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

Αυτόματα Συστήματα Αντίληψης και Εκμάθησης Ανθρώπινων Δράσεων

Διδακτορική Διατριβή

της

Νίκης Ευθυμίου

Διπλωματούχου Εφαρμοσμένων Μαθηματικών &
Φυσικών Επιστημών Ε.Μ.Π.

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

Αθήνα, Νοέμβριος 2023

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής

Αυτόματα Συστήματα Αντίληψης και Εκμάθησης Ανθρώπινων Δράσεων

Διδακτορική Διατριβή

της

Νίκης Ευθυμίου

Διπλωματούχου Εφαρμοσμένων Μαθηματικών & Φυσικών Επιστημών Ε.Μ.Π.

Συμβουλευτική Επιτροπή: Πέτρος Μαραγκός, Καθηγητής
Αλέξανδρος Ποταμιάνος, Αναπληρωτής Καθηγητής
Κωνσταντίνος Τζαφέστας, Αναπληρωτής Καθηγητής

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 2 Νοεμβρίου 2023.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*	*(Υπογραφή)*
……	……	……	……
Πέτρος Μαραγκός Καθηγητής Ε.Μ.Π.	Αλέξανδρος Ποταμιάνος Αναπληρωτής Καθηγητής Ε.Μ.Π.	Κωνσταντίνος Τζαφέστας Αναπληρωτής Καθηγητής Ε.Μ.Π.	Γεράσιμος Ποταμιάνος Αναπληρωτής Καθηγητής Πανεπιστήμιο Θεσσαλίας

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
……	……	……
Παναγιώτης Τσανάκας Καθηγητής Ε.Μ.Π.	Αθανάσιος Ροντογιάννης Αναπλ. Καθηγητής Ε.Μ.Π.	Νικόλαος Σμυρνής Καθηγητής Ε.Κ.Π.Α.

Αθήνα, Νοέμβριος 2023

(Υπογραφή)

.....
Νίκη Ευθυμίου
Διπλωματούχος Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών Ε.Μ.Π.

Περίληψη

Η μελέτη της ανθρώπινης δραστηριότητας έχει απασχολήσει και συνεχίζει να απασχολεί πολυπλεύρως την επιστημονική κοινότητα. Στην παρούσα διατριβή εστιάζουμε στην ανάπτυξη καινοτόμων αυτόματων συστημάτων ικανών να αντιληφθούν και να μάθουν τις διαφορετικές εκφάνσεις των ανθρώπινων δράσεων. Η διατριβή χωρίζεται σε δύο μέρη που διαφοροποιούνται κύρια από τον τρόπο αντίληψης των δράσεων. Στο πρώτος μέρος η καταγραφή των δράσεων γίνεται με οπτικούς αισθητήρες ενώ στο δεύτερο, με μη οπτικούς αισθητήρες καταγραφής βιοσημάτων.

Στο πρώτο μέρος της διατριβής ερευνούμε την ανάπτυξη ευφυών συστημάτων αναγνώρισης δράσεων παιδιών κατά τη διάρκεια αλληλεπιδράσεων τους με ρομπότ. Αρχικά μελετήσαμε την ανάπτυξη ενός διευρυμένου συστήματος αλληλεπίδρασης παιδιών με ρομπότ με πολλαπλές αντιληπτικές ικανότητες, λ.χ. αναγνώριση δράσεων, ομιλίας, εντοπισμού ομιλητών και αντικειμένων, μέσω πολλαπλών καμερών και δυνατότητα χρήσης και ενσωμάτωσης πολλαπλών ρομπότ σε ένα ενιαίο σύστημα. Συλλέξαμε και δημιουργήσαμε μια μεγάλη βάση παιδικών δραστηριοτήτων που περιλαμβάνει χειρονομίες, γενικευμένες κινήσεις του σώματος, ομιλία, έκφραση συναισθημάτων κ.α. μεμονωμένα αλλά και κατά τη διάρκεια των αλληλεπιδράσεων. Αξιολογήσαμε το σύστημα τόσο ως προς την ικανότητα των μονάδων αναγνώρισης όσο και ως προς την συνολική αλληλεπίδραση των παιδιών με τα ρομπότ. Στη συνέχεια αναπτύξαμε ένα πιο ελαφρύ σύστημα για τη σχεδίαση και εκτέλεση εκπαιδευτικών σεναρίων με βάση την οπτική πληροφορία για χρήση σε πιο ελεύθερα περιβάλλοντα, όπως η σχολική αίθουσα. Κατά την ανάπτυξη των παραπάνω συστημάτων, κάναμε εκτενή έρευνα ως προς την οπτική αναγνώριση των δράσεων και των χειρονομιών των παιδιών. Αναδείξαμε την ανάγκη ύπαρξης συστημάτων αναγνώρισης ειδικά προσαρμοσμένων σε δεδομένα παιδιών, μελετήσαμε την απόδοση των συστημάτων κατά τη χρήση πολλαπλών ή/και μεμονωμένων όψεων καταγραφής ενώ μελετήσαμε και εφαρμόσαμε μεθόδους επαυξημένης μάθησης για εύκολη επέκταση των δράσεων αναγνώρισης.

Στο δεύτερο μέρος της διατριβής ασχολούμαστε με την αναγνώριση δράσεων μέσω της χρήσης βιοσημάτων από έξυπνους φορετούς αισθητήρες, η χρήση των οποίων διευρύνεται καθημερινά και παρέχει νέες δυνατότητες για ανάπτυξη καινοτόμων συστημάτων αναγνώρισης της ανθρώπινης δραστηριότητας. Έτσι, ερευνούμε την αξιοποίηση σημάτων που συλλέγονται από έξυπνα ρολόγια, λ.χ. επιτάχυνσης, γωνιακής ταχύτητας, καρδιακού ρυθμού, ώστε να μελετήσουμε τη δυνατότητα δημιουργίας ψηφιακών φαινοτύπων των χρηστών, δηλαδή ψηφιακών προφίλ που μπορούν να αποδώσουν τα χαρακτηριστικά μοτίβα της καθημερινότητάς τους αλλά και να αποκαλύψουν σημαντικές μεταβολές από αυτά. Αρχικά, μελετήσαμε τη διαφορά των ψηφιακών αναπαραστάσεων-φαινοτύπων μεταξύ ενός δείγματος ελέγχου και ενός δείγματος ασθενών με ψυχωτικές διαταραχές μέσω εκτεταμένης στατιστικής ανάλυσης και αναπτύξαμε ένα ευφυές σύστημα ταυτοποίησης-ταξινόμησης του κάθε χρήστη με βάση τον ψηφιακό του φαινότυπο ερευνώντας εκτενώς τις επιμέρους παραμέτρους του συστήματος. Στη συνέχεια προσεγγίσαμε το πρόβλημα του εντοπισμού των ψυχωτικών υποτροπών που παρουσιάζονται συχνά σε άτομα με ψυχωτικές διαταραχές. Συγκεκριμένα, αντιμετωπίσαμε το πρόβλημα ως ένα πρόβλημα λανθασμένης ταξινόμησης υποθέτοντας πως κατά τη διάρκεια μιας υποτροπής της ασθένειας είναι δυνατόν να παρατηρήσουμε σημαντικές μεταβολές στον ψηφιακό φαινότυπο

και άρα να μειώνεται η πιθανότητα ταυτοποίησης του εκάστοτε χρήστη σε περιόδους υποτροπών ή λίγο πριν από αυτές. Τέλος, επεκτείναμε το σύστημα εντοπισμού υποτροπών συνδυάζοντας τεχνικές εντοπισμού ανωμαλιών με τις τεχνικές που αναπτύξαμε για την ταυτοποίηση των χρηστών ώστε να επιτύχουμε βελτιωμένη απόδοση στην εκτίμηση της φάσης διαταραχής των ασθενών.

