στο σύνολο δεδομένων αξιολόγησης.

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	73.0	63.0	67.0
Light	66.0	73.0	69.0
Deep	42.0	44.0	43.0
REM	53.0	54.0	53.0

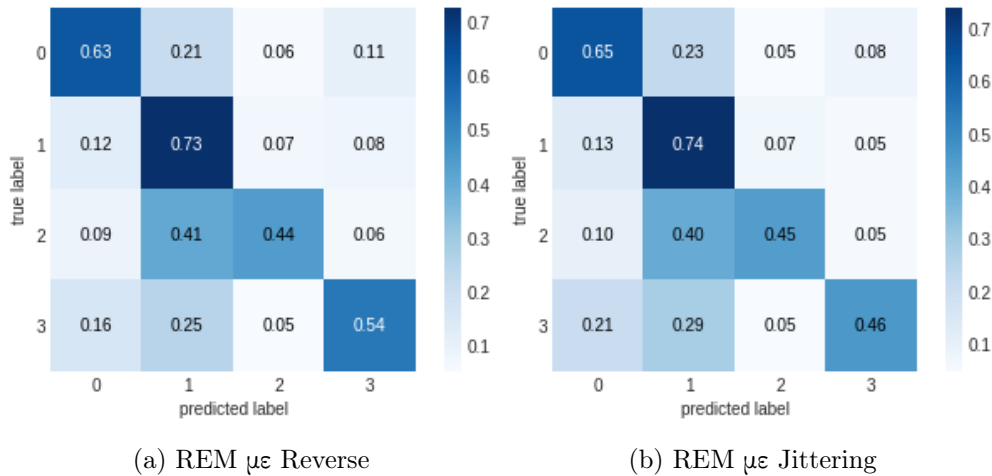
Πίνακας 3.9: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση.

Λεπτομερέστερα, από τον Πίνακα 3.9 και από τον Πίνακα Σύγχυσης (Σχήμα 3.3.4), παρατηρούμε ότι η κλάση Light αναγνωρίζεται πάρα πολύ καλά, γεγονός που αιτιολογείται από το πλήθος δεδομένων για αυτή την κλάση. Αντίστοιχα, η κλάση Deep όπου έχει και τον μικρότερο αριθμό δεδομένων, δεν





Σχήμα 3.3.4: Πίνακας Σύγκρισης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης.

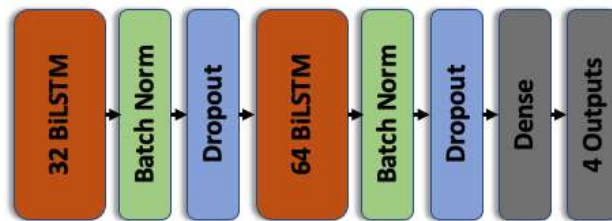


Σχήμα 3.3.5: Πίνακας Σύγκρισης (Confusion Matrix)

αναγνωρίζεται τόσο καλά και συγγέεται με την κλάση Light. Για αυτό τον λόγο, οι μετρικές Precision και Recall είναι χαμηλές σε ποσοστό. Λόγω της ανισορροπίας των δεδομένων (imbalance), το μοντέλο τείνει να ταξινομήσει τις ακολουθίες στη κλάση Light όπου έχει και τα περισσότερα δεδομένα ή και να συγγεί τις υπόλοιπες κλάσεις μεταξύ τους. Όμως, με τις τεχνικές Data Augmentation προσπαθήσαμε να αντιμετωπίσουμε αυτό το πρόβλημα ως εξής. Στο Σχήμα 3.3.5 απεικονίζεται η επίδραση του Jittering και του Reverse της κλάσης REM στην τελική ταξινόμηση. Παρατηρούμε, ότι με την τεχνική Jittering στην REM κλάση, το μοντέλο δυσκολεύεται να ξεχωρίσει την κλάση Wake από την κλάση REM, ενώ με την τεχνική Reverse στη κλάση REM, διαχωρίζονται οι δύο κλάσεις πολύ καλύτερα, δηλαδή η μετρική Recall ανεβαίνει κατά 8% (από 46% σε 54%). Το γεγονός αυτό αιτιολογείται διότι αν προσθέσουμε θόρυβο, η κλάση REM αρχίζει και ομοιάζει με την κλάση Wake. Από πλευράς ιατρικής, το Σχήμα 2.1.3 δείχνει ότι στο στάδιο REM, οι καρδιακοί παλμοί αυξάνονται οπότε τα RR-Intervals είναι μικρότερα σε χρόνο (ms) αφού τα διαδοχικά R-Points έχουν μικρότερη χρονική απόσταση μεταξύ τους. Επίσης, κατά το στάδιο REM, τα RR-Intervals έχουν **μεγαλύτερη διακύμανση** σε σχέση με το στάδιο Deep (Πίνακας 3.6 και 3.7), γεγονός που επιβεβαιώνει τη σύγκριση του μοντέλου να διαχωρίζει την κλάση REM καλά εξαιτίας της προσθήκης θορύβου. Οπότε, γι αυτό το λόγο κατά την τεχνική Data Augmentation δεν προσθέσαμε πολύ θόρυβο στα δεδομένα REM, επειδή τα επιπλέον δεδομένα που θα δημιουργούσαμε θα έμοιαζαν περισσότερο με τις κλάσεις Wake και Light όπου η διακύμανση των RR-Intervals είναι μεγαλύτερη συγκριτικά με τα υπόλοιπα στάδια, **συνεπώς τα μοντέλα θα συνέχισαν την κλάση REM με τις κλάσεις Wake και Light.**

### 3.3.2 Μοντελοποίηση με LSTM

Εκτός του συνελκτικτικού νευρωνικού δικτύου, δημιουργήσαμε και ένα LSTM μοντέλο που απεικονίζεται στο παρακάτω Σχήμα 3.3.6 για να παρατηρήσουμε τυχόν διαφορές στην ταξινόμηση των σταδίων ύπνου.

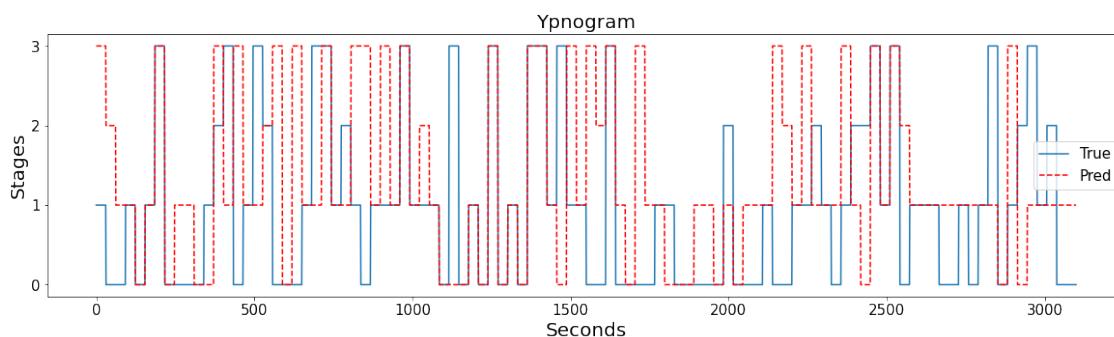


Σχήμα 3.3.6: Αρχιτεκτονική LSTM Δικτύου.

Αντίστοιχα με το CNN μοντέλο, το LSTM μοντέλο δέχεται ως είσοδο τις 30sec RR-Intervals ακολουθίες με μέγεθος (31,1) και ως έξοδο την κλάση στην οποία ανήκει σε δυαδική μορφή (One Hot Encoding). Συνοπτικά, τα LSTM κελιά είναι 32 και 64 για τα δύο LSTM επίπεδα και μετά από κάθε έξοδο των LSTM επιπέδων εφαρμόζεται κανονικοποίηση (Batch Normalization) και Dropout με regularization rate ίσο με 50%. Τα εξαγόμενα χαρακτηριστικά από τα επίπεδα LSTM τροφοδοτούνται σε ένα feed-forward νευρωνικό δίκτυο με 100 νευρώνες που καταλήγει στην έξοδο 4 νευρώνων. Πρέπει να επισημανθεί ότι κάθε επίπεδο LSTM επιστρέφει την έξοδο κάθε LSTM κελιού.

Η εκπαίδευση του μοντέλου LSTM γίνεται για 100 εποχές, χρησιμοποιώντας batch size ίσο με 1024. Επίσης, το 20% των δεδομένων εκπαίδευσης χρησιμοποιείται για validation ώστε να αξιολογηθεί η απόδοση του μοντέλου.

Αφού έχει εκπαιδευτεί το LSTM μοντέλο, γίνεται η πρόβλεψη στο σύνολο δεδομένων αξιολόγησης (test). Στο Σχήμα 3.3.7 παρατίθεται υπνόγραμμα για 100 διαδοχικές ακολουθίες του συνόλου αξιολόγησης και τις αντίστοιχες πραγματικές κλάσεις στις οποίες ανήκουν οι ακολουθίες.



