



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ  
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ  
ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ**

**ΑΝΑΓΝΩΡΙΣΗ ΑΝΘΡΩΠΙΝΗΣ ΚΙΝΗΤΙΚΗΣ  
ΔΡΑΣΤΗΡΙΟΤΗΤΑΣ ΓΙΑ ΤΗΝ ΜΟΝΤΕΛΟΠΟΙΗΣΗ  
ΚΙΝΗΤΙΚΩΝ ΣΥΜΠΕΡΙΦΟΡΩΝ**

**ΦΥΚΑ ΔΕΣΠΟΙΝΑ**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ  
Υπεύθυνος  
Τασουλής Σωτήριος  
Αναπληρωτής Καθηγητής**

**Λαμία, 2025**



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ  
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ  
ΒΙΟΙΑΤΡΙΚΗ**

**ΑΝΑΓΝΩΡΙΣΗ ΑΝΘΡΩΠΙΝΗΣ ΚΙΝΗΤΙΚΗΣ  
ΔΡΑΣΤΗΡΙΟΤΗΤΑΣ ΓΙΑ ΤΗΝ ΜΟΝΤΕΛΟΠΟΙΗΣΗ  
ΚΙΝΗΤΙΚΩΝ ΣΥΜΠΕΡΙΦΟΡΩΝ**

**ΦΥΚΑ ΔΕΣΠΟΙΝΑ**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Επιβλέπων  
Τασουλής Σωτήριος  
Αναπληρωτής Καθηγητής**

**Λαμία, 2025**

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

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.
2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.
3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια
4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία: 30/09/2025

Ο – Η Δηλ.

ΔΕΣΠΟΙΝΑ ΦΥΚΑ

(Υπογραφή)

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

**ΑΝΑΓΝΩΡΙΣΗ ΑΝΘΡΩΠΙΝΗΣ ΚΙΝΗΤΙΚΗΣ  
ΔΡΑΣΤΗΡΙΟΤΗΤΑΣ ΓΙΑ ΤΗΝ ΜΟΝΤΕΛΟΠΟΙΗΣΗ  
ΚΙΝΗΤΙΚΩΝ ΣΥΜΠΕΡΙΦΟΡΩΝ**

**Φύκα Δέσποινα**

**Τριμελής Επιτροπή:**

Τασουλής Σωτήριος, Αναπληρωτής Καθηγητής (επιβλέπων)

Δελήμπασης Κωνσταντίνος, Καθηγητής

Πλαγιανάκος Βασίλειος, Καθηγητής

# Περιεχόμενα

|                                                                        |           |
|------------------------------------------------------------------------|-----------|
| <b>Ευχαριστίες</b>                                                     | <b>9</b>  |
| <b>Περίληψη</b>                                                        | <b>10</b> |
| <b>Abstract</b>                                                        | <b>11</b> |
| <b>1 Εισαγωγή</b>                                                      | <b>12</b> |
| 1.1 Στόχοι της Έρευνας . . . . .                                       | 13        |
| 1.2 Ερευνητικά Ερωτήματα . . . . .                                     | 14        |
| <b>2 Θεωρητικό Υπόβαθρο</b>                                            | <b>15</b> |
| 2.1 Εισαγωγή στην Αναγνώριση Ανθρώπινης Δραστηριότητας (HAR) . . . . . | 15        |
| 2.1.1 Τύποι Συλλογής Δεδομένων στην HAR . . . . .                      | 16        |
| 2.2 Αναγνώριση Προτύπων και Μηχανική Μάθηση . . . . .                  | 17        |
| 2.3 Νευρωνικά Δίκτυα . . . . .                                         | 19        |
| 2.4 Βαθιά Μάθηση . . . . .                                             | 21        |
| 2.5 Συνελκτικά Νευρωνικά Δίκτυα (CNNs) . . . . .                       | 22        |
| 2.5.1 Συνελκτικό Επίπεδο . . . . .                                     | 25        |
| 2.5.2 Επίπεδο υποδειγματοληψίας . . . . .                              | 26        |
| 2.5.3 Επίπεδο Κανονικοποίησης . . . . .                                | 27        |
| 2.5.4 Πλήρως Συνδεδεμένα Επίπεδα . . . . .                             | 28        |
| 2.5.5 Συνάρτηση Ενεργοποίησης ReLU . . . . .                           | 28        |
| 2.6 Αναδρομικά Νευρωνικά Δίκτυα (RNNs) και Δίκτυα LSTM . . . . .       | 29        |
| 2.6.1 Δίκτυα Long Short Term Memory (LSTMs) . . . . .                  | 30        |
| <b>3 Σχετική Βιβλιογραφία</b>                                          | <b>32</b> |
| <b>4 Μεθοδολογία</b>                                                   | <b>35</b> |
| 4.1 Σχεδιασμός Έρευνας . . . . .                                       | 35        |
| 4.2 Σύνολο Δεδομένων . . . . .                                         | 36        |
| 4.3 Προεπεξεργασία και Ανάλυση Δεδομένων . . . . .                     | 38        |
| 4.4 Αλγόριθμοι και Μοντέλα . . . . .                                   | 41        |
| 4.5 Αξιολόγηση Μοντέλων . . . . .                                      | 42        |
| 4.5.1 Μετρικές Αξιολόγησης . . . . .                                   | 43        |
| <b>5 Ερευνητική Διαδικασία</b>                                         | <b>46</b> |
| 5.1 Τεχνολογικό Υπόβαθρο . . . . .                                     | 46        |

|          |                                                     |           |
|----------|-----------------------------------------------------|-----------|
| 5.2      | Ανάλυση Δεδομένων . . . . .                         | 47        |
| 5.3      | Ανάπτυξη και Εκπαίδευση Μοντέλων . . . . .          | 52        |
| 5.3.1    | Μοντέλο Logistic Regression . . . . .               | 53        |
| 5.3.2    | Μοντέλο CNN . . . . .                               | 55        |
| 5.3.3    | Μοντέλο LSTM . . . . .                              | 57        |
| 5.3.4    | Υβριδικό Μοντέλο CNN-LSTM . . . . .                 | 58        |
| <b>6</b> | <b>Αποτελέσματα</b>                                 | <b>61</b> |
| 6.1      | Αποτελέσματα Μοντέλου Logistic Regression . . . . . | 61        |
| 6.2      | Αποτελέσματα Μοντέλου CNN . . . . .                 | 63        |
| 6.3      | Αποτελέσματα Μοντέλου LSTM . . . . .                | 65        |
| 6.4      | Αποτελέσματα Υβριδικού Μοντέλου CNN-LSTM . . . . .  | 67        |
| 6.5      | Σύγκριση Μοντέλων . . . . .                         | 69        |
| <b>7</b> | <b>Συμπεράσματα</b>                                 | <b>73</b> |

# Κατάλογος σχημάτων

|     |                                                                                                                         |    |
|-----|-------------------------------------------------------------------------------------------------------------------------|----|
| 2.1 | Διάγραμμα Αναγνώρισης Ανθρώπινης Δραστηριότητας (πηγή: ιδία φωτογραφία) . . . . .                                       | 16 |
| 2.2 | Απλό Νευρωνικό Δίκτυο (πηγή: M. A. Nielsen (2015) [16]) . . . . .                                                       | 20 |
| 2.3 | Βαθύ Νευρωνικό Δίκτυο Αρχιτεκτονική (πηγή: M. A. Nielsen (2015) [16]) . . . . .                                         | 22 |
| 2.4 | Αρχιτεκτονική Τυπικού CNN (πηγή: Nicholas et al., 2022 [21]) . . . . .                                                  | 25 |
| 2.5 | Παράδειγμα max και average pooling (πηγή: Bhatti et al., 2023 [23]) . . . . .                                           | 27 |
| 2.6 | Συνάρτηση ReLU (πηγή: Ailephant AI Team (2025) [25]) . . . . .                                                          | 28 |
| 2.7 | Αναδρομικό Νευρωνικό Δίκτυο RNN (πηγή: Ailephant AI Team (2025) [27]) . . . . .                                         | 29 |
| 2.8 | Δίκτυο Long Short-Term Memory (LSTM) (πηγή: Medium.com [28]) . . . . .                                                  | 30 |
| 4.1 | Ταξινόμηση κατηγοριών ανθρώπινης δραστηριότητας για έναν τυχαίο χρήστη . . . . .                                        | 39 |
| 4.2 | Ταξινόμηση των 20 ετικετών με τις περισσότερες δηλωμένες τιμές από τους χρήστες (πηγή: ιδία φωτογραφία) . . . . .       | 40 |
| 4.3 | Οπτική απεικόνιση 5-fold Cross-Validation (Unseen Users) (πηγή: ιδία φωτογραφία) . . . . .                              | 42 |
| 5.1 | Διάγραμμα χρονικής κατανομής ενεργειών για δύο τυχαίους χρήστες (πηγή: ιδία φωτογραφία) . . . . .                       | 48 |
| 5.2 | Jaccard Heatmap συνύπαρξης ετικετών για έναν χρήστη (πηγή: ιδία φωτογραφία) . . . . .                                   | 49 |
| 5.3 | Γραφήματα raw_acc:magnitude_stats:mean για έναν χρήστη (πηγή: ιδία φωτογραφία) . . . . .                                | 50 |
| 5.4 | Γραφήματα watch_acceleration:3d:mean_z για έναν χρήστη (πηγή: ιδία φωτογραφία) . . . . .                                | 50 |
| 5.5 | Γραφήματα watch_acceleration:3d:mean_z για έναν χρήστη (πηγή: ιδία φωτογραφία) . . . . .                                | 52 |
| 6.1 | Διαγράμματα Jaccard για το μοντέλο Logistic Regression σε Predicted και True ετικέτες (πηγή: ιδία φωτογραφία) . . . . . | 63 |
| 6.2 | Διαγράμματα Jaccard για το μοντέλο CNN σε Predicted και True ετικέτες (πηγή: ιδία φωτογραφία) . . . . .                 | 65 |
| 6.3 | Διαγράμματα Jaccard για το μοντέλο LSTM σε Predicted και True ετικέτες (πηγή: ιδία φωτογραφία) . . . . .                | 67 |
| 6.4 | Διαγράμματα Jaccard για το υβριδικό μοντέλο CNN-LSTM σε Predicted και True ετικέτες (πηγή: ιδία φωτογραφία) . . . . .   | 68 |

## Κατάλογος πινάκων

|     |                                                                         |    |
|-----|-------------------------------------------------------------------------|----|
| 4.1 | Πίνακας αισθητήρων . . . . .                                            | 36 |
| 6.1 | Πίνακας τελικών τιμών μετρικών για το μοντέλο Logistic Regression . . . | 62 |
| 6.2 | Πίνακας τελικών τιμών μετρικών για το μοντέλο CNN . . . . .             | 64 |
| 6.3 | Πίνακας τελικών τιμών μετρικών για το μοντέλο LSTM . . . . .            | 66 |
| 6.4 | Πίνακας τελικών τιμών μετρικών για το υβριδικό μοντέλο CNN-LSTM . .     | 68 |
| 6.5 | Συγκριτικός πίνακας μέσων τιμών μετρικών από τους Πίνακες 6.1–6.4 . . . | 71 |

# Ευχαριστίες

Ξεκινώντας, θα ήθελα να ευχαριστήσω την οικογένεια μου και φίλους για την τόσο σημαντική στήριξη που μου παρείχαν όλα αυτά τα χρόνια σε αυτό το δύσβατο δρόμο της γνώσης. Από τις ευχαριστίες δεν θα μπορούσε να λείπει ο επιβλέπων Καθηγητής της πτυχιακής μου εργασίας κος Σωτήριος Τασουλής που ήταν εκεί όλο αυτόν τον καιρό να με καθοδηγεί σε κάθε δυσκολία που συνάντησα. Κλείνοντας, θα ήθελα να ευχαριστήσω το σύνολο των καθηγητών του τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική για την άριστη συνεργασία που είχαμε καθ' όλη την διάρκεια των σπουδών μου καθώς επίσης και για τις γνώσεις που μου μεταλαμπάδευσαν.

# Περίληψη

Η παρούσα μελέτη εστιάζει στην αναγνώριση ανθρώπινης κινητικής δραστηριότητας μέσω φορετών συσκευών, με σκοπό τη μοντελοποίηση συμπεριφορών που σχετίζονται με την υγεία και την καθημερινότητα. Χρησιμοποιείται το σύνολο δεδομένων ExtraSensory, το οποίο περιλαμβάνει πλούσια, ρεαλιστικά δεδομένα από αισθητήρες smartphone και smart-watch, καταγεγραμμένα σε φυσικές συνθήκες από 60 χρήστες. Η ανάλυση τέτοιων in-the-wild δεδομένων συνοδεύεται από προκλήσεις, όπως η απουσία ελέγχου στον τρόπο καταγραφής, η ετερογένεια των συσκευών και η υποκειμενικότητα στις επισημάνσεις.

Η μελέτη εξετάζει τη συμβολή των έξυπνων ρολογιών στην παρακολούθηση δραστηριοτήτων και αξιολογεί σύγχρονες τεχνικές μηχανικής μάθησης για την ταξινόμηση και πρόβλεψη συμπεριφοράς. Μέσω πειραματικής διαδικασίας, συγκρίνονται παραδοσιακές και σύγχρονες προσεγγίσεις, ενώ γίνεται χρήση πολυτροπικών αισθητηριακών δεδομένων, με στόχο την ανάπτυξη αποδοτικών και γενικεύσιμων μοντέλων για εξατομικευμένη ή καθολική πρόβλεψη δραστηριοτήτων.

**Λέξεις-κλειδιά:** Αναγνώριση ανθρώπινης δραστηριότητας, φορετοί αισθητήρες, δεδομένα αισθητήρων, μηχανική μάθηση, βαθειά μάθηση, νευρωνικό δίκτυο, CNN, LSTM, κινητική συμπεριφορά

# Abstract

This study focuses on human activity recognition using wearable devices, aiming to model behaviors related to health and everyday life. It utilizes the ExtraSensory dataset, which contains rich and realistic sensor data collected from smartphones and smartwatches worn by 60 users in natural, uncontrolled conditions. Analyzing such in-the-wild data presents significant challenges, including a lack of control over data collection, device heterogeneity, and the subjectivity of user-reported activity labels.

The study explores the role of smartwatches in monitoring human activity and evaluates modern machine learning techniques for the classification and prediction of behavioral patterns. Through a structured experimental process, traditional and deep learning approaches are compared using multimodal sensor data, aiming to develop efficient and generalizable models for both personalized and population-level behavior prediction.

**Keywords:** Human Activity Recognition (HAR), wearable sensors, smartwatches, ExtraSensory dataset, in-the-wild data, multimodal sensing, machine learning, behavioral modeling.

**Keywords:** Human activity recognition, wearable sensors, sensor data, machine learning, deep learning, neural network, CNN, LSTM, movement pattern

# 1 Εισαγωγή

Ο σύγχρονος τρόπος ζωής, χαρακτηριζόμενος από απαιτητικά εργασιακά ωράρια, παρατεταμένη έκθεση σε οθόνες και αυξανόμενη εξάρτηση από την τεχνολογία, έχει μειώσει σημαντικά τον χρόνο που τα άτομα αφιερώνουν σε δραστηριότητες σωματικής άσκησης. Η επικράτηση καθιστικών επαγγελμάτων, η αστικοποίηση και η ευκολία των ψηφιακών υπηρεσιών έχουν οδηγήσει σε αισθητή μείωση της καθημερινής φυσικής δραστηριότητας σε σύγκριση με το παρελθόν. Πολλοί άνθρωποι περνούν παρατεταμένα διαστήματα καθιστοί — όσο μελετούν, μετακινούνται ή ψυχαγωγούνται ψηφιακά — καλλιεργώντας συνήθειες που δίνουν προτεραιότητα στην ευκολία εις βάρος της σωματικής ευεξίας [1].

Αυτή η μεταβολή έχει σημαντικές επιπτώσεις στη δημόσια υγεία, συμβάλλοντας στην αυξημένη εμφάνιση παχυσαρκίας, καρδιαγγειακών νοσημάτων και μεταβολικών διαταραχών. Στις Ηνωμένες Πολιτείες, για παράδειγμα, η παιδική παχυσαρκία έχει αναδειχθεί σε σοβαρό ζήτημα δημόσιας υγείας, άρρηκτα συνδεδεμένο με την έλλειψη τακτικής σωματικής δραστηριότητας [2]. Παράλληλα, η ανεπαρκής φυσική άσκηση σχετίζεται με αυξημένα επίπεδα άγχους, στρες και γνωστικής έκπτωσης (MCI), εντείνοντας τις μακροπρόθεσμες αρνητικές συνέπειες σε ατομικό και κοινωνικό επίπεδο [3].

Η Αναγνώριση Ανθρώπινης Δραστηριότητας (Human Activity Recognition – HAR) αναφέρεται στη διαδικασία ερμηνείας της ανθρώπινης κίνησης μέσω τεχνικών μηχανικής μάθησης, αναλύοντας πρότυπα κίνησης για την αναγνώριση δραστηριοτήτων, χειρονομιών ή συμπεριφορών [4]. Η HAR αποτελεί κρίσιμο πεδίο έρευνας χάρη στις πολυάριθμες εφαρμογές της: στην υγεία (έγκαιρη διάγνωση ασθενειών, υποστήριξη ηλικιωμένων), στον αθλητισμό (ανάλυση αποδόσεων), σε συστήματα έξυπνων πόλεων (παρακολούθηση συμπεριφορών πεζών) και σε πολλές άλλες περιοχές.

Αν και οι τεχνικές που βασίζονται σε οπτικά δεδομένα (π.χ. κάμερες) αποδίδουν ικανοποιητικά σε ελεγχόμενα εργαστηριακά περιβάλλοντα, η εφαρμογή τους σε πραγματικές συνθήκες παρουσιάζει σημαντικές δυσκολίες λόγω παραγόντων όπως η οπτική πολυπλοκότητα του περιβάλλοντος και ο μεταβαλλόμενος φωτισμός [5]. Αντίθετα, η χρήση φορετών αισθητήρων — ενσωματωμένων σε συσκευές όπως smartwatches και smartphones — επιτρέπει τη συνεχή και ακριβή συλλογή δεδομένων κίνησης σε φυσικά περιβάλλοντα.

Τα πρώιμα συστήματα HAR χρησιμοποιούσαν κανόνες βάσει χαρακτηριστικών σημμάτων ή κλασικά μοντέλα μηχανικής μάθησης, όπως τα Δέντρα Απόφασης (Decision Trees). Η έλευση των τεχνικών βαθιάς μάθησης (Deep Learning) επέφερε σημαντικές βελτιώσεις στην ταξινόμηση δραστηριοτήτων, αξιοποιώντας Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks – CNNs) και Δίκτυα Μακροπρόθεσμης-Βραχυπρόθεσμης Μνήμης (Long Short-Term Memory – LSTMs) [6], [7].

Η ευρεία διάδοση φορετών συσκευών έχει καταστήσει τη HAR πιο προσιτή και ακριβή. Φορητοί αισθητήρες όπως επιταχυνσιόμετρα και γυροσκόπια επιτρέπουν τη συνεχή παρα-

κολούθηση φυσικής δραστηριότητας και φυσιολογικών παραμέτρων, όπως ο καρδιακός ρυθμός (HRV), διευρύνοντας τις δυνατότητες εφαρμογών σε τομείς όπως η ανίχνευση πτώσεων ηλικιωμένων, η πρόωμη διάγνωση νευροεκφυλιστικών ασθενειών και η παρακολούθηση στρες σε πραγματικό χρόνο [8].

Παρά την πρόοδο, εξακολουθούν να υπάρχουν σημαντικές προκλήσεις, όπως η υψηλή μεταβλητότητα δεδομένων μεταξύ διαφορετικών ατόμων, ο θόρυβος αισθητήρων και ζητήματα προστασίας προσωπικών δεδομένων, τα οποία περιορίζουν τη διαθεσιμότητα ποιοτικών συνόλων δεδομένων για εκπαίδευση [9].

Στην παρούσα μελέτη, χρησιμοποιείται το ExtraSensory Dataset [10], το οποίο περιλαμβάνει δεδομένα από smartphones και smartwatches, τα οποία έχουν συλλεχθεί κατά την καθημερινή ζωή χρηστών, προσφέροντας ένα πλούσιο και ρεαλιστικό πλαίσιο για την ανάπτυξη και αξιολόγηση μεθόδων HAR. Στις επόμενες ενότητες υλοποιούνται, αξιολογούνται και συγκρίνονται ως προς την αποδοτικότητα τους μοντέλα ευρέως χρησιμοποιούμενα στην HAR. Πιο συγκεκριμένα, αναλύονται ένα κλασικό μοντέλο ταξινόμησης (Logistic Regression ή LR), καθώς επίσης και μοντέλα βαθιάς μάθησης, όπως Συνελικτικά Νευρωνικά Δίκτυα (CNNs) και Δίκτυα Μακροπρόθεσμης Βραχυπρόθεσμης Μνήμης (LSTMs), ενώ παράλληλα προτείνεται ένα υβριδικό μοντέλο που συνδυάζει τις τεχνικές αυτές με στόχο τη βελτίωση της ακρίβειας ταξινόμησης δραστηριοτήτων.

## 1.1 Στόχοι της Έρευνας

Οι βασικοί στόχοι της παρούσας μελέτης είναι οι εξής:

1. Ανάπτυξη και αξιολόγηση μοντέλων βαθιάς μάθησης για την Αναγνώριση Ανθρώπινης Δραστηριότητας μέσω δεδομένων από φορητές συσκευές, εστιάζοντας στη σύγκριση μεταξύ Συνελικτικών Νευρωνικών Δικτύων (CNNs) και Δικτύων Μακροπρόθεσμης Βραχυπρόθεσμης Μνήμης (LSTMs).
2. Σχεδίαση και υλοποίηση ενός νέου υβριδικού μοντέλου, το οποίο συνδυάζει χαρακτηριστικά διαφόρων τεχνικών με στόχο τη βελτιστοποίηση της ακρίβειας ταξινόμησης.
3. Ανάλυση των αποτελεσμάτων με χρήση καθιερωμένων μετρικών αξιολόγησης και σύγκριση της προτεινόμενης προσέγγισης με υπάρχουσες μεθόδους στη διεθνή βιβλιογραφία.
4. Ανάδειξη των προκλήσεων που σχετίζονται με τη χρήση δεδομένων συλλεγμένων στον πραγματικό κόσμο, όπως η μεταβλητότητα και ο θόρυβος, καθώς και προτάσεις για μελλοντικές βελτιώσεις.

## 1.2 Ερευνητικά Ερωτήματα

Στο πλαίσιο της παρούσας μελέτης, διατυπώνονται τα ακόλουθα βασικά ερευνητικά ερωτήματα:

- RQ1: Πόσο αποτελεσματικά είναι τα Συνελικτικά Νευρωνικά Δίκτυα (CNNs) και τα Δίκτυα Μακροπρόθεσμης Βραχυπρόθεσμης Μνήμης (LSTMs) στην ταξινόμηση δραστηριοτήτων που προέρχονται από δεδομένα φορετών αισθητήρων σε πραγματικές συνθήκες;
- RQ2: Μπορεί ένα υβριδικό μοντέλο, που συνδυάζει τεχνικές βαθιάς μάθησης, να υπερτερήσει σε απόδοση σε σχέση με κλασικά μοντέλα HAR;
- RQ3: Ποιες είναι οι κύριες προκλήσεις που ανακύπτουν κατά την αναγνώριση ανθρώπινων δραστηριοτήτων μέσω φορετών συσκευών σε δεδομένα πραγματικού κόσμου;
- RQ4: Πώς μπορούν να βελτιωθούν τα μοντέλα HAR ώστε να αντιμετωπίζουν ζητήματα όπως ο θόρυβος των αισθητήρων και η ετερογένεια των χρηστών;

Για την καλύτερη κατανόηση του επιστημονικού πλαισίου στο οποίο εντάσσεται η παρούσα μελέτη, ακολουθεί στο επόμενο κεφάλαιο ανασκόπηση της υπάρχουσας βιβλιογραφίας γύρω από την Αναγνώριση Ανθρώπινης Δραστηριότητας (HAR), τις τεχνικές μηχανικής και βαθιάς μάθησης που έχουν εφαρμοστεί στον τομέα αυτό, καθώς και τις βασικές προκλήσεις που εξακολουθούν να υφίστανται.

## 2 Θεωρητικό Υπόβαθρο

Η παρούσα ενότητα παρουσιάζει την υφιστάμενη επιστημονική γνώση και ερευνητική δραστηριότητα στον τομέα της Αναγνώρισης Ανθρώπινης Δραστηριότητας (Human Activity Recognition – HAR), δίνοντας έμφαση στις τεχνικές που βασίζονται σε δεδομένα από φορητές συσκευές. Αρχικά, παρουσιάζονται βασικές έννοιες του πεδίου της HAR και οι κύριες προσεγγίσεις που έχουν υιοθετηθεί διαχρονικά, όπως η αναγνώριση με χρήση αισθητήρων (sensor-based) και μέσω οπτικών δεδομένων (vision-based). Στη συνέχεια, αναλύονται μοντέλα μηχανικής μάθησης που έχουν χρησιμοποιηθεί για την ταξινόμηση ανθρώπινων δραστηριοτήτων, καθώς και σύγχρονες προσεγγίσεις βαθιάς μάθησης, όπως τα CNNs και τα LSTMs, τα οποία έχουν βελτιώσει σημαντικά την ακρίβεια αναγνώρισης.

Ακολουθεί μια επισκόπηση των κυριότερων προκλήσεων που έχουν εντοπιστεί στη βιβλιογραφία, όπως η μεταβλητότητα των δεδομένων μεταξύ χρηστών, ο θόρυβος των αισθητήρων και τα ηθικά ζητήματα που σχετίζονται με την ιδιωτικότητα των δεδομένων. Τέλος, παρουσιάζονται και συγκρίνονται τα σημαντικότερα διαθέσιμα σύνολα δεδομένων που χρησιμοποιούνται σε έρευνες HAR, με ιδιαίτερη έμφαση στο ExtraSensory Dataset [10], το οποίο αξιοποιείται στην παρούσα μελέτη.

### 2.1 Εισαγωγή στην Αναγνώριση Ανθρώπινης Δραστηριότητας (HAR)

Η Αναγνώριση Ανθρώπινης Δραστηριότητας (Human Activity Recognition – HAR) αναφέρεται στη διαδικασία αυτόματης ανίχνευσης, ταξινόμησης ή πρόβλεψης της σωματικής ή συμπεριφορικής κατάστασης ενός ατόμου με βάση δεδομένα που προέρχονται από αισθητήρες ή άλλες καταγραφικές πηγές [11]. Οι δραστηριότητες μπορεί να είναι χαμηλού επιπέδου (π.χ. περπάτημα, κάθισμα, τρέξιμο) ή υψηλότερης πολυπλοκότητας (π.χ. φαγητό, εργασία, κοινωνική συναναστροφή).

Στόχος της HAR είναι η εξαγωγή νοήματος από ακατέργαστα δεδομένα χρόνου και η αντιστοίχισή τους σε καταστάσεις ή πρότυπα συμπεριφοράς. Το πεδίο αυτό συνδυάζει γνώσεις από την επιστήμη υπολογιστών, τη μηχανική μάθηση, τη φυσική δραστηριότητα και την υγειονομική παρακολούθηση.

Η τυπική ροή σε ένα HAR σύστημα περιλαμβάνει τα εξής στάδια:

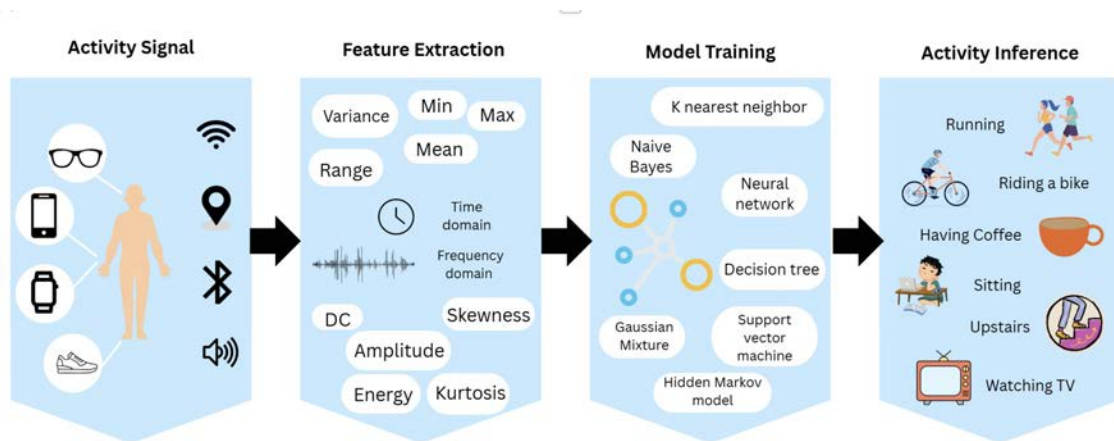
1. Καταγραφή δεδομένων από αισθητήρες (π.χ. επιταχυνσιόμετρο, γυροσκόπιο, μικρόφωνο),
2. Επεξεργασία και εξαγωγή χαρακτηριστικών (feature extraction),

3. Μοντελοποίηση και ταξινόμηση των δραστηριοτήτων μέσω αλγορίθμων μηχανικής μάθησης ή βαθιάς μάθησης,
4. Αξιολόγηση απόδοσης με χρήση κατάλληλων μετρικών (π.χ. ακρίβεια, F1-Score, Balanced Accuracy).

Η HAR βρίσκει εφαρμογή σε ένα ευρύ φάσμα περιοχών, όπως:

- έξυπνα συστήματα υγείας και απομακρυσμένη παρακολούθηση ασθενών,
- έξυπνα σπίτια και περιβάλλοντα διαβίωσης (ambient intelligence),
- ανάλυση συμπεριφοράς για αθλητική προπόνηση ή εργονομία,
- φορετές τεχνολογίες (wearables) και προσωποποιημένες ειδοποιήσεις.

Η παρούσα μελέτη εστιάζει στη HAR με χρήση δεδομένων από φορετές συσκευές σε πραγματικές συνθήκες (in-the-wild), αντικατοπτρίζοντας πιο ρεαλιστικά την ανθρώπινη συμπεριφορά αλλά συνοδεύεται και από σημαντικές προκλήσεις. Μια απεικόνιση για τα βήματα που ακολουθούνται κατά τη μελέτη της HAR για μια πρώτη κατανόηση, φαίνεται στο Σχήμα 2.1.



Σχήμα 2.1: Διάγραμμα Αναγνώρισης Ανθρώπινης Δραστηριότητας (πηγή: ίδια φωτογραφία)

### 2.1.1 Τύποι Συλλογής Δεδομένων στην HAR

Στην έρευνα γύρω από την Αναγνώριση Ανθρώπινης Δραστηριότητας (HAR), ο τρόπος με τον οποίο συλλέγονται τα δεδομένα διαδραματίζει καθοριστικό ρόλο στην ποιότητα των μοντέλων, την αξιοπιστία των συμπερασμάτων και τη δυνατότητα γενίκευσης σε πραγματικές συνθήκες. Η βιβλιογραφία διακρίνει τρεις βασικούς τύπους συλλογής δεδομένων: εργαστηριακή (lab-based), δομημένη ή ημι-δομημένη (scripted ή semi-controlled) και αδόμητη σε πραγματικές συνθήκες (in-the-wild) [12], [13].

Η εργαστηριακή προσέγγιση περιλαμβάνει τη συλλογή δεδομένων σε αυστηρά ελεγχόμενο περιβάλλον, όπου οι συμμετέχοντες καλούνται να εκτελέσουν προκαθορισμένες δραστηριότητες (π.χ. περπάτημα, κάθισμα, τρέξιμο), υπό την καθοδήγηση των ερευνητών. Αυτή η μέθοδος διασφαλίζει υψηλή ποιότητα επισημειωμένων δεδομένων (labels), αλλά περιορίζει την ποικιλία των κινητικών μοτίβων και συχνά παράγει δεδομένα που δεν αντανακλούν την πολυπλοκότητα της καθημερινής ζωής.

Η δομημένη ή ημι-δομημένη συλλογή αντιπροσωπεύει μια ενδιάμεση προσέγγιση. Οι συμμετέχοντες εκτελούν δραστηριότητες είτε μερικώς καθοδηγούμενες είτε εντός ελεγχόμενου περιβάλλοντος, διατηρώντας έναν βαθμό ελευθερίας. Αν και τα δεδομένα που προκύπτουν είναι πιο φυσικά, η εν λόγω μέθοδος εξακολουθεί να μην αποτυπώνει πλήρως τις ιδιαιτερότητες της καθημερινής συμπεριφοράς.

Η πλέον ρεαλιστική προσέγγιση είναι η συλλογή δεδομένων in-the-wild, όπου οι συμμετέχοντες χρησιμοποιούν τις συσκευές τους χωρίς περιορισμούς ή προγραμματισμένα σεναρία. Σε αυτό το πλαίσιο, οι δραστηριότητες καταγράφονται σε πραγματικό χρόνο και περιλαμβάνουν φυσική κίνηση, κοινωνική αλληλεπίδραση, μεταβαλλόμενα περιβάλλοντα και ανομοιογενή σήματα. Το ExtraSensory Dataset αποτελεί χαρακτηριστικό παράδειγμα τέτοιου τύπου συλλογής, προσφέροντας δεδομένα υψηλής ποικιλότητας και οικολογικής εγκυρότητας. Μπορεί κανείς να πει λοιπόν, ότι η HAR δεν χρειάζεται ακριβώς εργαστηριακά συστήματα, αλλά μπορεί να δουλέψει με δεδομένα από απλούς αισθητήρες από smartphone [14] ή smartwatch όπως αυτά που προσφέρονται στο ExtraSensory Dataset.