Λέξεις-Κλειδιά: Αναγνώριση Δράσεων, Αλληλεπίδραση Παιδιών - Ρομπότ, Ρομποτική Αντίληψη, Ψηφιακός Φαινότυπος, Εντοπισμός Ψυχωτικών Υποτροπών, Ταυτοποίηση Ατόμων, Βιοσήματα

Abstract

The study of human activity is a subject of great interest to the scientific community. The primary objective of this dissertation is to develop novel automatic systems with the ability to perceive and acquire knowledge regarding diverse facets of human behavior. The dissertation is structured into two sections, which are primarily distinguished by the method of perception employed; in the initial section, actions are perceived through visual sensors while in the subsequent section, non-visual biometric sensors are employed.

In the first section, we explore the development of intelligent action recognition systems for children during their interactions with robots. Initially, we study the development of a robust integrated system for child-robot interactions with multiple perceptual capabilities, such as action recognition, speech recognition, speaker localization and object tracking, through multiple cameras and leveraging multiple robots. We collect and create a large database of children's activities, including gestures, generalized body movements, speech, expression of emotions, etc., both individually and during interactions. We evaluated the system both in terms of the performance of the recognition modules and the overall interaction of children with robots. Subsequently, we developed a lightweight system for designing and executing educational scenarios based on visual information, for use in more open environments, such as classrooms. During the development of the above systems, we conducted extensive research on visual action and gesture recognition of children activities. We highlighted the need for recognition systems specifically adapted to children's data, assessed the impact of employing both multiple and single recording perspectives on the system's efficiency in recognition, and experimented with incremental learning methods for easier expansion of action recognition capabilities.

In the second section, we delve into the recognition of activity through the use of biometrics from smart wearable sensors. Wearable sensors such as smartwatches are becoming increasingly more popular, creating exciting new opportunities for the design of unique human activity recognition systems. Thus, we explore the use of signals collected from smartwatches, such as acceleration, angular velocity, and heart rate, to study the users' digital phenotypes, i.e., digital profiles that can reflect the characteristic patterns of their daily lives. These patterns are especially useful since they allow us to detect significant deviations in the daily life of a user. Initially we used comprehensive statistical analysis to compare the differences in digital representations-patterns between a control sample and a sample of individuals with psychiatric disorders. We developed an intelligent identification-classification system based on their digital patterns, thoroughly investigating the system's individual attributes. We then addressed the problem of detecting psychotic relapses that often occur in individuals with psychotic disorders. Specifically, we treated the problem as a misclassification problem, assuming that during a relapse of the disorder significant changes in the digital pattern could be observed, thus reducing the probability of identifying the respective user during relapse periods or shortly before them. Finally, we extended the relapse detection system by combining anomaly detection techniques with the techniques we developed for user identification to achieve improved and robust performance in estimating relapses.

Keywords: Action Recognition, Child-Robot Interaction, Multi-robot Perception, Digital Phenotyping, Psychotic Relapse Detection, Activity Recognition, Person Identification, Biosignals

Ευχαριστίες

Θα ήθελα να ευχαριστήσω από καρδιάς τον επιβλέποντα αυτής της διατριβής Καθ. Πέτρο Μαραγκό που μου άνοιξε την πόρτα του εργαστηρίου CVSP, αποτέλεσε το έναυσμα να μάθω τι σημαίνει όραση υπολογιστών και μηχανική μάθηση. Η καθοδήγηση του και η συνεργασία μας αυτά τα χρόνια, με πραγματική έγνοια και ενδιαφέρον, ήταν στήριγμα και οδηγός για την ερευνητική μου πορεία. Στη συνέχεια, θέλω να ευχαριστήσω τον Καθ. Γεράσιμο Ποταμιάνο για την πολύτιμη συνεργασία του στα πρώτα χρόνια της διατριβής όταν ακόμα άρχιζα να μαθαίνω να γράφω paper και να φτιάχνω references. Ακόμα, ευχαριστώ θερμά τον Καθ. Νικόλαο Σμυρνή για την στενή καθοδήγησή του και συνεργασία ως προς το δεύτερο μέρος της διατριβής, σε αυτή τη δύσκολη και πρωτότυπη έρευνα της διερεύνησης των βιοσημάτων. Θέλω να ευχαριστήσω τον Δρ. Γιώργο Ρετσινά για την καίριας σημασίας καθοδήγηση του και το παράδειγμά του ως ερευνητής, που ποτέ δεν αρκείται σε ένα αποτέλεσμα και ψάχνει την επιστημονική εξήγηση πίσω από το κάθε τι.

Επίσης, θα ήθελα να ευχαριστήσω τους ιατρούς Μανώλη Καλησπεράκη και τη Βάσια Γαρυφαλλή για την πολύτιμη συνεργασία και την προσπάθειά τους να μας εξηγήσουν πολύπλευρα τα ζητήματα των ψυχωτικών διαταραχών και των δυσκολιών που αντιμετωπίζουν οι ασθενείς.

Σημαντικό κομμάτι στην ερευνητική μου πορεία αλλά και στη ζωή μου όλα αυτά τα χρόνια είχαν ο Πέτρος και ο Παναγιώτης. Ο Πέτρος με πήρε από το χέρι και μου έδειξε το χώρο της έρευνας ανάμεσα σε γέλια, πειράγματα και πολλά πειράματα. Ο Παναγιώτης, ο αδερφός μου πια, ήταν πάντα εκεί, να κάνουμε μαζί αυτό το ταξίδι, να στηρίζουμε ο ένας τον άλλο με κάθε τρόπο, να μου μαθαίνει πως μιλάνε στους υπολογιστές και να πιστεύει πάντα σε μένα πιο πολύ απ' ότι εγώ. Το ευχαριστώ είναι λίγο και για τους δύο.

Η ερευνητική μου πορεία ξεκίνησε δειλά όταν η φίλη μου Αντιγόνη με ρώτησε αν με ενδιαφέρει να εργαστώ για λίγο στο εργαστήριο CVSP και συνεχίστηκε με την αρχή αυτή του διδακτορικού, κατά την οποία ήταν εκεί κανοντάς την πιο όμορφη και πιο σημαντική. Όλα αυτά τα χρόνια στο εργαστήριο, οι συνεργασίες και οι φιλίες που δημιουργήθηκαν μέσα από ατελείωτες ώρες δουλειάς, συζήτησης, πλάκας, έγνοιας και άλλων τόσων πραγμάτων γίναν ανεκτίμητες μέσα μου. Ο Χρήστος, ο Νίκος, η Ξανθή, η Δάφνη, ο Πάρης, η Γεωργία, η Βίκυ, η Φωτεινή, η Δέσποινα, η Τζο, η Νάνσυ, ο Θανάσης, ο Παναγιώτης όλοι μέσα μου σημαίνουν πολλά περισσότερα από απλοί συνάδελφοι και απλοί συνεργάτες, δίνοντας δύναμη και χαμόγελα στις εύκολες και δύσκολες στιγμές που είχε αυτό το διδακτορικό.

Φυσικά δεν μπορώ να παραλείψω να ευχαριστήσω τους φίλους μου που τόσα χρόνια ήταν δίπλα μου και δεν μου κάκιωναν τις αμέτρητες φορές που τους είπα ότι δεν μπορώ να τους δω γιατί έχω να ασχοληθώ με το διδακτορικό...