Σχήμα 3.3.7: Υπνόγραμμα 100 διαδοχικών ακολουθιών του συνόλου αξιολόγησης και των αντίστοιχων πραγματικών κλάσεων. Wake:0, Light:1, Deep:2, REM:3

Εποπτικά, στο Σχήμα 3.3.7 παρατηρούμε ότι η κλάση Deep δεν μπορεί να αναγνωριστεί καλά από το LSTM, για τους ίδιους λόγους που αναφέραμε και στο CNN, ενώ οι υπόλοιπες κλάσεις μπορούν να αναγνωριστούν καλύτερα όπως και στο CNN μοντέλο. Το συμπέρασμα αυτό μπορεί να διαγνωσθεί και από τον πίνακα μετρικών του συνόλου αξιολόγησης και από τον πίνακα σύγχυσης (Confusion Matrix). Γενικά, είναι εμφανής η επίδραση του μη ισορροπημένου συνόλου δεδομένων (imbalance) στο τελικό αποτέλεσμα ταξινόμησης. Στο Πίνακα 3.10 απεικονίζονται τα αποτελέσματα στο σύνολο δεδομένων

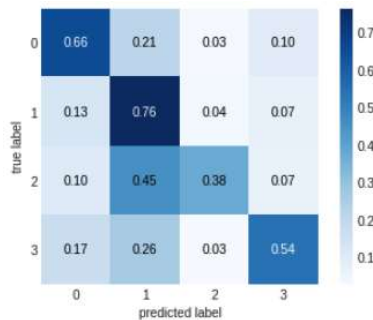
αξιολόγησης (test). Επίσης, στον Πίνακα 3.11 απεικονίζονται οι μετρικές ανά κλάση για το σύνολο δεδομένων αξιολόγησης.

<i>Metrics%</i>	<i>Values</i>
Accuracy	65.8
Precision	61.9
Recall	58.7
F1-Score	59.8
Cohen-Kappa	46.9
Specificity	87.3

Πίνακας 3.10: Μετρικές στο σύνολο δεδομένων αξιολόγησης.

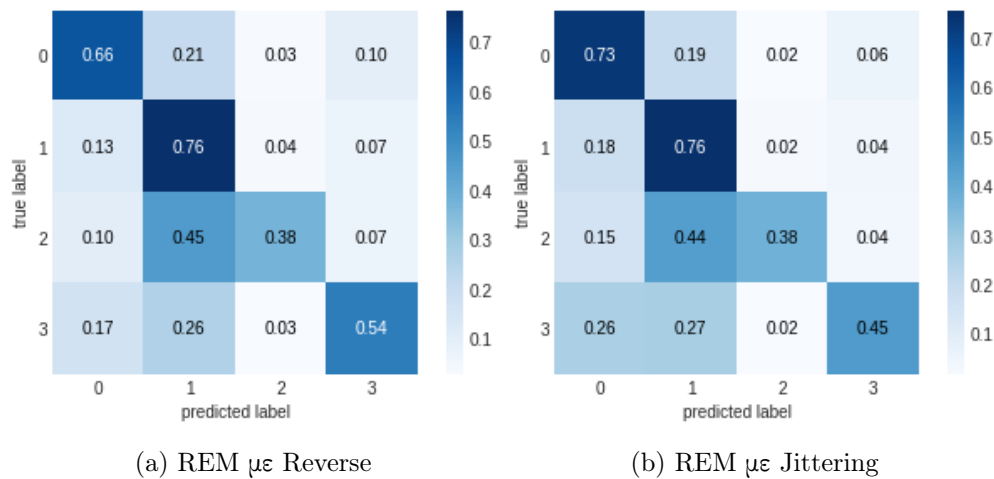
<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	73.0	66.0	69.0
Light	66.0	76.0	71.0
Deep	53.0	38.0	44.0
REM	55.0	54.0	55.0

Πίνακας 3.11: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανά κλάση.



Σχήμα 3.3.8: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης.

Παρατηρούμε από τον Πίνακα 3.11 και από τον Πίνακα Σύγχυσης, ότι η κλάση Light αναγνωρίζεται πάρα πολύ καλά, γεγονός που αιτιολογείται από το πλήθος δεδομένων για αυτή την κλάση, αλλά αυτό είναι εις βάρος της κλάσης Deep η οποία συγχέεται με την κλάση Light. Αυτή η παρατήρηση αντικατοπτρίζεται στο χαμηλό ποσοστό της μετρικής Recall. Λόγω της ανισορροπίας των δεδομένων (imbalance) το μοντέλο τείνει να ταξινομήσει τις ακολουθίες στη κλάση Light όπου έχει και τα περισσότερα δεδομένα ή και να συγχέει τις υπόλοιπες κλάσεις μεταξύ τους. Όμως, με τις τεχνικές Data Augmentation προσπαθήσαμε να αντιμετωπίσουμε αυτό το πρόβλημα με τον ίδιο τρόπο όπως και το μοντέλο CNN. Στο Σχήμα 3.3.9 απεικονίζεται η επίδραση του Jittering και του Reverse της κλάσης REM στην τελική ταξινόμηση. Παρατηρούμε, ότι με την τεχνική Jittering στην REM κλάση, το μοντέλο δυσκολεύεται να ξεχωρίσει την κλάση Wake από την κλάση REM, ενώ με την τεχνική Reverse στη κλάση REM, διαχωρίζονται οι δύο κλάσεις πολύ καλύτερα και η μετρική Recall ανεβαίνει κατά 10%. Τα συμπεράσματα δηλαδή είναι ίδια με αυτά του μοντέλου CNN, με κύρια διαφορά ότι η μετρική Precision της κλάσης Deep στο μοντέλο LSTM πετυχαίνει καλύτερα αποτελέσματα κατά 11% και ότι γενικά το μοντέλο LSTM πετυχαίνει καλύτερα αποτελέσματα, σε όλες τις μετρικές, σε σύγκριση με το μοντέλο CNN. Περαιτέρω

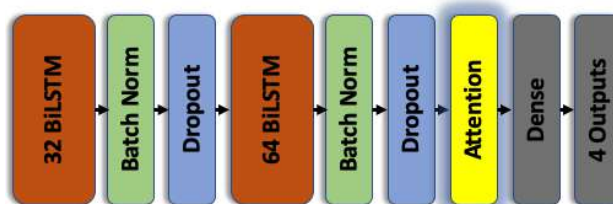


Σχήμα 3.3.9: Πίνακας Σύγχυσης (Confusion Matrix)

ανάλυση και σύγκριση θα πραγματοποιηθεί στο τέλος του κεφαλαίου. Άρα, τα καλύτερα αποτελέσματα ταξινόμησης στις δημοφιλείς μετρικές με βάση τον Πίνακα Σύγχυσης 3.3.9, τα δίνει η τεχνική Reverse της κλάσης REM και όχι η τεχνική Jittering. Η κύρια διαφορά μεταξύ των δύο τεχνικών βρίσκεται στην ταξινόμηση των κλάσεων Wake και REM, όπως διακρίνεται από το Σχήμα 3.3.9. Παρατηρούμε δηλαδή ότι με τη τεχνική Jittering προκύπτουν καλύτερα αποτελέσματα για την κλάση Wake αλλά χειρότερα για τη κλάση REM, με την μετρική Recall να δίνει τιμές 73% και 45% αντίστοιχα. Αντίθετα συμβαίνει με την τεχνική Reverse όπου προκύπτουν καλύτερα αποτελέσματα για την κλάση REM και χειρότερα για τη κλάση Wake, με 66% και 54% αντίστοιχα. Όμως, αυτό που επιθυμούμε είναι όσο το δυνατόν υψηλότερα ποσοστά των όλων των δημοφιλών μετρικών για όλες τις κλάσεις, οπότε για αυτό τον λόγο η τεχνική Reverse αποδεικνύεται ότι είναι καλύτερη.

### 3.3.3 Μοντελοποίηση με LSTM και μηχανισμό Προσοχής (Attention)

Σαν τελευταία δοκιμή στο πρόβλημα της ταξινόμησης των RR-Intervals, δημιουργήσαμε ένα LSTM μοντέλο με την προσθήκη ενός επιπέδου Προσοχής ώστε να παρατηρήσουμε αν μπορεί να βελτιωθεί η απόδοση του LSTM. Στόχος είναι η καλύτερη ταξινόμηση κυρίως της κλάσης Deep ώστε να βελτιωθούν και όλες οι γενικές μετρικές. Η αρχιτεκτονική του μοντέλου απεικονίζεται στο παρακάτω Σχήμα 3.3.10.



Σχήμα 3.3.10: Αρχιτεκτονική LSTM + Attention Δικτύου.

Το προτεινόμενο μοντέλο έχει την ίδια αρχιτεκτονική με το προηγούμενο LSTM μοντέλο, με τη διαφορά ότι έγινε η προσθήκη του επιπέδου Προσοχής (Attention Layer). Το μοντέλο δέχεται ως είσοδο τις 30sec RR-Intervals ακολουθίες με μέγεθος (31,1) και ως έξοδο, την κλάση στην οποία ανήκει σε δυαδική μορφή (One Hot Encoding). Συνοπτικά, τα LSTM κελιά είναι 32 και 64 για τα δύο επίπεδα και μετά από κάθε έξοδο των LSTM επιπέδων εφαρμόζεται κανονικοποίηση (Batch Normalization) και Dropout με regularization rate ίσο με 50%. Τα εξαγόμενα χαρακτηριστικά από τα επίπεδα LSTM

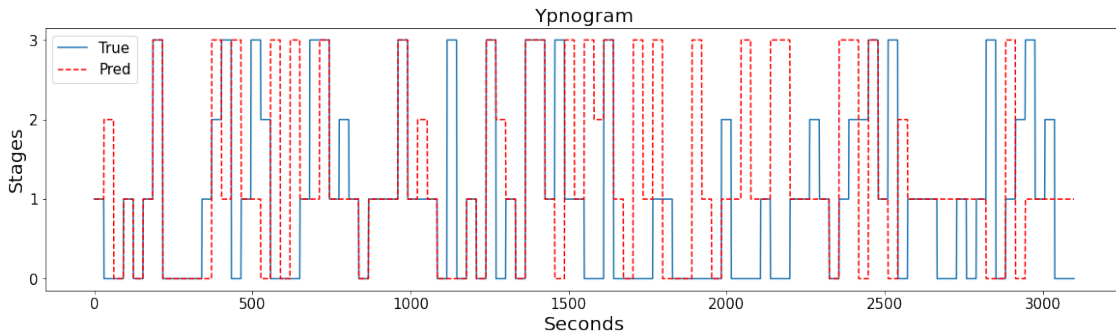
τροφοδοτούνται στο μηχανισμό Προσοχής και μετέπειτα σε ένα Feed-Forward νευρωνικό δίκτυο με 100 νευρώνες που καταλήγει στην έξοδο 4 νευρώνων. Πρέπει να επισημανθεί ότι κάθε επίπεδο LSTM επιστρέφει την έξοδο κάθε LSTM κελιού.