Παρότι η συλλογή δεδομένων in-the-wild ενισχύει την εφαρμοσιμότητα των μοντέλων σε πραγματικά σεναρία, ενέχει προκλήσεις, όπως η αβεβαιότητα στις ετικέτες λόγω αυτοαναφοράς, η εμφάνιση θορύβου και η ετερογένεια στις συσκευές και στους χρήστες. Ωστόσο, θεωρείται σήμερα ως η πλέον κατάλληλη στρατηγική για την εκπαίδευση μοντέλων που προορίζονται για πρακτική χρήση και γενίκευση.

## 2.2 Αναγνώριση Προτύπων και Μηχανική Μάθηση

Η αναγνώριση προτύπων (pattern recognition) αποσκοπεί στη δημιουργία αλγορίθμων που ανακαλύπτουν αυτόματα μοτίβα και κανονικότητες μέσα σε δεδομένα. Στόχος των αλγορίθμων αυτών είναι η εκτέλεση ευφυών και ανθρωπόμορφων εργασιών, όπως ο εντοπισμός ενός αντικειμένου σε μια εικόνα, η αναγνώριση χειρόγραφου ψηφίου ή η αναγνώριση ανθρώπινης δραστηριότητας σε βίντεο – αντικείμενο που εξετάζεται και στην παρούσα εργασία.

Η εμφάνιση της Μηχανικής Μάθησης (Machine Learning) συνέβαλε στη βελτίωση της διαδικασίας ανάπτυξης αλγορίθμων αναγνώρισης προτύπων. Κατά συνέπεια, οι τεχνικές μηχανικής μάθησης έχουν υιοθετηθεί σε πολλές σύγχρονες προσεγγίσεις για την επίλυση προβλημάτων αναγνώρισης προτύπων. Στη βασική τους μορφή, οι αλγόριθμοι αυτοί επεξεργάζονται ένα σύνολο δεδομένων και δημιουργούν ένα πιθανοτικό μοντέλο, με σκοπό να αποφασίζουν το πιθανότερο αποτέλεσμα με ελάχιστη ανθρώπινη παρέμβαση [11].

Για παράδειγμα, στο πρόβλημα της αναγνώρισης ανθρώπινης δραστηριότητας από δεδομένα αισθητήρων, ένας αλγόριθμος μηχανικής μάθησης λειτουργεί ως εξής: του παρέχεται

ένα εκτεταμένο σύνολο δεδομένων που έχει συλλεχθεί από αισθητήρες τοποθετημένους σε κινητές ή φορετές συσκευές (π.χ. επιταχυνσιόμετρα, γυροσκόπια, αισθητήρες μαγνητόμετρου). Το σύνολο αυτό – γνωστό ως σύνολο εκπαίδευσης (training set) – περιλαμβάνει χρονικές ακολουθίες μετρήσεων, που έχουν καταγραφεί καθώς ο χρήστης εκτελεί διάφορες δραστηριότητες, όπως βάδισμα, καθιστή στάση, ανεβοκατέβασμα σκάλας κ.λπ.

Κάθε δείγμα δεδομένων αναπαρίσταται ως πολυδιάστατο διάνυσμα που περιέχει τις τιμές των μετρήσεων από όλους τους διαθέσιμους αισθητήρες, σε διακριτές χρονικές στιγμές (π.χ. ένα παράθυρο 5 δευτερολέπτων με δεδομένα 50Hz παράγει 250 χρονικά βήματα για κάθε κανάλι εισόδου). Οι δραστηριότητες που εκτελούνται σε κάθε χρονική περίοδο είναι προκαθορισμένες μέσω επισημείωσης (labelling), η οποία μπορεί να έχει γίνει είτε αυτόματα είτε από τον ίδιο τον συμμετέχοντα.

Ο αλγόριθμος επεξεργάζεται το σύνολο εκπαίδευσης ώστε να μάθει χαρακτηριστικά μοτίβα στις μετρήσεις, τα οποία σχετίζονται με διαφορετικούς τύπους δραστηριότητας. Η διαδικασία αυτή ονομάζεται φάση εκπαίδευσης (training phase), κατά την οποία το μοντέλο προσπαθεί να ελαχιστοποιήσει το σφάλμα μεταξύ των προβλέψεών του και των πραγματικών ετικετών, διαμορφώνοντας μια λειτουργική απεικόνιση μεταξύ εισόδου και εξόδου.

Μόλις ολοκληρωθεί η φάση εκπαίδευσης, ο αλγόριθμος μπορεί να εφαρμοστεί για την αναγνώριση δραστηριότητας σε νέα, άγνωστα δεδομένα αισθητήρων (δεδομένα ελέγχου – test data). Λαμβάνει ως είσοδο μια χρονική ακολουθία μετρήσεων από αισθητήρες και παράγει ως έξοδο είτε μια πιθανότητα ανά κατηγορία είτε μια κατηγορία δραστηριότητας, η οποία θεωρείται ως η πλέον πιθανή βάσει των τιμών εισόδου. Η έξοδος αυτή αφορά την εκτίμηση της δραστηριότητας που εκτελείται κατά τη διάρκεια του εν λόγω παραθύρου δεδομένων.

Η ικανότητα του μοντέλου να ταξινομεί σωστά νέα παραδείγματα – τα οποία προέρχονται από χρήστες, χρονικές στιγμές ή συνθήκες που δεν είχαν εμφανιστεί κατά την εκπαίδευση – αποτελεί ένδειξη της ικανότητάς του για γενίκευση (generalization). Η γενίκευση είναι κρίσιμη σε εφαρμογές HAR, καθώς τα δεδομένα αισθητήρων παρουσιάζουν υψηλή ετερογένεια λόγω διαφορετικών κινήσεων, φορέων συσκευής ή περιβαλλοντικών παραμέτρων.

Το παραπάνω παράδειγμα εντάσσεται στην οικογένεια προβλημάτων ταξινόμησης (classification) στη μηχανική μάθηση. Σκοπός κάθε προβλήματος ταξινόμησης είναι να αντιστοιχίσει μια άγνωστη είσοδο σε μία από ένα πεπερασμένο σύνολο διακριτών κατηγοριών.

Άλλοι συνήθεις τύποι αλγορίθμων μηχανικής μάθησης περιλαμβάνουν:

- την παλινδρόμηση (regression), όπου η επιθυμητή έξοδος αποτελείται από μία ή περισσότερες συνεχείς μεταβλητές,
- και την ομαδοποίηση (clustering), όπου τα δεδομένα χωρίζονται σε ομάδες, προσδιορίζοντας τη δομή κατανομής τους στον χώρο εισόδου.

Οι προσεγγίσεις μηχανικής μάθησης διακρίνονται συνήθως σε τρεις βασικές κατηγορίες [15], ανάλογα με τη φάση εκπαίδευσης:

- Εποπτευόμενη μάθηση (Supervised learning): Το σύνολο εκπαίδευσης αποτελείται από παραδείγματα μαζί με τις αντίστοιχες ετικέτες τους. Τυπικές εφαρμογές της εποπτευόμενης μάθησης είναι η ταξινόμηση και η παλινδρόμηση.

- Μη εποπτευόμενη μάθηση (Unsupervised learning): Το σύνολο εκπαίδευσης περιλαμβάνει μόνο δεδομένα εισόδου χωρίς αντίστοιχες εξόδους. Η μη εποπτευόμενη μάθηση χρησιμοποιείται κυρίως για προβλήματα ομαδοποίησης, οπτικοποίησης και μείωσης διαστάσεων.
- Ενισχυτική μάθηση (Reinforcement learning): Ο αλγόριθμος δεν λαμβάνει άμεσα παραδείγματα με τις σωστές εξόδους, αλλά πρέπει να τις ανακαλύψει μέσω μιας διαδικασίας δοκιμής και σφάλματος, με στόχο τη μεγιστοποίηση μιας επιβράβευσης που προκύπτει από την αλληλεπίδρασή του με το περιβάλλον. Για παράδειγμα, μπορεί να χρησιμοποιηθεί ενισχυτική μάθηση για την εκπαίδευση ενός νευρωνικού δικτύου που παίζει το παιχνίδι του τάβλι σε υψηλό επίπεδο.

## 2.3 Νευρωνικά Δίκτυα

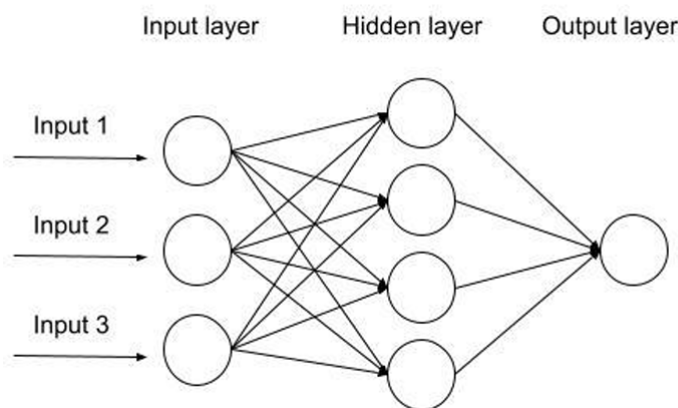
Τα νευρωνικά δίκτυα αποτελούν μια κατηγορία αλγορίθμων της μηχανικής μάθησης, η οποία στοχεύει στην αναγνώριση υποκείμενων σχέσεων μέσα σε ένα σύνολο δεδομένων. Αυτό πρακτικά σημαίνει ότι αναζητούνται αριθμητικά μοτίβα σε δεδομένα όπως εικόνες, ήχος, κείμενο κ.α., τα οποία έχουν αναπαρασταθεί ως διανύσματα. Ο όρος νευρωνικό δίκτυο είναι εμπνευσμένος από τους βιολογικούς νευρώνες του ανθρώπινου εγκεφάλου. Τα τεχνητά νευρωνικά δίκτυα προσομοιώνουν τη νευρωνική και ηλεκτρική δραστηριότητα του εγκεφάλου και του νευρικού συστήματος. Με παρόμοιο τρόπο, οι τεχνητοί νευρώνες μπορούν να λαμβάνουν, να επεξεργάζονται και να μεταδίδουν σήματα προς γειτονικούς νευρώνες.

Κάθε νευρωνικό δίκτυο αποτελείται από νευρώνες και τις μεταξύ τους συνδέσεις (edges). Οι συνδέσεις αυτές χαρακτηρίζονται από βάρη (weights), τα οποία προσδιορίζουν τη «δύναμη» της σύνδεσης μεταξύ δύο νευρώνων. Οι νευρώνες οργανώνονται συνήθως σε επίπεδα (layers): το επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα (hidden layers), και το επίπεδο εξόδου. Η έξοδος ενός επιπέδου λειτουργεί ως είσοδος για το επόμενο. Ένας νευρώνας μπορεί να συνδέεται είτε με όλους είτε με ένα υποσύνολο των νευρώνων του επόμενου επιπέδου. Το Σχήμα 2.2 παρουσιάζει ένα απλό παράδειγμα νευρωνικού δικτύου.

Κάθε νευρώνας υλοποιεί στην πραγματικότητα μία μαθηματική συνάρτηση, της οποίας η έξοδος καθορίζει την επιρροή του νευρώνα στην έξοδο του επιπέδου. Συνήθως, η συνάρτηση αυτή είναι ένας σταθμισμένος μέσος όρος των εισόδων του, δηλαδή:

$$f(x_1, x_2, \dots, x_n) = w_1x_1 + w_2x_2 + \dots = w_nx_n \quad (2.1)$$

όπου τα  $x_i$  αποτελούν τις εισόδους και τα  $w_i$  τα αντίστοιχα βάρη. Οι είσοδοι μπορεί να είναι είτε τα αρχικά δεδομένα εισόδου είτε οι εξόδοι νευρώνων προηγούμενων επιπέδων [16].



Σχήμα 2.2: Απλό Νευρωνικό Δίκτυο (πηγή: M. A. Nielsen (2015) [16])

Όπως αναφέρθηκε και στην Ενότητα 2.2, η εκπαίδευση ενός αλγορίθμου μηχανικής μάθησης μπορεί να είναι εποπτευόμενη, μη εποπτευόμενη ή ενισχυτική. Στο πλαίσιο της παρούσας μελέτης, η αναγνώριση ανθρώπινης δραστηριότητας προσεγγίζεται ως ένα πρόβλημα εποπτευόμενης ταξινόμησης, και ως εκ τούτου, το υπόλοιπο της ενότητας επικεντρώνεται στην εποπτευόμενη εκπαίδευση.

Στην εποπτευόμενη εκπαίδευση, τα νευρωνικά δίκτυα εκπαιδεύονται με τη χρήση επισημειωμένων δεδομένων (labeled data). Κατά τη διάρκεια της εκπαίδευσης, το δίκτυο επιχειρεί επανειλημμένα να προβλέψει τις σωστές ετικέτες για τα δεδομένα εισόδου και σχηματίζει συσχετίσεις βασισμένες σε πιθανότητες ανάμεσα στις προβλέψεις και τις πραγματικές ετικέτες. Μέσω αυτών των συσχετίσεων προσδιορίζονται τα διακριτικά χαρακτηριστικά κάθε κατηγορίας.

Οι συσχετίσεις αυτές ενσωματώνονται στο δίκτυο μέσω της τροποποίησης των βαρών. Κάθε βάρος προσαρμόζεται ανάλογα με το υπολογιζόμενο σφάλμα, το οποίο ορίζεται ως η διαφορά μεταξύ της προβλεπόμενης εξόδου του δικτύου και της πραγματικής επιθυμητής τιμής. Το δίκτυο ενημερώνει τα βάρη του λαμβάνοντας υπόψη αυτό το σφάλμα μέσω μιας συγκεκριμένης διαδικασίας που ονομάζεται backpropagation.

Σε εννοιολογικό επίπεδο, το backpropagation χρησιμοποιείται για να διορθώσει και να προσαρμόσει τις εσωτερικές αναπαραστάσεις που οδηγούν στον τελικό συσχετισμό εισόδου–εξόδου. Το συνθετικό “back” υποδηλώνει ότι οι υπολογισμοί πραγματοποιούνται από τα ανώτερα επίπεδα προς τα κατώτερα, μεταφέροντας το υπολογισμένο σφάλμα από την έξοδο προς την είσοδο. Ο στόχος είναι η προσαρμογή των βαρών έτσι ώστε η διαφορά μεταξύ της πραγματικής και της επιθυμητής εξόδου να ελαχιστοποιηθεί ή να φτάσει σε ένα αποδεκτό όριο σφάλματος.

Σε τεχνικό επίπεδο, το backpropagation υπολογίζει τη βαθμίδα (gradient) της συνάρτησης απώλειας ως προς τα βάρη του δικτύου, για κάθε δείγμα εισόδου–εξόδου. Η πιο διαδεδομένη προσέγγιση είναι η στοχαστική καθοδική πορεία βαθμίδας (stochastic gradient descent – SGD), η οποία προσαρμόζει σταδιακά τα βάρη με βάση μεμονωμένα ή μικρές ομάδες δειγμάτων [17].

Ένα από τα πιο σημαντικά πλεονεκτήματα των νευρωνικών δικτύων είναι η προσαρμοστικότητα τους σε μεταβαλλόμενα δεδομένα εισόδου, καθώς και η ικανότητά τους να μαθαίνουν χαρακτηριστικά που ταυτοποιούν τις δομές των δεδομένων χωρίς να απαιτείται προηγούμενη γνώση ή ανθρώπινη παρέμβαση. Ως εκ τούτου, το προ-επεξεργαστικό στάδιο των δεδομένων καθίσταται σημαντικά απλούστερο, μειώνοντας σημαντικά την ανάγκη για χειροκίνητο σχεδιασμό χαρακτηριστικών (feature engineering).

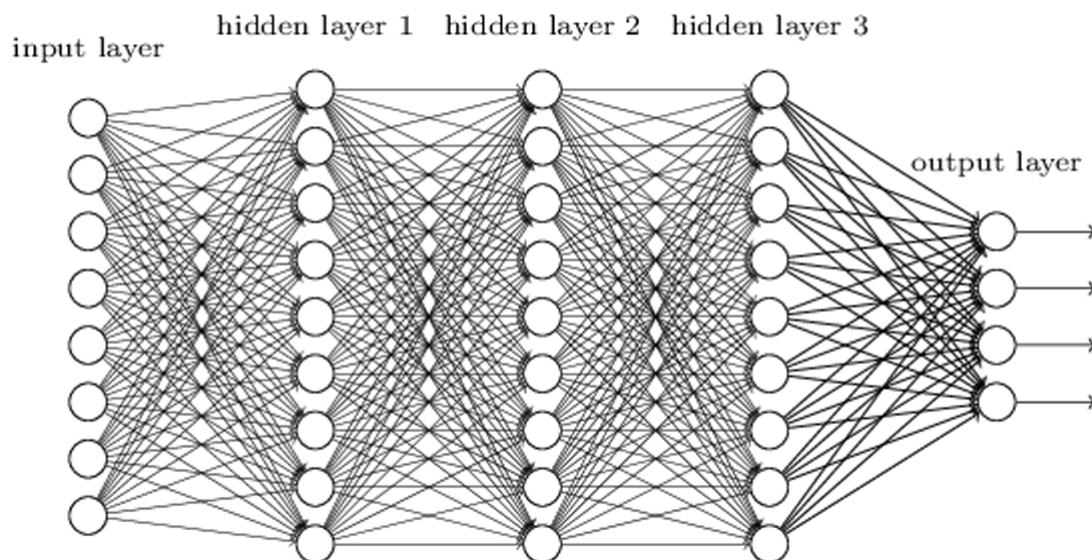
## 2.4 Βαθιά Μάθηση

Η βαθιά μάθηση (Deep Learning) αποτελεί υποκατηγορία της μηχανικής μάθησης και βασίζεται στην εκμάθηση αναπαραστάσεων (representation learning) μέσω πολυεπίπεδων νευρωνικών δικτύων. Σκοπός της εκμάθησης αναπαραστάσεων είναι ο μετασχηματισμός των αρχικών δεδομένων σε μια νέα, πιο αποδοτική μορφή, η οποία διατηρεί τα ουσιώδη χαρακτηριστικά των αντικειμένων ενδιαφέροντος και απορρίπτει περιττές πληροφορίες. Το δίκτυο μαθαίνει σταδιακά όλο και πιο σημασιολογικά πλούσιες αναπαραστάσεις των δεδομένων. Αρχικά, οι αναπαριστώμενες έννοιες είναι αφηρημένες, αλλά προοδευτικά εξελίσσονται σε πιο συγκεκριμένες.

Για παράδειγμα, στο πεδίο της αναγνώρισης ανθρώπινης δραστηριότητας από δεδομένα αισθητήρων, τα πρώτα επίπεδα ενός βαθιού νευρωνικού δικτύου τείνουν να μαθαίνουν χαμηλού επιπέδου χρονικά χαρακτηριστικά, όπως διακυμάνσεις της επιτάχυνσης ή βασικούς ρυθμούς κίνησης, οι οποίοι σχετίζονται με απλά μοτίβα δραστηριότητας. Καθώς το δίκτυο γίνεται βαθύτερο, τα επόμενα επίπεδα είναι ικανά να συνδυάζουν και να ερμηνεύουν αυτά τα σήματα με τρόπο που επιτρέπει τη διάκριση πιο σύνθετων κινητικών μοτίβων, όπως η εναλλαγή στάσης (π.χ. από καθιστή σε όρθια), η ρυθμική κίνηση που υποδηλώνει βάδισμα ή τρέξιμο, ή ακόμα και η αναγνώριση συμφραζομένων δραστηριοτήτων (π.χ. περπάτημα ενώ κάποιος μιλά στο τηλέφωνο).

Τα νευρωνικά δίκτυα αποτελούν τη βασική υποδομή (backbone) των αλγορίθμων βαθιάς μάθησης. Ουσιαστικά, η βαθιά μάθηση χρησιμοποιεί νευρωνικά δίκτυα για την εκμάθηση αναπαραστάσεων των δεδομένων σε πολλαπλά επίπεδα αφαίρεσης. Οι αναπαραστάσεις αυτές μαθαίνονται (σχεδόν πάντοτε) μέσω δικτύων με διαδοχικά επίπεδα (literal layers), τα οποία στοιβάζονται το ένα πάνω στο άλλο. Ο αριθμός των επιπέδων (layers) είναι αυτός που διαφοροποιεί ένα απλό νευρωνικό δίκτυο από έναν αλγόριθμο βαθιάς μάθησης: ένα νευρωνικό δίκτυο που αποτελείται από περισσότερα από τρία επίπεδα θεωρείται δίκτυο βαθιάς μάθησης (deep neural network). Το «βάθος» ενός τέτοιου δικτύου ορίζεται από το πλήθος των επιπέδων που συμβάλλουν ενεργά στη διαδικασία μετασχηματισμού των δεδομένων.

Ένα παράδειγμα βαθιού νευρωνικού δικτύου παρουσιάζεται στο Σχήμα 2.3.



Σχήμα 2.3: Βαθύ Νευρωνικό Δίκτυο Αρχιτεκτονική (πηγή: M. A. Nielsen (2015) [16])

Οι αρχιτεκτονικές βαθιάς μάθησης μαθαίνουν αναπαραστάσεις απευθείας από τα δεδομένα, χωρίς να απαιτούν χειροκίνητα σχεδιασμένα χαρακτηριστικά (hand-crafted features). Αυτή η ανεξαρτησία από την ανθρώπινη προσπάθεια στον σχεδιασμό των χαρακτηριστικών συνιστά ένα από τα μεγαλύτερα πλεονεκτήματα των βαθιών αρχιτεκτονικών έναντι των παραδοσιακών μεθόδων μηχανικής μάθησης.

Η βαθιά μάθηση προσφέρει υποσχόμενες λύσεις σε πλήθος προβλημάτων της επιστήμης των υπολογιστών, όπως η αναγνώριση δραστηριότητας (action recognition), η ανίχνευση αντικειμένων (object detection), η ταξινόμηση εικόνων (image classification), η επεξεργασία φυσικής γλώσσας (natural language processing) και η αναγνώριση φωνής (speech recognition) [18].

## 2.5 Συνελικτικά Νευρωνικά Δίκτυα (CNNs)

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks – CNNs) αποτελούν μια κατηγορία μοντέλων βαθιάς μάθησης που έχουν εφαρμοστεί με ιδιαίτερη επιτυχία σε προβλήματα ταξινόμησης δεδομένων υψηλής διαστασιμότητας, όπως χρονικές ακολουθίες από αισθητήρες κινητών και φορητών συσκευών. Αν και παραδοσιακά χρησιμοποιούνται σε εφαρμογές υπολογιστικής όρασης, τα CNNs έχουν επίσης αποδειχθεί κατάλληλα για την επεξεργασία μονοδιάστατων ή πολυκαναλικών χρονικών σημάτων, όπως εκείνα που παράγονται από επιταχυνσιόμετρα, γυροσκόπια και άλλους βιοαισθητήρες. Εφαρμογές CNNs στην αναγνώριση ανθρώπινης δραστηριότητας που έχουν ήδη διεξαχθεί έχουν αναδείξει την ικανότητά τους να εξάγουν αποδοτικά χρονικά χαρακτηριστικά [19]. Τα δίκτυα αυτά βασίζονται σε μια μαθηματική πράξη που ονομάζεται συνέλιξη (convolution), η οποία τους επιτρέπει να εξάγουν αποδοτικά χαρακτηριστικά από τοπικές περιοχές της εισόδου.

Όπως αναφέρθηκε στην Ενότητα 2.3, τα νευρωνικά δίκτυα αποτελούνται από νευρώνες με εκπαιδευσιμα βάρη. Αρχικά, δέχονται δεδομένα εισόδου, τα οποία επεξεργάζονται μέσω ενός αριθμού κρυφών επιπέδων (hidden layers). Κάθε κρυφό επίπεδο περιλαμβάνει ένα σύνολο νευρώνων, όπου κάθε νευρώνας συνδέεται πλήρως με όλους τους νευρώνες του προηγούμενου και του επόμενου επιπέδου.

Σε αντίθεση με τα παραδοσιακά νευρωνικά δίκτυα, τα CNNs παρουσιάζουν τοπική συνδεσιμότητα (local connectivity). Αυτό σημαίνει ότι, για παράδειγμα, στο πλαίσιο επεξεργασίας χρονικών σημάτων από αισθητήρες, το δίκτυο μπορεί να υποθέσει ότι η συσχέτιση μεταξύ διαδοχικών χρονικών δειγμάτων είναι πιθανώς πιο σημαντική από τη συσχέτιση μεταξύ χρονικά απομακρυσμένων σημείων της ακολουθίας. Ως εκ τούτου, οι συνδέσεις που σχετίζονται με χρονικά απομακρυσμένα δεδομένα μπορούν να παραληφθούν, περιορίζοντας έτσι την πολυπλοκότητα του μοντέλου χωρίς απώλεια κρίσιμης πληροφορίας.

Σε ένα CNN, κάθε νευρώνας «βλέπει» μόνο ένα τοπικό χρονικό παράθυρο της εισόδου, δηλαδή μια υποακολουθία συγκεκριμένου μήκους. Το μήκος αυτής της υποακολουθίας αποτελεί υπερπαραμέτρο (hyperparameter) και ονομάζεται πεδίο υποδοχής (receptive field) του νευρώνα. Τεχνικά, η τοπική συνδεσιμότητα επιτυγχάνεται με περιορισμό των συνδέσεων μεταξύ διαδοχικών επιπέδων, διατηρώντας μόνο αυτές που μεταφέρουν την πιο χρήσιμη πληροφορία. Έτσι, οι νευρώνες ενός επιπέδου συνδέονται αποκλειστικά με νευρώνες του επόμενου επιπέδου που αντιστοιχούν σε χρονικά γειτονικά σημεία του αρχικού σήματος, εντός του ίδιου πεδίου υποδοχής.

Αυτή η διαφορετική προσέγγιση επιτυγχάνει ταυτόχρονα δύο στόχους: μειώνει την υπολογιστική πολυπλοκότητα και καθιστά δυνατή την αποδοτική επεξεργασία δεδομένων υψηλής διάστασης, όπως οι χρονικές ακολουθίες από αισθητήρες ή οι εικόνες.

Επιπλέον, τα CNNs εισάγουν την έννοια των κοινών βαρών (shared weights) [20], επιβάλλοντας στους νευρώνες ενός επιπέδου που ανήκουν στο ίδιο πεδίο υποδοχής να μοιράζονται τα ίδια βάρη. Αυτό πρακτικά σημαίνει ότι το ίδιο φίλτρο (ανιχνευτής χαρακτηριστικού – feature detector) εφαρμόζεται σε πολλά σημεία της χρονικής ακολουθίας εισόδου. Για παράδειγμα, αν ένα συγκεκριμένο φίλτρο ανιχνεύει μια χαρακτηριστική διακύμανση σε κάποιο σημείο του σήματος – όπως μια απότομη μεταβολή στην επιτάχυνση ή μια περιοδικότητα – είναι πιθανό αυτό το μοτίβο να είναι εξίσου σημαντικό και σε άλλα χρονικά σημεία του ίδιου σήματος ή ακόμη και σε άλλα δείγματα.

Ως αποτέλεσμα, το δίκτυο είναι σε θέση, κατά τη διάρκεια της εκπαίδευσης, να μάθει γενικεύσιμα και επαναλαμβανόμενα χρονικά χαρακτηριστικά, ενώ ταυτόχρονα αποκτά την ικανότητα να αναγνωρίζει μοτίβα που εμφανίζονται σε διαφορετικές χρονικές θέσεις ή παρουσιάζουν μικρές παραλλαγές. Με άλλα λόγια, τα CNNs επιτυγχάνουν αυτό που ονομάζεται χρονική μεταθετική αμεταβλητότητα (temporal translational invariance).

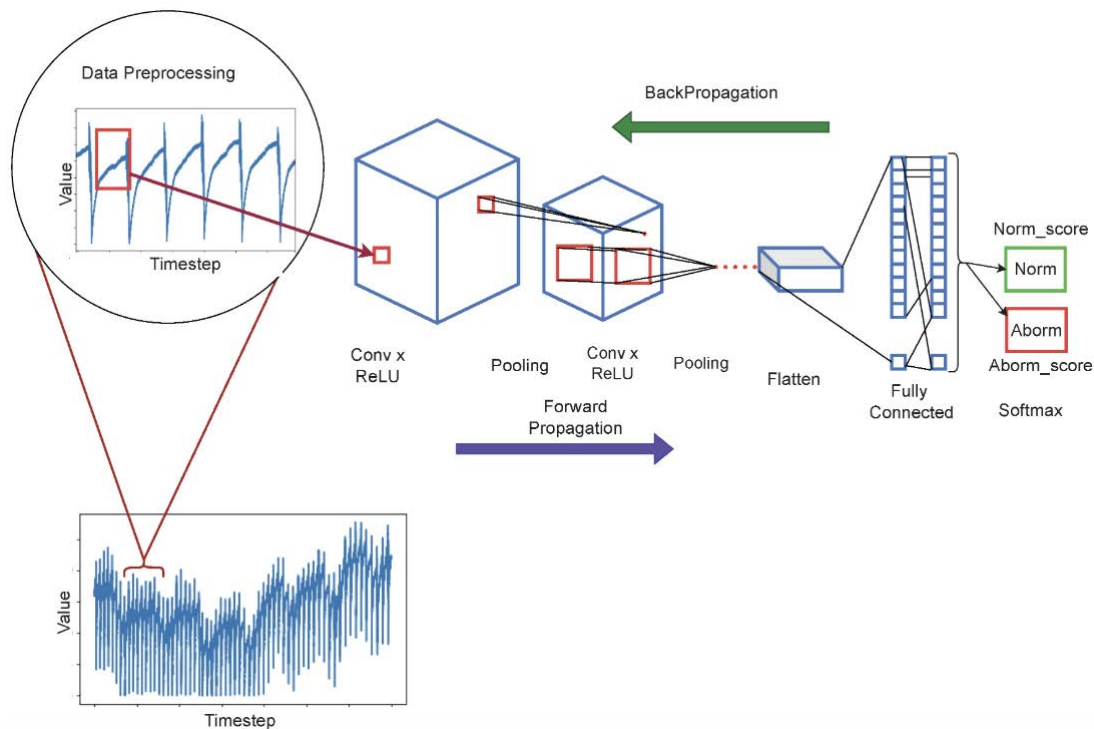
Η αρχιτεκτονική ενός CNN περιλαμβάνει ένα επίπεδο εισόδου, μια σειρά ενδιάμεσων (κρυφών) επιπέδων και ένα επίπεδο εξόδου. Η πράξη της συνέλιξης εφαρμόζεται σε κάθε επίπεδο μέσω της μετακίνησης ενός φίλτρου (kernel) πάνω από τη χρονική ακολουθία εισόδου – μπορούμε να το φανταστούμε ως ένα «παράθυρο» σταθερού μήκους που μετακινείται κατά μήκος του σήματος. Παραδοσιακά, τα πρώτα συνελκτικά επίπεδα μαθαίνουν χαμηλού επιπέδου χρονικά μοτίβα, όπως μεταβολές σήματος ή παλμούς, ενώ στα βαθύτερα επίπεδα το δίκτυο μαθαίνει πιο αφηρημένα και συνδυαστικά χαρακτηριστικά, έως ότου σχηματίσει

μια ολοκληρωμένη κατανόηση της κινητικής συμπεριφοράς.

Συνήθως, ένα CNN περιλαμβάνει τις εξής βασικές κατηγορίες επιπέδων:

- Συνελικτικά επίπεδα (convolutional layers)
- Επίπεδα υποδειγματοληψίας (pooling layers)
- Επίπεδα κανονικοποίησης (normalization layers)
- Πλήρως συνδεδεμένα επίπεδα (fully connected layers)
- Επίπεδα ενεργοποίησης ReLU

Η συνδυαστική χρήση των παραπάνω επιπέδων συνιστά τη δομική βάση ενός τυπικού CNN, όπως απεικονίζεται στο Σχήμα 2.4.



Σχήμα 2.4: Αρχιτεκτονική Τυπικού CNN (πηγή: Nicholas et al., 2022 [21])

### 2.5.1 Συνελικτικό Επίπεδο

Τα συνελικτικά επίπεδα (convolutional layers ή CONV layers) αποτελούν το βασικό δομικό στοιχείο ενός Συνελικτικού Νευρωνικού Δικτύου (CNN). Κάθε συνελικτικό επίπεδο αποτελείται από εκπαιδευσιμους πυρήνες (kernels) ή φίλτρα (filters), τα οποία ορίζονται από τις παραμέτρους του πλάτους, του ύψους και του βάθους. Τυπικά, τα φίλτρα έχουν μικρό μέγεθος στον χώρο (π.χ.  $3 \times 3$  ή  $5 \times 5$ ), αλλά εκτείνονται καθ' όλο το βάθος του όγκου εισόδου. Για παράδειγμα, στην περίπτωση χρονικών δεδομένων από πολλαπλούς αισθητήρες, μπορεί να χρησιμοποιηθεί φίλτρο διαστάσεων  $5 \times 1 \times 6$ , όπου το «5» αντιστοιχεί στο μήκος του χρονικού παραθύρου (π.χ. 5 διαδοχικά χρονικά βήματα) και το «6» στο βάθος της

εισόδου, δηλαδή στον αριθμό των διαφορετικών καναλιών αισθητήρων, όπως επιτάχυνση, γυροσκόπιο, προσανατολισμός κ.λπ. [22]

Κατά τη φάση εκπαίδευσης, τα φίλτρα των συνελκτικών επιπέδων εφαρμόζονται με συνέλιξη κατά μήκος των διαστάσεων του πλάτους και του ύψους του όγκου εισόδου, υπολογίζοντας το εσωτερικό γινόμενο μεταξύ των τιμών του φίλτρου και των αντίστοιχων τιμών του εισερχόμενου σήματος. Στην περίπτωση δεδομένων από αισθητήρες, ο όγκος εισόδου μπορεί να αναπαριστά χρονικά παράθυρα (sliding windows) που περιλαμβάνουν μετρήσεις από πολλαπλούς τύπους αισθητήρων (π.χ. επιταχυνσιόμετρο, γυροσκόπιο), οργανωμένους ως ξεχωριστά κανάλια εισόδου. Καθώς το φίλτρο μετακινείται κατά μήκος του παραθύρου εισόδου, παράγει έναν διδιάστατο χάρτη χαρακτηριστικών (feature map), ο οποίος αποτυπώνει χρονικά και διακαναλικά μοτίβα, όπως επαναλαμβανόμενες μεταβολές στην επιτάχυνση ή στην περιστροφή. Παρόμοιοι χάρτες εξάγονται για όλα τα φίλτρα και, με τη στοίβαξή τους στη διάσταση του βάθους, συγκροτείται ο τελικός όγκος εξόδου του συνελκτικού επιπέδου.