Πάνω από όλα θέλω να ευχαριστήσω την οικογένειά μου για την αγάπη τους, τη στήριξή τους, την υπομονή τους όλα αυτά τα χρόνια: τον πατέρα μου Βασίλη που με έμαθε να μαθαίνω, την μητέρα μου Αγγελική που με έκανε να πιστεύω ότι μπορώ να καταφέρω ό,τι θελήσω, την αδερφή μου Δώρα που πάντα μου έδειχνε καινούριους δρόμους, το Στέλιο που μου κρατούσε σφιχτά το χέρι για να μη χάσω το δρόμο μου. Η διατριβή είναι αφιερωμένη σ' αυτούς.

Το τελευταίο μεγάλο ευχαριστώ είναι στους πάνω από 100 ανθρώπους που συμμετείχαν στην ερευνά μας και μας έδωσαν πολύτιμα πειραματικά δεδομένα σε πραγματικές συνθήκες. Σας ευχαριστώ.

Περιεχόμενα

1

Εισαγωγή

1.1 Αντίληψη

Η αντίληψη μπορεί να οριστεί ως η διαδικασία με την οποία ερμηνεύουμε και κατανοούμε τις αισθητηριακές πληροφορίες από το περιβάλλον μας. Περιλαμβάνει όχι μόνο τη φυσική διέγερση των αισθήσεών μας αλλά και την ερμηνεία και την ενσωμάτωση αυτών των πληροφοριών με τις προηγούμενες εμπειρίες και γνώσεις μας. Η αντίληψη είναι απαραίτητη για την επιβίωσή μας, καθώς επιτρέπει να περιηγηθούμε και να αλληλεπιδράσουμε με τον κόσμο γύρω μας. Μας βοηθά να αναγνωρίζουμε αντικείμενα, ανθρώπους, τον περιβάλλον αλλά και να κατανοούμε τις κοινωνικές αλληλεπιδράσεις.

Για να αντιληφθούμε όλα όσα μας περιβάλλουν χρησιμοποιούμε τις αισθήσεις μας, μέσω των οποίων λαμβάνονται, επεξεργάζονται και ενσωματώνονται οι αισθητηριακές πληροφορίες. Οι πέντε αισθήσεις είναι η όραση, η ακοή, η αφή, η γεύση και η όσφρηση. Κάθε μία από αυτές παρέχει διαφορετικές πληροφορίες για το περιβάλλον και επεξεργάζεται από διαφορετικά αισθητήρια συστήματα στον εγκέφαλο. Μαζί, επιτρέπουν στους οργανισμούς να αντιλαμβάνονται και να αλληλεπιδρούν με το περιβάλλον τους με πολύπλοκο τρόπο.

Η αντίληψη είναι σημαντική γιατί επηρεάζει τις σκέψεις, τις πεποιθήσεις και τις συμπεριφορές μας. Οι αντιλήψεις μας διαμορφώνουν το πώς βλέπουμε τον εαυτό μας και τους άλλους και πώς κατανοούμε τον κόσμο. Διαδραματίζει επίσης κρίσιμο ρόλο στη λήψη αποφάσεων, καθώς οι αντιλήψεις μας επηρεάζουν τις επιλογές που κάνουμε και τον τρόπο με τον οποίο αντιδρούμε σε διαφορετικές καταστάσεις. Επιπλέον, η αντίληψη μπορεί να επηρεαστεί από παράγοντες όπως η κουλτούρα, τα συναισθήματα και οι προσωπικές προκαταλήψεις, οι οποίες μπορούν να επηρεάσουν την κατανόησή μας για τα γεγονότα και τις αλληλεπιδράσεις.

Τη σχέση μεταξύ αντίληψης και αντικειμενικής πραγματικότητας και το πως η αντίληψή μας διαμορφώνεται από τις υλικές συνθήκες του κόσμου γύρω μας συζητά ο Ένγκελς στη Διαλεκτική της φύσης [Engels; 1960] Ο Ένγκελς υποστηρίζει ότι η αντίληψή μας δεν είναι μια παθητική αντανάκλαση του κόσμου, αλλά είναι μια ενεργή διαδικασία ερμηνείας που διαμορφώνεται από το κοινωνικό και ιστορικό μας πλαίσιο. Υποστηρίζει ότι η επιστήμη παρέχει μια πιο ακριβή και αντικειμενική κατανόηση του κόσμου από τις υποκειμενικές μας αντιλήψεις και τονίζει τη σημασία της επιστημονικής έρευνας για την κατανόηση της σχέσης μεταξύ αντίληψης και αντικειμενικής πραγματικότητας.

1.2 Αυτόματα Συστήματα Αντίληψης: Είδη και Εφαρμογές

Το θέμα που πραγματεύεται η παρούσα διατριβή είναι η ανάπτυξη συστημάτων ικανών να αντιληφθούν τις ανθρώπινες δράσεις. Ο στόχος των αυτόματων συστημάτων αντίληψης είναι

να επιτρέψουν στις μηχανές να λαμβάνουν τεκμηριωμένες αποφάσεις με βάση την κατανόησή τους για το περιβάλλον γύρω τους. Αυτή η τεχνολογία είναι ζωτικής σημασίας σε διάφορες εφαρμογές, όπως η ρομποτική, τα αυτόνομα οχήματα και τα συστήματα όρασης υπολογιστών. Γενικά, τα τελευταία χρόνια με την ανάπτυξη της τεχνητής νοημοσύνης (Artificial Intelligence - AI) τα αυτόματα συστήματα αντίληψης χρησιμοποιούνται σε ολοένα και περισσότερες εφαρμογές βοηθώντας ή και αποκαθιστώντας την ανθρώπινη παρέμβαση, π.χ. διαλογικά συστήματα παροχής υπηρεσιών [Ni et al.; 2023], γλωσσικά μοντέλα ικανά για παραγωγή κειμένου [Brown et al.; 2020].

Τα αυτόματα συστήματα αντίληψης προσφέρουν τεράστιες δυνατότητες σε πολλούς τομείς της ζωής μας π.χ. βιομηχανία, υγεία, ασφάλεια. Πρώτον, ενισχύουν την αποδοτικότητα και την παραγωγικότητα. Με τις μηχανές να μπορούν να αντιλαμβάνονται το περιβάλλον τους, να το κατανοούν και να λαμβάνουν αποφάσεις, μειώνεται η ανάγκη για ανθρώπινη παρέμβαση, επιταχύνοντας έτσι τη διαδικασία. Για παράδειγμα, στα εργοστάσια παραγωγής, τα συστήματα αυτόματης αντίληψης μπορούν να ανιχνεύσουν ελαττώματα σε προϊόντα πιο γρήγορα από τον άνθρωπο [Villalba-Diez et al.; 2019], επιτρέποντας έγκαιρες παρεμβάσεις, μειώνοντας τη σπατάλη και αυξάνοντας την παραγωγικότητα.

Επιπρόσθετα, τα συστήματα αυτόματης αντίληψης ενισχύουν την ασφάλεια σε πολλές εφαρμογές. Για παράδειγμα, στα αυτόνομα οχήματα, τα αυτόματα συστήματα αντίληψης επιτρέπουν στο αυτοκίνητο να ανιχνεύει εμπόδια, πεζούς [Liu et al.; 2019b] και άλλα οχήματα στο δρόμο και να λαμβάνει τεκμηριωμένες αποφάσεις για την αποφυγή ατυχημάτων. Στον κλάδο της υγειονομικής περίθαλψης, τα αυτόματα συστήματα αντίληψης μπορούν να ανιχνεύσουν ασθένειες και ανωμαλίες στις ιατρικές εικόνες [Liu et al.; 2021a], επιτρέποντας έγκαιρες παρεμβάσεις.