Αξίζει να επισημανθεί ότι ο μηχανισμός Προσοχής (Attention Mechanism), βοηθάει το μοντέλο LSTM να εστιάσει την "προσοχή" του σε συγκεκριμένα σημεία των ακολουθιών RR-Intervals δίνοντας κάποιο βάρος (weight) σε αυτά τα σημεία ώστε το μοντέλο να τα "προσέξει" καθώς γίνεται η εκπαίδευση του [107]. Οπότε, λόγω της δυναμικής των φυσιολογικών αλληλεπιδράσεων στον εγκέφαλο και των αυτόνομων ελέγχων κατά τη διάρκεια κάθε σταδίου ύπνου [108], πιστεύουμε ότι οι διαδοχικοί φορείς χαρακτηριστικών από τα στρώματα LSTM συμβάλλουν διαφορετικά στην ταξινόμηση των τεσσάρων σταδίων ύπνου. Για να βελτιώσουμε την απόδοση των δικτύων LSTM εστιάζοντας σε ορισμένα χρονικά βήματα με πιο διακριτικά χαρακτηριστικά που σχετίζονται με το στάδιο ύπνου, αποφασίσαμε να προσθέσουμε ένα μηχανισμό προσοχής ώστε το μοντέλο να μαθαίνει αυτόματα τη σημασία των διανυσμάτων χαρακτηριστικών σε κάθε χρονικό βήμα της ακολουθίας.

Συγκεκριμένα, ο μηχανισμός προσοχής βοηθά τα επερχόμενα επίπεδα (Dense) του δικτύου να έχουν περισσότερη πληροφορία για να μάθουν αφού η τελευταία έχει κωδικοποιηθεί (encoded) από τα 2 επίπεδα LSTM (δηλαδή έχει γίνει εξαγωγή πληροφορίας από τα επίπεδα LSTM). Οπότε περιμένουμε να έχουμε λίγο καλύτερα αποτελέσματα από τον απλό αλγόριθμο LSTM του προηγούμενου κεφαλαίου 3.3.2 εφόσον, οι ακολουθίες δεν είναι τόσο μεγάλες σε μήκος ώστε ο μηχανισμός Προσοχής να έχει μεγαλύτερη επίδραση.[109]

Η εκπαίδευση του μοντέλου LSTM γίνεται για 100 εποχές, χρησιμοποιώντας batch size ίσο με 1024 και την μέθοδο ReduceLROnPlateau με ποσοστό 0.1. Επίσης, το 20% των δεδομένων εκπαίδευσης χρησιμοποιείται για validation ώστε να αξιολογηθεί η απόδοση του μοντέλου.

Αφού έχει εκπαιδευτεί το LSTM μοντέλο, γίνεται η πρόβλεψη στο σύνολο δεδομένων αξιολόγησης (test). Στο Σχήμα 3.3.11 παρατίθεται υπνόγραμμα για 100 διαδοχικές ακολουθίες του συνόλου αξιολόγησης και τις αντίστοιχες πραγματικές κλάσεις στις οποίες ανήκουν οι ακολουθίες.



Σχήμα 3.3.11: Υπνόγραμμα 100 διαδοχικών ακολουθιών του συνόλου αξιολόγησης και των αντίστοιχων πραγματικών κλάσεων.

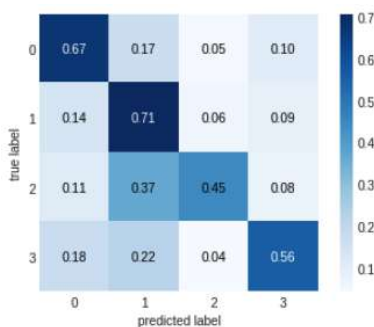
Εποπτικά, στο Σχήμα 3.3.11 παρατηρούμε ότι η κλάση Deep μπορεί να αναγνωριστεί καλύτερα από το απλό μοντέλο LSTM, και εστιάζουμε σε αυτή τη κλάση επειδή είναι η κλάση με τη χαμηλότερη απόδοση σε όλους τους πειραματισμούς μας. Αυτό αποδεικνύει εποπτικά ότι ο ισχυρισμός της βελτίωσης της κλάσης Deep είναι σωστός. Το συμπέρασμα αυτό μπορεί να διαγνωσθεί και από τον πίνακα μετρικών του συνόλου αξιολόγησης και από τον πίνακα σύγχυσης (Confusion Matrix). Στο Πίνακα 3.12 απεικονίζονται τα αποτελέσματα στο σύνολο δεδομένων αξιολόγησης (test). Επίσης, στον Πίνακα 3.13 απεικονίζονται οι μετρικές ανά κλάση για το σύνολο δεδομένων αξιολόγησης.

<i>Metrics%</i>	<i>Values</i>
Accuracy	66.0
Precision	62.1
Recall	59.0
F1-Score	60.0
Cohen-Kappa	47.9
Specificity	87.3

Πίνακας 3.12: Μετρικές στο σύνολο δεδομένων αξιολόγησης.

<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
Wake	71.0	67.0	69.0
Light	69.0	71.0	70.0
Deep	44.0	45.0	45.0
REM	53.0	56.0	55.0

Πίνακας 3.13: Μετρικές στο σύνολο δεδομένων αξιολόγησης ανα κλάση.

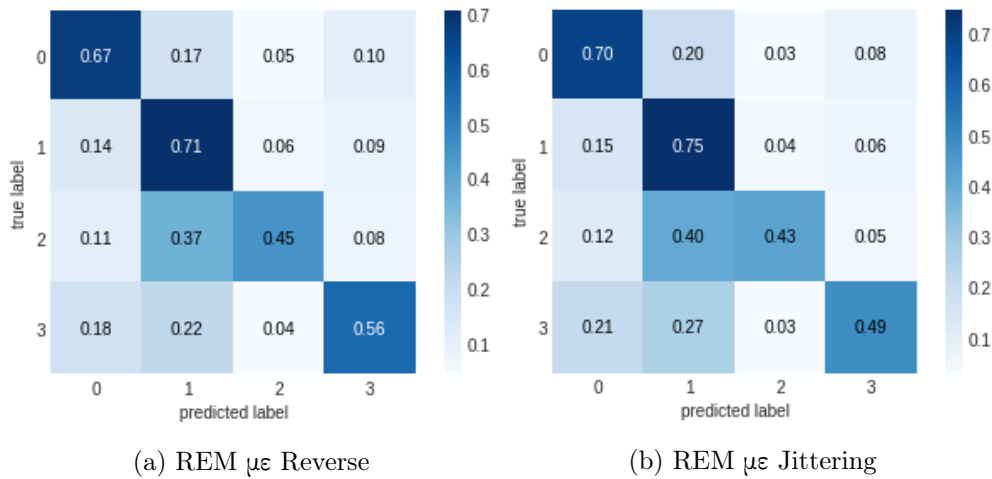


Σχήμα 3.3.12: Πίνακας Σύγχυσης (Confusion Matrix) στο σύνολο δεδομένων αξιολόγησης.

Από τον Πίνακα 3.13 και από τον Πίνακα Σύγχυσης παρατηρούμε ότι η κλάση Light αναγνωρίζεται πάρα πολύ καλά, γεγονός που αιτιολογείται από το πλήθος δεδομένων για αυτή την κλάση. Επίσης, το κέρδος που έχουμε με την προσθήκη του μηχανισμού Προσοχής είναι η αύξηση της μετρικής Recall για τις κλάσεις Deep και REM. Στο Σχήμα 3.3.13 απεικονίζεται η επίδραση του Jittering και του Reverse της κλάσης REM στην τελική ταξινόμηση. Παρατηρούμε, ότι με την τεχνική Jittering στην REM κλάση, το μοντέλο δυσκολεύεται να ξεχωρίσει την κλάση Wake από την κλάση REM, ενώ με την τεχνική Reverse στη κλάση REM, διαχωρίζονται οι δύο κλάσεις πολύ καλύτερα. Τα συμπεράσματα δηλαδή είναι ίδια με αυτά του μοντέλου LSTM, που αναπτύχθηκε στην υποενότητα 3.3.2, με κύρια διαφορά την **αύξηση της μετρικής Recall της κλάσης Deep με την προσθήκη μηχανισμού προσοχής**.

Περαιτέρω ανάλυση, σύγκριση όλων των αποτελεσμάτων και συμπερασμάτων από όλα τα μοντέλα Deep Learning θα πραγματοποιηθεί στο Κεφάλαιο 3.4.