Ένα από τα βασικά πλεονεκτήματα των συνελκτικών επιπέδων είναι ότι κάθε νευρώνας συνδέεται μόνο με μια τοπική περιοχή του όγκου εισόδου, η οποία είναι ισοδύναμη με το μέγεθος του φίλτρου. Η προσέγγιση αυτή μειώνει σημαντικά τον αριθμό των ελεύθερων παραμέτρων και επιτρέπει στο δίκτυο να αξιοποιήσει αποτελεσματικά μεγαλύτερο αριθμό επιπέδων, διατηρώντας παράλληλα τη δυνατότητα επεξεργασίας δεδομένων υψηλής διαστασιμότητας.

Τέλος, το μέγεθος της εξόδου ενός συνελκτικού επιπέδου εξαρτάται από τρεις υπερπαραμέτρους (hyperparameters):

- Βάθος (depth): Αντιστοιχεί στον αριθμό των φίλτρων που χρησιμοποιούνται. Κάθε φίλτρο μαθαίνει να ανιχνεύει διαφορετικό χαρακτηριστικό στο σήμα εισόδου.
- Βήμα μετακίνησης (stride): Καθορίζει πόσο προχωρά το φίλτρο κάθε φορά κατά τη συνέλιξη (π.χ. βήμα 1 σημαίνει ότι το φίλτρο μετακινείται κατά 1 θέση κάθε φορά).
- Μηδενική συμπλήρωση (zero-padding): Προσθέτει μηδενικές τιμές περιμετρικά του όγκου εισόδου, ώστε να ελέγχεται το τελικό μέγεθος των εξόδων και να αποφεύγεται η υπερβολική μείωση του χωρικού διαστήματος.

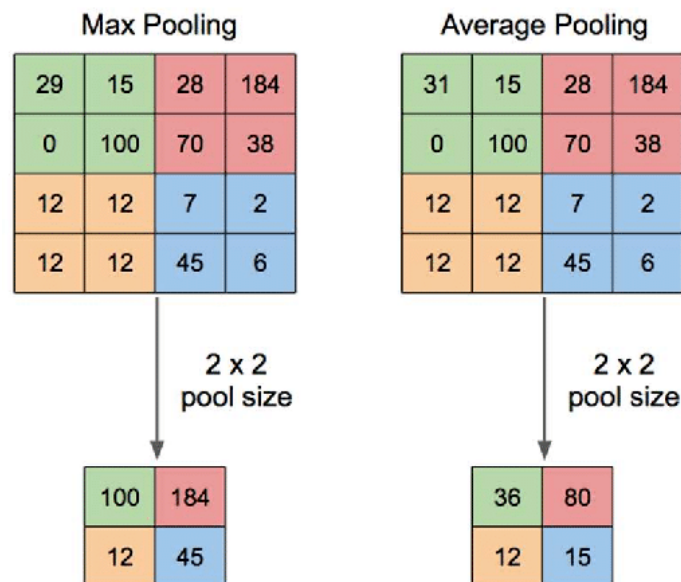
## 2.5.2 Επίπεδο υποδειγματοληψίας

Τα επίπεδα υποδειγματοληψίας (pooling layers) λειτουργούν ως μηχανισμός μείωσης της διαστασιμότητας (dimensionality reduction). Σκοπός τους είναι να υποδειγματοληπτούν (down-sample) τους χάρτες χαρακτηριστικών, συνοψίζοντας την παρουσία χαρακτηριστικών μέσα σε τοπικές περιοχές (patches). Η είσοδος χωρίζεται σε μικρές περιοχές, συνήθως διαστάσεων  $2 \times 2$ , από τις οποίες εξάγεται είτε ο μέγιστος είτε ο μέσος όρος της ενεργοποίησης του χαρακτηριστικού.

Αυτή η προσέγγιση μειώνει σημαντικά τον αριθμό των παραμέτρων και τις υπολογιστικές απαιτήσεις που απαιτούνται για την επεξεργασία των δεδομένων. Συνηθίζεται η περιοδική εισαγωγή επιπέδων υποδειγματοληψίας ανάμεσα σε διαδοχικά συνελκτικά επίπεδα

μέσα στην αρχιτεκτονική ενός CNN, με σκοπό τόσο τη μείωση της πολυπλοκότητας όσο και την αποφυγή υπερεκπαίδευσης (overfitting).

Οι δύο πιο κοινές τεχνικές υποδειγματοληψίας είναι η μέγιστη υποδειγματοληψία (max pooling) και η μέση υποδειγματοληψία (average pooling). Ένα παράδειγμα τους απεικονίζεται στο Σχήμα 2.5. Η μέγιστη υποδειγματοληψία λειτουργεί ταυτόχρονα και ως μηχανισμός καταστολής θορύβου, πέραν της μείωσης διαστάσεων, γεγονός που την καθιστά την πλέον διαδεδομένη επιλογή στην πράξη.



Σχήμα 2.5: Παράδειγμα max και average pooling (πηγή: Bhatti et al., 2023 [23])

### 2.5.3 Επίπεδο Κανονικοποίησης

Τα επίπεδα κανονικοποίησης (normalization layers) αποτελούν τεχνική που διευκολύνει την ταχύτερη και πιο σταθερή εκπαίδευση βαθιών νευρωνικών δικτύων, μέσω της τυποποίησης των εισόδων ενός επιπέδου. Στόχος είναι η επιτάχυνση της εκπαίδευσης και η βελτίωση της αριθμητικής σταθερότητας του μοντέλου.

Η πιο διαδεδομένη τεχνική κανονικοποίησης, η οποία εφαρμόζεται και στην παρούσα μελέτη, είναι η Κανονικοποίηση Κατά Παρτίδες (batch normalization). Η μέθοδος εφαρμόζεται σε δίκτυα όπου τα δεδομένα εκπαίδευσης χωρίζονται σε mini-batches σταθερού μεγέθους. Η Batch Normalization εφαρμόζεται σε όλα τα δείγματα της κάθε παρτίδας και τυποποιεί τις ενεργοποιήσεις των νευρώνων, διατηρώντας τη μέση τιμή κοντά στο 0 και τη τυπική απόκλιση κοντά στο 1.

## 2.5.4 Πλήρως Συνδεδεμένα Επίπεδα

Τα πλήρως συνδεδεμένα επίπεδα (fully connected – FC layers) σε ένα CNN είναι εκείνα όπου κάθε νευρώνας διατηρεί σύνδεση με όλες τις ενεργοποιήσεις του προηγούμενου επιπέδου. Αξίζει να σημειωθεί ότι η βασική διαφορά τους από τα συνελκτικά επίπεδα (CONV) έγκειται στο γεγονός ότι στους CONV νευρώνες η συνδεσιμότητα περιορίζεται τοπικά και αρκετοί εξ αυτών μοιράζονται παραμέτρους.

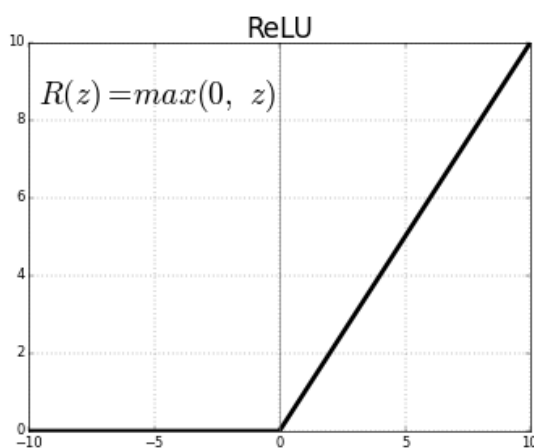
Συνήθως, τα τελικά επίπεδα ενός CNN είναι πλήρως συνδεδεμένα, με σκοπό να συσχετίσουν τις πληροφορίες που εξήχθησαν από τα προηγούμενα επίπεδα με την επιθυμητή τελική έξοδο του μοντέλου.

## 2.5.5 Συνάρτηση Ενεργοποίησης ReLU

Η συνάρτηση ενεργοποίησης ReLU (Rectified Linear Unit) είναι μια μη γραμμική συνάρτηση που επιστρέφει την ίδια την είσοδο αν είναι θετική, διαφορετικά επιστρέφει μηδέν. Η ReLU είναι εξαιρετικά απλή και υπολογιστικά αποδοτική, γεγονός που την καθιστά ιδιαίτερα χρήσιμη για την ταχεία εκπαίδευση βαθιών δικτύων.

Επιπλέον, με την εξάλειψη των αρνητικών τιμών στους χάρτες χαρακτηριστικών, η ReLU δημιουργεί αραιές αναπαραστάσεις (sparse representations). Η αραιώση αυτή οδηγεί σε συμπαγή μοντέλα, τα οποία συχνά εμφανίζουν καλύτερη γενίκευση και λιγότερο θόρυβο.

Η ReLU αποτελεί τη δημοφιλέστερη συνάρτηση ενεργοποίησης σε βαθιές αρχιτεκτονικές, καθώς προσφέρει υψηλή αποδοτικότητα σε σύγκριση με παραδοσιακές συναρτήσεις όπως η sigmoid ή η υπερβολική εφαιτομένη (tanh) [24]. Η συνάρτηση απεικονίζεται στο Σχήμα 2.6.



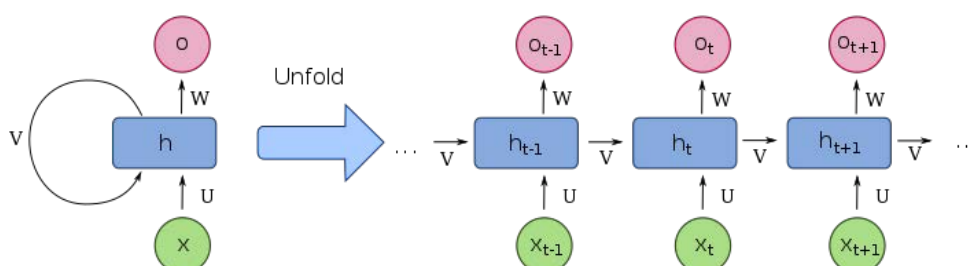
Σχήμα 2.6: Συνάρτηση ReLU (πηγή: Ailephant AI Team (2025) [25])

## 2.6 Αναδρομικά Νευρωνικά Δίκτυα (RNNs) και Δίκτυα LSTM

Η χρήση Συνελκτικών Νευρωνικών Δικτύων (CNNs) δεν παρέχει μέχρι σήμερα τον βέλτιστο τρόπο αντιμετώπισης των χρονικών εξαρτήσεων που μπορεί να είναι κρίσιμες για την αναγνώριση συνεχόμενων δραστηριοτήτων. Αντίθετα, τα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks – RNNs) αντιμετωπίζουν αυτό το ζήτημα μέσω της εισαγωγής βρόχων (loops) στη δομή τους, οι οποίοι επιτρέπουν στη πληροφορία να διατηρείται στον χρόνο.

Τα RNNs λειτουργούν σαν να διατηρούν μνήμη προηγούμενων εισόδων, και αξιοποιούν αυτή τη μνήμη για τη διαμόρφωση των προβλέψεων. Η ικανότητά τους να χειρίζονται σειριακές πληροφορίες τα καθιστά ιδιαίτερα κατάλληλα και για εργασίες όπως η αναγνώριση από βίντεο, η αναγνώριση φωνής ή η αναγνώριση χειρόγραφου κειμένου. Μπορούμε να τα αντιληφθούμε ως νευρωνικά δίκτυα με εσωτερικό βρόχο, τα οποία επεξεργάζονται σειρές δεδομένων διαδοχικά, διατηρώντας σε κάθε χρονική στιγμή μια εσωτερική κατάσταση (state) που συνοψίζει ό,τι έχει προηγηθεί [26].

Αν θελήσουμε να οπτικοποιήσουμε την εξέλιξη ενός RNN στον χρόνο (unrolled RNN), μπορούμε να την φανταστούμε ως διαδοχικά αντίγραφα του ίδιου υποδικτύου, όπου το κάθε υποδίκτυο μεταφέρει πληροφορία στο επόμενο, όπως παρουσιάζεται στο Σχήμα 2.7. Η αλυσιδωτή αυτή δομή καθιστά τα RNNs ιδιαίτερα κατάλληλα για τη μοντελοποίηση σειριακών δεδομένων.



Σχήμα 2.7: Αναδρομικό Νευρωνικό Δίκτυο RNN (πηγή: Ailephant AI Team (2025) [27])

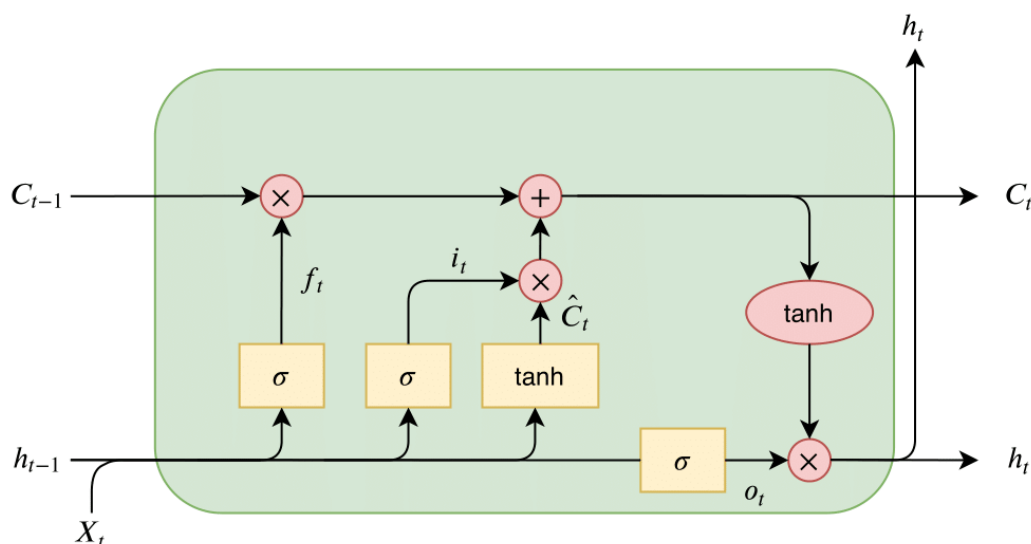
Ωστόσο, ένα βασικό μειονέκτημα των κλασικών RNNs είναι ότι δυσκολεύονται να διαχειριστούν μακροχρόνιες εξαρτήσεις. Όταν η κρίσιμη πληροφορία προέρχεται από απομακρυσμένες χρονικά καταστάσεις, το δίκτυο αδυνατεί να μάθει να συσχετίζει την απαραίτητη πληροφορία. Για την αντιμετώπιση αυτού του προβλήματος, σχεδιάστηκε η αρχιτεκτονική Long Short-Term Memory (LSTM), η οποία είναι ειδικά κατασκευασμένη για την αποδοτική εκμάθηση εξαρτήσεων μακράς διάρκειας σε χρονικές ακολουθίες.

## 2.6.1 Δίκτυα Long Short Term Memory (LSTMs)

Τα δίκτυα Long Short-Term Memory (LSTMs) αποτελούν μια παραλλαγή των αναδρομικών νευρωνικών δικτύων (RNNs), η οποία έχει σχεδιαστεί ειδικά για την αντιμετώπιση της αδυναμίας διατήρησης μακροχρόνιων εξαρτήσεων. Ουσιαστικά, προσφέρουν έναν μηχανισμό μεταφοράς πληροφορίας κατά μήκος πολλών χρονικών βημάτων, αποτρέποντας τη σταδιακή εξασθένιση των σημάτων κατά την επεξεργασία [26].

Τα LSTMs διατηρούν την αλυσιδωτή δομή των κλασικών RNNs, ωστόσο το επαναλαμβανόμενο υποδίκτυο διαθέτει πιο σύνθετη εσωτερική αρχιτεκτονική. Αντί για ένα απλό επίπεδο νευρώνων, κάθε μονάδα LSTM περιλαμβάνει τέσσερα διακριτά στοιχεία που αλληλεπιδρούν με συγκεκριμένο τρόπο: την κυψέλη (cell) και τρεις πύλες ελέγχου (gates) – την πύλη εισόδου (input gate), την πύλη εξόδου (output gate) και την πύλη λήθης (forget gate).

Η κατάσταση της κυψέλης (cell state) αποτελεί το θεμελιώδες στοιχείο του LSTM και λειτουργεί ως γραμμή μεταφοράς πληροφορίας κατά μήκος της χρονικής ακολουθίας. Οι πύλες ελέγχου μαθαίνουν ποιες πληροφορίες πρέπει να διατηρηθούν και ποιες να ξεχαστούν, επιτρέποντας στο δίκτυο να μεταδίδει σχετικές πληροφορίες μέσα σε μεγάλες χρονικές ακολουθίες χωρίς να εμφανίζεται το φαινόμενο της εξαφάνισης του gradient (vanishing gradient problem).



Σχήμα 2.8: Δίκτυο Long Short-Term Memory (LSTM) (πηγή: Medium.com [28])

Μια τυπική μονάδα LSTM περιλαμβάνει κυψέλη (cell), πύλη εισόδου (input gate), πύλη εξόδου (output gate) και πύλη λήθης (forget gate) όπως φαίνεται στο Σχήμα 2.8. Μπορούμε να αναλύσουμε τη λειτουργία του LSTM μέσω των εξής βασικών μηχανισμών:

- Συνάρτηση ενεργοποίησης tanh: Ρυθμίζει τις τιμές που διακινούνται στο δίκτυο. Επιστρέφει τιμές στο διάστημα  $[-1, 1]$  και χρησιμοποιείται για τη δημιουργία υποψηφίων τιμών εισόδου και εξόδου.

- Συνάρτηση ενεργοποίησης sigmoid: Χρησιμοποιείται από τις πύλες για να «αποφασίσουν» ποιες τιμές είναι σημαντικές και ποιες πρέπει να αγνοηθούν. Επιστρέφει τιμές στο διάστημα  $[0, 1]$ , όπου τιμές κοντά στο 0 σημαίνουν απόρριψη, ενώ τιμές κοντά στο 1 σημαίνουν διατήρηση.
- Πύλη Λήθης (Forget Gate): Η πύλη λήθης αποφασίζει ποιες πληροφορίες από την προηγούμενη χρονική κατάσταση θα διαγραφούν. Η προηγούμενη κρυφή κατάσταση και η τρέχουσα είσοδος εισέρχονται σε μια συνάρτηση sigmoid. Τιμές κοντά στο 0 οδηγούν σε απόρριψη πληροφορίας, ενώ τιμές κοντά στο 1 σε διατήρηση.
- Πύλη Εισόδου (Input Gate): Αυτή η πύλη επιλέγει ποιες νέες πληροφορίες θα προστεθούν στην κυψέλη. Η προηγούμενη κρυφή κατάσταση και η τρέχουσα είσοδος περνούν από μια sigmoid, δίνοντας μια «μάσκα» μεταξύ 0 και 1. Παράλληλα, μέσω της συνάρτησης  $\tanh$  παράγεται ένα διάνυσμα υποψήφιων νέων τιμών. Οι δύο έξοδοι πολλαπλασιάζονται και το αποτέλεσμα προστίθεται στην κυψέλη, ενημερώνοντας την κατάστασή της.
- Κατάσταση Κυψέλης (Cell State): Η κυψέλη λειτουργεί ως φορέας πληροφορίας μακράς διάρκειας. Αρχικά, το τρέχον state πολλαπλασιάζεται με την έξοδο της πύλης λήθης, διαγράφοντας άχρηστη πληροφορία. Έπειτα, προστίθενται οι υποψήφιες νέες τιμές από την πύλη εισόδου, διαμορφώνοντας τη νέα κατάσταση της κυψέλης.
- Πύλη Εξόδου (Output Gate): Η πύλη εξόδου καθορίζει ποιο μέρος της κυψέλης θα αποκαλυφθεί ως η νέα κρυφή κατάσταση. Η προηγούμενη κρυφή κατάσταση και η τρέχουσα είσοδος περνούν από μια sigmoid που λειτουργεί ως μάσκα εξόδου. Η ενημερωμένη κυψέλη περνά από  $\tanh$ , και το αποτέλεσμα πολλαπλασιάζεται με την έξοδο της sigmoid, παράγοντας τη νέα κρυφή κατάσταση που θα μεταφερθεί στον επόμενο χρονικό βρόχο. [29]

### 3 Σχετική Βιβλιογραφία

Ο τομέας της έρευνας για την αναγνώριση ανθρώπινης δραστηριότητας είναι ιδιαίτερα ενεργός, καθώς χρησιμοποιείται ευρέως στην καθημερινότητά μας. Η ενσωμάτωση πολυαισθητηριακών δεδομένων και η γεικευμένη πρόβλεψη δραστηριοτήτων σε πραγματικές συνθήκες προϋποθέτει συνδυασμό τεχνικών προεπεξεργασίας, επιλογής χαρακτηριστικών και σύγχρονων αλγορίθμων μηχανικής μάθησης. Υπάρχει πληθώρα δημοσιευμένων άρθρων και εργασιών που παρουσιάζουν και αναλύουν διαφορετικές προσεγγίσεις και λύσεις στο υπάρχον πρόβλημα, καθώς επίσης πρόσφατες ανασκοπήσεις υπογραμμίζουν τις προκλήσεις και τις μελλοντικές ευκαιρίες της βαθιάς μάθησης στην HAR, προσφέροντας μια ολοκληρωμένη εικόνα των υφιστάμενων μεθόδων [30]. Παρακάτω παρουσιάζονται ορισμένα από αυτά, που θεωρήθηκαν πιο σχετικά με την παρούσα έρευνα και εργασία.

Στην πρώτη μελέτη τους, οι Vaizman et al. (2017a) προσεγγίζουν το πρόβλημα της αναγνώρισης ανθρώπινης δραστηριότητας με την εφαρμογή ανεξάρτητων γραμμικών ταξινομητών για κάθε ετικέτα (binary relevance approach). Αντί για κοινό πολυκατηγοριακό ταξινομητή, προτείνεται η εκπαίδευση ενός Logistic Regression μοντέλου ανά ετικέτα, ώστε να αντιμετωπιστεί το πρόβλημα της πολυετικετικής φύσης των δεδομένων. Αυτή η προσέγγιση επιτρέπει μεγαλύτερο έλεγχο σε προβλήματα όπου οι ετικέτες εμφανίζονται με διαφορετικές συχνότητες ή αφορούν έννοιες ασυσχέτιστες μεταξύ τους (π.χ. “walking” και “in a meeting”).

Επιπλέον, η μελέτη εισάγει τρεις τεχνικές συγχώνευσης δεδομένων από πολλαπλούς αισθητήρες: την Early Fusion (EF), όπου όλα τα χαρακτηριστικά συγχωνεύονται πριν την εκπαίδευση, την Late Fusion by Averaging (LFA), όπου η κάθε πηγή παρέχει ξεχωριστή πρόβλεψη και υπολογίζεται μέσος όρος πιθανοτήτων και την Late Fusion by Learned Weights (LFL), όπου το μοντέλο μαθαίνει ξεχωριστά βάρη ανά αισθητήρα και ετικέτα [31].

Η επιλογή της logistic regression έγινε με βάση την ανάγκη για ερμηνευσιμότητα, ενώ δώθηκε έμφαση στην αντιμετώπιση ανισόροπων κατανομών ετικετών μέσω στάθμισης των κλάσεων. Το μοντέλο αξιολογείται με βάση την macro-F1, Balanced Accuracy και άλλες μετρικές, σε επίπεδο χρήστη (user-level cross-validation). Η μελέτη αυτή συνιστά μια γερή τεχνική βάση για HAR σε in-the-wild συνθήκες.

Η δεύτερη μελέτη των Vaizman et al. (2017b) εστιάζει στην αναβάθμιση της προηγούμενης προσέγγισης μέσω της υλοποίησης ενιαίου μοντέλου (multi-task classifier) για την ταυτόχρονη πρόβλεψη πολλαπλών ετικετών. Αντί για ανεξάρτητη εκπαίδευση ανά ετικέτα, το μοντέλο προβλέπει όλες τις ετικέτες μαζί, αξιοποιώντας κοινό σύνολο παραμέτρων. Αυτή η προσέγγιση θεωρείται πιο αποδοτική, ειδικά όταν υπάρχουν συσχετίσεις μεταξύ ετικετών — π.χ. “standing” και “at work” τείνουν να συμβαδίζουν.

Κεντρική καινοτομία της μελέτης αποτελεί η χρήση μάσκας ετικετών (label masking) για την αντιμετώπιση ατελών επισημάνσεων (missing labels). Σε συνθήκες in-the-wild, οι

χρήστες συχνά παραλείπουν να δηλώσουν κάποιες δραστηριότητες. Αν δεν ληφθεί υπόψη αυτό το φαινόμενο, το μοντέλο μπορεί να εκπαιδευτεί λανθασμένα, θεωρώντας τις μη δηλωμένες ετικέτες ως αρνητικές. Με την χρήση μιας μάσκας, ο υπολογισμός της συνάρτησης κόστους περιορίζεται μόνο στις ετικέτες που είναι γνωστές, διατηρώντας την αξιοπιστία της εκπαίδευσης [32].

Η μελέτη αυτή δείχνει επίσης ότι το ενοποιημένο μοντέλο υπερέχει έναντι των ανεξάρτητων ταξινομητών, ιδιαίτερα σε σπάνιες ετικέτες, καθώς μπορεί να μάθει γενικές συσχετίσεις και να βελτιώσει τη γενίκευση.

Η προσέγγιση των Ordóñez & Roggen (2016) βασίζεται, σε αντίθεση με τις παραπάνω μελέτες, αποκλειστικά σε βαθιά μάθηση (deep learning) και εισάγει τη γνωστή πλέον αρχιτεκτονική DeepConvLSTM. Το μοντέλο συνδυάζει συνελκτικά επίπεδα (Conv1D) για εξαγωγή τοπικών χαρακτηριστικών από χρονικά σήματα με επόμενα επίπεδα LSTM, τα οποία μοντελοποιούν τη χρονική εξάρτηση εντός του παραθύρου παρατήρησης.

Σε αντίθεση με τα γραμμικά μοντέλα των Vaizman, η DeepConvLSTM είναι end-to-end, δεν απαιτεί χειροκίνητη επιλογή χαρακτηριστικών, και εκπαιδεύεται απευθείας σε ακατέργαστα δεδομένα αισθητήρων. Η αξιολόγηση πραγματοποιείται σε δημοφιλή datasets (π.χ. Opportunity, PAMAP2), και τα αποτελέσματα δείχνουν σημαντική υπεροχή των deep models έναντι των παραδοσιακών προσεγγίσεων, ειδικά σε δραστηριότητες με πολύπλοκα κινητικά μοτίβα [6].

Η συνεισφορά της μελέτης έγκειται στη συνδυασμένη αξιοποίηση χωρικής και χρονικής πληροφορίας, ενώ η αρχιτεκτονική της έχει αποτελέσει τη βάση για μεγάλο αριθμό επόμενων εργασιών. Σε σύγκριση με τις μελέτες των Vaizman, η προσέγγιση DeepConvLSTM απαιτεί μεγαλύτερη υπολογιστική ισχύ και περισσότερα δεδομένα για αποτελεσματική εκπαίδευση, αλλά προσφέρει σημαντικά πλεονεκτήματα στην αυτόματη μάθηση αναπαραστάσεων.

Γενικότερα, παρατηρείται ότι η σύγχρονη ερευνητική κατεύθυνση στην HAR έχει επικεντρωθεί στη χρήση αναδρομικών νευρωνικών δικτύων τύπου LSTM και σε ελαφριές αρχιτεκτονικές βαθιάς μάθησης για την ανάπτυξη μοντέλων κατάλληλων για συσκευές χαμηλής υπολογιστικής ισχύος. Χαρακτηριστικά παραδείγματα είναι και οι μελέτες που αναφέρονται στις επόμενες παραγράφους.

Ο Wang και ο Liu (2020) πρότειναν ένα ιεραρχικό δίκτυο LSTM για την ταξινόμηση δραστηριοτήτων, το οποίο αποτελείται από πολλαπλά επίπεδα χρονικής ανάλυσης, βελτιώνοντας την απόδοση σε δεδομένα με πολυεπίπεδη δυναμική [33]. Αντίστοιχα, οι Oluwalade et al. (2021) αξιολόγησαν την απόδοση διαφορετικών deep learning αρχιτεκτονικών (CNN, RNN, LSTM) σε δεδομένα από smartphones και smartwatches, καταλήγοντας στο συμπέρασμα ότι τα LSTM παρουσιάζουν υψηλότερη ακρίβεια σε δραστηριότητες που απαιτούν χρονική κατανόηση [oluwalade2021].

Οι Mekruksavanich και Jitpattanakul (2020) πρότειναν ένα υβριδικό LSTM μοντέλο ειδικά προσαρμοσμένο για δεδομένα smartwatch, ενώ σε επόμενη εργασία τους (2020), ενίσχυσαν το μοντέλο με αναγνώριση κινήσεων χεριού, επιτυγχάνοντας περαιτέρω αύξηση της ακρίβειας ταξινόμησης σε εξειδικευμένες περιπτώσεις [34],[35]. Οι μελέτες αυτές δείχνουν τη δυνατότητα των LSTM να προσαρμόζονται σε πιο εξειδικευμένα κινητικά μοτίβα και χρήσεις.

Τέλος, οι Agarwal και Alam (2020) επικεντρώνονται στη βελτιστοποίηση του μεγέθους και της υπολογιστικής πολυπλοκότητας μοντέλων HAR για ενσωμάτωσή τους σε edge συσκευές, προτείνοντας ένα “lightweight” deep learning μοντέλο που επιτυγχάνει ανταγωνιστικά αποτελέσματα με σημαντικά μικρότερο αριθμό παραμέτρων [36]. Η κατεύθυνση αυτή καθιστά τη HAR περισσότερο εφαρμόσιμη σε φορητά και IoT συστήματα με περιορισμένους πόρους.

## 4 Μεθοδολογία

Το παρόν κεφάλαιο περιγράφει το μεθοδολογικό πλαίσιο που ακολουθήθηκε για την υλοποίηση της μελέτης. Αναλύεται ο συνολικός σχεδιασμός της έρευνας, το σύνολο δεδομένων που χρησιμοποιείται, καθώς και τα βήματα προεπεξεργασίας που εφαρμόστηκαν ώστε να καταστούν τα δεδομένα κατάλληλα για εκπαίδευση. Παρουσιάζονται αναλυτικά τα μοντέλα βαθιάς μάθησης που υλοποιούνται —συμπεριλαμβανομένων των CNN, LSTM και ενός προτεινόμενου υβριδικού μοντέλου— καθώς και οι μετρικές αξιολόγησης που αξιοποιούνται για τη σύγκριση της απόδοσής τους. Τέλος, συζητούνται ζητήματα ηθικής φύσεως και οι περιορισμοί που ενδέχεται να επηρεάσουν την ερμηνεία των αποτελεσμάτων.

### 4.1 Σχεδιασμός Έρευνας

Η παρούσα μελέτη ακολουθεί πειραματικό υπολογιστικό σχεδιασμό με στόχο την αναγνώριση ανθρώπινων δραστηριοτήτων βάσει δεδομένων αισθητήρων από φορητές συσκευές. Η μελέτη επικεντρώνεται στη σύγκριση διαφορετικών προσεγγίσεων βαθιάς μάθησης για την ταξινόμηση δραστηριοτήτων, αξιοποιώντας ως σημείο αναφοράς το ExtraSensory Dataset [37], το οποίο περιλαμβάνει δεδομένα από smartphones και smartwatches σε πραγματικές συνθήκες.

Η ερευνητική προσέγγιση περιλαμβάνει τα εξής βασικά στάδια:

- Προεπεξεργασία των δεδομένων
- Σχεδιασμό και εκπαίδευση μοντέλων (Logistic Regression, CNN, LSTM και ενός προτεινόμενου υβριδικού μοντέλου) και
- Αξιολόγηση της απόδοσης κάθε μοντέλου με χρήση καθιερωμένων μετρικών, οι οποίες αναλύονται σε επόμενη υποενότητα.

Ο σχεδιασμός της έρευνας διαμορφώθηκε με σκοπό να εξετάσει την αποτελεσματικότητα κάθε προσέγγισης, να αναδείξει τις προκλήσεις της HAR σε ρεαλιστικά δεδομένα και να συμβάλει στην κατανόηση των τεχνικών που ενδείκνυνται για συστήματα έγκυρης αναγνώρισης δραστηριοτήτων στην καθημερινή ζωή.