Ακόμα, τα συστήματα αυτόματης αντίληψης επιτρέπουν στις μηχανές να λειτουργούν σε περιβάλλοντα που είναι επικίνδυνα ή απρόσιτα για τον άνθρωπο. Για παράδειγμα, στην εξόρυξη, τα αυτόματα συστήματα αντίληψης μπορούν να χρησιμοποιηθούν για την ανίχνευση και την εξόρυξη ορυκτών σε ορυχεία χωρίς να διακινδυνεύουν ανθρώπινες ζωές. Στην εξερεύνηση του διαστήματος, τα αυτόματα συστήματα αντίληψης μπορούν να χρησιμοποιηθούν για την εξερεύνηση του διαστήματος, επιτρέποντας νέες ανακαλύψεις και καινοτομίες [Bickel et al.; 2020].

Η ανάπτυξη των αυτόματων συστημάτων αντίληψης χρονολογείται από τη δεκαετία του 1960, με την ανάπτυξη της τεχνητής νοημοσύνης και των συστημάτων υπολογιστικής όρασης. Τα πρώτα συστήματα αυτόματης αντίληψης ήταν απλά συστήματα αναγνώρισης εικόνων που μπορούσαν να ανιχνεύσουν και να ταξινομήσουν αντικείμενα σε εικόνες. Με τα χρόνια, η τεχνολογία έχει προχωρήσει, με νέους αλγόριθμους και εργαλεία να αναπτύσσονται για να βελτιώσουν την ακρίβεια και την ταχύτητα των αυτόματων συστημάτων αντίληψης. Στις μέρες μας, τα συστήματα αναγνώρισης εικόνας έχουν εξελιχθεί ραγδαία, σε σημείο που η αναγνώριση αντικειμένων από εικόνες έχει ξεπεράσει την ακρίβεια αναγνώρισης που έχει ο άνθρωπος, έχει επεκταθεί σε δύσκολες εργασίες ενώ η προσπάθεια αντίληψης επιμέρους στοιχείων σε βίντεο είναι αυτή που είναι ακόμα από τα κορυφαία ζητήματα επίλυσης των μεθόδων όρασης υπολογιστών και μηχανικής μάθησης.

Μερικές από τις εφαρμογές των συστημάτων επεξεργασίας εικόνας και βίντεο αφορούν:

- την ιατρική απεικόνιση, για διάγνωση, σχεδιασμό θεραπείας και ερευνητικούς σκοπούς. Για παράδειγμα, οι εικόνες μαγνητικής τομογραφίας και αξονικής τομογραφίας υποβάλλονται σε επεξεργασία για την ενίσχυση της αντίθεσης και τη βελτίωση της οπτικοποίησης των εσωτερικών οργάνων και δομών [Wen et al.; 2018, Yin et al.; 2022].

- την επιτήρηση από κάμερες (π.χ. απλές ή θερμικές) για τον εντοπισμό και την παρακολούθηση αντικειμένων ή καταστάσεων, την αναγνώριση προσώπων και την ανάλυση

προτύπων συμπεριφοράς [Chen; 2020]. Αυτά τα συστήματα μπορούν επίσης να χρησιμοποιηθούν σε εφαρμογές δημόσιας ασφάλειας, π.χ. ξέσπασμα φωτιάς,.

- τη γεωργία για την παρακολούθηση της ανάπτυξης των καλλιεργειών [SoftGrip;], την ανίχνευση ασθενειών και την ανάλυση της σύστασης του εδάφους. Αυτά τα συστήματα μπορούν να παρέχουν στους αγρότες δεδομένα σε πραγματικό χρόνο για την υγεία και την απόδοση των καλλιεργειών, επιτρέποντάς τους να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με την άρδευση, τη λίπανση και τον έλεγχο των παρασίτων.

Άλλα συστήματα αυτόματης αντίληψης είναι τα συστήματα επεξεργασίας ήχου όπου με μικρόφωνα και αλγόριθμους επεξεργασίας ήχου αναλύουν ήχους και παρέχουν πληροφορίες για το περιβάλλον, π.χ. αναγνώρισης ομιλίας, εντοπισμός ομιλητή, διαχωρισμός ήχων και ομιλητών. Στην καθημερινότητα μπορούμε εύκολα να συνδιαλλαγούμε με τέτοια συστήματα όπως τους βοηθούς φωνής, π.χ. η Alexa της Amazon και η Google's Assistant, χρησιμοποιούν συστήματα επεξεργασίας ήχου για να αναγνωρίζουν προφορικές εντολές και να παρέχουν απαντήσεις.

Ένας ακόμα τύπος συστήματος αυτόματης αντίληψης είναι το σύστημα απτικής επεξεργασίας (haptics). Αυτά το συστήματα χρησιμοποιούν αισθητήρες που μπορούν να ανιχνεύσουν την πίεση, τη θερμοκρασία και άλλα φυσικά χαρακτηριστικά για να παρέχουν πληροφορίες για το περιβάλλον. Για παράδειγμα, τα ρομπότ που εκτελούν βιομηχανικές εργασίες χρησιμοποιούν συστήματα απτικής επεξεργασίας για να ανιχνεύσουν την υφή και τη σκληρότητα των αντικειμένων και να προσαρμόσουν τις κινήσεις τους ανάλογα.

Αναρίθμητα είναι τα αυτόματα συστήματα αντίληψης που έχουν ήδη αναπτυχθεί και αναπτύσσονται καθημερινά ακολουθώντας την ταχύτατη πρόοδο της επιστήμης στις μέρες μας. Τα περισσότερα από αυτά πλέον χρησιμοποιούν διάφορους τύπους αισθητήρων και αυτόματων συστημάτων αντίληψης για να παρέχουν μια πιο ολοκληρωμένη κατανόηση του περιβάλλοντος. Για παράδειγμα, τα αυτόνομα αυτοκίνητα χρησιμοποιούν ένα συνδυασμό συστημάτων επεξεργασίας εικόνας, επεξεργασίας ήχου και απτικής επεξεργασίας για τη συλλογή πληροφοριών σχετικά με το περιβάλλον και τη λήψη τεκμηριωμένων αποφάσεων με βάση αυτές τις πληροφορίες.

1.3 Αυτόματα Συστήματα Αντίληψης και Εκμάθησης Δράσεων

Κύριο ζήτημα που πραγματεύεται αυτή η διατριβή είναι η ανάπτυξη συστημάτων αναγνώρισης δράσεων τόσο με χρήση οπτικής πληροφορίας όσο και με χρήση βιοσημάτων, που στη διεθνή βιβλιογραφία τα συναντάμε με τους όρους action και activity recognition. Συνήθως ο όρος action recognition αναφέρεται στην αναγνώριση δράσεων του σώματος που κωδικοποιούνται μέσω της οπτικής πληροφορίας, ενώ ο όρος activity recognition συνήθως αφορά πιο πολύπλοκες δραστηριότητες και η καταγραφή τους συνήθως αξιοποιεί και άλλους αισθητήρες καταγραφής της ανθρώπινης δραστηριότητας, πέρα από τους οπτικούς.

Στην ενότητα αυτή, για λόγους διάκρισης, θα αναφερόμαστε στο **action recognition** ως αναγνώριση ενεργειών και στο **activity recognition** ως αναγνώριση δραστηριότητας, παρ' ότι γενικά στη βιβλιογραφία δεν είναι έννοιες που διαχωρίζονται με ακρίβεια. Οι δύο αυτοί όροι αφορούν συναφή πεδία στην όραση υπολογιστών και τη μηχανική μάθηση και έχουν λάβει σημαντική προσοχή την τελευταία δεκαπενταετία. Περιλαμβάνουν τη χρήση αλγορίθμων και μοντέλων για την αυτόματη αναγνώριση και ταξινόμηση διαφορετικών τύπων ανθρώπινων ενεργειών και δραστηριοτήτων με βάση οπτικά και άλλου είδους δεδομένα, με πολλές εφαρμογές στη ρομποτική, την υγειονομική περίθαλψη και την παρακολούθηση (surveillance).