Σχήμα 3.3.13: Πίνακας Σύγχυσης (Confusion Matrix)

### 3.4 Ανάλυση Αποτελεσμάτων και Συμπεράσματα

Από την ανάλυση που πραγματοποιήσαμε για την ταξινόμηση των RR-Intervals σε 4 στάδια με διαφορετικούς αλγόριθμους Deep Learning, παρατηρούμε ότι το κάθε μοντέλο που χρησιμοποιήσαμε έχει τα δικά του πλεονεκτήματα και μειονεκτήματα. Στους Πίνακες 3.14 και 3.15, απεικονίζονται συνολικά οι αποδόσεις όλων των μοντέλων Deep Learning.

<i>Metrics%/Models</i>	<i>CNN</i>	<i>LSTM</i>	<i>LSTM + Attention</i>
Accuracy	63.7	65.8	66
Precision	58.4	61.9	62.1
Recall	58.4	58.7	59
F1-Score	58.2	59.8	60
Cohen-Kappa	44.6	46.9	48
Specificity	86.8	87.3	87.3

Πίνακας 3.14: Μετρικές στο σύνολο δεδομένων αξιολόγησης.

Παρατηρούμε ότι τα μοντέλα έχουν παραπλήσια απόδοση αλλά το μοντέλο **LSTM+Attention ξεχωρίζει σε όλες τις μετρικές**. Επίσης, στον πίνακα 3.15 παρατηρούμε ότι για όλες τις κλάσεις, τα αποτελέσματα είναι παραπλήσια με έντονο στοιχείο να αποτελεί η χαμηλή απόδοση της κλάσης Deep γεγονός που οφείλεται στον πολύ μικρό αριθμό δεδομένων που περιλαμβάνεται για μοντελοποίηση άρα οι αλγόριθμοι δεν μπορούν να γενικεύσουν καλά. Τέλος, παρατηρούμε ότι η μοντελοποίηση με LSTM μοντέλα υπερτερεί έναντι του CNN μοντέλου για το δεδομένο πρόβλημα και τον τρόπο με τον οποίο επεξεργαστήκαμε τα δεδομένα. Είναι ξεκάθαρο, ότι οι κλάσεις που ανιχνεύονται καλύτερα είναι οι κλάσεις με τα περισσότερα δεδομένα, δηλαδή Wake και Light επειδή ο οι αλγόριθμοι Deep Learning έχουν στη διάθεση τους πολλές διαφορετικές μεταξύ τους ακολουθίες RR-Intervals να μάθουν από αυτές τις κλάσεις σε σύγκριση με τις κλάσεις Deep και REM. Οπότε, όπως είναι λογικό κατά την γενίκευση των μοντέλων στο σύνολο δεδομένων αξιολόγησης (test), τα μοντέλα μπορούν καλύτερα να αναγνωρίσουν τις κλάσεις Wake και Light.

Συνολικά, στο κεφάλαιο αυτό περιγράψαμε τη διαδικασία που ακολουθήσαμε για την ταξινόμηση των σταδίων του ύπνου με τη χρήση μόνο των ακατέργαστων 30sec RR-Intervals από τη βάση δεδομένων MESA. Χρησιμοποιήσαμε 3 διαφορετικά μοντέλα (CNN, LSTM και LSTM+Attention) και καταφέραμε



	<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
CNN	Wake	73.0	63.0	67.0
	Light	66.0	73.0	69.0
	Deep	42.0	44.0	43.0
	REM	53.0	54.0	53.0
	<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
LSTM	Wake	73.0	66.0	69.0
	Light	66.0	76.0	71.0
	Deep	53.0	38.0	44.0
	REM	55.0	54.0	55.0
	<i>Stages/Metrics%</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 – Score</i>
LSTM+Att.	Wake	71.0	67.0	69.0
	Light	69.0	71.0	70.0
	Deep	44.0	45.0	45.0
	REM	53.0	56.0	55.0

Πίνακας 3.15: Μετρικές στο σύνολο δεδομένων αξιολόγησης.

να επιτύχουμε ικανοποιητικά αποτελέσματα συγκριτικά με τις ήδη υπάρχουσες εργασίες/δημοσιεύσεις που έχουν γίνει με τη χρήση της βάσης δεδομένων MESA και με τη χρήση μόνο των RR-Intervals. Από την ανάλυση που πραγματοποιήσαμε, προέκυψε ότι το καλύτερο μοντέλο είναι το **LSTM+Attention που πέτυχε accuracy 66%, F1-Score 60% και Cohen-Kappa 48%** ενώ οι ερευνητές της πιο πρόσφατης δημοσίευσης που βρήκαμε [68] πέτυχαν με το CNN μοντέλο τους **accuracy 80%** και **Cohen-Kappa 69%** χρησιμοποιώντας όμως για την εκπαίδευση του μοντέλου και την βάση SHHS. Οπότε, τα αποτελέσματα μας δεν μπορούν να συγκριθούν άμεσα, διότι οι ερευνητές χρησιμοποίησαν επιπλέον και μια εξίσου πάρα πολύ μεγάλη βάση δεδομένων όποτε είχαν περισσότερη πληροφορία για να τροφοδοτήσουν το μοντέλο τους πέραν της βάσης MESA αλλά και άλλο σύνολο αξιολόγησης. Ο λόγος που δεν χρησιμοποιήθηκε η βάση SHHS είναι ότι θέλαμε να συγκρίνουμε τα αποτελέσματα μας, με την προηγούμενη ανάλυση του Κεφαλαίου 3, δηλαδή την ταξινόμηση με εξαγωγή χαρακτηριστικών από τις ακολουθίες RR-Intervals (Κεφάλαιο 3) και την ταξινόμηση με ακατέργαστα δεδομένα (raw) χωρίς εξαγωγή χαρακτηριστικών (Κεφάλαιο 4).

Τελικά, συγκρίνοντας το καλύτερο μοντέλο με εξαγωγή χαρακτηριστικών (*BiLSTM<sub>101</sub>*) και χωρίς εξαγωγή (raw) (*LSTM + Attention*), προκύπτει ότι στις δημοφιλείς μετρικές, η ταξινόμηση με εξαγωγή χαρακτηριστικών υπερτερεί. Τα αποτελέσματα φαίνονται συγκεντρωτικά στον Πίνακα 3.16, όπου παρατηρούμε παραπλήσιες αποδόσεις εκτός από τη μετρική Cohen-Kappa. Το χαμηλό ποσοστό της μετρικής Cohen-Kappa για το μοντέλο *LSTM + Attention* μας αποκαλύπτει ότι πολύ πιθανό αρκετές ταξινομήσεις δεδομένων να έγιναν κατά τύχη από το μοντέλο, κάτι που είναι μειονέκτημα για το μοντέλο αυτό. Επίσης, πρέπει να σημειωθεί ότι δεν είναι άμεση η σύγκριση των δύο μοντέλων, καθώς οι μετρικές στο πειραματισμό με εξαγωγή χαρακτηριστικών έγινε ανά χρήστη (user id), ενώ στο δεύτερο πειραματισμό χωρίς εξαγωγή χαρακτηριστικών, έγινε ανεξαρτήτου χρήστη. Άρα, συμπεραίνουμε ότι από τη μία, η προσθήκη διαφορετικού είδους δεδομένων (π.χ. ακτιγραφία) οδηγεί τα μοντέλα Βαθιάς Μάθησης να εξάγουν πιο πλούσια χαρακτηριστικά και άρα καλύτερα αποτελέσματα ταξινόμησης σε όλες τις μετρικές, αλλά από την άλλη η χρήση μόνο των δεδομένων καρδιακών παλμών δίνει ικανοποιητικά αποτελέσματα. Δηλαδή, σε

επίπεδο που η ταξινόμηση χρησιμοποιηθεί σε κάποια εφαρμογή (π.χ λογισμικό ανάλυσης ύπνου), και τα διαθέσιμα δεδομένα είναι μόνο οι καρδιακοί παλμοί, τότε τα αποτελέσματα είναι ικανοποιητικά στο βαθμό που μας επιτρέπει η διαθεσιμότητα των δεδομένων. **Τέλος, επειδή το σύνολο δεδομένων είναι μη ισορροπημένο οι αποδόσεις των μοντέλων πρέπει να εκτιμηθούν με βάσει κυρίως τις μετρικές Precision, Recall, F1-Score και Cohen-Kappa. Άρα, η μοντελοποίηση μόνο με ακατέργαστα δεδομένα, μας δίνει την δυνατότητα, με απλό και γρήγορο τρόπο να πετύχουμε ικανοποιητικά αποτελέσματα για την ταξινόμηση του ύπνου σε 4 στάδια, χωρίς την χρονοβόρα και πολύπλοκη εξαγωγή handcrafted χαρακτηριστικών, με μικρή όμως απόδοση στο στάδιο Deep όπου αναφέραμε τους λόγους σε προηγούμενα Κεφάλαια και το χαμηλό ποσοστό Cohen-Kappa.**

<i>Metrics%/Models</i>	<i>BiLSTM</i>	<i>LSTM + Attention</i>
Accuracy	71.7	66
Precision	63.3	62.1
Recall	61.2	59
F1-Score	58.7	60
Cohen-Kappa	58.9	48
Specificity	88.7	87.3

Πίνακας 3.16: Μετρικές στο σύνολο δεδομένων αξιολόγησης για τις δύο μεθοδολογίες που ακολουθήσαμε. 1) Εξαγωγή χαρακτηριστικών (**BiLSTM<sub>101</sub>**) και 2) Ακατέργαστα (raw) δεδομένα χωρίς εξαγωγή χαρακτηριστικών (**LSTM + Attention**). Στη μεθοδολογία με εξαγωγή χαρακτηριστικών, η αξιολόγηση έγινε ανά χρήστη (user id), ενώ στην μεθοδολογία με χρήση raw δεδομένων RR-Intervals μόνο η αξιολόγηση έγινε ανεξαρτήτου χρήστη. Άρα Η σύγκριση δεν είναι άμεση, μόνο διαισθητική.