## 4.2 Σύνολο Δεδομένων

Το ExtraSensory Dataset είναι ένα πρωτοποριακό και εκτενές σύνολο δεδομένων που δημιουργήθηκε για την αναγνώριση ανθρώπινων δραστηριοτήτων και συμφραζομένων σε φυσικές, καθημερινές συνθήκες. Πρόκειται για μία από τις πρώτες συστηματικές προσπάθειες συλλογής δεδομένων in-the-wild, όπου οι συμμετέχοντες δεν ακολουθούν προκαθορισμένα σενάρια, αλλά χρησιμοποιούν τις συσκευές τους φυσικά, κατά τη διάρκεια της καθημερινότητάς τους.

Η συλλογή των δεδομένων έγινε από 60 συμμετέχοντες που χρησιμοποίησαν το προσωπικό τους smartphone σε συνδυασμό με ένα smartwatch (Pebble) που τους παραχωρήθηκε. Κάθε συσκευή κατέγραφε αισθητηριακά δεδομένα σε παράθυρα των 20 δευτερολέπτων ανά λεπτό, δημιουργώντας έτσι περισσότερα από 300.000 παραδείγματα με ετικέτες.

Σε αντίθεση με προηγούμενα σύνολα που βασίζονταν σε σενάρια με οδηγίες και τοποθετήσεις συσκευών σε συγκεκριμένες θέσεις, το ExtraSensory επιδιώκει να αποτυπώσει την πολυπλοκότητα και τη μεταβλητότητα της ανθρώπινης συμπεριφοράς με τη μικρότερη δυνατή παρεμβολή, επιτυγχάνοντας δεδομένα που αντικατοπτρίζουν όσο το δυνατόν καλύτερα την πραγματική καθημερινότητα των χρηστών.

Οι συμμετέχοντες προέρχονταν κυρίως από την πανεπιστημιούπολη του University of California, San Diego (UCSD). Το δείγμα περιλάμβανε:

- 34 γυναίκες και 26 άνδρες.
- Ηλικίες από 18 έως 42 ετών (μέσος όρος  $24.7 \pm 5.6$ ).
- 34 χρήστες iOS και 26 χρήστες Android, με διάφορες συσκευές.
- Ετερόκλητα εθνικά υπόβαθρα.
- Οι περισσότεροι συμμετέχοντες (93%) ήταν δεξιόχειρες και φορούσαν το smartwatch στο αριστερό τους χέρι.

Η καταγραφή διήρκεσε από 3 έως 28 ημέρες ανά χρήστη (μέσος όρος ~7.6 ημέρες), με περισσότερα από 5000 παραδείγματα ανά άτομο.

Το σύνολο περιλαμβάνει δεδομένα από αισθητήρες τόσο του smartphone όσο και του smartwatch, με πολυδιάστατες μετρήσεις:

Πίνακας 4.1: Πίνακας αισθητήρων

| Αισθητήρας          | Πληροφορία                        | Συχνότητα   | Δείγμα   |
|---------------------|-----------------------------------|-------------|----------|
| Accelerometer (Acc) | 3D επιτάχυνση (τηλέφωνο)          | 40Hz        | ~308.000 |
| Gyroscope (Gyro)    | Γωνιακή ταχύτητα                  | 40Hz        | ~292.000 |
| Watch Accelerometer | 3D επιτάχυνση από smartwatch      | 25Hz        | ~210.000 |
| Location            | Συντεταγμένες, ταχύτητα, ακρίβεια | μεταβλητή   | ~274.000 |
| Audio (MFCC)        | Ήχος - 13 MFCC χαρακτηριστικά     | ~430 frames | ~302.000 |

| Αισθητήρας       | Πληροφορία                                 | Συχνότητα | Δείγμα   |
|------------------|--------------------------------------------|-----------|----------|
| Phone State (PS) | Κατάσταση μπαταρίας, WiFi, ώρας, εφαρμογής | 1x/λεπτό  | ~308.000 |

Κάθε χρονικό παράθυρο συνοδεύεται από πολλαπλές ετικέτες που επισημαίνονται από τον ίδιο τον χρήστη μέσω εύχρηστης διεπαφής. Οι ετικέτες διακρίνονται σε:

- Κύριες Ετικέτες (7 αμοιβαία αποκλειόμενες): lying down, sitting, standing, walking, running, bicycling, sleeping.
- Δευτερεύουσες Ετικέτες (100+): περιβάλλον (π.χ. at home, at gym), δραστηριότητα (π.χ. cooking, cleaning), μετακίνηση (π.χ. on a bus, driving), συνοδεία (π.χ. with friends), θέση τηλεφώνου (π.χ. in pocket), κ.λπ.

Ο κάθε συμμετέχων χρησιμοποιεί κατά μέσο όρο 15 διαφορετικές ετικέτες, με 3.8 ετικέτες ανά λεπτό.

Το βασικό σύστημα που περιγράφεται από τους συγγραφείς εφαρμόζει γραμμικούς ταξινομητές (logistic regression) ανά αισθητήρα και ανά ετικέτα. Για τη συγχώνευση των δεδομένων από πολλαπλούς αισθητήρες, δοκιμάζονται τρεις στρατηγικές:

1. Early Fusion (EF): Συνένωση όλων των χαρακτηριστικών πριν την ταξινόμηση.
2. Late Fusion by Averaging (LFA): Μέσος όρος των πιθανοτήτων ανά αισθητήρα.
3. Late Fusion by Learned Weights (LFL): Εκμάθηση βαρών ανά αισθητήρα για κάθε ετικέτα.

Η μέθοδος LFL αποδείχθηκε η πιο αποτελεσματική, αποδίδοντας υψηλότερες Balanced Accuracy μετρικές σε μεγάλο εύρος ετικετών, ειδικά σε περιπτώσεις όπου τα δεδομένα ήταν ετερογενή ή πολυδιάστατα.

Το ExtraSensory Dataset ξεχωρίζει για:

- Τη ρεαλιστικότητα των δεδομένων (χωρίς σκηνοθεσία, φυσική χρήση συσκευών).
- Τη μεγάλη ποικιλία συσκευών και συνηθειών.
- Την υποστήριξη για πολλαπλές ταυτόχρονες ετικέτες (multi-label classification).
- Την ευελιξία του να αξιοποιηθεί για εκπαίδευση μοντέλων σε unseen users.

Παράλληλα, η αυτοαναφορά από τους χρήστες προκαλεί αβεβαιότητα στα labels λόγω υποκειμενικότητας ή καθυστερημένης καταγραφής, γεγονός που καθιστά τις μετρικές όπως το F1 ευαίσθητες και λιγότερο αντιπροσωπευτικές από το Balanced Accuracy (BA).

Συμπερασματικά, το σύνολο ExtraSensory αποτελεί ιδανικό πρότυπο για έρευνα στην αναγνώριση δραστηριότητας σε πραγματικές συνθήκες. Επιτρέπει την ανάπτυξη και συγκριτική αξιολόγηση αλγορίθμων με ακριβές, πλούσια επισημασμένα και πολυτροπικά δεδομένα. Είναι ιδανικό για μελέτες που εστιάζουν:

- Στην πολυαισθητηριακή ενοποίηση (sensor fusion),
- Στην προσωποποιημένη αναγνώριση (user adaptation),
- Στη multi-label μάθηση, και
- Στην πρόβλεψη συμφραζομένων σε πραγματικές συνθήκες.

Ως δημόσια διαθέσιμο σύνολο, προσφέρει σε ερευνητές και φοιτητές μια εξαιρετική βάση για πειραματισμό και βελτίωση των μεθόδων HAR, ειδικά σε σενάρια όπου η γενίκευση είναι βασικό ζητούμενο.

### 4.3 Προεπεξεργασία και Ανάλυση Δεδομένων

Η προεπεξεργασία των δεδομένων αποτελεί κρίσιμο στάδιο στην ανάπτυξη κάθε συστήματος αναγνώρισης ανθρώπινης δραστηριότητας, καθώς εξασφαλίζει ότι τα εισερχόμενα δεδομένα είναι ποιοτικά, σταθεροποιημένα και κατάλληλα μορφοποιημένα για την εκπαίδευση μοντέλων μηχανικής μάθησης. Στην παρούσα μελέτη, η προετοιμασία των δεδομένων βασίστηκε στη δομή του συνόλου ExtraSensory, το οποίο παρέχει για κάθε χρήστη αρχεία τύπου .csv.gz με εξαγόμενα χαρακτηριστικά από αισθητήρες καθώς και συμφραζόμενες ετικέτες (context labels).

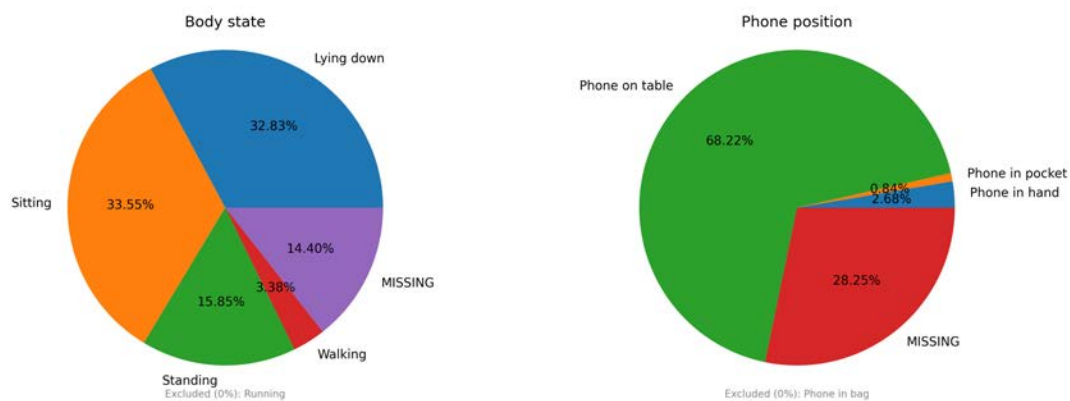
Κάθε αρχείο αντιστοιχεί σε έναν συμμετέχοντα και περιλαμβάνει τέσσερα βασικά στοιχεία: τον πίνακα χαρακτηριστικών  $X$ , τον πίνακα ετικετών  $Y$ , τον πίνακα ένδειξης ελλειπόντων ετικετών  $M$ , καθώς και ένα διανυσματικό πίνακα με χρονικές σφραγίδες (timestamps). Κάθε δείγμα συνοδεύεται από ετικέτες εισόδου που δηλώθηκαν από τον ίδιο τον χρήστη μέσω εφαρμογής. Τα δεδομένα συλλέχθηκαν σε χρονικά παράθυρα διάρκειας 20 δευτερολέπτων ανά λεπτό, οδηγώντας σε μεγάλο πλήθος δειγμάτων, μερικές φορές υπερβαίνοντας τις 300.000 εγγραφές ανά συμμετέχοντα.

Για την ανάγνωση των δεδομένων αρχικά αποσυμπιέστηκε ο φάκελος με τα αρχεία των 60 συμμετεχόντων. Στη συνέχεια έγινε ανάλυση της γραμμής επικεφαλίδας ώστε να εντοπιστούν τα ονόματα των χαρακτηριστικών και των ετικετών, και τέλος φορτώθηκε το σώμα του κάθε αρχείου σε αριθμητικούς πίνακες. Ιδιαίτερη προσοχή δώθηκε στη σωστή διαχείριση των ετικετών: η παρουσία μιας κατάστασης έχει αποδοθεί με τιμή 1, η απουσία με 0, ενώ οι μη επισημασμένες περιπτώσεις έχουν καταχωρηθεί ως NaN. Παρότι για λόγους συμβατότητας τα NaN μετατράπηκαν σε 0 στον πίνακα  $Y$ , το σύνολο των «απροσδιόριστων» τιμών είναι καταγεγραμμένο στον πίνακα  $M$ , ώστε να μπορεί να ληφθεί υπόψη κατά την εκπαίδευση και την αξιολόγηση των μοντέλων πως δεν πρόκειται για απουσία κατάστασης αλλά απουσία καταγραφής αυτής.

Όσον αφορά τα χαρακτηριστικά, το dataset περιλαμβάνει πληθώρα αισθητήρων, όπως επιταχυνσιόμετρο, γυροσκόπιο, μικρόφωνο, δεδομένα τοποθεσίας, καθώς και καταστάσεις

συσκευής (π.χ. τύπος σύνδεσης ή επίπεδο μπαταρίας). Για κάθε χρονικό σημείο έχουν καταγραφεί δεκάδες έως εκατοντάδες αριθμητικά χαρακτηριστικά που περιλαμβάνουν τόσο στατιστικά του χρόνου όσο και φασματικά αποσπάσματα – όπως για παράδειγμα οι συντελεστές MFCC από την ανάλυση ήχου. Αυτή η πολυαισθητηριακή φύση του σήματος καθιστά το πρόβλημα ιδιαίτερα σύνθετο και υψηλής διαστασιμότητας.

Όπως αναφέρθηκε και σε προηγούμενη ενότητα, τα δεδομένα συλλέχθηκαν in-the-wild, γεγονός που σημαίνει ότι καθώς η ετικετοποίηση χρειαζόταν είσοδο από το χρήστη χειροκίνητα, είναι συχνό αυτή να λείπει (NaN τιμές) για διαφορετικές ετικέτες και διαφορετικό πλήθος χρονικών στιγμών αναλόγως τον χρήστη. Χαρακτηριστικό παράδειγμα αποτελεί το Σχήμα 4.1 όπου για έναν τυχαίο χρήστη υπάρχουν δεδομένα που λείπουν (αρχικά NaN τιμές) και αν δεν υπήρχαν τα τμήματα “MISSING” τότε τα pie charts που εικονίζονται θα ήταν ημιτελή.



Σχήμα 4.1: Ταξινόμηση κατηγοριών ανθρώπινης δραστηριότητας για έναν τυχαίο χρήστη

Έτσι, η προεπεξεργασία ξεκίνησε με την απομάκρυνση χαρακτηριστικών που περιέχουν μεγάλο ποσοστό ελλিপών τιμών, δηλαδή χαρακτηριστικών που εμφάνιζαν NaN σε περισσότερο από το 50% των δεδομένων. Στη συνέχεια, εφαρμόστηκε τυποποίηση (standardization) σε κάθε χαρακτηριστικό ξεχωριστά, με στόχο την κανονικοποίηση της κλίμακας τιμών. Η προσέγγιση αυτή επέτρεψε στα μοντέλα να εκπαιδευτούν χωρίς να επηρεάζονται από μεταβλητότητες στη μεγέθυνση ή την ένταση των σημάτων διαφορετικών αισθητήρων.

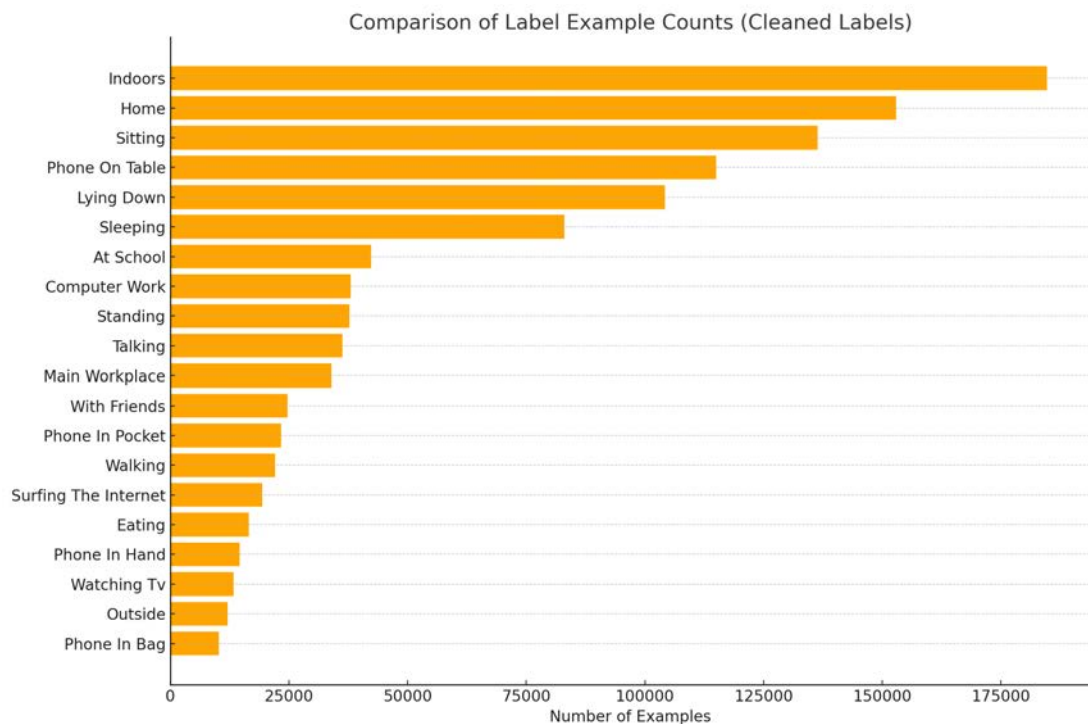
Για τα μοντέλα που βασίζονται σε χρονικές ακολουθίες, όπως τα LSTM και τα CNN, τα δεδομένα εισόδου αναδιαμορφώθηκαν μέσω σχηματισμού κινούμενων παραθύρων χρόνου (sliding windows). Η διαδικασία αυτή αφορά την αποκοπή διαδοχικών τμημάτων σταθερού μήκους από τη χρονική σειρά, προκειμένου να δημιουργηθούν υποδειγματικές ακολουθίες εισόδου. Κάθε τέτοιο παράθυρο συνοδεύεται από τον αντίστοιχο στόχο εξόδου, ο οποίος μπορεί να είναι το σύνολο των ετικετών που ενεργοποιούνται εντός του παραθύρου ή η κυρίαρχη ετικέτα.

Ένα επιπλέον κρίσιμο στοιχείο της ανάλυσης ήταν η διαχείριση του πολυετικετικού χαρακτήρα του συνόλου δεδομένων. Σε αντίθεση με τα μονοετικετικά προβλήματα, όπου κάθε δείγμα ανήκει αποκλειστικά σε μία κατηγορία, στο ExtraSensory κάθε δείγμα μπορεί να χαρακτηρίζεται από πολλαπλές ετικέτες ταυτόχρονα, όπως “sitting”, “in meeting”, και “at

work”. Η πολυπλοκότητα της επισημείωσης απαιτούσε μοντέλα που είναι σε θέση να χειριστούν πολλαπλές εξόδους και να εκπαιδευτούν με πολυετικετικές συναρτήσεις κόστους, όπως η binary cross-entropy ανά ετικέτα. Για την αξιολόγηση, χρησιμοποιήθηκαν ειδικές μετρικές που λάμβαναν υπόψη και την ταυτόχρονη παρουσία πολλαπλών τάξεων.

Η παραπάνω διαδικασία εξασφάλισε ότι τα δεδομένα που εισήχθησαν στα μοντέλα ήταν επαρκώς καθαρά, εξισορροπημένα και αντιπροσωπευτικά της πολυπλοκότητας της καθημερινής ανθρώπινης δραστηριότητας όπως αυτή καταγράφεται σε φυσικές συνθήκες.

Στην εκπαίδευση και δοκιμή των μοντέλων χρησιμοποιήθηκαν τα δεδομένα από τις ετικέτες “FIX\_walking”, “SITTING”, “LOC\_home”, “OR\_outside” και “OR\_indoors”. Αυτές λόγω μεγαλύτερης πληθώρας διαθέσιμων τιμών - όπως φαίνεται στο σχήμα 4.2 - θεωρήθηκαν καταλληλότερες για εξαγωγή σαφών αποτελεσμάτων για τις ανάγκες της παρούσας μελέτης. Από τα δεδομένα αισθητήρων χρησιμοποιήθηκαν αυτά που αφορούσαν τις ετικέτες της επιτάχυνσης “raw\_acc” και “with\_acceleration” που μετονομάστηκαν σε “Acc” και “WAcc” αντίστοιχα.



Σχήμα 4.2: Ταξινόμηση των 20 ετικετών με τις περισσότερες δηλωμένες τιμές από τους χρήστες (πηγή: ίδια φωτογραφία)

Ήδη στο παραπάνω σχήμα η εικοστή ετικέτα “Phone in Bag” έχει σημαντικά μικρότερο αριθμό δειγμάτων από τις πρώτες έξι ετικέτες. Δεν έγινε απεικόνιση των υπολοίπων καθώς ο αριθμός δειγμάτων μειώνεται ακόμα περισσότερο και πιο δραστικά.

Στόχος της μελέτης αυτής είναι να συγκριθούν οι αλγόριθμοι που αναφέρονται στην επόμενη ενότητα ως προς την απόδοση τους στην πρόβλεψη της κίνησης ή μη κίνησης του χρή-

στη σε συνδιασμό με δευτερεύουσες ετικέτες που αφορούσαν την τοποθεσία στην οποία βρισκόταν. Με αυτό τον τρόπο βγαίνουν συμπεράσματα για την πιθανή συμπεριφορά που μπορεί να έχει ένας χρήστης ανάλογα με το μέρος στο οποίο βρίσκεται.

## 4.4 Αλγόριθμοι και Μοντέλα

Η επιλογή των μοντέλων που χρησιμοποιούνται βασίστηκε στην ανάγκη να συγκριθούν προσεγγίσεις διαβαθμισμένης πολυπλοκότητας όσον αφορά την ικανότητά τους να αναγνωρίζουν ανθρώπινες δραστηριότητες βάσει δεδομένων αισθητήρων. Δοκιμάστηκαν τόσο παραδοσιακές μέθοδοι, όπως η λογιστική παλινδρόμηση, όσο και σύγχρονες αρχιτεκτονικές βαθιάς μάθησης, συμπεριλαμβανομένων συνελικτικών και αναδρομικών νευρωνικών δικτύων. Σκοπός είναι να αξιολογηθεί σε ποιο βαθμό οι πιο σύνθετες αρχιτεκτονικές προσφέρουν συγκριτικό πλεονέκτημα σε σχέση με απλούστερες γραμμικές μεθόδους, σε πραγματικά και πολύπλοκα δεδομένα πολλαπλών αισθητήρων.

Η λογιστική παλινδρόμηση (Logistic Regression) χρησιμοποιείται ως μοντέλο βάσης (baseline) και δεν περιγράφεται σε βάθος στην παρούσα ενότητα, καθώς αναλύεται διεξοδικά σε επόμενο κεφάλαιο.

Η συνελικτική προσέγγιση (CNN) αξιοποιεί την τοπική συσχέτιση μεταξύ διαδοχικών χρονικών δειγμάτων και επιτρέπει την ανίχνευση διακριτών κινητικών μοτίβων εντός των sliding windows. Η αρχιτεκτονική αποτελείται από επάλληλα συνελικτικά και πυκνά επίπεδα, και έχει σχεδιαστεί ώστε να επεξεργάζεται κανονικοποιημένες ακολουθίες σταθερού μήκους.

Το αναδρομικό νευρωνικό δίκτυο τύπου LSTM αξιοποιείται με στόχο τη μοντελοποίηση μακροχρόνιων χρονικών εξαρτήσεων. Η εσωτερική του αρχιτεκτονική, με πύλες εισόδου, εξόδου και λήθης, επιτρέπει στο δίκτυο να διατηρεί σχετική πληροφορία από προηγούμενα βήματα της ακολουθίας και να διαμορφώνει την πρόβλεψη ανάλογα με τη χρονική δομή των δεδομένων.

Τέλος, υλοποιήθηκε και ένα υβριδικό μοντέλο CNN και LSTM, το οποίο συνδυάζει τη δυνατότητα εξαγωγής τοπικών χαρακτηριστικών από το CNN με την ικανότητα χρονικής μοντελοποίησης του LSTM. Η συνελικτική ενότητα δρα ως feature extractor εντός κάθε παραθύρου, και η έξοδός της τροφοδοτεί το LSTM, το οποίο μαθαίνει τη δυναμική εξέλιξη των δραστηριοτήτων.

Και τα τρία μοντέλα βαθιάς μάθησης (CNN, LSTM, CNN+LSTM) έχουν διαμορφωθεί ώστε να υποστηρίζουν πολυετικετική έξοδο (multi-label prediction), με χρήση κατάλληλης συνάρτησης απώλειας (binary cross-entropy) και μάσκας για τη διαχείριση των ελλিপών ετικετών. Η εκπαίδευση πραγματοποιείται σε πολλούς χρήστες και η αξιολόγηση γίνεται σε unseen χρήστες, βάσει προκαθορισμένων folds.

## 4.5 Αξιολόγηση Μοντέλων

Η αξιολόγηση της απόδοσης των μοντέλων αποτελεί καθοριστικό στάδιο για την τεκμηριωμένη σύγκριση μεταξύ διαφορετικών προσεγγίσεων ταξινόμησης. Στην παρούσα μελέτη, υιοθετείται ένα αυστηρά διαχωρισμένο σχήμα εκπαίδευσης και ελέγχου ανά χρήστη, με στόχο τη μελέτη της ικανότητας γενίκευσης των μοντέλων σε μη-ορατούς χρήστες (unseen users). Η πειραματική διαδικασία βασίζεται σε 5-fold cross-validation, με διαχωρισμό των 60 χρηστών του ExtraSensory Dataset σε πέντε μη επικαλυπτόμενα υποσύνολα (folds), όπως προτείνεται και στο πρωτότυπο έργο των Vaizman et al. (2017) για το σύνολο ExtraSensory [31].

Κατά την εφαρμογή της μεθόδου, σε κάθε επανάληψη (fold) τα δεδομένα από τέσσερις ομάδες χρηστών χρησιμοποιούνται για εκπαίδευση και η πέμπτη για αξιολόγηση. Με αυτόν τον τρόπο εξασφαλίζεται ότι κάθε χρήστης συμμετέχει ακριβώς μία φορά στο σύνολο ελέγχου και ότι η αξιολόγηση προσομοιώνει ρεαλιστικά το σενάριο εφαρμογής σε νέους χρήστες. Το σχήμα 4.3 οπτικοποιεί τη λογική του 5-fold cross-validation.

|       |       |       |       |       |
|-------|-------|-------|-------|-------|
| Test  | Train | Train | Train | Train |
| Train | Test  | Train | Train | Train |
| Train | Train | Test  | Train | Train |
| Train | Train | Train | Test  | Train |
| Train | Train | Train | Train | Test  |

Σχήμα 4.3: Οπτική απεικόνιση 5-fold Cross-Validation (Unseen Users) (πηγή: ίδια φωτογραφία)

Πιο αναλυτικά και για να γίνει πιο κατανοητή η προσέγγιση, το 5-fold cross-validation είναι μια τεχνική αξιολόγησης που χρησιμοποιείται ευρέως στη μηχανική μάθηση για την εκτίμηση της γενικευσιμότητας ενός μοντέλου. Σύμφωνα με τη μέθοδο αυτή, το αρχικό σύνολο δεδομένων χωρίζεται σε πέντε ισομεγέθη και αμοιβαία αποκλειόμενα υποσύνολα (folds). Σε κάθε επανάληψη (fold), τέσσερα από τα πέντε υποσύνολα χρησιμοποιούνται για εκπαίδευση του μοντέλου, ενώ το πέμπτο χρησιμοποιείται για αξιολόγηση. Η διαδικασία επαναλαμβάνεται πέντε φορές, εναλλάσσοντας σε κάθε επανάληψη το fold που χρησιμοποιείται ως σύνολο ελέγχου, ώστε κάθε δείγμα να χρησιμοποιηθεί ακριβώς μία φορά για έλεγχο.

Η τελική επίδοση του μοντέλου υπολογίζεται συνήθως ως μέσος όρος των επιδόσεων σε όλα τα folds, παρέχοντας μια πιο σταθερή και αξιόπιστη εκτίμηση της απόδοσης σε άγνωστα δεδομένα, σε σύγκριση με μια απλή διάσπαση train/test. Στο πλαίσιο της παρούσας μελέτης, το 5-fold cross-validation εφαρμόζεται σε επίπεδο χρηστών, διασφαλίζοντας αξιολόγηση σε unseen users και την αποφυγή διαρροής πληροφορίας (information leakage) μεταξύ εκπαίδευσης και ελέγχου.

### 4.5.1 Μετρικές Αξιολόγησης

Σημαντικό κομμάτι της μηχανικής μάθησης, αποτελούν οι μετρικές αξιολόγησης. Χρησιμοποιώντας τις μετρικές αυτές, μπορεί να γίνει ευδιάκριτη η επιτυχία που έχει κάποιο μοντέλο μηχανικής μάθησης και αν αυτό επαρκεί ως προς τις απαιτήσεις που τίθενται. Παράλληλα, καθίσταται εφικτή η σύγκριση μεταξύ των μοντέλων.

Τα Confusion Matrices αποτελούν μία μέθοδο οπτικοποίησης των αποτελεσμάτων της κατηγοριοποίησης. Σε ένα πρόβλημα δυαδικής κατηγοριοποίησης, το Confusion Matrix αποτελεί ένα πίνακα 2x2, στον οποίο φαίνονται το πλήθος των True Positive (αληθώς θετικά - TP), True Negative (αληθώς αρνητικά - TN), False Positive (ψευδώς θετικά - FP) και False Negative (ψευδώς αρνητικά - FN).

Με βάση τα στοιχεία αυτά, μπορούν να υπολογιστούν μετρικές με τις οποίες γίνεται κατανοητή η ακρίβεια ή μη της κατηγοριοποίησης.

Το Accuracy, είναι ο λόγος των ορθών προβλέψεων δια του συνόλου των στοιχείων:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.1)$$

Σε σύνολα δεδομένων όπου τα αρνητικά παραδείγματα (True Negatives – TN) είναι εξαιρετικά λίγα, η συνολική ακρίβεια (Accuracy) μπορεί να εμφανιστεί τεχνητά αυξημένη, ακόμη και όταν ο ταξινομητής κατατάσσει όλα τα δείγματα ως θετικά. Σε τέτοιες περιπτώσεις, η μετρική αυτή ενδέχεται να παραπλανεί, καθώς δεν αποτυπώνει τη δυσκολία του προβλήματος ούτε τη σοβαρότητα του σφάλματος όταν ένα αρνητικό δείγμα κατηγοριοποιηθεί λανθασμένα. Ιδιαίτερα όταν η σωστή αναγνώριση των αρνητικών περιπτώσεων είναι σημαντική για την εφαρμογή, η απλή χρήση του Accuracy δεν επαρκεί ως δείκτης αξιολόγησης.

Για τον λόγο αυτό, υιοθετούνται μετρικές όπως το Precision και το Recall (ή Sensitivity), που παρέχουν πιο λεπτομερή πληροφόρηση για τη συμπεριφορά του μοντέλου ως προς τις θετικές προβλέψεις. Συγκεκριμένα, το Precision εκφράζει το ποσοστό των προβλέψεων μιας κατηγορίας που ήταν ορθές. Εάν το Precision ισούται με 1.0 για μια κατηγορία Y, τότε αυτό σημαίνει ότι όλα τα παραδείγματα που ταξινομήθηκαν ως Y πράγματι ανήκουν σε αυτή, ανεξαρτήτως του αν υπήρχαν και άλλα πραγματικά στοιχεία της κατηγορίας που δεν εντοπίστηκαν.

Το Precision υπολογίζεται με βάση την ακόλουθη σχέση:

$$Precision = \frac{TP}{TP + FP} \quad (4.2)$$

Εάν το Recall ισούται με 1.0 για μια κατηγορία, τότε αυτό σημαίνει ότι όλα τα παραδείγματα που πραγματικά ανήκουν στην κατηγορία Y εντοπίστηκαν σωστά από το μοντέλο, ανεξαρτήτως του αν ταξινομήθηκαν και κάποια λανθασμένα δείγματα ως Y.

Το Recall υπολογίζεται με βάση την ακόλουθη σχέση:

$$Recall = \frac{TP}{TP + FN} = Sensitivity \quad (4.3)$$

Λόγω του ότι οι δύο αυτές μετρικές αρκετές φορές έχουν αντίστροφη σχέση, συνήθως υπολογίζεται το F1-Score, το οποίο αποτελεί τον αρμονικό μέσο μεταξύ Precision και Recall, και υπολογίζεται με την συνάρτηση:

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4.4)$$

Συμπληρωματικό του Recall είναι το Specificity και μετρά το ποσοστό των αρνητικών δειγμάτων που αναγνωρίζονται σωστά ως αρνητικά. Δηλαδή, όταν το Specificity ισούται με 1.0, το μοντέλο δεν κάνει κανένα λάθος θετικής ταξινόμησης σε αρνητικά παραδείγματα. Είναι σημαντικό όταν το κόστος ενός ψευδώς θετικού (False Positive) είναι υψηλό, όπως σε ιατρικές εφαρμογές.

Ο ορισμός της Specificity δίνεται ως:

$$Specificity = \frac{TN}{TN + FP} \quad (4.5)$$

Τέλος, το Balanced Accuracy συνδυάζει το Sensitivity και το Specificity σε μία ενιαία μετρική, προσφέροντας μια πιο αντιπροσωπευτική αποτίμηση της απόδοσης του μοντέλου όταν τα δεδομένα είναι ανισόρροπα ως προς τις κλάσεις. Είναι ιδιαίτερα χρήσιμο όταν οι θετικές και οι αρνητικές περιπτώσεις εμφανίζονται με πολύ διαφορετική συχνότητα, καθώς δίνει ίσο βάρος σε κάθε μία.

Το Balanced Accuracy ορίζεται ως:

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2} \quad (4.6)$$

Επιπρόσθετα, ιδιαίτερη έμφαση δίνεται στη σύγκριση της απόδοσης του μοντέλου στο σύνολο εκπαίδευσης και στο σύνολο ελέγχου, με στόχο την εκτίμηση της ικανότητας γενίκευσης. Για τον σκοπό αυτό, χρησιμοποιείται το Balanced Accuracy τόσο στα δεδομένα εκπαίδευσης όσο και στα δεδομένα ελέγχου.