Αναγνώριση Ενεργειών - Action Recognition

Πιο συγκεκριμένα, η αναγνώριση ενεργειών αναφέρεται στη διαδικασία ταξινόμησης συγκεκριμένων ανθρώπινων ενεργειών κάθε φορά, όπως το περπάτημα, το τρέξιμο ή το άλμα, με βάση κυρίως οπτικά δεδομένα. Αυτό μπορεί να επιτευχθεί χρησιμοποιώντας διάφορες τεχνικές, συμπεριλαμβανομένων μοντέλων βαθιάς μάθησης, όπως συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNN) και αναδρομικά νευρωνικά δίκτυα (Recurrent Neural Networks - RNN), καθώς και πιο παραδοσιακούς αλγόριθμους μηχανικής μάθησης, όπως μηχανές υποστήριξης διανυσμάτων (Support Vector Machine - SVM).

Μία από τις βασικές προκλήσεις στην αυτόματη αναγνώριση ενεργειών είναι η αντιμετώπιση των παραλλαγών στον τρόπο με τον οποίο διαφορετικοί άνθρωποι εκτελούν την ίδια ενέργεια. Για παράδειγμα, διαφορετικοί άνθρωποι μπορεί να περπατούν με διαφορετικό βηματισμό, γεγονός που μπορεί να κάνει δύσκολη την ανάπτυξη μοντέλων αναγνώρισης μεγάλης ακρίβειας. Έτσι, οι αλγόριθμοι που χρησιμοποιούνται μαθαίνουν από μια πληθώρα παραδειγμάτων διαφορετικών ανθρώπων που εκτελούν την ίδια ενέργεια και είναι σε θέση να προσαρμοστούν σε αυτές τις παραλλαγές. Μια άλλη πρόκληση που συναντάται είναι η αντιμετώπιση πολύπλοκων περιβάλλοντων και ακατάστατων παρασκηνίων. Αυτό μπορεί να δυσκολέψει τους αλγόριθμους να προσδιορίσουν με ακρίβεια τις ενέργειες που λαμβάνουν χώρα σε μια σκηνή. Για την αντιμετώπιση αυτού, πολλές τεχνικές περιλαμβάνουν την κατάτμηση των οπτικών δεδομένων σε μικρότερες περιοχές και την ανάλυση κάθε περιοχής ξεχωριστά, καθώς και τη χρήση πιο εξελιγμένων μοντέλων όπως τα spatiotemporal CNN [Feichtenhofer et al.; 2017] που μπορούν να συλλάβουν τόσο χωρικές όσο και χρονικές πληροφορίες.

Σχετικά με την αναγνώριση ενεργειών από βίντεο, βλέπουμε πως αποτελεί είναι ένα δύσκολο αλλά δημοφιλές πρόβλημα στον τομέα της όρασης υπολογιστών, καθώς απαιτεί τον ακριβή εντοπισμό και την ταξινόμηση των ανθρώπινων ενεργειών από μια ακολουθία στιγμιοτύπων ενός βίντεο. Μία σημαντική εργασία που αξιολογεί και συγκρίνει διάφορους ανιχνευτές χαρακτηριστικών και τοπικούς περιγραφητές χρησιμοποιώντας οπτικές λέξεις και ταξινόμηση μέσω SVM σε τέσσερις δημοφιλείς βάσεις παρουσιάζεται στο [Wang et al.; 2009]. Οι συγγραφείς μεταξύ άλλων, αποδεικνύουν ότι η τακτική δειγματοληψία χωροχρονικών χαρακτηριστικών αποδίδει σταθερά καλύτερα σε σύγκριση με τη χρήση χωροχρονικών ανιχνευτών σημείων ενδιαφέροντος για το πρόβλημα της αναγνώρισης ανθρώπινων ενεργειών σε ρεαλιστικές συνθήκες.

Τα τελευταία χρόνια, οι προσεγγίσεις που βασίζονται σε βαθιά μάθηση έχουν επιτύχει πολύ καλές επιδόσεις σε διάφορα προβλήματα της αναγνώρισης ενεργειών. Ένα από τα πιο ευρέως χρησιμοποιούμενα μοντέλα βαθιάς μάθησης για την αναγνώριση ενεργειών βίντεο είναι το μοντέλο CNN δύο ροών που προτάθηκε από τους Simonyan και Zisserman στην εργασία τους [Simonyan and Zisserman; 2014]. Το μοντέλο που προτείνουν χρησιμοποιεί δύο ξεχωριστές ροές εισόδου, μία για χωρικές πληροφορίες (προέρχεται από στιγμιότυπα RGB) και μία για χρονικές πληροφορίες (προέρχεται από την οπτική ροή) και τις συνδυάζει για να ταξινομήσει τις ενέργειες. Αυτό το μοντέλο πέτυχε κορυφαία απόδοση σε πολλά σύνολα δεδομένων αναφοράς, συμπεριλαμβανομένων των UCF101 [Soomro et al.; 2012] και HMDB51 [Kuehne et al.; 2011], ενώ ακόμα η λογική της προτεινόμενης αρχιτεκτονικής συναντάται σε πολλές εργασίες.

Μια άλλη δημοφιλής προσέγγιση είναι τα 3D CNN, τα οποία επεκτείνουν τα παραδοσιακά 2D CNN στη χρονική διάσταση. Αυτό επιτρέπει στο μοντέλο να μαθαίνει χωροχρονικά χαρακτηριστικά απευθείας από ακατέργαστα καρέ βίντεο. Στην εργασία [Wang et al.; 2018], οι Wang et al. έχοντας ως αναφορά τις μεθόδους non-local means της όρασης υπολογιστών, πρότειναν ένα δομικό μπλοκ για την καταγραφή χρονικών εξαρτήσεων μεγάλης εμβέλειας σε τρισδιάστατα CNN.

Τα μοντέλα που βασίζονται στην προσοχή (attention models) έχουν επίσης δείξει πολλά

υποσχόμενα αποτελέσματα για την αναγνώριση ενεργειών βίντεο. Οι μηχανισμοί προσοχής επιτρέπουν στο μοντέλο να εστιάζει στα πιο σχετικά στιγμιότυπα ή σε περιοχές του βίντεο. Οι Lin et al. [Lin et al.; 2019] πρότειναν μια μονάδα χρονικής μετατόπισης που ενσωματώνει τη χρονική προσοχή σε ένα 2D CNN, που οδηγεί σε βελτιωμένη απόδοση και χαμηλότερο υπολογιστικό κόστος απ' ότι τα 3D CNN.

Πιο πρόσφατα, μοντέλα που βασίζονται σε μετασχηματιστές (transformers), τα οποία έχουν δείξει μεγάλη επιτυχία σε εργασίες επεξεργασίας φυσικής γλώσσας, έχουν επίσης εφαρμοστεί στην αναγνώριση ενεργειών βίντεο. Οι Liu et al. [Liu et al.; 2021b] πρότειναν ένα ιεραρχικό μοντέλο μετασχηματιστή που αποκαλούν Swin Transformer του οποίου η αναπαράσταση υπολογίζεται με μετακινούμενα παράθυρα. Οι συγγραφείς υποστηρίζουν πως το μοντέλο αυτό είναι πολύ αποδοτικό καθώς περιορίζει τους υπολογισμούς που σχετίζονται με το self-attention να γίνονται σε μη επικαλυπτόμενα παράθυρα ενώ παράλληλα επιτρέπει τη συσχέτιση της πληροφορίας ανάμεσα σε διαφορετικά παράθυρα. Η αρχιτεκτονική αυτή διακρίνεται ως προς την ευελιξία της στη μοντελοποίηση της πληροφορίας σε διάφορες κλίμακες και τη γραμμική πολυπλοκότητα της σε σχέση με το μέγεθος της εικόνας.