Επιπρόσθετα, καταλήγουμε ότι το μεγαλύτερο μειονέκτημα της προσέγγισης που αναπτύξαμε (ταξινόμηση με χρήση μόνο των RR-Intervals), είναι ότι χρησιμοποιήθηκαν μόνο οι ακολουθίες που είχαν μέγεθος 28 έως 35, οπότε χάνεται πληροφορία που θα έδιναν οι μικρότερες ή μεγαλύτερες σε μέγεθος ακολουθίες. Όμως, δεν μπορούσαμε να το αποφύγουμε αυτό, επειδή όπως αναφέραμε στην υποσημειωμένη 3.2.2 (όπου αναπτύχθηκαν οι τεχνικές Data Augmentation) και με βάση τα Σχήματα 3.2.2 και 3.3.3, **η τεχνική padding δεν έχει νόημα να εφαρμοστεί σε ακολουθίες με μικρό πλήθος σημείων RR-Intervals σε σύγκριση με το μέσο μήκος των RR-Intervals, επειδή τα τελευταία θα αποτελούσαν τεχνητά δεδομένα που δεν είναι πραγματικά (artificial data)**. Αυτό το πρόβλημα μπορεί να αντιμετωπιστεί με μια διαφορετική προσέγγιση ώστε να επιτευχθούν καλύτερα αποτελέσματα. Δηλαδή, το πρόβλημα που πρέπει να αντιμετωπιστεί είναι η μη σταθερή (μεταβαλλόμενη) συχνότητα δειγματοληψίας των RR-Intervals. Αυτό σημαίνει ότι επειδή τα RR-Intervals δεν πραγματοποιούνται σε σταθερό χρονικό διάστημα, τότε είναι λογικό πως ανά 30 δευτερόλεπτα δεν θα έχουμε κάθε φορά σταθερό πλήθος R-Points. Οι ερευνητές στο [68] χρησιμοποίησαν τους καρδιακούς παλμούς (Heart Rate) για την ταξινόμηση αλλά δεν αναφέρουν πολλά στοιχεία για το πως προ-επεξεργάζονται τα δεδομένα πριν την ταξινόμηση, (π.χ εφαρμόζουν τεχνικές resampling)

Τέλος, στόχος της προσέγγισης μας, ήταν με εύκολο και γρήγορο τρόπο να καταφέρουμε να πετύχουμε ικανοποιητικά αποτελέσματα χωρίς την διαδικασία εξαγωγής χαρακτηριστικών (handcrafted features) όπως παρουσιάστηκε στο Κεφάλαιο 3. Επιπλέον, είναι φανερό ότι η αδυναμία ταξινόμησης της κλάσης Deep, εξαιτίας του μικρού αριθμού δεδομένων, οδηγεί στην χαμηλή απόδοση των αλγορίθμων γεγονός που επιβεβαιώνεται και στα [68][1][104][88]. Επίσης, είναι λογικό να έχουμε μικρό αριθμό δεδομένων στην κλάση Deep αφού, όπως αναφέρθηκε και στην εισαγωγή, αποτελεί το 8% του ύπνου.

Τελικά, πιθανές βελτιώσεις προτείνονται στο μοντέλο ώστε να επιτευχθούν καλύτερα αποτελέσματα όπως για παράδειγμα η δοκιμή μοντέλων CNN-LSTM με Attention. Επίσης, η χρήση και επιπλέον δεδομένων όπως επιταχυνσιόμετρου (accelerometer data) ή οξυγόνου ( $SpO_2$ ) θα μπορούσε να καλύψει τις ανάγκες για καλύτερα αποτελέσματα ώστε τα μοντέλα να "μάθουν" και επιπλέον χαρακτηριστικά που διαθέτουν τα δεδομένα κίνησης (accelerometer), χωρίς εξαγωγή χαρακτηριστικών όπως έγινε στο Κεφάλαιο 2. Τέλος, σημαντική επίδραση στην βελτίωση των μετρικών μπορεί να αποτελέσει μια πιο στοχευμένη τεχνική Data Augmentation για κάθε κλάση, ώστε οι αλγόριθμοι να μπορούν να διαχωρίσουν καλύτερα τις κλάσεις μεταξύ τους.

Εποπτικά, η προσέγγιση που ακολουθήθηκε για την ταξινόμηση του ύπνου σε 4 στάδια, αποτελεί μια ικανοποιητική δουλειά όπου με βάση τις παραπάνω παρατηρήσεις αυτής της υποενότητας (3.4), θα μπορούσε να χρησιμοποιηθεί ως βάση για περαιτέρω ερευνητική διερεύνηση του προβλήματος ταξινόμησης των σταδίων ύπνου από την οπτική της Τεχνητής Νοημοσύνης.



## Κεφάλαιο 4

# Επίλογος

### 4.1 Σύνοψη και Μελλοντικές Επεκτάσεις

Η παρούσα μεταπτυχιακή εργασία μπορεί να χωριστεί σε 3 κύριες ενότητες όσον αφορά τον **τρόπο προσέγγισης του προβλήματος ταξινόμησης** δεδομένων από αισθητήρες στις 4 κλάσεις/στάδια του ύπνου. Το κύριο θέμα εστιάζεται στους διαφορετικούς πειραματισμούς που πραγματοποιήσαμε, ώστε να ανακαλύψουμε με ποιόν τρόπο μπορούμε να πετύχουμε καλύτερα αποτελέσματα ταξινόμησης, αντιμετωπίζοντας παράλληλα τα όποια προβλήματα αναδύονται με βάση το διαθέσιμο σύνολο δεδομένων (βάση δεδομένων MESA). **Σε πρώτο στάδιο**, προσπαθήσαμε να αναπαράγουμε τα αποτελέσματα της έρευνας [1] ώστε η τελευταία να αποτελέσει μια βάση για να βελτιώσουμε τα αποτελέσματα της. Κρατώντας τον ίδιο άξονα, στο **Κεφάλαιο 2**, προσπαθήσαμε να αντιμετωπίσουμε το πρόβλημα του μη ισορροπημένου συνόλου δεδομένων, γεγονός το οποίο αποτελεί κεντρικό πρόβλημα που επηρεάζει σημαντικά τα αποτελέσματα των μοντέλων Deep Learning. Δοκιμάσαμε διαφορετικά μοντέλα (Bidirectional LSTM, Transformer), πειραματιστήκαμε με τις υπερπαραμέτρους των ήδη υπαρχόντων μοντέλων, με διαφορετικό μήκος παραθύρωσης των δεδομένων, με τη προεπεξεργασία των δεδομένων (Feature Scaling), με τον αλγόριθμο εκμάθησης των μοντέλων (Adam optimizer) καθώς και γενικότερα με την εκπαίδευση των μοντέλων (αριθμός εποχών, κ.α). Οπότε, στα δύο πρώτα μέρη της εργασίας στοχεύσαμε στην **ανακάλυψη των προβλημάτων της έρευνας [1] και στη βελτίωση των αποτελεσμάτων ταξινόμησης** με διάφορους τρόπους αλλά το **κεντρικό συμπέρασμα** μας είναι ότι η περαιτέρω εξερεύνηση των τεχνικών τεχνητής αύξησης συνόλου δεδομένων (Data Augmentation) για την αντιμετώπιση του μη ισορροπημένου συνόλου δεδομένων, θα οδηγήσει σε σημαντική βελτίωση των αποτελεσμάτων. **Στο τελευταίο μέρος της εργασίας, Κεφάλαιο 3**, πειραματιστήκαμε μόνο με ακατέργαστα (raw) δεδομένα καρδιακών παλμών (RR-Intervals) χωρίς την εξαγωγή χαρακτηριστικών (feature extraction), βασιζόμενοι σε πρόσφατες έρευνες γύρω από την ταξινόμηση των σταδίων ύπνου από φορητές συσκευές (smartwatches) [68]. Συνδυάζοντας μια κατάλληλη προεπεξεργασία των δεδομένων RR-Intervals και τεχνικές Data Augmentation, ταξινομήσαμε τα δεδομένα με διαφορετικά μοντέλα Βαθιάς Μάθησης όπως CNN, LSTM και LSTM+Attention, καταφέροντας να επιτύχουμε ικανοποιητικά αποτελέσματα συγκριτικά με τις ήδη υπάρχουσες εργασίες/δημοσιεύσεις που έχουν γίνει με τη χρήση της βάσης δεδομένων MESA και με τη χρήση μόνο των RR-Intervals. Το μοντέλο με τις καλύτερες αποδόσεις στις δημοφιλείς μετρικές ήταν το LSTM+Attention. Σκοπός του πειραματισμού μας με ακατέργαστα δεδομένα καρδιακών παλμών ήταν να αξιολογήσουμε πόσο καλά αποτελέσματα πετυχαίνουν τα μοντέλα χωρίς την εξαγωγή χαρακτηριστικών και να παρατηρήσουμε αν μόνο με τους καρδιακούς παλμούς, που καταγράφονται από τις περισσότερες φορητές συσκευές (smartwatches), μπορούμε να πετύχουμε ικανοποιητικά αποτελέσματα. Τελικά, συμπεράναμε ότι με τη χρήση μόνο ακατέργαστων δεδομένων καρδιακών παλμών πετυχαίνουμε καλά αποτελέσματα, οπότε η εξαγωγή

πολύπλοκων χαρακτηριστικών μπορεί να αποφευχθεί.