Το Train Balanced Accuracy (Train\_BA) αναφέρεται στο Balanced Accuracy που επιτυγχάνει το μοντέλο στο σύνολο εκπαίδευσης. Υψηλές τιμές (>0.90) υποδηλώνουν ότι το μοντέλο έχει καταφέρει να μάθει αποτελεσματικά τα πρότυπα των δεδομένων στα οποία εκπαιδεύτηκε. Ωστόσο, αυτή η μετρική από μόνη της δεν επαρκεί για την αξιολόγηση της ποιότητας του μοντέλου, καθώς ενδέχεται να αντικατοπτρίζει υπερεκπαίδευση (overfitting).

Το Test Balanced Accuracy (Test\_BA) είναι το Balanced Accuracy που υπολογίζεται στο σύνολο ελέγχου, δηλαδή σε δεδομένα χρηστών που δεν χρησιμοποιήθηκαν κατά την εκπαίδευση. Το Test\_BA αποτελεί τον κύριο δείκτη αξιολόγησης της γενίκευσης του μοντέλου, καθώς αποτυπώνει την ικανότητά του να προβλέπει σωστά σε νέα, μη ορασμένα δεδομένα. Όσο υψηλότερη είναι η τιμή της Test\_BA, τόσο πιο αξιόπιστο θεωρείται το μοντέλο σε πραγματικές συνθήκες.

Για την ποσοτική αποτίμηση της διαφοράς μεταξύ της απόδοσης στο train set και στο test set χρησιμοποιείται το Balanced Accuracy Gap (BA\_Gap), που ορίζεται ως η εξής διαφορά:

$$BA_{Gap} = Train_{BA} - Test_{BA} \quad (4.7)$$

Μικρή τιμή του BA\_Gap ( $<0.05$ ) υποδηλώνει ότι το μοντέλο έχει σταθερή και συνεπή συμπεριφορά μεταξύ εκπαίδευσης και ελέγχου. Αντίθετα, μεγάλη τιμή του BA\_Gap ( $>0.10$ ) αποτελεί ένδειξη υπερεκπαίδευσης, όπου το μοντέλο έχει μάθει πολύ καλά τα δεδομένα εκπαίδευσης αλλά αδυνατεί να γενικεύσει σε νέα δείγματα. Επομένως, το BA\_Gap λειτουργεί ως ποσοτικός δείκτης αξιοπιστίας της γενίκευσης, ιδιαίτερα χρήσιμος για τη σύγκριση μεταξύ διαφορετικών αρχιτεκτονικών και παραμετροποιήσεων.

## 5 Ερευνητική Διαδικασία

Η παρούσα ενότητα παρουσιάζει αναλυτικά τη διαδικασία που ακολουθήθηκε για την υλοποίηση των πειραμάτων και την εξαγωγή των αποτελεσμάτων. Περιλαμβάνει τα βασικά τεχνολογικά εργαλεία που χρησιμοποιήθηκαν, τη διαδικασία ανάλυσης και επεξεργασίας των δεδομένων, καθώς και την ανάπτυξη και εκπαίδευση των αλγορίθμων.

### 5.1 Τεχνολογικό Υπόβαθρο

Στο πλαίσιο της παρούσας εργασίας, η υλοποίηση και εκτέλεση των πειραμάτων πραγματοποιήθηκε με τη χρήση της γλώσσας προγραμματισμού Python, σε περιβάλλον Visual Studio Code (VS Code). Η κύρια έκδοση Python που χρησιμοποιήθηκε είναι η 3.13.3, ενώ σε ορισμένα σημεία της υλοποίησης κρίθηκε απαραίτητη η χρήση της έκδοσης 3.11.8, λόγω ασυμβατότητας της βιβλιοθήκης TensorFlow με τις πιο πρόσφατες εκδόσεις.

Η ανάπτυξη της πειραματικής διαδικασίας βασίστηκε σε ένα ευρύ φάσμα βιβλιοθηκών ανοικτού κώδικα. Από τις πλέον βασικές είναι η NumPy, η οποία παρέχει δομές και συναρτήσεις για τη διαχείριση και εκτέλεση πράξεων σε πολυδιάστατους πίνακες, καθώς και εργαλεία για υπολογισμούς γραμμικής άλγεβρας, απαραίτητους για την επεξεργασία αριθμητικών χαρακτηριστικών από αισθητήρες.

Η βιβλιοθήκη Pandas, που βασίζεται στην υποδομή που προσφέρει το NumPy, χρησιμοποιήθηκε για τη διαχείριση των δεδομένων σε μορφή DataFrame, επιτρέποντας την ανάγνωση, φιλτράρισμα, συγχώνευση και μετασχηματισμό των δεδομένων εισόδου. Επιτρέπει τη γρήγορη ανάλυση μεγάλων πινάκων, γεγονός κρίσιμο σε δεδομένα υψηλής διαστασιμότητας, όπως αυτά του ExtraSensory dataset.

Για την οπτικοποίηση των δεδομένων και των αποτελεσμάτων χρησιμοποιήθηκαν οι βιβλιοθήκες Matplotlib και Seaborn. Η πρώτη προσφέρει γραφήματα και πλήρη έλεγχο στη μορφοποίηση, ενώ η δεύτερη προσφέρει πιο αφαιρετικές και αισθητικά βελτιστοποιημένες απεικονίσεις, όπως πυκνότητες κατανομών, συσχετίσεις και χάρτες θερμότητας.

Η διαχείριση αρχείων και χρονικών δεδομένων επιτεύχθηκε μέσω των βιβλιοθηκών os, sys, gzip, datetime, pytz και StringIO. Αυτές αξιοποιήθηκαν για την ανάγνωση συμπιεσμένων αρχείων .csv.gz, τη δημιουργία ροών δεδομένων από string buffers, καθώς και για τον χειρισμό χρονικών σφραγίδων (timestamps) με ακρίβεια ζώνης ώρας.

Στον τομέα της υλοποίησης μοντέλων βαθιάς μάθησης, αξιοποιήθηκε η πλατφόρμα TensorFlow και το υψηλού επιπέδου API της, Keras. Μέσω αυτών αναπτύχθηκαν συνελκτικά (CNN), αναδρομικά (LSTM) και υβριδικά (CNN+LSTM) μοντέλα, με τη χρήση επιπέδων όπως τα Conv1D, LSTM, Dense, Dropout, BatchNormalization, Bidirectional, και το εργαλείο EarlyStopping για την αποτροπή υπερεκπαίδευσης. Τα μοντέλα εκπαιδεύτηκαν σε

διαδοχικά χρονικά παράθυρα (sliding windows) και σχεδιάστηκαν να υποστηρίζουν έξοδο πολλαπλών ετικετών (multi-label classification).

Για την αξιολόγηση της απόδοσης των μοντέλων, χρησιμοποιήθηκαν συναρτήσεις από το πακέτο scikit-learn, όπως οι Accuracy\_score, Precision\_score, Recall\_score, f1\_score, confusion\_matrix, καθώς και το compute\_class\_weight για τη στάθμιση των ετικετών με βάση τη συχνότητα εμφάνισής τους. Οι μετρικές αυτές αξιοποιήθηκαν τόσο σε επίπεδο εκπαίδευσης όσο και σε επίπεδο δοκιμής, με στόχο την αντικειμενική σύγκριση των μοντέλων.

Συνολικά, η επιλογή των βιβλιοθηκών έγινε με γνώμονα τη συμβατότητα, την επεκτασιμότητα και τη σταθερότητα του πειραματικού περιβάλλοντος, ενώ δόθηκε έμφαση στη χρήση τεκμηριωμένων εργαλείων ευρείας αποδοχής στη διεθνή ερευνητική κοινότητα.

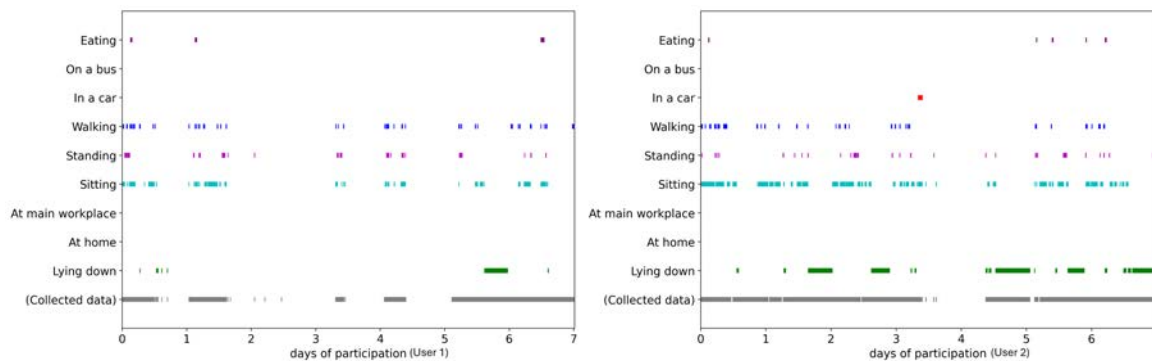
## 5.2 Ανάλυση Δεδομένων

Το στάδιο της ανάλυσης των δεδομένων αποτέλεσε δομικό στοιχείο για την πορεία της παρούσας μελέτης. Εκτός από όσα έχουν ήδη αναφερθεί στην προηγούμενη ενότητα για μια αρχική κατανόηση του συνόλου δεδομένων και πως προέκυψε το σύνολο των παραδειγμάτων που έχουν καταγραφεί, επόμενο σημαντικό βήμα αποτελεί η κατανόηση της κατανομής των δεδομένων χρονικά καθώς επίσης η εξαγωγή αποτελεσμάτων για τη συσχέτιση των δηλωμένων ετικετών με τα αριθμητικά δεδομένα που συλλέγονται από τους αισθητήρες του έκυπνου ρολογιού ή κινητού.

Για την καλύτερη κατανόηση της χρονικής κατανομής των δεδομένων των χρηστών, δημιουργήθηκαν χρονικά διαγράμματα κατανομής ενεργειών (action raster plots). Σε κάθε γραμμή του άξονα y εμφανίζεται μια ετικέτα (label). Ο άξονας x έχει αντιστοιχιστεί στον χρόνο (σε ημέρες). Για κάθε σημείο που εντοπίζεται η ετικέτα σε κάποιο δείγμα, σχεδιάζεται μια κάθετη γραμμή (marker). Η πρώτη σειρά, σημειωμένη με γκρι χρώμα, δείχνει όλα τα παραδείγματα που καταγράφηκαν, παρέχοντας πληροφορία και για το πόσο ενεργός ήταν ένας χρήστης. Γενικά, μέσα από αυτά τα γραφήματα δίνεται η δυνατότητα οπτικής αναγνώρισης μοτίβων συμμετοχής και περιοχών πυκνότητας καταγραφών. Υπάρχει συγκεκριμένη μέρα που ο χρήστης είναι πιο ενεργός ή έχει πιο καθιστική ρουτίνα; Φαίνεται να είναι κάποια συγκεκριμένη μέρα λιγότερο ή περισσότερο ενεργός στην καταγραφή ετικετών; Αυτές είναι κάποιες από τις ερωτήσεις στις οποίες επιχειρείται να δοθεί απάντηση.

Παρακάτω στο Σχήμα 5.1 εμφανίζονται δύο τέτοια διαγράμματα για δύο τυχαίους χρήστες. Μια πρώτη εμφανής διαφορά αφορά το πόσο ενεργοί ήταν οι χρήστες στην καταγραφή ετικετών, με τον χρήστη 2 να είναι πιο ενεργός εκτός από μια μέρα σε αντίθεση με τον χρήστη 1. Άλλη παρατήρηση είναι πως και στις 2 περιπτώσεις μπορούμε να διακρίνουμε στιγμές που έχουν καταγραφεί πολλαπλές ετικέτες. όταν οι ετικέτες “Lying Down” και “Sitting” συμπίπτουν, μπορούμε να θεωρήσουμε ότι προσφέρουν παρόμοιο εννοιολογικό αποτέλεσμα. Ωστόσο, μπορούν να γίνουν διακριτές και χρονικές στιγμές όπου μπορεί να συμπίπτουν ετικέτες αντίθετες, όπως “Sitting” και “Walking”, οι οποίες είναι λογικά ασύμβατες. Η συνύπαρξη αυτή αντανακλά είτε καθυστερημένη ανανέωση από τον χρήστη είτε σφάλμα επισημείωσης, και αποτελεί μία από τις προκλήσεις που σχετίζονται με τη

χρήση in-the-wild δεδομένων.

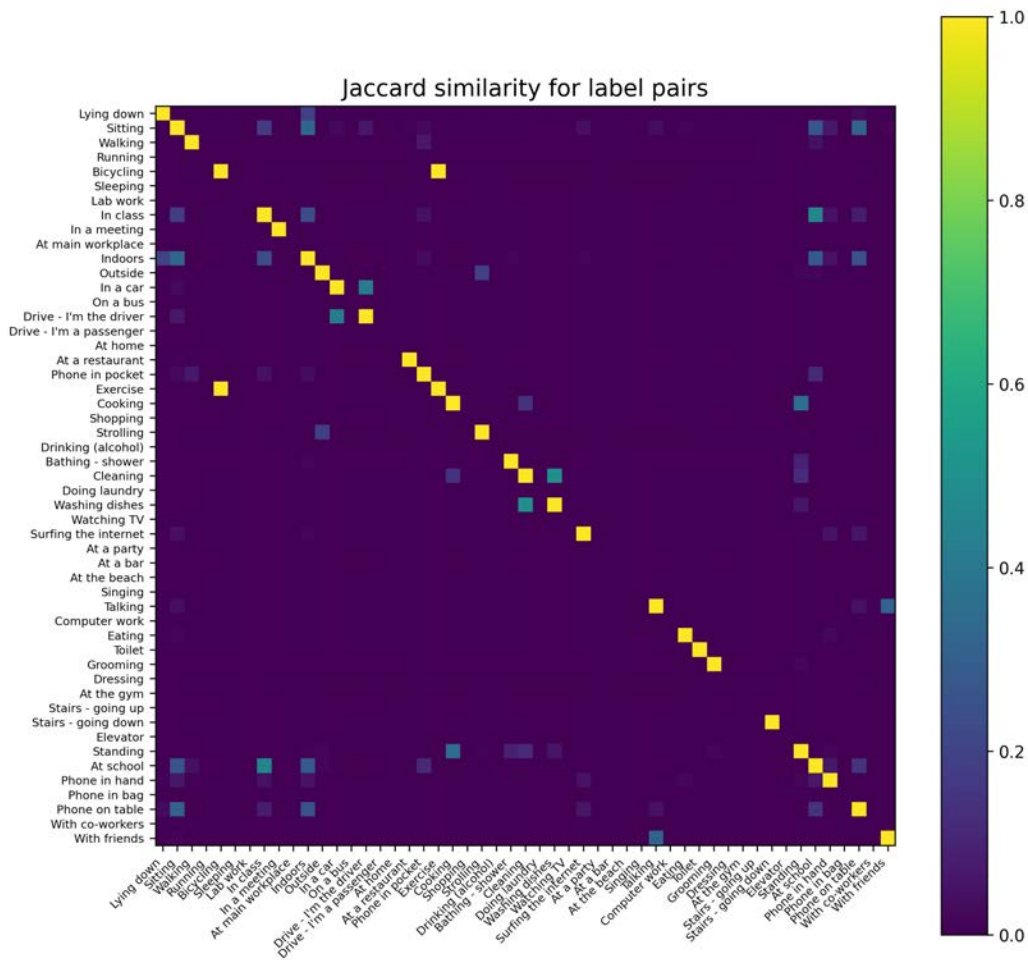


Σχήμα 5.1: Διάγραμμα χρονικής κατανομής ενεργειών για δύο τυχαίους χρήστες (πηγή: ίδια φωτογραφία)

Για την αποτύπωση της συνύπαρξης ετικετών χρησιμοποιήθηκε ο δείκτης ομοιότητας Jaccard (Jaccard similarity index). Ο δείκτης αυτός μετρά τον βαθμό αλληλοεπικάλυψης μεταξύ δύο συνόλων, υπολογίζοντας τον λόγο του πλήθους των κοινών στοιχείων προς το πλήθος των συνολικών (ένωσης) τους όπως φαίνεται στον τύπο:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (5.1)$$

Στην περίπτωση των ετικετών, για κάθε ζεύγος τους  $(i, j)$ , ο υπολογισμός βασίστηκε στην ταυτόχρονη ενεργοποίησή τους στα ίδια χρονικά παραδείγματα. Ο πίνακας που προκύπτει (Jaccard matrix) περιλαμβάνει τιμές μεταξύ 0 και 1, όπου οι τιμές κοντά στο 1 υποδεικνύουν ότι οι δύο ετικέτες συνυπάρχουν συχνά, ενώ οι τιμές κοντά στο 0 δείχνουν ότι εμφανίζονται ανεξάρτητα ή μία από την άλλη. Για την απεικόνιση του πίνακα χρησιμοποιήθηκε θερμικός χάρτης (heatmap), επιτρέποντας την οπτική αναγνώριση ισχυρά συσχετισμένων ετικετών. Η πληροφορία αυτή είναι χρήσιμη τόσο για την κατανόηση της συμπεριφοράς των χρηστών, όσο και για την κατασκευή μοντέλων ταξινόμησης που αξιοποιούν συσχετίσεις μεταξύ labels (multi-label ή multi-task learning). Ο θερμικός χάρτης αυτός παρατίθεται παρακάτω στο Σχήμα 5.2, το οποίο αντιστοιχεί σε όλα τα δεδομένα του χρήστη 2 από το Σχήμα 5.1.



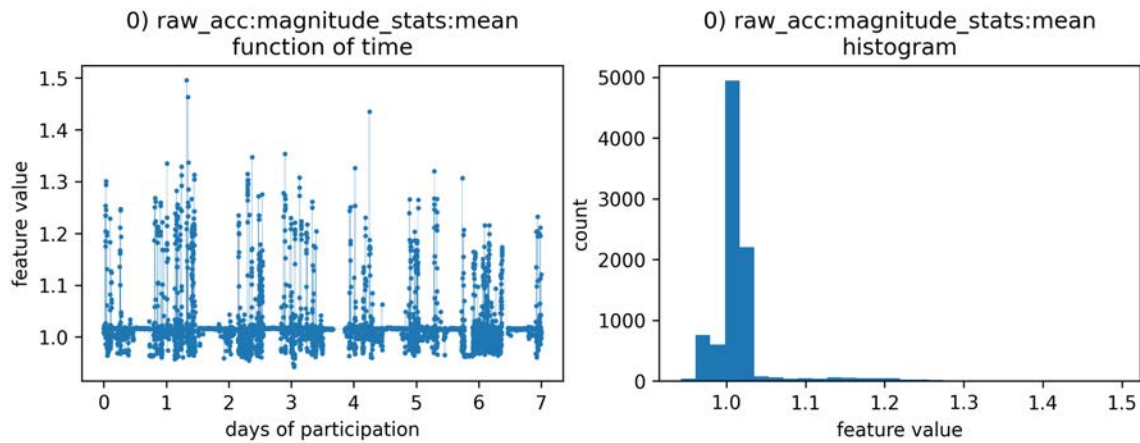
Σχήμα 5.2: Jaccard Heatmap συνύπαρξης ετικετών για έναν χρήστη (πηγή: ίδια φωτογραφία)

Στο Σχήμα 5.2 οι άξονες x και y αντιστοιχούν στις ετικέτες ενώ κάθε χρώμα του διαγράμματος αντιστοιχεί σε έναν αριθμό από το 0 έως το 1. Παρατηρείται ότι όταν x και y έχουν ίδια τιμή το χρώμα είναι αυτό που αντιστοιχεί στο 1 καθώς είναι η ίδια δραστηριότητα ενώ όταν τα σημεία αυτά έχουν το χρώμα 0 τότε φαίνεται πως δεν δηλώθηκε η δραστηριότητα από τον χρήστη και έτσι δεν υπήρχε η δυνατότητα συσχέτισης.

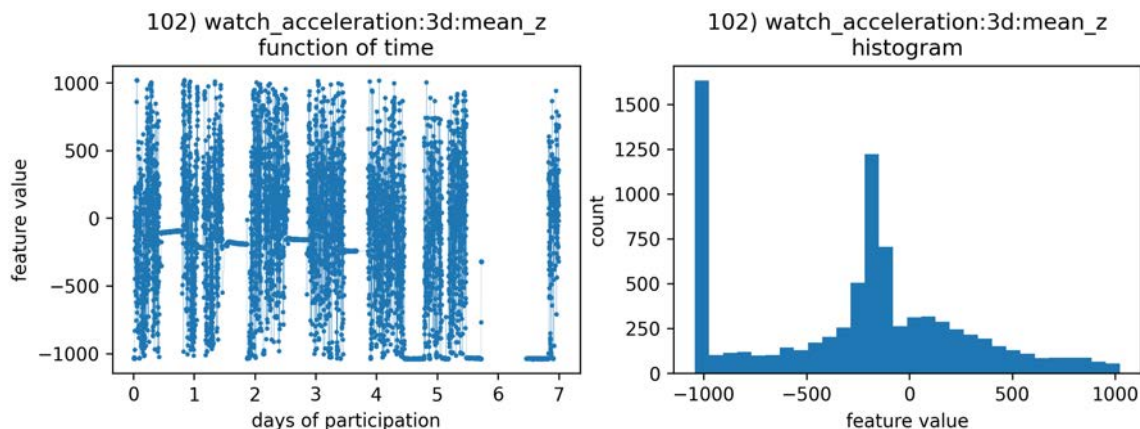
Η άλλη πτυχή ανάλυσης που αναφέρθηκε πιο πάνω αφορά τη σύνδεση μεταξύ των ετικετών και των αισθητήρων που παρείχαν τα δεδομένα κατά την εμφάνισή τους. Η ανάλυση αυτή απαντά στο πολύ βασικό ερώτημα ποιοι αισθητήρες είναι ενεργοί ή σημαντικοί κάθε φορά που παρατηρείται μια συγκεκριμένη ετικέτα. Καθώς η παρούσα μελέτη επικεντρώθηκε περισσότερο στο χαρακτηριστικό της επιτάχυνσης από το επιταχυνσιόμετρο, στα Σχήματα 5.3 και 5.4 δίνονται γραφήματα που αποτυπώνουν την εξέλιξη της τιμής κάθε χαρακτηριστικού κατά τη διάρκεια της συμμετοχής ενός τυχαίου χρήστη και την κατανομή των τιμών του.

Για κάθε επιλεγμένο χαρακτηριστικό του συνόλου δεδομένων, απεικονίστηκε η τιμή του

ως συνάρτηση του χρόνου συμμετοχής του χρήστη (αριστερό γράφημα), καθώς και η συνολική κατανομή συχνοτήτων των τιμών του χαρακτηριστικού (δεξί γράφημα). Η προσέγγιση αυτή επιτρέπει την αναγνώριση χρονικών προτύπων, την εντοπισμό εξάρσεων ή ακραίων τιμών (outliers), καθώς και την εκτίμηση της κατανομής των χαρακτηριστικών (συμμετρία, πολικότητα, κ.ά.).



Σχήμα 5.3: Γραφήματα raw\_acc:magnitude\_stats:mean για έναν χρήστη (πηγή: ίδια φωτογραφία)



Σχήμα 5.4: Γραφήματα watch\_acceleration:3d:mean\_z για έναν χρήστη (πηγή: ίδια φωτογραφία)

Και στα δύο Σχήματα 5.3 και 5.4, στο αριστερό διάγραμμα, ο άξονας x απεικονίζει τις ημέρες συμμετοχής του χρήστη στο πείραμα, ενώ ο άξονας y δείχνει τη μέση τιμή του χαρακτηριστικού σε κάθε δείγμα. Στο δεξί διάγραμμα ο άξονας x απεικονίζει τις τιμές που καταγράφηκαν, ενώ ο άξονας y δείχνει το πλήθος των δειγμάτων με την τιμή αυτή.

Στο Σχήμα 5.3, το χαρακτηριστικό “raw\_acc:magnitude\_stats:mean” εκφράζει τον μέσο όρο του μέτρου της επιτάχυνσης που κατέγραψε ο επιταχυνσιόμετρο του τηλεφώνου (raw

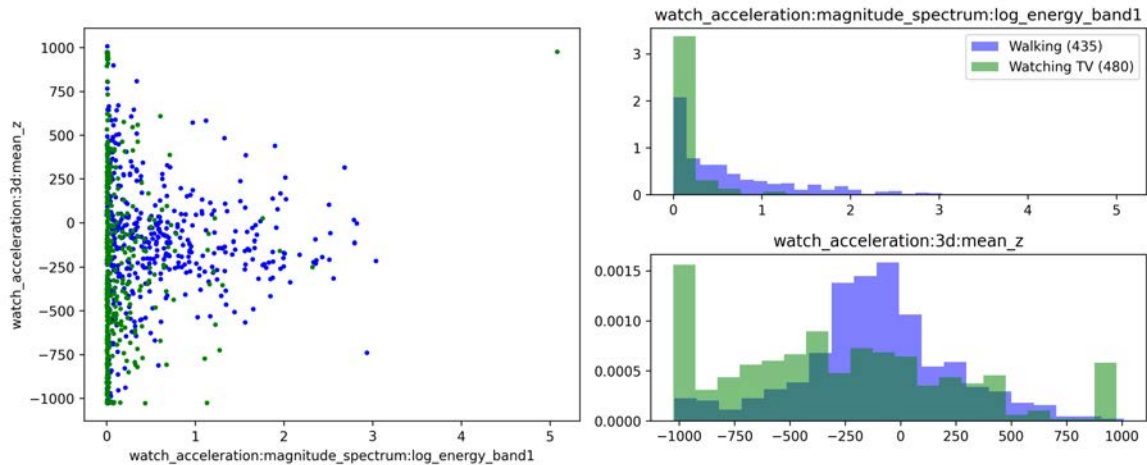
accelerometer) και αντιστοιχεί στην ένταση της κίνησης. Στο αριστερό διάγραμμα παρατηρούνται πολλές αιχμές (peaks), οι οποίες αντιστοιχούν σε φάσεις αυξημένης κινητικότητας (π.χ. περπάτημα, τρέξιμο) και αντίστοιχα υπάρχουν περίοδοι σχετικής σταθερότητας κοντά στο 1, που σχετίζονται με καθιστική ή αδρανή συμπεριφορά. Στο ιστόγραμμα δεξιά, η κατανομή είναι δεξιά-λοξή και η πλειοψηφία των τιμών συγκεντρώνεται λίγο πάνω από το 1, με μακριά ουρά προς προς υψηλότερες τιμές. Η μεγάλη συγκέντρωση κοντά στο 1 οδηγεί στην υπόθεση ότι ο χρήστης είναι συχνά αδρανής (π.χ. καθιστός ή ξαπλωμένος), με την πιο έντονη κίνηση να είναι σπάνια αλλά διακριτή.

Στο Σχήμα 5.4, το χαρακτηριστικό “watch\_acceleration:3d:mean\_z” αφορά τη μέση επιτάχυνση στον άξονα z του επιταχυνσιόμετρου του smartwatch του χρήστη. Η κατεύθυνση z συχνά σχετίζεται με την κάθετη θέση του καρπού. Στο αριστερό διάγραμμα παρατηρείται έντονη διακύμανση μεταξύ υψηλών θετικών και αρνητικών τιμών (π.χ. από +1000 έως -1000), δηλαδή οι τιμές αλλάζουν δραστικά μέσα στην ημέρα, υποδεικνύοντας ενεργές κινήσεις του χεριού. Η περιοδικότητα της κίνησης υποδηλώνει καθημερινή χρήση του ρολογιού, πιθανόν κατά τη διάρκεια εργασιών ή σωματικής δραστηριότητας. Στο ιστόγραμμα δεξιά, η κατανομή παρουσιάζει διπολικότητα, με δύο κυρίαρχες περιοχές: μία θετική και μία αρνητική. Υπάρχει συμμετρία γύρω από το 0, γεγονός που μπορεί να οφείλεται στη φυσική κίνηση του χεριού προς διαφορετικές κατευθύνσεις.

Γενικά, το χαρακτηριστικό “raw\_acc:magnitude\_stats:mean” παρουσιάζει σταθερές χαμηλές τιμές με περιοδικές εξάρσεις, γεγονός που αντιστοιχεί σε φυσική κινητικότητα. Αντίθετα, το χαρακτηριστικό watch\_acceleration:3d:mean\_z παρουσιάζει εντονότερες διακυμάνσεις και διπολική κατανομή, γεγονός που υποδεικνύει αυξημένη ενεργότητα στο χέρι του χρήστη, πιθανόν λόγω χειρονομιών ή μεταφοράς αντικειμένων.

Στη συνέχεια, δημιουργήθηκαν διαγράμματα διασποράς για επιλεγμένα ζεύγη χαρακτηριστικών, ομαδοποιημένα με βάση την ενεργή ετικέτα σε κάθε δείγμα. Τα διαγράμματα συνοδεύονται από κανονικοποιημένα ιστογράμματα για κάθε χαρακτηριστικό, διαχωρισμένα ανά κλάση. Η προσέγγιση αυτή ακολουθήθηκε για να αποκαλύψει διαχωρισιμότητα μεταξύ των ετικετών στον χώρο χαρακτηριστικών, επιτρέποντας την οπτική κατανόηση της συνεισφοράς κάθε αισθητήρα.

Παίρνοντας πιο συγκεκριμένα το ζευγάρι χαρακτηριστικών “Walking” και “Watching TV”, στο Σχήμα 5.5 για έναν τυχαίο χρήστη, το τριπλό αυτό γράφημα εξετάζει τη διαχωρισιμότητα μεταξύ των ετικετών αυτών. Στο διάγραμμα διασποράς (scatter plot) αριστερά, τα περισσότερα παραδείγματα “Watching TV” είναι συγκεντρωμένα στην περιοχή με πολύ χαμηλή φασματική ενέργεια και αρνητικές τιμές επιτάχυνσης στον z-άξονα. Αντίθετα, τα παραδείγματα “Walking” εμφανίζουν μεγαλύτερη διασπορά και φασματική δραστηριότητα. Στο ιστόγραμμα του log\_energy\_band1 (πάνω δεξιά) η ετικέτα “Walking” σχετίζεται με αυξημένη ενέργεια στον φασματικό χώρο, στοιχείο που υποστηρίζει την υπόθεση αυξημένης δραστηριότητας, ενώ στο ιστόγραμμα του mean\_z (κάτω δεξιά) ο διαχωρισμός είναι πιο ενδεικτικός της φυσικής στάσης του χεριού, στατική ή ενεργή.



Σχήμα 5.5: Γραφήματα watch\_acceleration:3d:mean\_z για έναν χρήστη (πηγή: ίδια φωτογραφία)

Γενικότερα λοιπόν, απεικονίζεται η διασπορά και η κατανομή των χρονικών δειγμάτων για τις ετικέτες Walking και Watching TV στον χώρο δύο χαρακτηριστικών του αισθητήρα επιτάχυνσης του smartwatch. Τα παραδείγματα της ετικέτας Watching TV παρουσιάζουν χαμηλή φασματική ενέργεια και σταθερή αρνητική επιτάχυνση στον άξονα z, ένδειξη ακινησίας του χεριού. Αντιθέτως, τα παραδείγματα Walking εμφανίζουν μεγαλύτερη φασματική δραστηριότητα και ευρεία διακύμανση επιτάχυνσης, κάτι που αποδίδεται στη φυσική κίνηση του καρπού κατά τη διάρκεια βάρδισης. Οι δύο κατηγορίες δεν είναι πλήρως διαχωρίσιμες, ωστόσο παρουσιάζουν διακριτά πρότυπα πυκνότητας, υποδηλώνοντας τη χρησιμότητα των επιλεγμένων χαρακτηριστικών για την ταξινόμηση κινητικών και καθιστικών δραστηριοτήτων.

Η παραπάνω ανάλυση ανέδειξε βασικά στατιστικά και χρονικά χαρακτηριστικά των αισθητηριακών δεδομένων, επιτρέποντας την εντοπισμό μοτίβων και συσχετίσεων μεταξύ επιμέρους χαρακτηριστικών και συμφραζομένων ετικετών. Μέσω της οπτικοποίησης και της εξερεύνησης της κατανομής των τιμών, εντοπίστηκαν κρίσιμα σήματα και διαφοροποιήσεις μεταξύ κατηγοριών, οι οποίες μπορούν να αξιοποιηθούν για τη βελτίωση της ακρίβειας στην αναγνώριση δραστηριοτήτων. Οι παρατηρήσεις αυτές καθοδηγούν την επιλογή χαρακτηριστικών και επιβεβαιώνουν τη δυνατότητα εφαρμογής μοντέλων μηχανικής μάθησης για την πρόβλεψη του συμφραζομένου. Στην επόμενη ενότητα, περιγράφεται η υλοποίηση και αξιολόγηση των αλγορίθμων ταξινόμησης που εφαρμόστηκαν στο εν λόγω σύνολο δεδομένων.