Αναγνώριση Δραστηριότητας - Activity Recognition

Η αναγνώριση δραστηριότητας συναντάται σε διάφορους τομείς όπως η υγειονομική περίθαλψη, ο αθλητισμός και η αλληλεπίδραση ανθρώπου-υπολογιστή. Ως δραστηριότητες ορίζονται πιο πολύπλοκες δράσεις, που είτε θα μπορούσαμε να πούμε ότι είναι μια ακολουθία ενεργειών (π.χ. μαγείρεμα) είτε ακόμα και πιο αφηρημένες καταστάσεις που δεν ορίζονται από μια συγκεκριμένη διαδοχή κινήσεων, όπως π.χ. η δραστηριότητα ενός ανθρώπου κατά τη διάρκεια της ημέρας. Στη διαδικασία της ανάπτυξης συστημάτων αναγνώρισης δραστηριοτήτων εισάγονται πλέον αισθητήρες κάθε είδους ώστε να έχουμε στη διάθεσή μας αρκετή πληροφορία για να λύσουμε το εκάστοτε πρόβλημα. Πολλές φορές μάλιστα προτιμώνται αισθητήρες εκτός των καμερών καθώς βοηθούν στη διαφύλαξη του απορρήτου των συμμετεχόντων και σε ορισμένες περιπτώσεις είναι πιο εύκολη η χρήση τους και λιγότερο παρεμβατική στη ζωή όσων καταγράφονται. Στην παρούσα διατριβή, η αναγνώριση δραστηριότητας αναφέρεται στη χρήση πολυτροπικών δεδομένων, όπως τα δεδομένα από φορητούς αισθητήρες, π.χ. επιταχυνσιόμετρα ή μόνιτορ καρδιακών παλμών.

Στην εργασία [Lv et al.; 2019], οι συγγραφείς προτείνουν ένα end-to-end βαθύ νευρωνικό δίκτυο για αναγνώριση σύνθετης δραστηριότητας χρησιμοποιώντας πολυτροπική πληροφορία από αισθητήρες (γυροσκόπιο, επιταχυνσιόμετρο και μαγνητόμετρο) και συνδυάζοντας CNN και RNN δίκτυα. Πιο συγκεκριμένα, χρησιμοποιούν ιεραρχικά δίκτυα CNN για να εκμεταλλευτούν τη σχέση μεταξύ των δεδομένων που καταγράφονται από παρόμοιους αισθητήρες, ενοποιούν τις σχέσεις των τροπικοτήτων διαφόρων αισθητήρων και χρησιμοποιούν RNN για να μοντελοποιήσουν τη χρονική σχέση των σημάτων που καταγράφονται. Τέλος, αξιολογούν το σύστημά τους σε δύο βάσεις με πραγματικά δεδομένα.

Οι έξυπνες συσκευές όπως τα έξυπνα ρολόγια και τα smartphone διαθέτουν επίσης ενσωματωμένους αισθητήρες που μπορούν να χρησιμοποιηθούν για την αναγνώριση δραστηριότητας. Στο [Wang et al.; 2019] παρουσιάζεται μια εκτενής έρευνα σχετικά με τους σύγχρονους αισθητήρες και τις τεχνικές που χρησιμοποιούνται για την εκμάθηση και την ταξινόμηση των δραστηριοτήτων (τεχνικές βαθιάς μάθησης και πιο κλασικές προσεγγίσεις).

Μια άλλη προσέγγιση στην αναγνώριση δραστηριοτήτων προτείνεται στην εργασία [Zou et al.; 2019] μια πολυτροπική προσέγγιση για αναγνώριση δραστηριότητας συνδυάζοντας σήματα WiFi και οπτικές πληροφορίες. Οι συγγραφείς προτείνουν ένα κοινό πολυτροπικό πλαίσιο μάθησης ώστε να βελτιώσει την ακρίβεια και την ευρωστία της αναγνώρισης που επιτυγχάνει κάθε τροπικότητα ξεχωριστά. Ενώ μετέπειτα οι Zhang et al. [Zhang et al.; 2020] προσπαθούν

μέσω της αύξησης των δεδομένων (data augmentation) και της αξιοποιήσης μιας Dense-LSTM αρχιτεκτονικής να επιτύχουν μια αυξημένη ακρίβεια αναγνώρισης για δέκα δραστηριότητες.

1.4 Δομή Διατριβής

Συνολικά, η αναγνώριση και εκμάθηση ανθρώπινων δράσεων έχει ένα ανεξάντλητο εύρος προβλημάτων προς επίλυση ενώ η διαρκής εξέλιξη των μεθόδων αναγνώρισης επιτρέπουν την αντιμετώπιση πολλών περιορισμών και την ανάπτυξη συστημάτων μεγάλης ακρίβειας. Στη συνέχεια της διατριβής θα μελετήσουμε διεξοδικά τα θέματα της αναγνώρισης παιδικών δράσεων κατά την αλληλεπίδραση παιδιών και ρομπότ, την ανάπτυξη συστημάτων αντίληψης και διαχείρισης ρομποτικών συστημάτων για τέτοιες αλληλεπιδράσεις καθώς και την αναγνώριση πιο αφηρημένων δραστηριοτήτων των ανθρώπων μέσα από πολυμεσικούς αισθητήρες.

Έτσι, η δομή των επόμενων κεφαλαίων της διατριβής οργανώνεται ως εξής:

- Στο κεφάλαιο 2 παρουσιάζεται μια επισκόπηση για τα συστήματα αλληλεπίδρασης παιδιών και ρομπότ (Child-Robot Interaction) που έχουν αναπτυχθεί τα τελευταία χρόνια μαζί με τους σκοπούς και τις εφαρμογές τους. Επίσης, εξηγούνται αναλυτικά τα δύο CRI συστήματα που αναπτύξαμε κατά τη διάρκεια της παρούσας διατριβής: α) το ChildBot, που είναι ένα σύστημα αντίληψης και αλληλεπίδρασης παιδιών με πολλαπλά ρομπότ, και β) το TeachBot που αποτελεί ένα ευφυές σύστημα αλληλεπίδρασης παιδιού-ρομπότ για την σχεδίαση και την εκτέλεση εκπαιδευτικών-ψυχαγωγικών σεναρίων με έμφαση στην οπτική πληροφορία.

- Στο κεφάλαιο 3 αναλύονται οι μέθοδοι που αναπτύχθηκαν για την αναγνώριση δράσεων από μία όψη, οι μέθοδοι σύμμειξης της πληροφορίας για αναγνώριση από πολλαπλούς αισθητήρες καθώς και οι τεχνολογίες που αξιοποιήθηκαν για την ανάπτυξη συστήματος επαυξημένης μάθησης. Οι παραπάνω μέθοδοι, που αφορούν τόσο κλασικές μεθόδους όρασης υπολογιστών όσο και μεθόδους βαθιάς μάθησης, αξιολογούνται σε δεδομένα παιδιών που εκτελούν κινήσεις και χειρονομίες.

- Στο κεφάλαιο 4 παρουσιάζουμε μια επισκόπηση των αντιληπτικών συστημάτων σε εφαρμογές ηλεκτρονικής υγείας αλλά και πιο συγκεκριμένα σε εφαρμογές ψυχικής υγείας. Στη συνέχεια παρουσιάζουμε το ερευνητικό έργο e-Prevention, την πολυμεσική βάση δεδομένων που δημιουργήθηκε για ερευνητικούς σκοπούς και περιλαμβάνει περισσότερες από 20.000 μέρες καταγραφών καθώς και μια επισκόπηση των ερευνητικών εργασιών που έχουν αξιοποιήσει τα βιοσήματα της βάσης.