Συνοπτικά, οι **κύριες συνεισφορές** αυτής της μελέτης συνοψίζονται παρακάτω:

- Για το πρώτο μέρος της εργασίας με τη χρήση εξαγόμενων χαρακτηριστικών από τα δεδομένα της βάσης MESA, από όλους τους πειραματισμούς μας **καταφέραμε την αύξηση της απόδοσης των δημοφιλών μετρικών ταξινόμησης (και ανά κλάση) για όλα τα μοντέλα και επιπλέον βγάλαμε πολύ χρήσιμα συμπεράσματα για την επίδραση των τεχνικών Data Augmentation, Undersampling και για τον μήκος παραθύρωσης των δεδομένων**. Επίσης, προτείνουμε διάφορες μεθόδους που θα μπορούσαν να βελτιώσουν ακόμα περισσότερο τις αποδόσεις των μοντέλων Βαθιάς Μάθησης (Deep Learning).
- Για το δεύτερο μέρος της εργασίας η συνεισφορά είναι ότι η χρήση μιας διαφορετικής προσέγγισης προετοιμασίας των δεδομένων για μοντελοποίηση και η χρήση μόνο των καρδιακών παλμών οδήγησε σε καλά αποτελέσματα σε σχέση με τα αποτελέσματα του πρώτου μέρους όπου χρησιμοποιήθηκε περισσότερη πληροφορία, και σε σχέση με τις υπάρχουσες πρόσφατες έρευνες που έχουν γίνει με τη χρήση της βάσης δεδομένων MESA και με τη χρήση μόνο των RR-Intervals.

Τέλος, οι **πιθανές μελλοντικές επεκτάσεις** της εργασίας μας συνοψίζονται παρακάτω:

- Για το πρώτο μέρος της εργασίας με τη χρήση εξαγόμενων χαρακτηριστικών από τα δεδομένα της βάσης MESA, ο πειραματισμός με επιπλέον εξαγόμενα χαρακτηριστικά (**features engineering**) από τα δεδομένα ακτιγραφίας και κυρίως των καρδιακών παλμών, θα μπορούσε πιθανώς να οδηγήσει σε καλύτερη απόδοση των μοντέλων. Επιπλέον, η **προσθήκη και άλλου είδους δεδομένων** μπορεί να βοηθήσει την ταξινόμηση, όπως για παράδειγμα ο κιρκάδιος κύκλος (circadian cycle) δεδομένα στάσης χεριών και σώματος (hand and body posture), δεδομένα αναπνοής (Respiration Signal), δεδομένα επιτάχυνσης χεριών τριών αξόνων (3-axis accelerometer) και οξυγόνου (**SpO<sub>2</sub>**) τα οποία μπορούν να εμπλουτίσουν περαιτέρω την πληροφορία που επεξεργάζονται τα μοντέλα, οπότε τα τελευταία θα μπορούν να διακρίνουν καλύτερα τις διαφορές ανάμεσα στα στάδια του ύπνου.
- Για το δεύτερο μέρος της εργασίας με τη χρήση ακατέργαστων δεδομένων καρδιακών παλμών της βάσης MESA, οι πιθανές βελτιώσεις ώστε να επιτευχθούν καλύτερα αποτελέσματα είναι η χρήση διαφορετικών μοντέλων όπως για παράδειγμα η δοκιμή μοντέλων CNN-LSTM με Attention. Επίσης, η χρήση και επιπλέον δεδομένων όπως επιταχυνσιόμετρου (accelerometer data) ή οξυγόνου (**SpO<sub>2</sub>**) θα μπορούσε να καλύψει τις ανάγκες για καλύτερα αποτελέσματα ώστε τα μοντέλα να "μάθουν" και επιπλέον χαρακτηριστικά. Τέλος, σημαντική επίδραση στην βελτίωση των μετρικών μπορεί να αποτελέσει μια πιο στοχευμένη τεχνική Data Augmentation για κάθε κλάση, ώστε τα μοντέλα να μπορούν να διαχωρίσουν καλύτερα τις κλάσεις μεταξύ τους.

# Παράρτημα Α

## Βιβλιογραφία

- [1] Zhai, B. “Making Sense of Sleep: Multimodal Sleep Stage Classification in a Large, Diverse Population Using Movement and Cardiac Sensing”. In: *ACM Journals, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.47 (June 2020), pp. 1–33. DOI: <https://doi.org/10.1145/3397325>.
- [2] Resources, I. B. and Network, I. *Brain Basics: Understanding Sleep*. URL: <https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Understanding-Sleep>.
- [3] Medicine (US) Committee on Sleep Medicine, I. of and Research. *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem, Chapter 2: Sleep Physiology*. Washington (DC), 2006.
- [4] Resources, I. B. and Network, I. *Circadian Rhythms*. DOI: <https://www.nigms.nih.gov/education/fact-sheets/Pages/circadian-rhythms.aspx#:~:text=Circadian%20rhythms%20are%20physical%2C%20mental, the%20study%20of%20circadian%20rhythms..>
- [5] Kumara, J. “Analysis of Electroencephalography (EEG) Signals and Its Categorization—A Study”. In: *Procedia Engineering, Volume 38, Pages 2525-2536* (Sept. 2012). DOI: <https://doi.org/10.1016/j.proeng.2012.06.298>.
- [6] Silber, M. H. “The Visual Scoring of Sleep in Adults”. In: *Journal of Clinical Sleep Medicine* (Mar. 2007). DOI: <https://doi.org/10.5664/jcsm.26814>.
- [7] R., C. “Modulation of the Sympatho-Vagal Balance during Sleep: Frequency Domain Study of Heart Rate Variability and Respiration”. In: *Frontiers in Physiology* (Mar. 2012). DOI: [10.3389/fphys.2012.00045](https://doi.org/10.3389/fphys.2012.00045).
- [8] N., M. “Biosleep: A comprehensive sleep analysis system”. In: *Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Engineering in Medicine and Biology Society*. (2001). DOI: [10.1109/IEMBS.2001.1020520](https://doi.org/10.1109/IEMBS.2001.1020520).
- [9] H., D.-H. “Interrater reliability for sleep scoring according to the Rechtschaffen Kales and the new AASM standard”. In: *Journal of Sleep Research* (Mar. 2009). DOI: [10.1111/j.1365-2869.2008.00700.x](https://doi.org/10.1111/j.1365-2869.2008.00700.x).
- [10] M., C. “Principles and Practice of Sleep Medicine”. In: *4th ed. Philadelphia: Elsevier Saunders*. (2005).
- [11] Varga, B. “Heart Rate and Heart Rate Variability during Sleep in Family Dogs (*Canis familiaris*). Moderate Effect of Pre-Sleep Emotions”. In: *Family Dogs (Canis familiaris). Moderate Effect of Pre-Sleep Emotions. Animals (Basel)* (July 2018). DOI: [10.3390/ani8070107](https://doi.org/10.3390/ani8070107).



- [12] Patel, A. “Physiology, Sleep Stages”. In: *National Library of Medicine* (2021). DOI: <https://pubmed.ncbi.nlm.nih.gov/30252388/>.
- [13] MD, R. B. “Introduction”. In: *Sleep Medicine Pearls (Third Edition), Pages 10-14* (2015). DOI: <https://doi.org/10.1016/B978-1-4557-7051-9.00002-4>.
- [14] Cash, S. S. “The Human K-Complex Represents an Isolated Cortical Down-State”. In: *Science* (May 2009). DOI: [10.1126/science.1169626](https://doi.org/10.1126/science.1169626).
- [15] Gais, S. “Learning-Dependent Increases in Sleep Spindle Density”. In: *Journal of Neuroscience* (Aug. 2002). DOI: [10.1523/JNEUROSCI.22-15-06830.2002](https://doi.org/10.1523/JNEUROSCI.22-15-06830.2002).
- [16] Dictionary, A. “APA Dictionary of Psychology: Delta wave”. In: *American Psychological Association* (). DOI: <https://dictionary.apa.org/delta-wave>.
- [17] Monica, C. della. “Rapid Eye Movement Sleep, Sleep Continuity and Slow Wave Sleep as Predictors of Cognition, Mood, and Subjective Sleep Quality in Healthy Men and Women, Aged 20–84 Years”. In: *Front in Psychiatry* (June 2018). DOI: [10.3389/fpsy.2018.00255](https://doi.org/10.3389/fpsy.2018.00255).
- [18] Ferri, R. “REM Sleep EEG Instability in REM Sleep Behavior Disorder and Clonazepam Effects”. In: *Sleep* (Aug. 2017). DOI: [10.1093/sleep/zsx080](https://doi.org/10.1093/sleep/zsx080).
- [19] Somers, V. K. “Sympathetic-nerve activity during sleep in normal subjects”. In: *New England journal of medicine* (Feb. 1993). DOI: [10.1056/NEJM199302043280502](https://doi.org/10.1056/NEJM199302043280502).
- [20] Madsen, P. L. “Cerebral O2 metabolism and cerebral blood flow in humans during deep and rapid-eye-movement sleep”. In: *Journal of Applied Physiology* (June 1991). DOI: [10.1152/jappl.1991.70.6.2597](https://doi.org/10.1152/jappl.1991.70.6.2597).
- [21] NHLBI (National Heart, L. and Institute), B. *Sleep, Sleep Disorders, and Biological Rhythms: NIH Curriculum Supplement Series*. Colorado Springs, 2003.
- [22] Catcheside, P. G. “Noninvasive cardiovascular markers of acoustically induced arousal from non-rapid-eye-movement sleep”. In: *Sleep* (Nov. 2002). DOI: [10.1093/sleep/25.7.797](https://doi.org/10.1093/sleep/25.7.797).
- [23] Tank, J. “Relationship between blood pressure, sleep K-complexes, and muscle sympathetic nerve activity in humans”. In: *American Journal of Physiology Regulatory Integrative Comparative Physiology* (July 2003). DOI: [10.1152/ajpregu.00013.2003](https://doi.org/10.1152/ajpregu.00013.2003).
- [24] NJ, D. “Respiratory physiology: Control of ventilation”. In: *Principles and Practice of Sleep Medicine. 4th ed. Philadelphia: Elsevier Saunders, pp. 224–229.* (2005).
- [25] Labarca, G. “CPAP in patients with obstructive sleep apnea and type 2 diabetes mellitus: Systematic review and meta-analysis”. In: *Clinical Respiratory Journal* (Aug. 2018). DOI: [10.1111/crj.12915](https://doi.org/10.1111/crj.12915).
- [26] Khattak, H. K. “Obstructive Sleep Apnea in Heart Failure: Review of Prevalence, Treatment with Continuous Positive Airway Pressure, and Prognosis”. In: *Texas heart Institute Journal* (June 2018). DOI: [10.14503/THIJ-15-5678](https://doi.org/10.14503/THIJ-15-5678).
- [27] OV, Y. “Sleep and wakefulness disorders in neurodegenerative diseases”. In: *Zh Nevrol Psikhiatr Im S S Korsakova* (2018). DOI: <https://doi.org/10.17116/jnevro20181184283>.
- [28] S, L. “Deterioration of executive dysfunction in elderly with REM sleep behavior disorder (RBD)”. In: *Neurobiology of aging* (Oct. 2018). DOI: [10.1016/j.neurobiolaging.2018.06.029](https://doi.org/10.1016/j.neurobiolaging.2018.06.029).
- [29] Liu, S. “Excessive daytime sleepiness in Chinese patients with sporadic amyotrophic lateral sclerosis and its association with cognitive and behavioural impairments”. In: *Journal of neurology, neurosurgery, psychiatry* (Oct. 2018). DOI: [10.1136/jnnp-2018-318810](https://doi.org/10.1136/jnnp-2018-318810).
- [30] Kayaba, M. “Clinical significance of social jetlag in patients with excessive daytime sleepiness”. In: *Chronobiology international* (Nov. 2018). DOI: [10.1080/07420528.2018.1499666](https://doi.org/10.1080/07420528.2018.1499666).
- [31] Handley, S. “Deformities of Nature: Sleepwalking and Non-Conscious States of Mind in Late Eighteenth-Century Britain”. In: *Journal of the history of ideas* (July 2017). DOI: [10.1353/jhi.2017.0023](https://doi.org/10.1353/jhi.2017.0023).