### 5.3 Ανάπτυξη και Εκπαίδευση Μοντέλων

Μετά την διερεύνηση και ανάλυση των δεδομένων και την κατανόηση των ιδιοτήτων των αισθητηριακών χαρακτηριστικών, ακολουθεί η υλοποίηση των αλγορίθμων ταξινόμησης με στόχο την σύγκριση της αποδοτικότητά τους στην αυτόματη αναγνώριση δραστηριοτή-

των. Η παρούσα ενότητα παρουσιάζει τα μοντέλα που αναπτύχθηκαν για την πρόβλεψη των ετικετών, τόσο με παραδοσιακές προσεγγίσεις όσο και με τεχνικές βαθιάς μάθησης. Συγκεκριμένα, αξιοποιήθηκαν: (α) η λογιστική παλινδρόμηση (Logistic Regression) ως βασικός γραμμικός ταξινομητής, (β) ένα συνελκτικό νευρωνικό δίκτυο (CNN) για αξιοποίηση τοπικών χωροχρονικών προτύπων, (γ) ένα επαναλαμβανόμενο νευρωνικό δίκτυο τύπου LSTM, σχεδιασμένο να διαχειρίζεται ακολουθίες δεδομένων, και (δ) ένα υβριδικό μοντέλο συνδιασμού των CNN και LSTM, που συνδυάζει τα πλεονεκτήματα των δύο προηγούμενων αρχιτεκτονικών. Για κάθε μοντέλο περιγράφεται η αρχιτεκτονική, η διαδικασία εκπαίδευσης και η αξιολόγηση επί του συνόλου δοκιμής, με στόχο τη σύγκριση των αποτελεσμάτων και την ανάδειξη των καταλληλότερων προσεγγίσεων για το πρόβλημα αναγνώρισης ανθρώπινης δραστηριότητας.

Για όλα τα μοντέλα ακολουθήθηκε η ίδια διαδικασία προετοιμασίας των δεδομένων όπως αναφέρθηκε και στο κεφάλαιο 4. Τα δεδομένα κάθε χρήστη ανακτώνται ξεχωριστά και οργανώνονται σε πίνακες χαρακτηριστικών (X), ετικετών (Y), και μασκών απουσίας ετικετών (M), μαζί με τα αντίστοιχα χρονικά στιγμιότυπα και ονόματα χαρακτηριστικών και ετικετών.

Για όλα τα μοντέλα οι επιλεγμένες ετικέτες είναι οι ίδιες ['FIX\_walking', 'SITTING', 'LOC\_home', 'OR\_outside', 'OR\_indoors'], όπως επίσης χρησιμοποιήθηκαν τα δεδομένα από τους ίδιους αισθητήρες ['Acc', 'WAcc'] που αφορούν την επιτάχυνση. Για κάθε επανάληψη και για κάθε ετικέτα που πρόκειται να προβλεφθεί, ενοποιούνται τα δεδομένα εκπαίδευσης όλων των χρηστών του fold σε ένα κοινό σύνολο δειγμάτων. Η προσέγγιση αυτή επιτρέπει την εκπαίδευση ενός καθολικού μοντέλου ανά ετικέτα, αξιοποιώντας δεδομένα από πολλούς χρήστες, με στόχο τη βελτιστοποίηση της ικανότητας γενίκευσης του μοντέλου και σε δείγματα από άγνωστα άτομα. Στα δεδομένα που προκύπτουν από τους αισθητήρες εφαρμόζεται στάθμιση (standardization) με αφαίρεση του μέσου όρου και διαίρεση με την τυπική απόκλιση των χαρακτηριστικών εκπαίδευσης, τόσο στα δεδομένα που αποτελούν το σύνολο εκπαίδευσης όσο και το σύνολο αξιολόγησης.

Σε όλα τα μοντέλα, για την αξιολόγησή τους, οι μετρικές που υπολογίζονται για κάθε δραστηριότητα και κάθε fold είναι οι Accuracy, Sensitivity, Specificity, Balanced Accuracy, Precision, Recall, F1-Score και Confusion Matrix με τους τύπους που έχουν αναλυθεί και στην υποενότητα 4.5.1.

### 5.3.1 Μοντέλο Logistic Regression

Η λογιστική παλινδρόμηση χρησιμοποιείται ως το πιο απλό γραμμικό μοντέλο που ερευνάται στην παρούσα εργασία και μπορεί να εφαρμοστεί στο πρόβλημα της δυαδικής ταξινόμησης. Αποτελεί ένα εύκολο, κατανοητό και ερμηνεύσιμο μοντέλο που λειτουργεί και ως η κύρια βάση για τη σύγκριση με τις πιο πολύπλοκες προσεγγίσεις που ακολουθούν. Η υλοποίησή του στηρίζεται σε ένα συνεκτικό pipeline που περιλαμβάνει τόσο την προετοιμασία των δεδομένων όσο και την αξιολόγηση της απόδοσης με κατάλληλες μετρικές [38].

Τα βασικά βήματα της προεπεξεργασίας, όπως η επιλογή αισθητήρων (Acc και WAcc), η τυποποίηση χαρακτηριστικών και η διαχείριση των NaN τιμών, έχουν ήδη περιγραφεί εκτε-

νέστερα σε προηγούμενη ενότητα. Αξίζει να επισημανθεί ότι η ομοιογένεια στα δεδομένα που αποτελούν την είσοδο για όλα τα μοντέλα εξασφαλίζει και τη δικαιότερη σύγκριση μεταξύ τους. Η μορφοποίηση εφαρμόζεται με βάση τα στατιστικά του συνόλου εκπαίδευσης, ενώ η επιλογή αισθητήρων περιορίζει τον όγκο των δεδομένων σε ένα υποσύνολο πληροφοριακά πλούσιων σημάτων, πιο σχετικών μεταξύ τους.

Η λογιστική παλινδρόμηση χρησιμοποιεί ως είσοδο τα δεδομένα του συνόλου ExtraSensor, τα οποία αποτελούνται από χρονοσειρές που περιγράφουν τη συμπεριφορά 60 συμμετεχόντων. Κάθε χρήστης διαθέτει  $N$  χρονικά δείγματα, όπου κάθε δείγμα αντιστοιχεί σε διάρκεια ενός λεπτού και συνοδεύεται από  $D \approx 1700$  εξαγόμενα χαρακτηριστικά (features) από αισθητήρες του κινητού τηλεφώνου και του smartwatch.

Τα δεδομένα οργανώνονται μαθηματικά ως:

$$X \in \mathbb{R}^{N \times D}, \quad Y \in \{0, 1\}^{N \times L}, \quad M \in \{0, 1\}^{N \times L} \quad (5.2)$$

όπου  $X$  είναι ο πίνακας χαρακτηριστικών,  $Y$  ο πίνακας των δυαδικών ετικετών (labels) για  $L$  δραστηριότητες και  $M$  ο πίνακας που υποδεικνύει τις τιμές που λείπουν (missing values). Κάθε γραμμή του  $X$  αντιστοιχεί σε ένα χρονικό στιγμιότυπο ενός λεπτού, ενώ κάθε στήλη σε ένα αισθητηριακό χαρακτηριστικό. Οι τιμές του πίνακα  $Y$  λαμβάνουν την τιμή 1 όταν η αντίστοιχη δραστηριότητα ισχύει, 0 όταν δεν ισχύει και NaN όταν η τιμή λείπει (οπότε το  $M_{ij} = 1$ ).

Από τα συνολικά διαθέσιμα χαρακτηριστικά, επιλέγονται μόνο εκείνα που αντιστοιχούν στους αισθητήρες επιταχυνσιόμετρο κινητού (Acc) και επιταχυνσιόμετρο smartwatch (WAcc), μέσω της συνάρτησης `project_features_to_selected_sensors( X, feat_sensor_names, sensors_to_use)` και στη συνέχεια εφαρμόζεται κανονικοποίηση (standardization) ώστε όλα τα χαρακτηριστικά να αποκτήσουν μηδενικό μέσο και μοναδιαία διασπορά. Οι ελλείπουσες τιμές αντικαθίστανται με μηδέν (μέση τιμή).

Η βασική λογική που υλοποιείται στον κώδικα της εργασίας εστιάζει στην απομόνωση του προβλήματος ταξινόμησης για κάθε ετικέτα, π.χ. “FIX\_walking”, και στη συνέχεια στον μετασχηματισμό της σε πρόβλημα δυαδικής πρόβλεψης. Από τον πολυετικετικό πίνακα ετικετών ( $Y$ ), εξάγεται ο σχετικός δυαδικός πίνακας (vector) για την αντίστοιχη ετικέτα. Παράλληλα αγνοούνται τα δείγματα στα οποία η πληροφορία για την συγκεκριμένη ετικέτα θεωρείται ελλιπής, σύμφωνα με τον πίνακα ελλείψεων που προκύπτει ( $M$ ). Με τον τρόπο αυτό εξασφαλίζεται ότι το μοντέλο εκπαιδεύεται μόνο με παραδείγματα που παρέχουν αξιόπιστη πληροφόρηση.

Η εκπαίδευση του μοντέλου πραγματοποιείται με τη χρήση της έτοιμης υλοποίησης της βιβλιοθήκης scikit-learn, την LogisticRegression. Μια βασική σχεδιαστική επιλογή είναι η χρήση της παραμέτρου `class_weight='balanced'`, ώστε να αντιμετωπιστεί το πρόβλημα της έντονης ανισορροπίας κλάσεων, που χαρακτηρίζει τις περισσότερες ετικέτες του συνόλου δεδομένων της έρευνας. Έτσι, δίνεται αυξημένη σημασία στα σπάνια θετικά παραδείγματα, χωρίς να χρειάζεται περαιτέρω εξισορρόπηση του dataset μέσω oversampling ή undersampling. Το εκπαιδευμένο μοντέλο, μαζί με τις παραμέτρους μορφοποίησης και τις σχετικές πληροφορίες, επιστρέφεται ως λεξικό για να εξασφαλιστεί η απαραίτητη ευκολία και στην αξιολόγηση που πραγματοποιείται μέσω 5-fold cross-validation με διαχωρισμό

χρηστών (user-independent folds), όπου κάθε fold περιλαμβάνει διαφορετικό υποσύνολο συμμετεχόντων ως σύνολο ελέγχου.

Συνολικά, το σύνολο δεδομένων περιλαμβάνει περίπου  $N \approx 3 \times 10^6$  λεπτά συμπεριφοράς, με περίπου 400-500 χαρακτηριστικά μετά την επιλογή αισθητήρων και πέντε δραστηριότητες-στόχους.

Κατά το στάδιο του testing, τα δεδομένα επεξεργάζονται με τις ίδιες παραμέτρους (μέσος όρος και διασπορά από το train set) και γίνεται πρόβλεψη με τη χρήση της μεθόδου predict της κλάσης LogisticRegression από τη βιβλιοθήκη sklearn. Οι προβλέψεις συγκρίνονται με τις πραγματικές τιμές και αξιολογούνται μέσω μετρικών που εστιάζουν όχι μόνο στην ακρίβεια αλλά και στην ικανότητα εντοπισμού των θετικών παραδειγμάτων. Η Balanced Accuracy, δηλαδή ο μέσος όρος ανάκλησης και ειδικότητας, προτιμάται ως η κύρια μετρική καθώς λαμβάνει υπόψη της και τις δύο κλάσεις ισάξια, ακόμη και σε περιπτώσεις έντονης ανισορροπίας θετικών/αρνητικών δειγμάτων. Η μετρική F1-Score δίνει συνολική την εικόνα για την ισορροπία ανάμεσα στα Precision και Recall. Η απεικόνιση των προβλέψεων σε σχέση με τα πραγματικά χρονικά στιγμιότυπα ολοκληρώνει την αξιολόγηση, προσφέροντας οπτική επιβεβαίωση της συμπεριφοράς του μοντέλου στον χρόνο.

Η λογιστική παλινδρόμηση, παρά την απλότητά της ως μοντέλο, έχει αποδειχθεί πολύ χρήσιμο εργαλείο για την αρχική αξιολόγηση της πληροφορίας που φέρουν τα χαρακτηριστικά. Με την ευκολία που εμφανίζει στην υλοποίηση, καθώς και την ταχύτητα και την ερμηνευσιμότητά της, αποτελεί αναπόσπαστο μέρος του συστήματος και σημαντικό σημείο αναφοράς για τις επόμενες και πιο σύνθετες προσεγγίσεις.

### 5.3.2 Μοντέλο CNN

Η επιλογή ενός Συνελικτικού Νευρωνικού Δικτύου (Convolutional Neural Network – CNN) για την παρούσα εργασία βασίστηκε στην ανάγκη για αξιοποίηση των χωρικών σχέσεων που μπορούν να βρεθούν στα δεδομένα αισθητήρων (π.χ. τρεις διαστάσεις (x, y, z), κλίση τηλεφώνου κ.α.). Αν και τα CNN χρησιμοποιούνται κυρίως σε προβλήματα εικόνας, θεωρήθηκε αρκετά ενδιαφέρουσα η έρευνα για την απόδοσή του σε χρονικά δεδομένα. Εδώ λοιπόν, προσαρμόζονται επιτυχώς στα πολυδιάστατα χρονικά δεδομένα, επιτρέποντας την ανίχνευση τοπικών μοτίβων στα σήματα από τους αισθητήρες. Η υλοποίηση δίνει έμφαση στη διαχείριση πολλαπλών χρηστών, στον διαχωρισμό σε folds και στη γενική στρατηγική εκπαίδευσης και αξιολόγησης [38].

Το pipeline παραμένει συμβατό με τις υπόλοιπες αρχιτεκτονικές που ερευνήθηκαν, ακολουθώντας τα γνωστά πλέον προκαθορισμένα βήματα προεπεξεργασίας ώστε να διατηρείται η απαραίτητη συνέπεια. Πρώτα επιλέγονται συγκεκριμένοι αισθητήρες (Acc και WAcc), εφαρμόζεται προσαρμογή των δεδομένων και στη συνέχεια αγνοούνται τα παραδείγματα με μη διαθέσιμη πληροφορία για την εκάστοτε ετικέτα. Ωστόσο, σε αντίθεση με τη λογιστική παλινδρόμηση που αναφέρθηκε νωρίτερα, η είσοδος του CNN πρέπει να μορφοποιηθεί ως τανυστής κατάλληλης διάστασης για χρήση από συνελκτικά φίλτρα. Αυτό σημαίνει ότι τα δεδομένα χρειάστηκε να αναδιαμορφωθούν (reshape) ώστε να προσομοιάζουν ακολουθίες (temporal sequences) με κανάλια που αντιστοιχούν στους επιλεγμένους αισθητήρες.

Για την αξιοποίηση της χρονικής πληροφορίας, τα δεδομένα οργανώνονται σε παράθυρα

διαδοχικών δειγμάτων (sliding windows). Η διαδικασία αυτή ομαδοποιεί τις χρονοσειρές κάθε χρήστη σε τμήματα σταθερού μήκους  $W = 20$  χρονικών στιγμών (20 λεπτών), τα οποία μετακινούνται με βήμα  $S = 10$ , δημιουργώντας επικάλυψη 50% μεταξύ διαδοχικών παραθύρων. Αν ο συνολικός αριθμός δειγμάτων είναι  $N$ , τότε ο αριθμός των παραθύρων δίνεται από:

$$M = \frac{N - W}{S} + 1 \quad (5.3)$$

Για παράδειγμα, εάν  $N = 10,000$ , τότε  $M = 999$  παράθυρα. Κάθε παράθυρο έχει διαστάσεις:

$$X_{\text{win}} \in \mathbb{R}^{W \times D'} \quad (5.4)$$

όπου  $D'$  είναι ο αριθμός χαρακτηριστικών μετά την επιλογή των αισθητήρων, περίπου  $D' \approx 450$ . Η ετικέτα κάθε παραθύρου ορίζεται βάσει του κανόνα “at least one”, δηλαδή το παράθυρο λαμβάνει τιμή 1 εφόσον τουλάχιστον ένα από τα 20 χρονικά δείγματα έχει θετική ετικέτα για τη συγκεκριμένη δραστηριότητα. Με αυτόν τον τρόπο διατηρείται η πληροφορία ύπαρξης μιας συμπεριφοράς μέσα σε κάθε χρονικό τμήμα.

Η συνάρτηση `train_cnn_model` είναι αυτή που περιγράφει την εκπαίδευση του συγκεκριμένου νευρωνικού δικτύου, χρησιμοποιώντας τα δεδομένα που προέρχονται από πολυάριθμους χρήστες οι οποίοι έχουν επιλεγεί σύμφωνα με τις λίστες των UUIDs ανά fold. Τα δεδομένα όλων των χρηστών στο training set ενοποιούνται, με στόχο τη δημιουργία ενός συνεκτικού συνόλου εκπαίδευσης. Κατά την κατασκευή του μοντέλου, χρησιμοποιείται ένα απλό CNN architecture με δύο επίπεδα convolution και pooling, ακολουθούμενα από fully connected layers και μία τελική έξοδο με ενεργοποίηση sigmoid για δυαδική πρόβλεψη.

Το νευρωνικό δίκτυο εκπαιδεύεται για κάθε ετικέτα ξεχωριστά, καθώς το πρόβλημα ορίζεται ως σειρά από δυαδικές ταξινομήσεις. Η απώλεια υπολογίζεται με binary cross-entropy, ενώ η εκπαίδευση υλοποιείται μέσω του optimizer Adam (Adaptive Moment Estimation), ο οποίος προσαρμόζει δυναμικά τον ρυθμό μάθησης για κάθε παράμετρο. Εφαρμόζεται dropout για την αποτροπή overfitting και early stopping ανάλογα με την απόδοση σε validation set, το οποίο εξάγεται ως υποσύνολο του training set. Το μοντέλο εκπαιδεύεται για έως 30 epochs με batch size = 64, ώστε να παραμένει ελαφρύ και να αποφεύγεται η υπερεκμάθηση για το συγκεκριμένο σύνολο, αφού πρέπει να ληφθεί υπόψιν και ο περιορισμένος αριθμός θετικών δειγμάτων για κάποιες από τις ετικέτες.

Η αρχιτεκτονική του δικτύου περιλαμβάνει δύο συνελκτικά επίπεδα (Conv1D) με 64 και 32 φίλτρα αντίστοιχα, μέγεθος πυρήνα 3, και ενεργοποίηση ReLU, ακολουθούμενα από επίπεδο Global Max Pooling και δύο fully connected layers με ενεργοποίηση ReLU και dropout 0.3. Η έξοδος του δικτύου έχει έναν νευρώνα με ενεργοποίηση sigmoid.

Στο στάδιο της αξιολόγησης, η συνάρτηση `test_cnn_model` εφαρμόζει την ίδια στρατηγική μετασχηματισμού στα δεδομένα του testing set, χρησιμοποιώντας τα στατιστικά (μέσο και διασπορά) του συνόλου εκπαίδευσης, και τροφοδοτεί τα κανονικοποιημένα χαρακτηριστικά στο εκπαιδευμένο νευρωνικό δίκτυο. Οι έξοδοι του μοντέλου μετατρέπονται σε δυαδικές προβλέψεις με χρήση κατωφλίου (threshold = 0.5) και συγκρίνονται με τις πραγματικές ετικέτες. Οι μετρικές που υπολογίζονται είναι ίδιες με εκείνες της logistic regression:

Recall, Specificity, Balanced Accuracy, F1-Score κ.α., παρέχοντας ένα πλούσιο σύνολο για συγκριτική αξιολόγηση.

Τα αποτελέσματα συλλέγονται ανά fold και αποθηκεύονται τόσο σε αρχεία CSV όσο και ως πίνακες NumPy, διατηρώντας έτσι τη δυνατότητα περαιτέρω ανάλυσης ή απεικόνισης. Επιπλέον, για κάθε ετικέτα, παράγονται συνολικά predictions και ground truths επιτρέποντας μετά τη μελέτη χρονικών μοτίβων.

Το CNN μοντέλο λοιπόν αποδεικνύεται ικανό να συλλάβει δομές στα δεδομένα εισόδου που παραδοσιακά μοντέλα δεν μπορούν να εντοπίσουν, ενώ παραμένει πιο αποδοτικό σε σχέση με τα LSTM ως προς την ταχύτητα εκπαίδευσης. Λόγω της συνελκτικής φύσης του μπορεί να λειτουργεί ως ενδιάμεση λύση μεταξύ απλών και χρονικά συνειδητών αρχιτεκτονικών, καθιστώντας το κατάλληλο για συνδυασμό με άλλες μορφές δικτύων, όπως θα φανεί παρακάτω όπου μελετάται ένα υβριδικό μοντέλο CNN-LSTM.

### 5.3.3 Μοντέλο LSTM

Η έρευνα για τη χρήση ενός Long Short-Term Memory (LSTM) δικτύου στηρίζεται στην ανάγκη αξιοποίησης της χρονικής διαδοχής και της εξάρτησης μεταξύ στιγμών που παρατηρήθηκε ότι υπάρχουν στο πρόβλημα της αναγνώρισης δραστηριοτήτων. Σε αντίθεση με τις συνελκτικές και γραμμικές μεθόδους που αναφέρθηκαν νωρίτερα, τα LSTM δίκτυα έχουν την ικανότητα να διατηρούν πληροφορία από προηγούμενα χρονικά βήματα, κάτι που μπορεί να είναι καθοριστικό σε δεδομένα αισθητήρων που μεταβάλλονται με τον χρόνο. Η υλοποίηση που επιλέχθηκε προσαρμόζει την αρχιτεκτονική του LSTM στις απαιτήσεις του συγκεκριμένου dataset, διατηρώντας παράλληλα τη συνοχή με τα υπόλοιπα μοντέλα της εργασίας [38].

Όπως και στα προηγούμενα μοντέλα, η είσοδος στο LSTM προκύπτει από την επιλογή συγκεκριμένων αισθητήρων (Acc και WAcc), καθώς και από την προσαρμογή των χαρακτηριστικών σε επίπεδο training set. Αν κάτι μπορεί να θεωρηθεί πρωτότυπο, αυτό αφορά τον τρόπο με τον οποίο μετασχηματίζεται το dataset ώστε να μπορεί να υποστηρίξει χρονικές ακολουθίες. Για κάθε χρονική στιγμή δημιουργείται ένα παράθυρο δεδομένων (sliding window), το οποίο αποτελεί ακολουθία χαρακτηριστικών από διαδοχικές χρονικές στιγμές. Ο αριθμός των βημάτων προς τα πίσω (window size) είναι παραμετροποιήσιμος και καθορίζει πόση «μνήμη» διαθέτει το μοντέλο για την πρόβλεψη της τρέχουσας ετικέτας. Στην παρούσα εργασία χρησιμοποιήθηκε μήκος παραθύρου  $W = 20$  χρονικών στιγμών και βήμα μετατόπισης  $S = 10$ , ώστε να υπάρχει επικάλυψη 50% μεταξύ διαδοχικών παραθύρων. Κάθε παράθυρο περιγράφεται μαθηματικά όπως φαίνεται στην παρακάτω σχέση:

$$X_{\text{win}} \in \mathbb{R}^{W \times D'} \quad (5.5)$$

όπου  $D' \approx 450$  είναι ο αριθμός χαρακτηριστικών μετά την επιλογή αισθητήρων. Κάθε παράθυρο χαρακτηρίζεται ως θετικό εάν τουλάχιστον ένα από τα δείγματά του αντιστοιχεί σε θετική ετικέτα. Τα δεδομένα όλων των χρηστών στο training fold ενοποιούνται, δημιουργώντας το τελικό σύνολο εκπαίδευσης.

Η συνάρτηση `train_lstm_model` περιγράφει τη διαχείριση της διαδικασίας εκπαίδευσης. Αρχικά, ενώνει τα δεδομένα από όλους τους χρήστες του εκάστοτε training fold και τα μετατρέπει σε μια μορφή κατάλληλη για χρονική επεξεργασία. Αυτό σημαίνει ότι κάθε δείγμα δεν είναι πια ένα στιγμιότυπο με χαρακτηριστικά, αλλά μια ακολουθία στιγμιότυπων με χρονική δομή. Το ίδιο ισχύει και για τις ετικέτες, οι οποίες αντιστοιχούν στη χρονική στιγμή που ακολουθεί την ακολουθία.

Το μοντέλο LSTM αποτελείται από ένα LSTM επίπεδο (LSTM layer με διάσταση 64), το οποίο ακολουθείται από πλήρως συνδεδεμένα επίπεδα (Dense) και ένα τελικό επίπεδο πρόβλεψης με ενεργοποίηση sigmoid. Η επιλογή της binary cross-entropy ως loss function και η χρήση ξανά του Adam optimizer διατηρούν τη συμβατότητα με τις υπόλοιπες αρχιτεκτονικές. Κατά την εκπαίδευση εφαρμόζεται dropout = 0.3 για αποτροπή υπερεκπαίδευσης, που θα μπορούσε να οδηγήσει σε λανθασμένα αποτελέσματα προς σύγκριση, και χρησιμοποιείται επίσης early stopping (μια τεχνική regularization) με βάση την επίδοση στο validation set, ιδιαίτερα για ετικέτες με πιο αραιή εμφάνιση, έτσι ώστε να σταματήσει η εκπαίδευση πριν το μοντέλο αρχίσει να υπερεκπαιδεύεται. Το δίκτυο εκπαιδεύεται και πάλι για έως 30 epochs, με batch size = 64, ώστε να επιτυγχάνεται ισορροπία μεταξύ υπολογιστικού κόστους και γενίκευσης.

Κατά τη φάση του testing, τα δεδομένα χρήστη του test set υποβάλλονται στις ίδιες μετατροπές όπως παραπάνω στα προηγούμενα μοντέλα και τροφοδοτούνται στο εκπαιδευμένο δίκτυο. Η συνάρτηση `test_lstm_model` εξάγει τις προβλέψεις, οι οποίες μετατρέπονται σε δυαδικές τιμές μέσω ενός κατωφλίου (threshold = 0.5). Η σύγκριση με τις πραγματικές τιμές γίνεται με τις ίδιες μετρικές αξιολόγησης που χρησιμοποιούνται σε όλα τα μοντέλα: Recall, Specificity, Balanced Accuracy, F1-Score κ.α. Τα αποτελέσματα αποθηκεύονται για κάθε fold ξεχωριστά σε αρχεία .csv και .npy, διατηρώντας τη δυνατότητα επαναληψιμότητας και σύγκρισης, με τον ίδιο τρόπο όπως και στα προηγούμενα μοντέλα.

Ιδιαίτερη σημασία έχει το γεγονός ότι το LSTM προσφέρει αυξημένη ικανότητα μοντελοποίησης της χρονικής εξάρτησης, κάτι που φαίνεται να έχει τη δυνατότητα να βελτιώσει αρκετά την επίδοση σε ετικέτες όπου η χρονική συνέχεια έχει σημασία (όπως “walking”, “sitting”, ή “at home”). Παράλληλα, αποτελεί και δομικό στοιχείο για το υβριδικό μοντέλο CNN-LSTM που ακολουθεί, όπου η συνελικτική επεξεργασία συνδυάζεται με την χρονική δυναμική των LSTM. Έτσι, το LSTM αναδεικνύεται όχι μόνο ως αυτόνομο εργαλείο πρόβλεψης, αλλά και ως κρίσιμο δομικό μέρος για πιο σύνθετες αρχιτεκτονικές.

### 5.3.4 Υβριδικό Μοντέλο CNN-LSTM

Το υβριδικό μοντέλο CNN και LSTM σχεδιάστηκε με σκοπό να συνδυάσει τα πλεονεκτήματα δύο διαφορετικών τύπων νευρωνικών δικτύων, από τη μία πλευρά τις συνελικτικές ιδιότητες του CNN, που ανιχνεύουν τοπικά πρότυπα στα χαρακτηριστικά, και από την άλλη, τη χρονική μνήμη του LSTM, που μπορεί να κάνει αντιληπτές εξαρτήσεις στο χρόνο. Το τελικό αποτέλεσμα είναι μια αρχιτεκτονική που επιτρέπει την ταυτόχρονη εξαγωγή χωρικής και χρονικής πληροφορίας, κάτι κρίσιμο σε δεδομένα αισθητήρων, όπου η συμπεριφορά του χρήστη εκφράζεται μέσω πολύπλοκων χρονικά εξελισσόμενων μοτίβων.

Η βασική λογική πίσω από την υλοποίηση ξεκινά με τη μετατροπή των δεδομένων εισόδου σε χρονικές ακολουθίες [38]. Όπως στο απλό δίκτυο LSTM, τα δεδομένα ομαδοποιούνται σε ακολουθίες παραθύρων (sliding windows), όπου κάθε παράθυρο περιλαμβάνει μια σειρά διαδοχικών στιγμών με τα αντίστοιχα χαρακτηριστικά. Το μήκος του παραθύρου έχει οριστεί σε  $W = 20$  χρονικές στιγμές και το βήμα μετατόπισης σε  $S = 10$ , δημιουργώντας επικάλυψη 50% μεταξύ διαδοχικών παραθύρων και σε αυτό το μοντέλο. Η είσοδος κάθε παραθύρου έχει διαστάσεις:

$$X_{\text{win}} \in \mathbb{R}^{W \times D'} \quad (5.6)$$

όπου  $D' \approx 450$  είναι ο αριθμός χαρακτηριστικών μετά την επιλογή αισθητήρων (Acc και WAcc). Η ετικέτα κάθε παραθύρου χαρακτηρίζεται ως θετική αν τουλάχιστον ένα από τα δείγματά του αντιστοιχεί σε θετική τιμή της συγκεκριμένης δραστηριότητας. Τα δεδομένα των χρηστών του training fold ενοποιούνται σε ένα κοινό σύνολο εκπαίδευσης.

Πριν όμως η ακολουθία εισαχθεί στο LSTM, περνά πρώτα από συνελκτικά φίλτρα, τα οποία εφαρμόζονται χρονικά, δηλαδή κατά μήκος του άξονα της ακολουθίας. Αυτό επιτρέπει την ανίχνευση σύντομων, τοπικών μεταβολών στο σήμα, όπως για παράδειγμα αλλαγές σε επιταχύνσεις ή ενδείξεις από τον αισθητήρα ρολογιού που υποδηλώνουν μετακίνηση ή παύση.

Η συνελκτική φάση αποτελείται από δύο επίπεδα 1D convolution με 64 και 32 φίλτρα αντίστοιχα, μέγεθος πυρήνα 3, ενεργοποίηση ReLU και dropout 0.3. Στη συνέχεια εφαρμόζεται ένα επίπεδο MaxPooling1D για μείωση της διάστασης, και τα feature maps που προκύπτουν εισάγονται σε ένα επίπεδο LSTM με 64 κρυφά επίπεδα, το οποίο επεξεργάζεται πλέον τη «συμπυκνωμένη» και εννοιολογικά εμπλουτισμένη χρονική ακολουθία. Με αυτόν τον τρόπο, το LSTM δεν λειτουργεί πάνω σε ακατέργαστα δεδομένα, αλλά σε αναπαραστάσεις ανώτερου επιπέδου, κάτι που ενισχύει την αποδοτικότητά του.

Η συνολική αρχιτεκτονική του μοντέλου μπορεί να περιγραφεί μαθηματικά ως:

$$\hat{y} = \sigma(f_{\text{LSTM}}(f_{\text{CNN}}(X_{\text{win}}; \theta_1); \theta_2)) \quad (5.7)$$

όπου  $f_{\text{CNN}}$  και  $f_{\text{LSTM}}$  αντιπροσωπεύουν τις επιμέρους υπομονάδες του υβριδικού μοντέλου, και  $\theta_1, \theta_2$  τα σύνολα παραμέτρων τους. Η έξοδος του LSTM τροφοδοτείται σε ένα πλήρως συνδεδεμένο επίπεδο (Dense) με 32 νευρώνες και ενεργοποίηση ReLU, ενώ η τελική έξοδος είναι ένας νευρώνας με ενεργοποίηση sigmoid για δυαδική ταξινόμηση. Η εκπαίδευση του δικτύου πραγματοποιείται με συνάρτηση κόστους binary cross-entropy και optimizer Adam, ενώ εφαρμόζονται τεχνικές dropout και early stopping για αποτροπή υπερεκμάθησης. Ο αριθμός των epochs περιορίστηκε σε 30, με batch size = 64, ώστε να επιτυγχάνεται ισορροπία μεταξύ ακρίβειας και υπολογιστικής αποδοτικότητας.

Η συνάρτηση train\_cnn\_lstm\_model ενώνει τα δεδομένα των χρηστών του training set, εφαρμόζει τις απαραίτητες μετατροπές και τροφοδοτεί τις ακολουθίες στο υβριδικό δίκτυο. Το μοντέλο κατασκευάζεται δυναμικά για κάθε ετικέτα, επιτρέποντας εκπαίδευση σε ξεχωριστά δυαδικά προβλήματα. Το validation set εξάγεται από το training set και χρησιμοποιείται για την παρακολούθηση της απόδοσης ανά epoch. Στο στάδιο της αξιολόγησης, η συνάρτηση test\_cnn\_lstm\_model εφαρμόζει την ίδια διαδικασία προεπεξεργασίας στα

δεδομένα του test set και εξάγει τις προβλέψεις. Οι έξοδοι μετατρέπονται σε δυαδικές τιμές μέσω κατωφλίου 0.5 και συγκρίνονται με τις πραγματικές ετικέτες. Τα αποτελέσματα αποθηκεύονται ξανά σε αρχεία ανά fold, επιτρέποντας τη μεταγενέστερη στατιστική ανάλυση. Όπως και στα υπόλοιπα μοντέλα, η βασική μετρική αξιολόγησης είναι η Balanced Accuracy, ενώ καταγράφονται και τα Recall, Specificity, F1-Score κ.α.

Η υλοποίηση του CNN και LSTM αποτελεί το πιο σύνθετο αρχιτεκτονικά σημείο της εργασίας, αλλά ανταμείβει με την ικανότητά του να μοντελοποιεί τόσο τις στιγμιαίες διαφοροποιήσεις των σημάτων όσο και την ευρύτερη χρονική δυναμική. Η προσέγγιση αυτή είναι ιδανική για δραστηριότητες που παρουσιάζουν τόσο σταθερότητα στο χρόνο όσο και εσωτερικές μικρομεταβολές, όπως το περπάτημα ή η εναλλαγή θέσης τηλεφώνου. Επομένως, το CNN-LSTM δεν είναι απλώς ένας συνδυασμός δύο τεχνικών, αλλά μια ενοποιημένη αρχιτεκτονική που αξιοποιεί τα δεδομένα αισθητήρων σε πολλαπλά επίπεδα, συνδυάζοντας τη δύναμη της συνελκτικής ανάλυσης με τη χρονική μνήμη των LSTM.