- Στο κεφάλαιο 5 διερευνούμε αναπαραστάσεις βιοσημάτων που καταγράφονται μέσω έξυπνων ρολογιών ως προς τη σημαντικότητά τους και τη διαφοροποίησή τους ανάμεσα σε δύο σύνολα, ελέγχου και ασθενών με ψυχωτικές διαταραχές. Στη συνέχεια, μελετάμε περαιτέρω τη δημιουργία ψηφιακών φαινοτύπων μέσω των εξαχθέντων βιοδεικτών καθώς και την ανάπτυξη αλγορίθμων ικανών να ταυτοποιούν τους χρήστες των ρολογιών μέσω των ψηφιακών τους φαινοτύπων.

- Στο κεφάλαιο 6 εστιάζουμε στον εντοπισμό των διαφορετικών περιόδων των ψυχωτικών διαταραχών, ύφεσης και υποτροπής, μέσα από τις αλλαγές που μπορεί να παρουσιάζονται στους ψηφιακούς φαινότυπους ασθενών. Η διάκριση αυτή μελετάται τόσο μέσω του πρίσματος της μεταβολής της πιθανότητας ταυτοποίησης ενός ατόμου από τα βιοσήματά του όσο και ως συνδυασμός του εντοπισμού ανωμαλιών κατά την ανακατασκευή των βιοσημάτων και της ταυτοποίησης των ασθενών.

- Στο κεφάλαιο 7 παρουσιάζουμε συνοπτικά τις συνεισφορές της παρούσας διατριβής καθώς και κάποιες μελλοντικές κατευθύνσεις που θα μπορούσαν να επιλεγούν για την επέκταση της έρευνας.

Μέρος Ι

Αναγνώριση Δράσεων Με Χρήση Οπτικής Πληροφορίας

2

Συστήματα Αλληλεπίδρασης Παιδιών και Ρομπότ

2.1 Επισκόπηση Συστημάτων Αλληλεπίδρασης Παιδιών και Ρομπότ

Στον τομέα της αλληλεπίδρασης ανθρώπου-ρομπότ (Human-Robot Interaction - HRI) η αλληλεπίδραση μεταξύ παιδιών και ρομπότ κατέχει ιδιαίτερη θέση. Η αλληλεπίδραση παιδιών-ρομπότ (Child-Robot Interaction - CRI) είναι διαφορετική από την αλληλεπίδραση μεταξύ ενηλίκων και ρομπότ, καθώς τα παιδιά έχουν διαφορετικό, ατελώς ανεπτυγμένο, γνωστικό επίπεδο. Τα παιδιά πολλές φορές δεν βλέπουν ένα ρομπότ ως μια μηχανή που εκτελεί ένα υπολογιστικό πρόγραμμα, αλλά του αποδίδουν χαρακτηριστικά που συνήθως αποδίδονται σε έμβιους οργανισμούς [Belpaeme et al.; 2013]. Ο ανθρωπομορφισμός βέβαια παρατηρείται τόσο σε ενήλικες όσο και σε παιδιά, ήδη από την ηλικία των τριών ετών ή πιθανά ακόμα και σε νεότερες ηλικίες [Berry and Springer; 1993]. Τα παιδιά δείχνουν πιο πρόθυμα από τους ενήλικες να διατηρήσουν την ψευδαίσθηση ότι τα ρομπότ είναι ζωντανά [Turkle et al.; 2006]. Άλλωστε η ικανότητα προσποίησης, αλλά και απόδοσης ανθρώπινων στοιχείων σε αντικείμενα είναι στοιχεία που δεν συναντώνται συνήθως σε συγγενή ανθρωποειδή, π.χ. γορίλες, χιμπατζήδες, αλλά σχετίζονται αποκλειστικά με την εξέλιξη του ανθρώπου. Πιστεύεται ότι και οι δύο ικανότητες έχουν νευροψυχολογική βάση και ότι αποτελούν αναπόσπαστο μέρος της ανάπτυξης των κοινωνικών, γνωστικών και γλωσσικών δεξιοτήτων, ενώ ταυτόχρονα παρατηρούνται ιδιαίτερα κατά την προσχολική και πρωτοβάθμια εκπαίδευση ενός παιδιού, δηλαδή μεταξύ τριών και έντεκα ετών [Smith; 2009]. Αυτή η τάση για κοινωνικό παιχνίδι διαφαίνεται και στην στάση των παιδιών σχετικά με την τεχνολογία. Τα παιχνίδια και συγκεκριμένα τα ρομπότ αντιμετωπίζονται εύκολα ως ζωντανά πλάσματα που έχουν «πιστεύω, επιθυμίες και προθέσεις» [Rao and Georgeff; 1995].

Μία από τις κύριες προκλήσεις στην αλληλεπίδραση παιδιού-ρομπότ είναι να αποτυπωθούν οι αυθόρμητες και γνήσιες αντιλήψεις των παιδιών για τα ρομπότ και για την αλληλεπίδραση γενικότερα. Ουσιώδης διαφοροποίηση ανάμεσα στην αλληλεπίδραση των παιδιών και των ενηλίκων με τα ρομπότ έγκειται στη συμπεριφορά και τα φυσικά χαρακτηριστικά των παιδιών, π.χ. την άρθρωση, τον αυθορμητισμό και το ύψος του σώματος τους που διαφέρουν σημαντικά από εκείνα των ενηλίκων. Έτσι, είναι ανάγκη τα αυτόματα συστήματα αντίληψης που προορίζονται για τη χρήση με παιδιά, να αναπτυχθούν με ειδικό τρόπο ώστε να λαμβάνουν υπ' όψιν αυτές τις διαφορές και τελικώς να αντιλαμβάνονται επιτυχώς όσα κάνει το παιδί, όπως τις δράσεις του και την ομιλία του [Kennedy et al.; 2017].

Λόγω του μεγάλου εύρους των ρομποτικών εφαρμογών, ο τομέας της αλληλεπίδρασης παι-

διού - ρομπότ (CRI) αφορά όχι μόνο μηχανικούς αλλά και επιστήμονες από διάφορους τομείς όπως θεραπευτές, ψυχολόγους, παιδαγωγούς, δασκάλους κ.α.. Πολλές μελέτες σε αυτόν τον τομέα επικεντρώνονται στο πως η πνευματική και γνωστική ανάπτυξη των παιδιών επηρεάζεται από τέτοιου είδους αλληλεπιδράσεις [Boccanfuso et al.; 2016, Belpaeme et al.; 2018, Davison et al.; 2020]. Επίσης, οι θεραπευτικοί σκοποί συναντώνται ιδιαιτέρως συχνά κατά τη χρήση των ρομπότ σε παιδιά, π.χ. πολυάριθμες έρευνες βρίσκουμε για τη χρήση των ρομπότ με παιδιά του αυτιστικού φάσματος [Wood et al.; 2017], την παιδιατρική αποκατάσταση [Pulido et al.; 2017] και τη διαχείριση του διαβήτη [Robinson et al.; 2020]. Μεταξύ του τεράστιου αριθμού τεχνολογιών που έχουν αναπτυχθεί τις τελευταίες δεκαετίες για εκπαιδευτικούς και ψυχαγωγικούς σκοπούς, τα κοινωνικά ρομπότ έχουν εξέχουσα θέση λόγω του μεγάλου φάσματος εφαρμογών στις οποίες μπορούν να χρησιμοποιηθούν, π.χ. να μάθουν στα παιδιά να γράφουν [Chandra et al.; 2019], να τους διδάξουν μια δεύτερη γλώσσα [Kennedy et al.; 2016] ή ακόμα να βοηθήσουν στην κοινωνική και συναισθηματική ανάπτυξή τους [Wolfe et al.; 2018].