- 
- [32] Zambotti, M. de. “Wearable Sleep Technology in Clinical and Research Settings”. In: *Med Sci Sports Exerc* (July 2019), pp. 1538–1557. DOI: [10.1249/MSS.0000000000001947](https://doi.org/10.1249/MSS.0000000000001947).
- [33] Shelgikar, A. V. “Sleep Tracking, Wearable Technology, and Opportunities for Research and Clinical Care”. In: *Chest* (Sept. 2016). DOI: [10.1016/j.chest.2016.04.016](https://doi.org/10.1016/j.chest.2016.04.016).
- [34] Khosla, S. “Consumer sleep technology: accuracy and impact on behavior among healthy individuals”. In: *Journal of Clinical Sleep Medicine* (May 2020). DOI: <https://doi.org/10.5664/jcsm.8450>.
- [35] Baron, K. G. “Feeling validated yet? A scoping review of the use of consumer-targeted wearable and mobile technology to measure and improve sleep”. In: *Sleep Medicine Review* (Aug. 2018). DOI: [10.1016/j.smr.2017.12.002](https://doi.org/10.1016/j.smr.2017.12.002).
- [36] C, L. “Sleep apps: what role do they play in clinical medicine?” In: *Current opinion in pulmonary medicine* (Nov. 2017). DOI: [10.1097/MCP.0000000000000425](https://doi.org/10.1097/MCP.0000000000000425).
- [37] Dunn, J. “Wearables and the medical revolution”. In: *Personalized medicine* (Sept. 2018). DOI: [10.2217/pme-2018-0044](https://doi.org/10.2217/pme-2018-0044).
- [38] Roomkham, S. “Promises and Challenges in the Use of Consumer-Grade Devices for Sleep Monitoring”. In: *IEEE Reviews in Biomedical Engineering* (Mar. 2018). DOI: [10.1109/RBME.2018.2811735](https://doi.org/10.1109/RBME.2018.2811735).
- [39] Bianchi, M. T. “Sleep devices: wearables and nearables, informational and interventional, consumer and clinical”. In: *Metabolism: clinical and experimental* (July 2018). DOI: [10.1016/j.metabol.2017.10.008](https://doi.org/10.1016/j.metabol.2017.10.008).
- [40] D., R. “Electromyographic Kinesiology, Research Methods in Biomechanics”. In: *Medicine Science in Sports Exercise* (2014).
- [41] Lilly, L. S. *Pathophysiology of Heart Disease: A Collaborative Project of Medical Students and Faculty, 6th Edition*. 2016.
- [42] C., I. “The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications”. In: *American Academy of Sleep Medicine* (2007).
- [43] Koudikelis, D. Z. *PSG Polysomnography image*. URL: [https://www.dentalsleepmed.gr/english/sleep\\_study - diag - 4.html](https://www.dentalsleepmed.gr/english/sleep_study - diag - 4.html).
- [44] Markun, L. C. “Clinician-Focused Overview and Developments in Polysomnography”. In: *Sleep and Neurological Conditions (A Avidan, Section Editor)* (Nov. 2020).
- [45] Ibáñez, V. “A survey on sleep assessment methods”. In: *Peer J* (2018). DOI: [10.7717/peerj.4849](https://doi.org/10.7717/peerj.4849).
- [46] Balli, S. “The Usage of Statistical Learning Methods on Wearable Devices and a Case Study: Activity Recognition on Smartwatches”. In: *Advances in Statistical Methodologies and Their Application to Real Problems* (Apr. 2017). DOI: [10.5772/66213](https://doi.org/10.5772/66213).
- [47] ActiGraph. *ActiGraph*. URL: <https://actigraphcorp.com/academic-research/>.
- [48] Palotti, J. “Benchmark on a large cohort for sleep-wake classification with machine learning techniques”. In: *NPJ Digital Medicine* (June 2019). DOI: <https://doi.org/10.1038/s41746-019-0126-9>.
- [49] Goldstone, A. “Actigraphy in the digital health revolution: still asleep?” In: *Sleep* (Sept. 2018). DOI: [10.1093/sleep/zsy120](https://doi.org/10.1093/sleep/zsy120).
- [50] Bild, D. E. “Multi-Ethnic Study of Atherosclerosis: objectives and design”. In: *American Journal of Epidemiology* (Nov. 2002). DOI: [10.1093/aje/kwf113](https://doi.org/10.1093/aje/kwf113).
- [51] Burke, G. “The Multiethnic Study of Atherosclerosis”. In: *Global Heart, Volume 11, Issue 3, Pages 267-268* (Sept. 2016). DOI: <https://doi.org/10.1016/j.gheart.2016.09.001>.
- [52] MESA. *MESA Information for Researchers*. URL: <https://www.mesa-nhlbi.org/researchers.aspx>.
- [53] actigraphy, M. *MESA Actigraphy Introduction*. URL: <https://sleepdata.org/datasets/mesa/pages/actigraphy-introduction.md>.
-