## 6 Αποτελέσματα

Η απόδοση όλων των μοντέλων αξιολογήθηκε ανά ετικέτα με χρήση των μετρικών που αναφέρθηκαν στην παρούσα εργασία, όπως ακρίβεια (Accuracy), ανάκληση (Recall), ειδικότητα (Specificity), Precision, F1-Score και Balanced Accuracy. Για τη διερεύνηση των σχέσεων μεταξύ των προβλεπόμενων ετικετών και πόσο κοντά βρίσκονται στις αντίστοιχες πραγματικές ετικέτες, υπολογίστηκε ο δείκτης ομοιότητας Jaccard σε επίπεδο ζευγών, τόσο για τις προβλεπόμενες όσο και για τις πραγματικές ετικέτες.

Κατά την εκπαίδευση των μοντέλων εφαρμόστηκε σε όλα 5-Fold Cross Validation έτσι ώστε να διασφαλιστεί η ασφαλής γενίκευση των αποτελεσμάτων και να μετρηθεί η σταθερότητά τους σε διαφορετικές κατανομές των δεδομένων. Για κάθε fold καταγράφηκαν οι μετρικές απόδοσης ανά ετικέτα και η διακύμανση των αποτελεσμάτων ανά fold ήταν περιορισμένη στις περισσότερες ετικέτες, το οποίο δείχνει ότι διατηρείται μια συνεπής απόδοση ανεξαρτήτως της επιλογής training και validation συνόλων. Εξαίρεση αποτελούν ορισμένες ετικέτες οι οποίες περιγράφουν καταστάσεις λιγότερο ξεκάθαρες, όπως το “Walking” και το “Outside”, όπου παρατηρούνται ελαφρώς αυξημένες αποκλίσεις, κάτι που μπορεί να αποδοθεί στη φυσική μεταβλητότητα αυτών των δραστηριοτήτων και στην επικαλυπτόμενη πληροφορία που καταγράφουν οι αισθητήρες. Η χρήση της Balanced Accuracy ως τη βασική μετρική επέτρεψε την ισορροπημένη αξιολόγηση μεταξύ πιο συχνών και σπάνιων ετικετών, περιορίζοντας την πιθανότητα υπερβολικής προσαρμογής των μοντέλων στο συγκεκριμένο σύνολο δεδομένων [38].

Στις ενότητες που ακολουθούν οι τιμές των μετρικών έχουν στρογγυλοποιηθεί σε δύο δεκαδικά ψηφία για λόγους ομοιομορφίας και ευκολίας σύγκρισης κατά την ανάγνωση του κειμένου, όμως οι ακριβείς τιμές αναγράφονται στον εκάστοτε πίνακα αποτελεσμάτων για το κάθε μοντέλο.

### 6.1 Αποτελέσματα Μοντέλου Logistic Regression

Το μοντέλο Logistic Regression εφαρμόστηκε ως ένας γραμμικός ταξινομητής για το πρόβλημα της πολυετικετικής πρόβλεψης στα δεδομένα δραστηριοτήτων και καταστάσεων των χρηστών και παρακάτω παρατίθενται τα αποτελέσματα του τα οποία στη συνέχεια θα συγκριθούν με τα υπόλοιπα μοντέλα.

Τα αποτελέσματα δείχνουν μια υψηλή διακύμανση ανά ετικέτα, γεγονός που αντανακλά τις διαφορετικές προκλήσεις που μπορεί να παρουσιάζει η κάθε δραστηριότητα για το μοντέλο. Η ετικέτα “Indoors” φαίνεται να έχει τις καλύτερες επιδόσεις, με τιμές περίπου Balanced Accuracy  $\approx 0.78$ , Recall  $\approx 0.86$  και εντυπωσιακό Precision  $\approx 0.98$ . Αυτό σημαίνει ότι το μοντέλο μπορεί να προβλέπει με ιδιαίτερη ακρίβεια τις περιπτώσεις που ο χρή-

στης βρίσκεται εντός κτιρίου, με ελάχιστα ψευδώς θετικά, αποδίδοντας και ένα εξαιρετικό F1-Score  $\approx 0.91$ . Παρόμοια καλή απόδοση καταγράφεται στην ετικέτα “Sitting”, με τιμές Balanced Accuracy  $\approx 0.67$  και F1-Score  $\approx 0.64$ , δείχνοντας ότι το μοντέλο καταφέρνει να διακρίνει ικανοποιητικά την κατάσταση “Sitting”, με Precision  $\approx 0.64$  και Recall  $\approx 0.63$ .

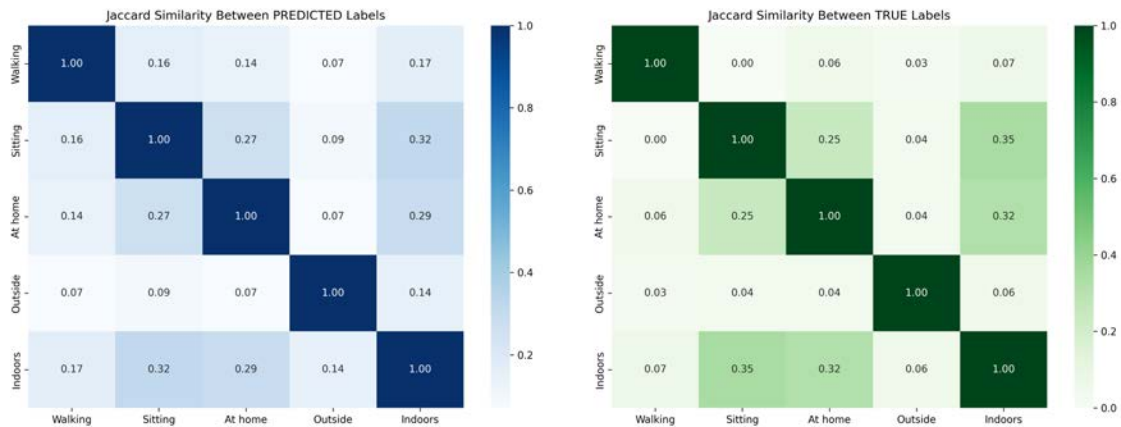
Πίνακας 6.1: Πίνακας τελικών τιμών μετρικών για το μοντέλο Logistic Regression

| Ετικέτα | Accuracy | Balanced_Acc | Precision | Recall  | F1-Score | Specificity |
|---------|----------|--------------|-----------|---------|----------|-------------|
| Walking | 0.83636  | 0.77152      | 0.26576   | 0.69532 | 0.38132  | 0.84772     |
| Sitting | 0.67650  | 0.67092      | 0.64246   | 0.62962 | 0.63554  | 0.71220     |
| At home | 0.62296  | 0.62916      | 0.54808   | 0.67010 | 0.60054  | 0.58820     |
| Outside | 0.83464  | 0.77330      | 0.28734   | 0.70122 | 0.40306  | 0.84542     |
| Indoors | 0.84682  | 0.78018      | 0.97852   | 0.85512 | 0.91260  | 0.70524     |

Αντιθέτως, οι ετικέτες “Walking” και “Outside” αποτελούν σημεία αδυναμίας για το μοντέλο. Για την ετικέτα “Walking”, η τιμή για τη μετρική Recall είναι σχετικά υψηλή (0.70), όμως η τιμή της Precision βρίσκεται στο πολύ χαμηλό νούμερο του 0.27. Αυτό συνεπάγεται μεγάλη ποσότητα ψευδώς θετικών προβλέψεων με αποτέλεσμα και χαμηλό F1-Score  $\approx 0.38$ . Παρόμοια προβληματική είναι η εικόνα για την ετικέτα “Outside”, όπου η τιμή Precision φτάνει μόλις στο 0.29, παρά το υψηλό Recall  $\approx 0.70$ . Αυτή η συμπεριφορά υποδηλώνει ότι το μοντέλο εντοπίζει πολλές πιθανές στιγμές των δραστηριοτήτων αυτών, αλλά συχνά μπερδεύει τα πρότυπά τους με άλλες ετικέτες.

Για την ετικέτα “At home”, η απόδοση είναι μέτρια, με Balanced Accuracy  $\approx 0.63$  και F1-Score  $\approx 0.60$ . Η τιμή του Recall αγγίζει το 0.67, κάτι που δείχνει ότι το μοντέλο καταφέρνει να εντοπίσει τις περισσότερες στιγμές παραμονής στο σπίτι, όμως η τιμή της Precision είναι πιο περιορισμένη (0.55), που σημαίνει πως και εδώ υπάρχει ένα μέσο ποσοστό λανθασμένων θετικών προβλέψεων.

Στους πίνακες Jaccard, που εμφανίζονται στο Σχήμα 6.1, οι προβλέψεις του μοντέλου δείχνουν πιο έντονες συσχετίσεις απ’ ό,τι οι πραγματικές ετικέτες. Συγκεκριμένα, η μεγαλύτερη ομοιότητα εντοπίζεται μεταξύ “Sitting” και “Indoors” (0.32), καθώς και μεταξύ “At home” και “Indoors” (0.29). Αυτό επιβεβαιώνει τη λογική συνύπαρξη αυτών των καταστάσεων, όπως αυτές καταγράφονται και στο ground truth. Ωστόσο, παρατηρείται και μια μέτρια συσχέτιση μεταξύ “Walking” και “Indoors” (0.17) η οποία δεν υπάρχει στον πίνακα των πραγματικών ετικετών, κάτι που υποδηλώνει μια σύγχυση του μοντέλου, πιθανότατα σε περιπτώσεις μετακίνησης εντός του σπιτιού που δεν ταξινομούνται ως “walking”.



Σχήμα 6.1: Διαγράμματα Jaccard για το μοντέλο Logistic Regression σε Predicted και True ετικέτες (πηγή: ίδια φωτογραφία)

Συγκρίνοντας με τον πίνακα Jaccard των πραγματικών ετικετών, βλέπουμε ότι η μέγιστη συσχέτιση υπάρχει ξανά μεταξύ “Sitting” και “Indoors” (0.35), και “At home” με “Indoors” (0.32). Οι πραγματικές ετικέτες παρουσιάζουν λιγότερες επικαλύψεις με την ετικέτα “Walking”, γεγονός που οδηγεί στο συμπέρασμα ότι το μοντέλο Logistic Regression υπερεκτιμά την παρουσία αυτής της ετικέτας.

Συνολικά, το μοντέλο Logistic Regression επιτυγχάνει ικανοποιητικές επιδόσεις σε σαφώς καθορισμένες και στατικές δραστηριότητες, όπως το “Indoors” και το “Sitting”, ενώ δυσκολεύεται με πιο δυναμικές ή ενδιάμεσες περιπτώσεις όπως “Walking” και “Outside”. Η χρήση του δείκτη Jaccard αποδεικνύεται χρήσιμη για την ανάλυση των επικαλύψεων μεταξύ ετικετών και για την ανάδειξη πιθανών συγχύσεων που προκύπτουν στη διαδικασία πρόβλεψης.

## 6.2 Αποτελέσματα Μοντέλου CNN

Το μοντέλο Συνελκτικού Νευρωνικού Δικτύου (CNN) εφαρμόστηκε με στόχο να εκμεταλλευτεί τοπικές χρονικές και σχέσεις στις διαστάσεις στα δεδομένα των αισθητήρων. Η αρχιτεκτονική του μοντέλου βασίστηκε σε διαδοχικά convolutional layers, τα οποία εξήγαγαν χωρικά μοτίβα από τα χρονικά παράθυρα των σημάτων, ακολουθούμενα από πλήρως συνδεδεμένα (dense) επίπεδα που πραγματοποίησαν την τελική πολυετικετική ταξινόμηση. Το μοντέλο εκπαιδεύτηκε με χρήση cross entropy loss, ενώ χρησιμοποιήθηκαν και τεχνικές regularization για την αποφυγή υπερεκπαίδευσης.

Η αποδοτικότητα του μοντέλου αξιολογήθηκε με βάση τις βασικές μετρικές απόδοσης: ακρίβεια (Accuracy), Balanced Accuracy, Precision, Recall, F1-Score και Specificity, για κάθε μία από τις πέντε ετικέτες δραστηριότητας που επιλέχθηκαν. Τα αποτελέσματα συνοψίζονται στον πίνακα που ακολουθεί:

Πίνακας 6.2: Πίνακας τελικών τιμών μετρικών για το μοντέλο CNN

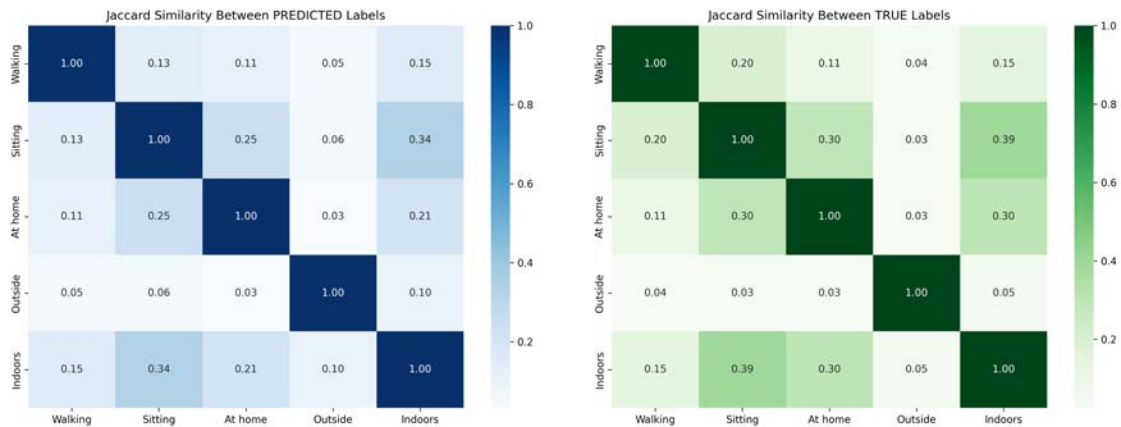
| Ετικέτα | Accuracy | Balanced_Acc | Precision | Recall  | F1-Score | Specificity |
|---------|----------|--------------|-----------|---------|----------|-------------|
| Walking | 0.84562  | 0.73734      | 0.52882   | 0.58012 | 0.54056  | 0.89456     |
| Sitting | 0.73920  | 0.74036      | 0.77606   | 0.71154 | 0.74182  | 0.76920     |
| At home | 0.62726  | 0.62354      | 0.63552   | 0.53134 | 0.57626  | 0.71568     |
| Outside | 0.89428  | 0.80982      | 0.39264   | 0.71128 | 0.50166  | 0.90832     |
| Indoors | 0.91586  | 0.78006      | 0.98648   | 0.92596 | 0.95502  | 0.63416     |

Σύμφωνα λοιπόν με τον πίνακα, η ετικέτα “Indoors” παρουσιάζει την υψηλότερη απόδοση σχεδόν σε όλες τις μετρικές. Το μοντέλο φαίνεται να έχει μάθει πολύ αποτελεσματικά τα χαρακτηριστικά αυτής της ετικέτας, με F1-Score που φτάνει το 0.95 και Precision άνω του 98%. Αυτό φανερώνει ότι το “Indoors” διαθέτει ένα πιο σαφές και αναγνωρίσιμο μοτίβο στα δεδομένα εισόδου, πιθανώς λόγω της στατικότητας των σημάτων ή της παρουσίας επαναλαμβανόμενων συνθηκών εντός εσωτερικών χώρων.

Αντίθετα, η ετικέτα “At home” εμφανίζει τις χαμηλότερες επιδόσεις, με τιμές Accuracy μόλις 0.62 και Recall στο 0.53. Το CNN μοντέλο δυσκολεύεται να διακρίνει αυτή τη δραστηριότητα από άλλες, πιθανώς λόγω της επικάλυψης της με καταστάσεις όπως το “Sitting” ή το “Indoors”, η οποία όμως είναι και λογική. Η τιμή του Precision διατηρείται σχετικά υψηλή (0.63), δείχνοντας ότι οι προβλέψεις που έγιναν ως “At home” ήταν σε μεγάλο βαθμό σωστές, αλλά χάθηκαν πολλές αληθινές περιπτώσεις (γι αυτό και το χαμηλό Recall).

Η ετικέτα “Walking” έχει μέτρια τιμή Recall (0.58), αλλά χαμηλό Precision (0.52), γεγονός που δείχνει ότι το μοντέλο προβλέπει αρκετές περιπτώσεις ως “Walking” που δεν είναι. Παρόλα αυτά, η ακρίβεια παραμένει υψηλή (0.84), γεγονός που μπορεί να οφείλεται στην άνιση κατανομή των ετικετών και την σχετική σπανιότητα των false positive δειγμάτων.

Η εικόνα γίνεται πιο ξεκάθαρη και μέσω των πινάκων Jaccard Similarity, όπως αυτοί παρουσιάζονται στο Σχήμα 6.2, που καταγράφηκαν για τις προβλεφθείσες και τις πραγματικές ετικέτες. Οι πίνακες δείχνουν όπως και στο προηγούμενο μοντέλο την επικάλυψη μεταξύ των ετικετών, προσδιορίζοντας πόσο συχνά συνυπάρχουν ανά δύο σε ένα δείγμα.



Σχήμα 6.2: Διαγράμματα Jaccard για το μοντέλο CNN σε Predicted και True ετικέτες (πηγή: ίδια φωτογραφία)

Στον πίνακα των προβλεφθέντων ετικετών, παρατηρείται σημαντική και αναμενόμενη συσχέτιση μεταξύ “Sitting” και “Indoors” (0.34), καθώς και μεταξύ “At home” και “Sitting” (0.25). Αντίστοιχες συσχετίσεις υπάρχουν και στον πίνακα των πραγματικών ετικετών, αλλά πιο έντονες (π.χ. το “Indoors” με το “Sitting” φτάνει το 0.39). Αυτό επιβεβαιώνει ότι οι ετικέτες αυτές παρουσιάζουν επικάλυψη και στο επίπεδο του ground truth, γεγονός που καθιστά πιο απαιτητική την ταξινόμηση.

Το CNN φαίνεται να διαχειρίζεται αυτή την πολυπλοκότητα καλύτερα από το Logistic Regression, ειδικά όσον αφορά πιο σαφώς ορισμένες καταστάσεις, όπως είναι το “Indoors” ή το “Outside”. Παρόλο που οι τιμές του Precision σε ετικέτες όπως το “Outside” παραμένουν χαμηλές (0.39), το μοντέλο διατηρεί υψηλό Recall, κάτι που είναι ιδιαίτερα χρήσιμο σε εφαρμογές όπου η μη ανίχνευση μίας δραστηριότητας (false negatives) είναι πιο επιβλαβής από την εσφαλμένη ανίχνευση (false positives).

Συνολικά, το CNN απέδωσε καλά σε περιπτώσεις όπου οι δραστηριότητες παρουσιάζουν διακριτά μοτίβα στους αισθητήρες, ενώ αντιμετώπισε μεγαλύτερη δυσκολία σε περιβάλλοντα με ασαφή ή επικαλυπτόμενα χαρακτηριστικά. Η επιτυχία του στις συγκεκριμένες κατηγορίες ενισχύει την άποψη ότι τα CNN είναι ιδιαίτερα αποτελεσματικά όταν υπάρχουν δομικά επαναλαμβανόμενα πρότυπα στα δεδομένα εισόδου.

## 6.3 Αποτελέσματα Μοντέλου LSTM

Το μοντέλο LSTM (Long Short-Term Memory) εφαρμόστηκε με στόχο την αξιοποίηση της χρονικής δυναμικής των δεδομένων, κάτι που το καθιστά ιδιαίτερος κατάλληλο για τη συγκεκριμένη περίπτωση, όπου οι δραστηριότητες εξελίσσονται με φυσική αλληλουχία στον χρόνο. Η απόδοσή του αξιολογήθηκε με βάση μετρικές ανά ετικέτα και μέσω της Jaccard Similarity τόσο για τις προβλεπόμενες όσο και για τις πραγματικές ετικέτες, ακριβώς όπως και παραπάνω.

Αναλύοντας τα αποτελέσματα που παρατίθενται και στον ακόλουθο πίνακα, παρατηρείται ότι το LSTM πετυχαίνει μια ιδιαίτερα υψηλή συνολική απόδοση. Για την ετικέτα “Indoors”, το Precision αγγίζει το 0.99, ενώ το Recall βρίσκεται στο 0.92, με αποτέλεσμα ένα F1-Score που αγγίζει την τιμή 0.95. Η Balanced Accuracy ανέρχεται στο 0.79 και η Accuracy στο 0.86, καθιστώντας αυτή την ετικέτα την πιο εύστοχα προβλεπόμενη από το μοντέλο. Η “Walking” καταγράφει επίσης πολύ καλά αποτελέσματα, με Accuracy  $\approx 0.86$  και F1-Score  $\approx 0.70$ , αποδεικνύοντας ότι η LSTM αρχιτεκτονική είναι ικανή να εντοπίζει καθαρά μοτίβα κινητικότητας.

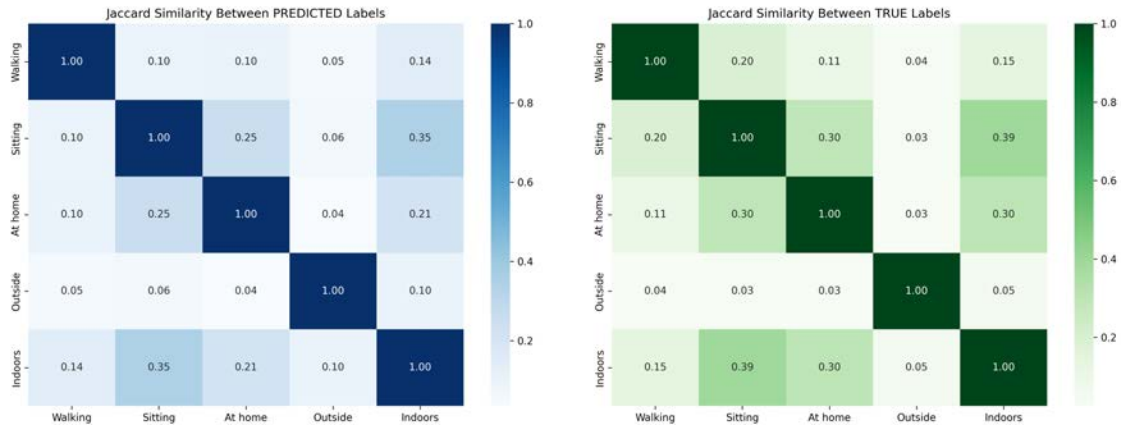
Πίνακας 6.3: Πίνακας τελικών τιμών μετρικών για το μοντέλο LSTM

| Ετικέτα | Accuracy | Balanced_Acc | Precision | Recall  | F1-Score | Specificity |
|---------|----------|--------------|-----------|---------|----------|-------------|
| Walking | 0.86061  | 0.78612      | 0.42698   | 0.86270 | 0.56968  | 0.70954     |
| Sitting | 0.72041  | 0.71048      | 0.78454   | 0.68274 | 0.72037  | 0.73822     |
| At home | 0.64215  | 0.63104      | 0.53166   | 0.68170 | 0.54872  | 0.58037     |
| Outside | 0.88984  | 0.78278      | 0.38374   | 0.86226 | 0.51994  | 0.70330     |
| Indoors | 0.86421  | 0.79024      | 0.98786   | 0.92052 | 0.95051  | 0.66036     |

Αντίστοιχα θετικά είναι τα αποτελέσματα και για την ετικέτα “Outside”, με Accuracy  $\approx 0.89$ , Balanced Accuracy  $\approx 0.78$  και F1-Score  $\approx 0.52$ . Σε αυτή την περίπτωση, το σχετικά χαμηλό Precision (0.38) δείχνει ότι υπάρχουν κάποιες λανθασμένες προβλέψεις ως προς αυτή την κατηγορία, ωστόσο το Recall παραμένει ικανοποιητικό (0.86), γεγονός που δείχνει ότι το μοντέλο εντοπίζει επιτυχώς τις περισσότερες περιπτώσεις εξωτερικών δραστηριοτήτων, ακόμη και αν μερικές ετικέτες που δεν είναι “Outside” ταξινομούνται λανθασμένα ως τέτοιες.

Η κατηγορία “Sitting” έχει Balanced Accuracy  $\approx 0.71$ , Precision  $\approx 0.78$  και F1-Score  $\approx 0.72$ , που δείχνουν ότι οι δραστηριότητες στατικής φύσης αναγνωρίζονται με αξιοπιστία. Από την άλλη πλευρά, η “At home” ετικέτα αποτελεί για ακόμη μία φορά την πιο απαιτητική περίπτωση. Με Accuracy  $\approx 0.64$ , Balanced Accuracy  $\approx 0.63$  και F1-Score  $\approx 0.55$ , φαίνεται ότι το μοντέλο δυσκολεύεται να διαχωρίσει την έννοια του “At home” από τις υπόλοιπες, πιθανώς επειδή η συγκεκριμένη ετικέτα επικαλύπτεται νοηματικά με άλλες όπως το “Indoors” ή το “Sitting”.

Η σύγχυση αυτή αποτυπώνεται ξεκάθαρα και στους πίνακες Jaccard Similarity που απεικονίζονται στο Σχήμα 6.3. Στον πίνακα των προβλεπόμενων ετικετών εμφανίζεται υψηλή ομοιότητα μεταξύ των κατηγοριών “Sitting” και “Indoors” (0.35), καθώς και μεταξύ “Sitting” και “At home” (0.25). Αντίστοιχα, η τιμή του “Indoors” παρουσιάζει επικάλυψη με αυτή του “At home” (0.21), ποσοστά που δείχνουν ότι οι προβλέψεις του μοντέλου αντικατοπτρίζουν την συνύπαρξη αυτών των δραστηριοτήτων. Στον πίνακα των πραγματικών ετικετών, οι ίδιες συσχετίσεις παρατηρούνται με ακόμη πιο έντονο τρόπο, όπως μεταξύ “Sitting” και “Indoors” (0.39), αλλά και “Sitting” – “At home” (0.30), γεγονός που ενισχύει την άποψη ότι πρόκειται για συσχετίσεις που αντικατοπτρίζουν και την πραγματική συμπεριφορά και δεν αποτελούν αποκλειστικά προβλήματα πρόβλεψης προς επίλυση.



Σχήμα 6.3: Διαγράμματα Jaccard για το μοντέλο LSTM σε Predicted και True ετικέτες (πηγή: ίδια φωτογραφία)

Οι ετικέτες “Walking” και “Outside” παραμένουν σταθερά διακριτές τόσο στις προβλέψεις όσο και στην πραγματική κατανομή, με πολύ χαμηλές τιμές Jaccard ως προς τις υπόλοιπες (κάτω από 0.10) που αποδεικνύει ότι αντιπροσωπεύουν μοναδικά και μη επικαλυπτόμενα μοτίβα στο σύνολο δεδομένων. Αυτό αποτελεί και έναν από τους κύριους λόγους για την αυξημένη ακρίβεια των προβλέψεων για αυτές τις κατηγορίες.

Συνολικά, το LSTM επιβεβαιώνει την καταλληλότητά του για εφαρμογή στην παρούσα περίπτωση. Αξιοποιώντας τη χρονική πληροφορία, επιτυγχάνει υψηλή αναγνωρισιμότητα για τις περισσότερες από τις δραστηριότητες, ενώ παράλληλα διαχειρίζεται με σχετική επιτυχία και τη σύνθετη φύση των αλληλεπικαλυπτόμενων ετικετών. Η απόδοσή του υπερέρχει σε σχέση με τα παραδοσιακά μοντέλα, προσφέροντας αποτελέσματα με μεγαλύτερη ακρίβεια και συνέπεια στην ταξινόμηση των δραστηριοτήτων.

## 6.4 Αποτελέσματα Υβριδικού Μοντέλου CNN-LSTM

Το υβριδικό μοντέλο CNN-LSTM συνδυάζει την ικανότητα των συνελκτικών νευρωνικών δικτύων να εντοπίζουν τοπικά πρότυπα στα δεδομένα με την ισχυρή ικανότητα των LSTM να απομνημονεύουν μακροπρόθεσμες χρονικές εξαρτήσεις. Αυτό καθιστά το συγκεκριμένο μοντέλο ιδιαίτερα κατάλληλο για προβλήματα ταξινόμησης με χρονοσειρές, όπως στην παρούσα εργασία όπου τα σήματα των αισθητήρων εξετάζονται σε πολυδιάστατο χρονικό πλαίσιο.

Σύμφωνα με τα αποτελέσματα που προέκυψαν από την εφαρμογή του CNN-LSTM μοντέλου στα δεδομένα, και τα οποία καταγράφονται στον πίνακα που ακολουθεί, παρατηρούνται σημαντικές βελτιώσεις σε αρκετές κατηγορίες, ιδίως σε σχέση με απλούστερα μοντέλα όπως η λογιστική παλινδρόμηση (Logistic Regression). Ειδικότερα, η κατηγορία “Indoors” πέτυχε εξαιρετικά υψηλά ποσοστά σε όλους τους δείκτες απόδοσης, με Accuracy σχεδόν 0.97 και F1-Score  $\approx$  0.96, κάτι που υποδηλώνει την ικανότητα του μοντέλου να αναγνωρίζει με σαφήνεια τις συγκεκριμένες συνθήκες. Παρόμοια, η κατηγορία “Sitting” παρουσίασε

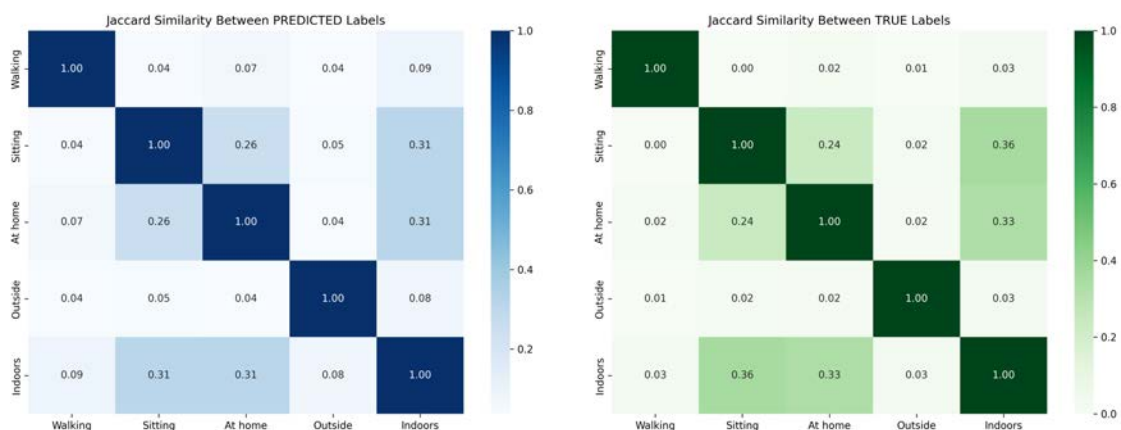
σταθερές επιδόσεις με Balanced Accuracy  $\approx 0.77$  και F1-Score  $\approx 0.74$ , γεγονός που δείχνει ικανότητα του μοντέλου να διαχωρίζει επιτυχώς και αυτή τη δραστηριότητα.

Πίνακας 6.4: Πίνακας τελικών τιμών μετρικών για το υβριδικό μοντέλο CNN-LSTM

| Ετικέτα | Accuracy | Balanced_Acc | Precision | Recall  | F1-Score | Specificity |
|---------|----------|--------------|-----------|---------|----------|-------------|
| Walking | 0.91070  | 0.80734      | 0.23844   | 0.69660 | 0.33212  | 0.91806     |
| Sitting | 0.77868  | 0.77034      | 0.74738   | 0.73828 | 0.74146  | 0.80242     |
| At home | 0.63674  | 0.63480      | 0.62604   | 0.65100 | 0.63774  | 0.61860     |
| Outside | 0.91216  | 0.78280      | 0.29650   | 0.64132 | 0.40154  | 0.92426     |
| Indoors | 0.92600  | 0.78698      | 0.98714   | 0.93548 | 0.96036  | 0.63842     |

Από την άλλη πλευρά, οι κατηγορίες “Walking” και “Outside” εμφανίζουν χαμηλότερα F1-Scores (0.33 και 0.40 αντίστοιχα), κάτι που συνδέεται με χαμηλά επίπεδα Precision (0.24 για Walking και κάτι λιγότερο από 0.30 για Outside). Η χαμηλή ακρίβεια υποδεικνύει ότι το μοντέλο συχνά συγχέει και μπερδεύει τις συγκεκριμένες δραστηριότητες με άλλες, παρά το ότι το Recall παραμένει σε ικανοποιητικά επίπεδα (πάνω από 0.64 και για τις δύο ετικέτες). Αυτό το φαινόμενο μπορεί οφείλεται σε ομοιότητες στα κινητικά μοτίβα ή στα χαρακτηριστικά των σημάτων μεταξύ των δραστηριοτήτων αυτών και άλλων ετικετών, κάτι που δυσκολεύει το μοντέλο να διακρίνει ξεκάθαρα τα όρια μεταξύ τους.

Η ετικέτα “At home” επιτυγχάνει μέτρια αποτελέσματα με Accuracy και F1-Score περίπου στο 0.64, κάτι που αντικατοπτρίζει μια ισορροπία μεταξύ Precision και Recall, χωρίς ωστόσο ιδιαίτερες ακραίες τιμές. Είναι αξιοσημείωτο ότι για όλες τις κατηγορίες εκτός από τις “Indoors” και “Sitting”, η τιμή του Specificity τείνει να είναι υψηλότερη από αυτή του Recall, κάτι που σημαίνει ότι το μοντέλο είναι περισσότερο ικανό στο να αναγνωρίζει σωστά τα αρνητικά δείγματα παρά τα θετικά.



Σχήμα 6.4: Διαγράμματα Jaccard για το υβριδικό μοντέλο CNN-LSTM σε Predicted και True ετικέτες (πηγή: ίδια φωτογραφία)

Η ανάλυση ομοιότητας Jaccard μεταξύ των προβλεπόμενων και των πραγματικών ετικετών που φαίνεται στο Σχήμα 6.4 υποστηρίζει όλες τις παραπάνω παρατηρήσεις. Στον συνδυαστικό χάρτη Jaccard, παρατηρείται έντονη ομοιότητα ανάμεσα στις κατηγορίες “Sitting”, “At home” και “Indoors”, τόσο στους πραγματικούς όσο και στους προβλεπόμενους συνδυασμούς, με τις τιμές να βρίσκονται όλες εντός του εύρους 0.25–0.36. Αυτό υποδηλώνει ότι οι συγκεκριμένες ετικέτες εμφανίζουν επικάλυψη χαρακτηριστικών ή χρονική συνάφεια. Αντίθετα, η κατηγορία “Outside” παραμένει διακριτή, καθώς παρουσιάζει πολύ χαμηλή Jaccard ομοιότητα με τις υπόλοιπες, ενισχύοντας την άποψη ότι πρόκειται για μια ξεχωριστή συνθήκη η οποία ωστόσο εντοπίζεται με περιορισμένη ακρίβεια.

Συνολικά, το CNN-LSTM μοντέλο παρουσιάζει ισχυρές επιδόσεις σε ετικέτες με ξεκάθαρα μοτίβα και καλές χρονικές εξαρτήσεις, όπως “Indoors” και “Sitting”, ενώ αντιμετωπίζει δυσκολίες με κατηγορίες που είτε παρουσιάζουν μικρότερη διάρκεια είτε αλληλοεπικαλυπτόμενα χαρακτηριστικά, όπως “Walking” και “Outside”. Ωστόσο, το γεγονός ότι πρόκειται για ένα μοντέλο με πλεονεκτήματα και από τις δύο κατηγορίες αρχιτεκτονικών (συνελικτικές και επαναλαμβανόμενες) το καθιστά ιδιαίτερα αξιόπιστο και σταθερό σε μια γενική αποτίμηση.

## 6.5 Σύγκριση Μοντέλων

Η πειραματική αξιολόγηση των τεσσάρων μοντέλων ταξινόμησης (Logistic Regression, CNN, LSTM και CNN-LSTM) κατέδειξε ουσιαστικές διαφορές ως προς τη συμπεριφορά τους απέναντι σε πολυδιάστατα, πολυετικετικά και χρονικά εξαρτώμενα δεδομένα, όπως αυτά του ExtraSensory dataset. Η αξιολόγηση βασίστηκε σε πέντε επιλεγμένες ετικέτες δραστηριοτήτων (Walking, Sitting, At home, Outside, Indoors), ενώ εφαρμόστηκαν οι ίδιες συνθήκες εκπαίδευσης, μέθοδοι επικύρωσης και μετρικές σύγκρισης, προκειμένου να διασφαλιστεί ότι οι συγκρίσεις θα έχουν την απαραίτητη εγκυρότητα. Οι δείκτες που λήφθηκαν υπόψη ήταν κυρίως το Accuracy, η Balanced Accuracy, η Precision, το Recall, το F1-Score και η Specificity, ενώ η χρήση των πινάκων Jaccard similarity επέτρεψε μια πιο ποιοτική αποτύπωση της σχέσης μεταξύ πραγματικών και προβλεπόμενων ετικετών.

Το μοντέλο λογιστικής παλινδρόμησης (Logistic Regression), παρότι είναι υπολογιστικά ελαφρύ και εύκολα ερμηνεύσιμο και κατανοητό, εμφάνισε τους χαμηλότερους δείκτες απόδοσης συνολικά. Παρά την ικανοποιητική συμπεριφορά του σε ετικέτες όπως “Sitting” και “Indoors”, όπου καταγράφηκαν F1-Scores άνω του 0.63, παρουσίασε σημαντικές δυσκολίες στην αναγνώριση πιο σύνθετων ή δυναμικών δραστηριοτήτων, όπως το “Walking” και το “Outside”. Οι τιμές Precision για τις συγκεκριμένες ετικέτες ήταν ιδιαίτερα χαμηλές (περίπου 0.26 και 0.29 αντίστοιχα), ενώ και οι πίνακες Jaccard αποκάλυψαν σημαντική επικάλυψη μεταξύ άσχετων κατηγοριών, ιδίως μεταξύ των “At home” και “Sitting”. Αυτά τα αποτελέσματα συμφωνούν με τις επιστημονικές της βιβλιογραφίας (Vaizman et al., 2017) [31], σύμφωνα με την οποία η λογιστική παλινδρόμηση υπολείπεται σημαντικά σε απόδοση όταν χρησιμοποιείται σε δεδομένα πολυετικετικής φύσης χωρίς μετασχηματισμό ή χωρική/χρονική ενίσχυση.

Το συνελικτικό νευρωνικό δίκτυο (CNN) έδωσε αποτελέσματα αισθητά βελτιωμένα σε

σύγκριση με το μοντέλο λογιστικής παλινδρόμησης. Η δυνατότητα του CNN να αναγνωρίζει τοπικά πρότυπα επέτρεψε στο μοντέλο να επιτυγχάνει αυξημένη απόδοση σε ετικέτες με ισχυρή χωρική διαφοροποίηση, όπως “Indoors” (F1-Score  $\approx 0.91$ ) και “Sitting”. Ωστόσο, οι ετικέτες που περιλάμβαναν διαχρονική δυναμική (όπως “Walking” και “Outside”) εξακολούθησαν να εμφανίζουν προβλήματα, κυρίως λόγω της απουσίας μοντελοποίησης της χρονικής πληροφορίας. Η παρατήρηση αυτή επιβεβαιώνει και τα ευρήματα από τη σχετική βιβλιογραφία (Ordóñez & Roggen, 2016) [6], όπου αναφέρεται ότι τα CNN είναι ιδιαίτερα αποδοτικά σε tasks με ισχυρή τοπική δομή, αλλά περιορίζονται όταν απουσιάζει η παράλληλη επεξεργασία χρονικών εξαρτήσεων.

Το LSTM μοντέλο επιλέχθηκε για να ενισχύσει τη διαχείριση της χρονικής διάστασης και ανέδειξε μεγαλύτερη σταθερότητα στην απόδοση του σε ετικέτες όπως “Walking” και “Indoors”, με F1-Scores της τάξης του 0.73 και 0.95 αντίστοιχα. Η ευαισθησία του LSTM στη μεταβολή των σημάτων στο χρόνο συνέβαλε σημαντικά στη βελτίωση της ακρίβειας, ιδίως σε δραστηριότητες που εξελίσσονται σταδιακά. Ωστόσο, παρότι υπερείχε σε σχέση με το CNN στις δυναμικές κατηγορίες, παρουσίασε οριακή υποχώρηση στην “Sitting” και “At home”, πιθανόν λόγω αδυναμίας εντοπισμού στατικών ή στιγμιαίων μοτίβων. Τα αποτελέσματα αυτά συνάδουν με τα συμπεράσματα της Shoaib et al. (2016) [39], όπου καταγράφεται η ανάγκη ενίσχυσης των LSTM με επιπλέον μηχανισμούς ανίχνευσης χωρικών συσχετισμών για μέγιστη απόδοση.

Η καλύτερη συνολική συμπεριφορά προέκυψε από το υβριδικό μοντέλο CNN-LSTM, το οποίο αξιοποίησε αποτελεσματικά τα πλεονεκτήματα των δύο αρχιτεκτονικών. Το CNN-LSTM πέτυχε τα υψηλότερα F1-Scores σε κρίσιμες κατηγορίες, με κορυφαία επίδοση στην “Indoors” (0.96), ενώ διατήρησε σταθερά υψηλές τιμές Recall και Specificity σε όλες σχεδόν τις ετικέτες. Η κατηγορία “Sitting” αναγνωρίστηκε με σαφήνεια (F1-Score  $\approx 0.74$ ), ενώ ακόμη και σε δύσκολες κατηγορίες, όπως “Walking”, το Recall παρέμεινε υψηλό (0.69), υποδεικνύοντας την ικανότητα του μοντέλου στον εντοπισμό πραγματικών περιστατικών. Οι πίνακες Jaccard similarity, σε επίπεδο συνολικής ομοιότητας μεταξύ ετικετών, ανέδειξαν στην περίπτωση του CNN-LSTM τη μικρότερη απόκλιση μεταξύ προβλεπόμενων και πραγματικών συνδυασμών ετικετών, στοιχείο που υποδηλώνει ισχυρότερη μοντελοποίηση των πραγματικών σχέσεων ταυτόχρονης εμφάνισης.

Η παραπάνω υπεροχή του CNN-LSTM ευθυγραμμίζεται με τη βιβλιογραφία. Στη μελέτη των Ronao & Cho (2016) [40] τεκμηριώνεται ότι οι υβριδικές αρχιτεκτονικές επιτυγχάνουν ανώτερη απόδοση στην αναγνώριση δραστηριοτήτων όταν τα σήματα χαρακτηρίζονται από πολυπλοκότητα στην χωροχρονική τους δομή. Παρομοίως, οι Ordóñez και Roggen (2016) [6] προτείνουν τη συνδυαστική χρήση CNN και LSTM ως ιδανική λύση για την ταξινόμηση σήματος που μεταβάλλεται τόσο χωρικά όσο και χρονικά, επισημαίνοντας την αυξημένη απόδοση στις ίδιες κατηγορίες που παρατηρήθηκε και στην παρούσα εργασία.

Συνοψίζοντας, η πειραματική ανάλυση έκανε σαφή την υπεροχή της υβριδικής προσέγγισης έναντι των επιμέρους μοντέλων. Το CNN-LSTM απέδωσε με συνέπεια σε όλες τις ετικέτες, συνδυάζοντας υψηλές τιμές σε κρίσιμες μετρικές και ισχυρή ερμηνευσιμότητα των προβλέψεων μέσω των πινάκων Jaccard. Η προσέγγιση αυτή, εδραιωμένη τόσο εμπειρικά όσο και θεωρητικά, κρίνεται ως η πλέον κατάλληλη για εφαρμογές πολυετικετικής αναγνώρισης δραστηριότητας σε δεδομένα αισθητήρων.

Για να γίνει καλύτερα κατανοητή η σύγκριση, στον Πίνακα 6.5 που ακολουθεί, γίνεται η σύγκριση των τεσσάρων μοντέλων που υλοποιήθηκαν: Logistic Regression, CNN, LSTM και CNN-LSTM. Οι τιμές προέκυψαν από τον υπολογισμό του μέσου όρου των πέντε δραστηριοτήτων (Walking, Sitting, At Home, Indoors, Outside) για καθεμία από τις βασικές μετρικές αξιολόγησης: Accuracy, Balanced Accuracy, Precision, Recall και F1-Score κ.α.

Πίνακας 6.5: Συγκριτικός πίνακας μέσων τιμών μετρικών από τους Πίνακες 6.1–6.4

| Μοντέλο             | Accuracy | Balanced_Acc | Precision | Recall | F1-Score |
|---------------------|----------|--------------|-----------|--------|----------|
| Logistic Regression | 0.76     | 0.731        | 0.55      | 0.70   | 0.59     |
| CNN                 | 0.80     | 0.74         | 0.66      | 0.69   | 0.66     |
| LSTM                | 0.80     | 0.74         | 0.62      | 0.74   | 0.66     |
| CNN-LSTM            | 0.83     | 0.76         | 0.58      | 0.73   | 0.61     |

Η ανάλυση των μέσων τιμών του Πίνακα 6.5 επιβεβαιώνει με ποσοτικό τρόπο τα ευρήματα που περιγράφηκαν ποιοτικά στις προηγούμενες παραγράφους. Παρατηρείται ότι τα μοντέλα βαθιάς μάθησης (CNN, LSTM και CNN-LSTM) υπερέχουν σε όλες τις μετρικές σε σχέση με το γραμμικό baseline της Logistic Regression. Η διαφορά αυτή είναι εντονότερη στη μετρική Balanced Accuracy (0.76 έναντι 0.73), η οποία λαμβάνει υπόψη την ισορροπία ανάμεσα στις δύο κλάσεις, γεγονός που υποδεικνύει ότι τα πιο σύνθετα μοντέλα έχουν μεγαλύτερη ικανότητα γενίκευσης και δεν επηρεάζονται σε τόσο μεγάλο βαθμό από την ανισορροπία των δεδομένων.

Το CNN και το LSTM παρουσιάζουν παρόμοια συνολική επίδοση, με μέση ακρίβεια γύρω στο 0.80, διαφοροποιούμενα κυρίως ως προς το Recall. Το LSTM επιτυγχάνει ελαφρώς υψηλότερη τιμή Recall (0.74 έναντι 0.69), γεγονός που δείχνει ότι εντοπίζει περισσότερα πραγματικά θετικά δείγματα, αξιοποιώντας αποτελεσματικά τη χρονική συσχέτιση των παρατηρήσεων. Το CNN, αντίθετα, υπερέχει οριακά στην Precision (0.66 έναντι 0.62), αποτυπώνοντας την ικανότητά του να αποφεύγει λανθασμένες ταξινομήσεις. Η συμπληρωματική φύση των δύο αυτών μοντέλων αποτέλεσε τη βάση για τη σχεδίαση του υβριδικού CNN-LSTM.

Το υβριδικό μοντέλο CNN-LSTM καταγράφει τη συνολικά υψηλότερη απόδοση, με Accuracy 0.83 και Balanced Accuracy 0.76, επιβεβαιώνοντας ότι ο συνδυασμός συνελκτικών και επαναλαμβανόμενων δομών αυξάνει σημαντικά την ικανότητα του δικτύου να αποτυπώνει χωροχρονικά μοτίβα. Παρότι η Precision του CNN-LSTM παραμένει συγκρίσιμη με εκείνη των άλλων δύο νευρωνικών μοντέλων, το δίκτυο επιτυγχάνει σταθερά υψηλό Recall, κάτι που σημαίνει ότι αναγνωρίζει με συνέπεια τις θετικές περιπτώσεις ανεξαρτήτως κατηγορίας. Το αποτέλεσμα αυτό αντικατοπτρίζει τη θεωρητική υπόθεση που διατυπώνεται στη βιβλιογραφία από τους Ordóñez & Roggen (2016) [6], σύμφωνα με την οποία η συνέλιξη απομονώνει τοπικά πρότυπα, ενώ το LSTM διατηρεί τη χρονική τους συνέχεια.

Συνολικά, ο Πίνακας 6.5 αναδεικνύει μια προοδευτική βελτίωση καθώς η αρχιτεκτονική μεταβαίνει από απλές γραμμικές σε πιο πολύπλοκες βαθιές μορφές. Το Logistic Regression μοντέλο λειτουργεί ως αποτελεσματικό baseline, το CNN και το LSTM αποδεικνύονται

επαρκή για επιμέρους κατηγορίες, ενώ το CNN–LSTM επιτυγχάνει τη βέλτιστη συνολική απόδοση, ισορροπώντας μεταξύ ακρίβειας, ευαισθησίας και ικανότητας γενίκευσης. Η υπεροχή αυτή καθιστά την υβριδική προσέγγιση την πλέον κατάλληλη για αναγνώριση δραστηριοτήτων με χωροχρονικά εξαρτώμενα δεδομένα αισθητήρων.

## 7 Συμπεράσματα

Η παρούσα εργασία είχε ως βασικό στόχο την ανάπτυξη, υλοποίηση και συγκριτική αξιολόγηση μοντέλων μηχανικής μάθησης και βαθιάς μάθησης για την πολυετικετική ταξινόμηση δραστηριοτήτων, αξιοποιώντας δεδομένα από φορετούς αισθητήρες του ExtraSensory dataset. Η μεθοδολογία που ακολουθήθηκε βασίστηκε στη διαδοχική εφαρμογή διαφορετικών μοντέλων – ξεκινώντας από ένα κλασικό μοντέλο λογιστικής παλινδρόμησης και στη συνέχεια εισάγοντας συνελκτικά (CNN), επαναλαμβανόμενα (LSTM) και τελικά υβριδικά (CNN-LSTM) νευρωνικά δίκτυα.

Αρχικά, πραγματοποιήθηκε διεξοδική και επιμελής ανάλυση των δεδομένων και των χαρακτηριστικών των ετικετών, ώστε να εντοπιστούν οι όποιες ενδογενείς ανισορροπίες και να επιλεγούν οι κατάλληλες μετρικές αξιολόγησης πέραν της απλής ακρίβειας, όπως η Balanced Accuracy, το Recall και το F1-Score. Κατά τη διαδικασία εκπαίδευσης, εφαρμόστηκε επικύρωση με διασταυρούμενη μέθοδο (k-fold cross-validation) και αναλύθηκαν τόσο τα αριθμητικά αποτελέσματα των μοντέλων όσο και η ποιοτική συμπεριφορά τους μέσω των πινάκων ομοιότητας Jaccard. Ο συνδυασμός των εργαλείων αυτών παρείχε μια ολοκληρωμένη εικόνα για τη συμπεριφορά κάθε μοντέλου σε κάθε δραστηριότητα.

Ο αρχικός στόχος ήταν να διαπιστωθεί κατά πόσο ένα υβριδικό μοντέλο CNN-LSTM μπορεί να αποδώσει καλύτερα σε ένα ρεαλιστικό, πολυετικετικό πρόβλημα αναγνώρισης δραστηριότητας, σε σύγκριση με πιο απλές ή μεμονωμένες αρχιτεκτονικές. Τα αποτελέσματα επιβεβαίωσαν την αρχική υπόθεση: το CNN-LSTM υπερέβη όλα τα υπόλοιπα μοντέλα σε κρίσιμους δείκτες, παρουσιάζοντας υψηλή ακρίβεια, Recall και F1-Score σε σχεδόν όλες τις εξεταζόμενες ετικέτες. Η επίδοση αυτή ενισχύθηκε και από τη μικρότερη απόκλιση μεταξύ των προβλεπόμενων και των πραγματικών ετικετών, όπως φάνηκε στις απεικονίσεις Jaccard. Συνεπώς, η κατάληξη στα ευρήματα αυτά θεωρείται θετική για την πορεία της έρευνας, αφού όχι μόνο επαληθεύτηκε η ερευνητική υπόθεση, αλλά αναδείχθηκε και η προστιθέμενη αξία που έχει ο συνδυασμός των χωρικών και χρονικών προτύπων.

Η έρευνα αυτή διαφοροποιείται από προηγούμενες σε δύο κύρια σημεία. Πρώτον, χειρίζεται το πρόβλημα της ταξινόμησης ως γνήσια πολυετικετικό (multi-label), χωρίς να το μετατρέπει σε σύνολο από ανεξάρτητα δυαδικά προβλήματα (π.χ. binary relevance), όπως επιχειρείται στις περισσότερες μελέτες για απλοποίηση. Δεύτερον, εφαρμόζει μια ενοποιημένη αξιολόγηση με τη χρήση των Jaccard Similarity πινάκων, η οποία εντοπίζει επιτυχώς αλληλοεπικαλύψεις μεταξύ ετικετών, γεγονός ιδιαίτερα σημαντικό σε προβλήματα όπου οι δραστηριότητες συνυπάρχουν ή εναλλάσσονται ταχύτατα.

Η εγκυρότητα των ευρημάτων ενισχύεται από τη συμφωνία τους με τη διεθνή βιβλιογραφία. Οι Ordóñez και Roggen (2016) [6], αλλά και οι Shoaib et al. (2016) [39], αναφέρουν με σαφήνεια την υπεροχή των υβριδικών μοντέλων σε περιβάλλοντα με σύνθετες δομές σήματος. Αντίστοιχα, οι επιδόσεις των CNN και LSTM ακολουθούν τις τάσεις που κατα-

γράφονται και σε προηγούμενες έρευνες, με τα LSTM να αποδίδουν καλύτερα σε χρονικά συνεκτικές δραστηριότητες και τα CNN σε τοπικά εντοπίσιμα μοτίβα. Η σύγκριση με την εργασία των Vaizman et al. (2017) [31], που χρησιμοποιεί το ίδιο dataset, υποδεικνύει ότι η ενσωμάτωση βαθιών μοντέλων και κυρίως η επιλογή συνδυαστικών αρχιτεκτονικών μπορεί να οδηγήσει σε υψηλότερες επιδόσεις, ιδίως όταν εφαρμόζονται πιο αυστηρές μετρικές αξιολόγησης και λεπτομερής χειρισμός της πολυετικετικότητας.

Ως προς την επιστημονική συνεισφορά, η παρούσα εργασία ενισχύει το πεδίο της αναγνώρισης δραστηριοτήτων, παρουσιάζοντας μια ολοκληρωμένη προσέγγιση που αξιοποιεί τόσο την πληροφορία της χωρικής κατανομής όσο και τη χρονική συνέχεια των δεδομένων. Η χρήση πολυετικετικών δεδομένων χωρίς απλοποίηση, η συγκριτική μελέτη διαφορετικών κατηγοριών μοντέλων, και η εφαρμογή προηγμένων μετρικών αξιολόγησης αποτελούν στοιχεία που διαφοροποιούν την εργασία αυτή από μεγάλο μέρος της βιβλιογραφίας, προσφέροντας πρακτικά και θεωρητικά συμπεράσματα που μπορούν να αξιοποιηθούν σε μελλοντικές εφαρμογές.

Ωστόσο, δεν παραβλέπονται ορισμένοι περιορισμοί που εντοπίστηκαν σε όλη τη διάρκεια της έρευνας. Ο όγκος των δεδομένων, αν και εκτενής, περιείχε ετικέτες με αρκετά έντονα άνιση κατανομή, γεγονός που επηρέασε τις επιδόσεις σε κατηγορίες με λιγότερα δείγματα ποσοτικά. Επίσης, το μοντέλο αξιολογήθηκε σε δεδομένα από συγκεκριμένο περιβάλλον και πληθυσμό, κάτι που περιορίζει την γενικευσιμότητα των ευρημάτων. Επιπλέον, η υλοποίηση βασίστηκε σε σχετικά απλές αρχιτεκτονικές χωρίς hyperparameter tuning μεγάλης κλίμακας ή χρήση τεχνικών όπως είναι οι attention μηχανισμοί.

Με βάση τα παραπάνω, προκύπτουν προτάσεις για χρήσιμη μελλοντική έρευνα. Αρχικά, η αξιολόγηση των μοντέλων σε πραγματικό χρόνο ή σε δεδομένα από διαφορετικούς πληθυσμούς και περιβάλλοντα θα μπορούσε να ενισχύσει τη γενικευσιμότητα των συμπερασμάτων. Επιπλέον, η ενσωμάτωση attention μηχανισμών ή transformer-based αρχιτεκτονικών θα μπορούσε να προσφέρει περαιτέρω βελτιώσεις, ιδιαίτερα σε ετικέτες με επικάλυψη. Τέλος, η διερεύνηση εναλλακτικών μεθόδων χειρισμού της πολυετικετικότητας (π.χ. sequence-to-sequence ταξινόμηση ή graph-based μοντέλα) θα μπορούσε να οδηγήσει σε ακόμα πιο ακριβή και ερμηνεύσιμα αποτελέσματα.

Συμπερασματικά, η εργασία αυτή αποδεικνύει ότι η πολυετικετική αναγνώριση δραστηριοτήτων είναι ένα πεδίο με ιδιαίτερες απαιτήσεις, αλλά και σημαντικές ευκαιρίες για καινοτομία. Η χρήση συνδυαστικών αρχιτεκτονικών όπως το CNN-LSTM αναδεικνύεται ως μία από τις πιο υποσχόμενες κατευθύνσεις, ικανή να υποστηρίξει εφαρμογές σε πραγματικές συνθήκες, με υψηλά ποσοστά ακρίβειας και αξιοπιστίας.

# Βιβλιογραφία

- [1] J. Durnin, *Physical Activity and Health*. CRC Press, 1992.
- [2] S. A. Zengaro και F. Zengaro, “Global Perspectives on Physical Activity: Navigating Cultural Influences for Health and Wellness”, στο *Global Innovations in Physical Education and Health*, 2025, σσ. 22–35. doi: 10.4018/979-8-3693-3952-7.ch020
- [3] S. Bauman κ.ά., “Correlates of physical activity: why are some people physically active and others not?”, *The Lancet*, τόμ. 380, αρθμ. 9838, σσ. 258–271, 2012.
- [4] A. Subasi, K. Khateeb, T. Brahimi και A. Sarirete, “Human Activity Recognition Using Machine Learning Methods in a Smart Healthcare Environment”, στο *Next-Generation Technology-Driven Personalized Medicine and Smart Healthcare*, Academic Press, 2020, σσ. 123–144.
- [5] J. Wang, Y. Chen, S. Hao, X. Peng και L. Hu, “Deep learning for sensor-based activity recognition: A survey”, *Pattern Recognition Letters*, τόμ. 119, σσ. 3–11, 2019.
- [6] F. J. Ordóñez και D. Roggen, “Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition”, *Sensors*, τόμ. 16, αρθμ. 1, σ. 115, 2016, Accessed: 2025-04-20. doi: 10.3390/s16010115
- [7] H. Hammerla, S. Halloran και T. Plotz, “Deep, convolutional, and recurrent models for human activity recognition using wearables”, στο *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI)*, 2016.
- [8] M. Gjoreski κ.ά., “Monitoring stress with a wrist device using context”, *Journal of Biomedical Informatics*, τόμ. 73, σσ. 159–170, 2017.
- [9] A. Reiss και D. Stricker, “Introducing a new benchmarked dataset for activity monitoring”, στο *Proceedings of the 16th International Symposium on Wearable Computers*, 2012, σσ. 108–109.
- [10] O. Weiss, L. Minnen και T. Starner, “ExtraSensory: A Large-Scale Dataset for Human Context Recognition”, *ACM Transactions on Interactive Intelligent Systems*, τόμ. 8, αρθμ. 3, 2018.
- [11] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [12] A. Bulling, U. Blanke και B. Schiele, “A tutorial on human activity recognition using body-worn inertial sensors”, *ACM Computing Surveys (CSUR)*, τόμ. 46, αρθμ. 3, σσ. 1–33, 2014. doi: 10.1145/2499621 διεύθν.: <https://doi.org/10.1145/2499621>

- [13] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou και Y. Amirat, “Physical Human Activity Recognition Using Wearable Sensors”, *Sensors*, τόμ. 15, αρθμ. 12, σσ. 31 314–31 338, 2015. DOI: 10.3390/s151229858 διεύθυν.: <https://doi.org/10.3390/s151229858>
- [14] S. Tasoulis, C. Doukas, V. Plagianakos και I. Maglogiannis, “Statistical data mining of streaming motion data for activity and fall recognition in assistive environments”, *Neurocomputing*, τόμ. 107, σσ. 87–96, 2013, Timely Neural Networks Applications in Engineering, ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2012.08.036> διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S0925231212007606>
- [15] R. S. Sutton και A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2018.
- [16] M. A. Nielsen, *Neural Networks and Deep Learning*. 2015, Available at: <http://neuralnetworksanddeeplearning.com>.
- [17] D. E. Rumelhart, G. E. Hinton και R. J. Williams, “Learning representations by back-propagating errors”, *Nature*, τόμ. 323, σσ. 533–536, 1986.
- [18] I. Goodfellow, Y. Bengio και A. Courville, *Deep Learning*. MIT Press, 2016.
- [19] S. K. Tasoulis, C. N. Doukas, I. Maglogiannis και V. P. Plagianakos, “Statistical data mining of streaming motion data for fall detection in assistive environments”, στο *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, σσ. 3720–3723. DOI: 10.1109/IEMBS.2011.6090632
- [20] A. Krizhevsky, I. Sutskever και G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, στο *Advances in Neural Information Processing Systems (NeurIPS)*, τόμ. 25, 2012, σσ. 1097–1105.
- [21] I. T. Nicholaus, J.-S. Lee και D.-K. Kang, “One-Class Convolutional Neural Networks for Water-Level Anomaly Detection”, *Sensors*, τόμ. 22, αρθμ. 22, 2022, ISSN: 1424-8220. DOI: 10.3390/s22228764 διεύθυν.: <https://www.mdpi.com/1424-8220/22/22/8764>
- [22] M. Zeng κ.ά., “Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors”, στο *Proceedings of the 6th International Conference on Mobile Computing, Applications and Services (MobiCASE)*, ICST (Institute for Computer Sciences, Social-Informatics και Telecommunications Engineering), 2014. DOI: 10.4108/icst.mobibase.2014.257786
- [23] U. A. Bhatti κ.ά., “Deep Learning-Based Trees Disease Recognition and Classification Using Hyperspectral Data”, *Computers, Materials and Continua*, τόμ. 77, αρθμ. 1, σσ. 681–697, 2023, ISSN: 1546-2218. DOI: <https://doi.org/10.32604/cmc.2023.037958> διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S1546221823000929>
- [24] X. Glorot, A. Bordes και Y. Bengio, “Deep Sparse Rectifier Neural Networks”, στο *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011, σσ. 315–323.

- [25] Ailephant AI Team, *ReLU Function – Glossary*, Accessed: 2025-09-16, 2025. διεύθν.: <https://ailephant.com/glossary/relu-function/>
- [26] C. Olah, *Understanding LSTM Networks*, <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>, Accessed: 2025-04-19, 2015.
- [27] Ailephant AI Team. “Recurrent Neural Network – Glossary”, επίσκεψη 16 Σεπτ. 2025. διεύθν.: <https://ailephant.com/glossary/recurrent-neural-network/>
- [28] I. BOUZENIA. “LSTM-Based Malware Detection | Python & TensorFlow”. Medium blog post, επίσκεψη 16 Σεπτ. 2025. διεύθν.: <https://medium.com/nerd-for-tech/lstm-based-malware-detection-python-tensorflow-9d047cfe8766>
- [29] H. H. Pham, H. Salmane, L. Khoudour, A. Crouzil, P. Zegers και S. A. Velastin, “Spatio-temporal Image Representation of 3D Skeletal Movements for View-Invariant Action Recognition with Deep Convolutional Neural Networks”, *Sensors*, τόμ. 19, αρθμ. 8, 2019. DOI: 10.3390/s19081932
- [30] S. V. Georgakopoulos, S. K. Tasoulis, G. I. Mallis κ.ά., “Change Detection and Convolution Neural Networks for Fall Recognition”, *Neural Computing and Applications*, τόμ. 32, σσ. 17 245–17 258, 2020. DOI: 10.1007/s00521-020-05208-8 διεύθν.: <https://doi.org/10.1007/s00521-020-05208-8>
- [31] Y. Vaizman, K. Ellis και G. Lanckriet, “Recognizing Detailed Human Context In-the-Wild from Smartphones and Smartwatches”, *IEEE Pervasive Computing*, τόμ. 16, αρθμ. 4, σσ. 62–74, Δεκ. 2017. DOI: 10.1109/MPRV.2017.3971131 διεύθν.: <https://doi.org/10.1109/MPRV.2017.3971131>
- [32] Y. Vaizman, N. Weibel και G. Lanckriet, “Context Recognition In-the-Wild: Unified Model for Multi-Modal Sensors and Multi-Label Classification”, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, τόμ. 1, αρθμ. 4, Δεκ. 2017. DOI: 10.1145/3161192 διεύθν.: <https://doi.org/10.1145/3161192>
- [33] L. Wang και R. Liu, “Human Activity Recognition Based on Wearable Sensor Using Hierarchical Deep LSTM Networks”, *Circuits, Systems, and Signal Processing*, Φεβ. 2020.
- [34] S. Mekruksavanich και A. Jitpattanakul, “Smartwatch-Based Human Activity Recognition Using Hybrid LSTM Network”, στο *2020 IEEE SENSORS*, IEEE, 2020, σσ. 1–4.
- [35] S. Mekruksavanich, A. Jitpattanakul, P. Youplao και P. Yupapin, “Enhanced Hand-Oriented Activity Recognition Based on Smartwatch Sensor Data Using LSTMs”, *Symmetry*, τόμ. 12, αρθμ. 9, 2020, Accessed: 2025-04-19. διεύθν.: <https://www.mdpi.com/2073-8994/12/9/1570>
- [36] P. Agarwal και M. Alam, “A Lightweight Deep Learning Model for Human Activity Recognition on Edge Devices”, *Procedia Computer Science*, 2020, International Conference on Computational Intelligence and Data Science, Accessed: 2025-04-19. διεύθν.: <https://www.sciencedirect.com/science/article/pii/S1877050920307559>

- [37] U. S. D. C. R. Institute, *ExtraSensory Dataset Website*, Accessed: 2025-05-07, 2016. διεύθυν.: <http://extrasensory.ucsd.edu/>
- [38] D. Fyka, *HarProjectPublic*, GitHub repository. Accessed: 2025-09-16, 2025. διεύθυν.: <https://github.com/DespoinaFyka/HarProjectPublic>
- [39] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten και P. J. M. Havinga, “Complex Human Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors”, *Sensors*, τόμ. 16, αρθμ. 4, 2016, ISSN: 1424-8220. DOI: 10.3390/s16040426 διεύθυν.: <https://www.mdpi.com/1424-8220/16/4/426>
- [40] C. A. Ronao και S.-B. Cho, “Human activity recognition with smartphone sensors using deep learning neural networks”, *Expert Systems with Applications*, τόμ. 59, σσ. 235–244, 2016, ISSN: 0957-4174. DOI: <https://doi.org/10.1016/j.eswa.2016.04.032> διεύθυν.: <https://www.sciencedirect.com/science/article/pii/S0957417416302056>