- [54] Philipps. *Philipps Actiwatch Spectrum*. URL: <https://www.usa.philips.com/healthcare/product/HC1046964/spectrum-activity-monitorx>.
- [55] PSG, M. *MESA Polysomnography Introduction*. URL: <https://sleepdata.org/datasets/mesa/pages/polysomnography-introduction.md>.
- [56] device, M. P. *MESA PSG device*. URL: <https://www.compumedics.com.au/en/products/somte/>.
- [57] Brown, M. “ISCEV Standard for Clinical Electro-oculography (EOG)”. In: *Springer Link* (Nov. 2006). DOI: [10.1007/s10633-006-9030-0](https://doi.org/10.1007/s10633-006-9030-0).
- [58] B., K. “European data format ‘plus’ (EDF+), an EDF alike standard format for the exchange of physiological data”. In: *Clinical Neurophysiology* 114 (9): 1755–61 (Sept. 2003). DOI: [10.1016/S1388-2457\(03\)00123-8](https://doi.org/10.1016/S1388-2457(03)00123-8).
- [59] Perez-Pozuelo, I. “The future of sleep health: a data-driven revolution in sleep science and medicine”. In: *NPJ Digital Medicine* (Mar. 2020). DOI: <https://doi.org/10.1038/s41746-020-0244-4>.
- [60] Ambale-Venkatesh, B. “Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis”. In: *Circulation research* (Aug. 2017). DOI: [10.1161/CIRCRESAHA.117.311312](https://doi.org/10.1161/CIRCRESAHA.117.311312).
- [61] Hathaway, Q. A. “Deep Neural Survival Networks for Cardiovascular Risk Prediction: The Multi-Ethnic Study of Atherosclerosis (MESA)”. In: *Circulation research* (Apr. 2021). DOI: <https://doi.org/10.1101/2021.04.12.21255286>.
- [62] Ward, A. “Machine learning and atherosclerotic cardiovascular disease risk prediction in a multi-ethnic population”. In: *medRxiv* (2020). DOI: <https://doi.org/10.1038/s41746-020-00331-1>.
- [63] Chen, T. “XGBoost: A Scalable Tree Boosting System”. In: *arXiv* (Mar. 2016). DOI: <https://doi.org/10.48550/arXiv.1603.02754>.
- [64] Breiman, L. “Random Forests”. In: *Springer Link, Machine Learning volume 45, pages5–32* (Oct. 2001). DOI: <https://doi.org/10.1023/A:1010933404324>.
- [65] Vallat, R. “An open-source, high-performance tool for automated sleep staging”. In: *eLife* (Oct. 2021). DOI: [10.7554/eLife.70092](https://doi.org/10.7554/eLife.70092).
- [66] Perslev, M. “U-Sleep: resilient high-frequency sleep staging”. In: *NPJ Digital Medicine* (Apr. 2021). DOI: [10.1038/s41746-021-00440-5](https://doi.org/10.1038/s41746-021-00440-5).
- [67] Stephansen, J. B. “Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy”. In: *NPJ Digital Medicine* (Dec. 2018). DOI: [10.1038/s41467-018-07229-3](https://doi.org/10.1038/s41467-018-07229-3).
- [68] Sridhar, N. “Deep learning for automated sleep staging using instantaneous heart rate”. In: *npj Digital Medicine* (Aug. 2020). DOI: [10.1038/s41746-020-0291-x](https://doi.org/10.1038/s41746-020-0291-x).
- [69] Emmert-Streib, F. “An Introductory Review of Deep Learning for Prediction Models With Big Data”. In: *arXiv* (Feb. 2020). DOI: <https://doi.org/10.3389/frai.2020.00004>.
- [70] Hochreiter, S. “Long Short-term Memory”. In: *PubMed* (Dec. 1997). DOI: <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [71] Sherstinsky, A. “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network”. In: *arXiv* (Jan. 2021). DOI: <https://doi.org/10.48550/arXiv.1808.03314>.
- [72] Colah. *Understanding LSTM Networks*. URL: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- [73] Alzubaidi, L. “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions”. In: *Journal of Big Data* (2021). DOI: <https://doi.org/10.1186/s40537-021-00444-8>.
- [74] Education, I. C. *Convolutional Neural Networks*. URL: <https://www.ibm.com/cloud/learn/convolutional-neural-networks>.
- [75] Rumelhart, D. “The back-propagation algorithm”. In: *Nature Vol 323* (Oct. 1986).

- 
- [76] Wikipedia. *Gradient descent*. URL: [https://en.wikipedia.org/wiki/Gradient\\_descent](https://en.wikipedia.org/wiki/Gradient_descent).
- [77] Fukushima, K. "Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position". In: *Biol. Cybernetics* 36, 193-202 (1980).
- [78] LeCun, Y. "Backpropagation Applied to Handwritten Zip Code Recognition". In: *Neural Computation* 1, 541-551 (1989).
- [79] LeCun, Y. "Gradient Based Learning Applied to Document Recognition". In: *IEEE* (1998).
- [80] Krizhevsky, A. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Communications of the ACM Volume 60, Issue 6* (2017). DOI: <https://doi.org/10.1145/3065386>.
- [81] Simonyan, K. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *arXiv* (2014). DOI: <https://doi.org/10.48550/arXiv.1409.1556>.
- [82] Szegedy, C. "Going Deeper with Convolutions". In: *arXiv* (2014). DOI: <https://doi.org/10.48550/arXiv.1409.4842>.
- [83] He, K. "Deep Residual Learning for Image Recognition". In: *arXiv* (2015). DOI: <https://doi.org/10.48550/arXiv.1512.03385>.
- [84] Vaswani, A. "Attention Is All You Need". In: *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA* (Dec. 2017). DOI: <https://doi.org/10.48550/arXiv.1706.03762>.
- [85] Wikipedia. *Gated Recurrent Unit*. URL: [https://en.wikipedia.org/wiki/Gated\\_recurrent\\_unit](https://en.wikipedia.org/wiki/Gated_recurrent_unit).
- [86] Fedorin, I. "Sleep Stages Classification in a Healthy People Based on Optical Plethysmography and Accelerometer Signals via Wearable Devices". In: *IEEE 2nd Ukraine Conference on Electrical and Computer Engineering* (July 2019). DOI: [10.1109/UKRCON.2019.8879875](https://doi.org/10.1109/UKRCON.2019.8879875).
- [87] Radha, M. "Sleep stage classification from heart-rate variability using long short-term memory neural networks". In: *Nature* (Oct. 2019). DOI: <https://doi.org/10.1038/s41598-019-49703-y>.
- [88] Zhang, Y. "Sleep Stage Classification Using Bidirectional LSTM in Wearable Multi-sensor Systems". In: *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops* (Sept. 2019). DOI: [10.1109/INFCOMW.2019.8845115](https://doi.org/10.1109/INFCOMW.2019.8845115).
- [89] Radu, V. "Multimodal Deep Learning for Activity and Context Recognition". In: *ACM Journals, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1.157 (Jan. 2018), pp. 1-27. DOI: <https://doi.org/10.1145/3161174>.
- [90] Malik, M. "Age-predicted maximal heart rate revisited". In: *Annals of Noninvasive Electrocardiology* (1996). DOI: <https://doi.org/10.1111/j.1542-474X.1996.tb00275.x>.
- [91] Pichot, V. "HRVanalysis: A Free Software for Analyzing Cardiac Autonomic Activity". In: *Frontiers Physiology* (Nov. 2016). DOI: <https://doi.org/10.3389/fphys.2016.00557>.
- [92] Shaffer, F. "An Overview of Heart Rate Variability Metrics and Norms". In: *Front Public Health* (Sept. 2017). DOI: [10.3389/fpubh.2017.00258](https://doi.org/10.3389/fpubh.2017.00258).
- [93] Eikermann, M. and Zaremba, S. *Sleep Medicine Miller's Anesthesia*, 10. 2020. DOI: [267-293.e8](https://doi.org/10.1016/j.annarc.2020.07.008).
- [94] Walch, O. "Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device". In: *Sleep* 42.12 (Aug. 2019). DOI: <https://doi.org/10.1093/sleep/zsz180>.
- [95] Um, T. T. "Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring using Convolutional Neural Networks". In: *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (Nov. 2017). DOI: <https://doi.org/10.1145/3136755.3136817>.
-

- 
- [96] Iwana, B. K. “An empirical survey of data augmentation for time series classification with neural networks”. In: *PLoS ONE* (July 2021). DOI: <https://doi.org/10.1371/journal.pone.0254841>.
- [97] Chang, L. “SleepGuard: Capturing Rich Sleep Information Using Smartwatch Sensing”. In: *Data. Proc. ACM Interact.Mob. Wearable Ubiquitous Technol.* (Sept. 2018). DOI: <https://doi.org/10.1145/3264908>.
- [98] Slyusarenko, K. “Poster: Sleep Stages Classifier with Eliminated Apnea Impact”. In: *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2019 International Symposium on Wearable Computers (UbiComp/ISWC’19 Adjunct)* (Sept. 2019). DOI: <https://dl.acm.org/doi/pdf/10.1145/3341162.3343851>.
- [99] Scalise, L. “Wearables for health and fitness: Measurement characteristics and accuracy”. In: *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)* (May 2018). DOI: [10.1109/I2MTC.2018.8409635](https://doi.org/10.1109/I2MTC.2018.8409635).
- [100] Resource, T. N. S. R. *Multi-Ethnic Study of Atherosclerosis*. URL: <https://sleepdata.org/datasets/ mesa>.
- [101] GQ, Z. “The National Sleep Research Resource: towards a sleep data commons”. In: *J Am Med Inform Assoc.* (Oct. 2018). DOI: [10.1093/jamia/ocy064](https://doi.org/10.1093/jamia/ocy064).
- [102] X, C. “Differences in Sleep Disturbances: The Multi-Ethnic Study of Atherosclerosis (MESA)”. In: *Sleep* (June 2015). DOI: [10.5665/sleep.4732](https://doi.org/10.5665/sleep.4732).
- [103] Mitsukura, Y. “Sleep stage detection using only heart rate”. In: *Health Informatics Journal* (Feb. 2019). DOI: <https://doi.org/10.1177/1460458219827349>.
- [104] Zhang, X. “Sleep Stage Classification Based on Multi-level Feature Learning and Recurrent Neural Networks via Wearable Device”. In: *Comput Biol Med* (Dec. 2018). DOI: [10.1016/j.combiomed.2018.10.010](https://doi.org/10.1016/j.combiomed.2018.10.010).
- [105] QRS, W. *QRS Complex*. URL: [https://en.wikipedia.org/wiki/QRS\\_complex](https://en.wikipedia.org/wiki/QRS_complex).
- [106] Morelli, D. “Analysis of the Impact of Interpolation Methods of Missing RR-Intervals Caused by Motion Artifacts on HRV Features Estimations”. In: *Sensors* (2019). DOI: [10.3390/s19143163](https://doi.org/10.3390/s19143163).
- [107] Kwon, H. B. “Attention-Based LSTM for Non-Contact Sleep Stage Classification Using IR-UWB Radar”. In: *IEEE Journal of Biomedical and Health Informatics, Volume: 25, Issue: 10* (Oct. 2021). DOI: [10.1109/JBHI.2021.3072644](https://doi.org/10.1109/JBHI.2021.3072644).
- [108] Srivastava, N. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *Journal of Machine Learning Research 15 1929-1958* (June 2014).
- [109] Bahdanau, D. “NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE”. In: *ICLR 2015* (Sept. 2014). DOI: <https://doi.org/10.48550/arXiv.1409.0473>.