Όσον αφορά στις εκπαιδευτικές εφαρμογές, πολλές εργασίες επικεντρώνονται στη θεωρητική διερεύνηση διαφορετικών συμπεριφορών κοινωνικών ρομπότ στη μαθησιακή εμπειρία, χωρίς να εμβαθύνουν στις τεχνικές πτυχές, αλλά κυρίως χρησιμοποιώντας έτοιμες λύσεις - μη ερευνητικές για την αντίληψη του περιβάλλοντος και των δράσεων. Αυτό έχει ως αποτέλεσμα έναν περιορισμένο χώρο αλληλεπίδρασης. Στην εργασία των Kennedy et al. [Kennedy et al.; 2015], τα παιδιά έπαιζαν με ένα ρομπότ NAO σε ένα εκπαιδευτικό σενάριο μαθηματικών. Στο [Saerbeck et al.; 2010], οι Saerbeck et al. μελέτησαν την επίδραση της συμπεριφοράς των κοινωνικών ρομπότ στη μαθησιακή απόδοση των παιδιών στο πλαίσιο μιας εργασίας εκμάθησης γλώσσας. Παρόμοιες μελέτες μπορούν να βρεθούν στο [Gordon et al.; 2015].

Έχει παρατηρηθεί πως σε μια τάξη όπου ο μαθητής βρίσκεται στο επίκεντρο της εκπαιδευτικής διαδικασίας, η παρουσία των ρομποτικών πρακτόρων κάνει πιο ευχάριστη τη διαδικασία μάθησης και παρακινεί τους μαθητές να συμμετέχουν [Kanda et al.; 2012]. Τα ρομπότ ενθαρρύνοντας την αλληλεπίδραση μαζί τους, επιτυγχάνουν να εξάπτουν την περιέργεια των παιδιών [Kanda et al.; 2007]. Μέχρι στιγμής ένας μεγάλος όγκος εργασιών επικεντρώνονται σε θέματα που αφορούν στην ένταξη και στη χρήση των ρομπότ στην σχολική αίθουσα και για το λόγο αυτό οι αλληλεπιδράσεις σχεδιάζονται με πολύ καθορισμένους σκοπούς-στόχους. Τέτοιοι στόχοι είναι η μελέτη του επιπέδου στοχοπροσήλωσης και συνεργασίας των παιδιών με τα ρομπότ [Filippini et al.; 2020, Komatsubara et al.; 2019], η ανάλυση των παραμέτρων που επηρεάζουν μια μακροχρόνια αλληλεπίδραση [Davison et al.; 2020], και η μελέτη για την κατάλληλη προσαρμογή σε πραγματικό χρόνο του περιεχομένου του μαθήματος και της δυσκολίας της διάδρασης [Senft et al.; 2018].

Αν και σημειώνεται μια συνεχιζόμενη αύξηση στη χρήση των κοινωνικών ρομπότ σε εκπαιδευτικές - ψυχαγωγικές διαδικασίες εδώ και αρκετά χρόνια, υπάρχει μια στασιμότητα ως προς την ευρύτερη ένταξη της ρομποτικής αλληλεπίδρασης σε ένα ελεύθερο πλαίσιο μέσα στη σχολική τάξη. Μια εργασία που κινείται εν μέρει στην κατεύθυνση του συγκεκριμένου project ως προς αυτό το χαρακτηριστικό, είναι η δουλειά των Shiomi et al. [Shiomi et al.; 2015]. Οι συγγραφείς τοποθέτησαν ένα κοινωνικό ρομπότ, το Robovie, στην αίθουσα των φυσικών επιστημών σε ένα δημοτικό σχολείο με σκοπό να επιτρέψουν την ελεύθερη αλληλεπίδραση των παιδιών κατά τη διάρκεια των διαλειμμάτων. Τα παιδιά είχαν τη δυνατότητα να κάνουν ελεύθερα ερωτήσεις στο ρομπότ σχετικά με τις επιστήμες ενώ συγχρόνως το ρομπότ τα ενθάρρυνε προς την κατεύθυνση αυτή. Το Robovie ελεγχόταν από μακριά και δεν ήταν εφοδιασμένο με έξυπνες τεχνολογίες που θα μπορούσαν να κάνουν αυτόματη την αλληλεπίδραση παιδιού-ρομπότ όμως η εργασία των Shiomi et al. αποτελεί μια από τις λίγες εργασίες ελεύθερης διάδρασης που στοχεύει στην ένταξη των ρομποτικών πρακτόρων στην εκπαιδευτική διαδικασία χωρίς να τροποποιεί το υφιστάμενο πλαίσιο στο οποίο διεξάγεται αλλά να το ενισχύει.

Τέλος, μια ακόμα πλευρά που πρέπει να εξεταστεί είναι το κομμάτι της ηθικής δεοντολογίας

ως προς τη χρήση των ρομπότ, όχι μόνο στην εκπαιδευτική διαδικασία αλλά και γενικότερα στην αλληλεπίδραση με τα παιδιά. Στην εργασία της Sharkey [Sharkey; 2016] παρουσιάζεται μια εκτενής έρευνα και αξιολόγηση σχετικά με τη χρήση των ρομπότ στις σχολικές αίθουσες ανάλογα με το ρόλο τους στην εκπαιδευτική διαδικασία. Οι κατηγορίες-σενάρια που διακρίνονται είναι οι εξής: α) το ρομπότ έχει το ρόλο του δασκάλου για όλη την τάξη, β) το ρομπότ λειτουργεί ως συνεργάτης-σύμβουλος του παιδιού, γ) το ρομπότ έχει το ρόλο του μαθητή και το παιδί έχει το ρόλο του δασκάλου, δ) το ρομπότ χρησιμοποιείται για τηλεκπαίδευση. Η Sharkey καταλήγει πως τα σενάρια (α) και (δ) που αντικαθιστούν τον καθηγητή δίνουν πολύ περιορισμένες δυνατότητες ως προς τις συνολικές ικανότητες που πρέπει να έχει ο εκπαιδευτικός, κρίση για τις μαθησιακές ανάγκες των παιδιών, δυνατότητα να ελέγξουν και να καθοδηγήσουν τους μαθητές, και για το λόγο αυτό αποθαρρύνεται η χρήση τους. Αντίθετα, τα σενάρια (β) και (γ) που λειτουργούν ενισχυτικά με το ρόλο του δασκάλου στην τάξη, ενθαρρύνουν και ενισχύουν τη μάθηση ενώ λαμβάνοντας τις κατάλληλες προφυλάξεις για τη διασφάλιση της ιδιωτικότητας των παιδιών δίνουν νέες εκπαιδευτικές δυνατότητες. Ένα ακόμα σημείο που υπογραμμίζεται πως πρέπει να δοθεί προσοχή είναι αυτό της «κοινωνικής σχέσης» που μπορεί να θεωρήσει λανθασμένα το παιδί ότι δημιουργεί με το ρομπότ. Παρόμοια, οι Di Dio et al. [Di Dio et al.; 2020] πραγματεύονται το θέμα της δημιουργίας σχέσεων εμπιστοσύνης μεταξύ παιδιού και ρομπότ, συγκριτικά με τη σχέση μεταξύ παιδιού και ενήλικα κατά τη διάρκεια κάποιων αλληλεπιδράσεων. Συνοπτικά οι συγγραφείς επισημαίνουν πως λειτουργούν οι ίδιοι ψυχολογικοί μηχανισμοί στα παιδιά και στις δυο περιπτώσεις ανάπτυξης των μεταξύ τους σχέσεων. Επιπρόσθετα παρατηρείται πως η σχέση εμπιστοσύνη προς το ρομποτικό πράκτορα είναι πιο ισχυρή στην ηλικία των εφτά ετών απ' ότι στην προσχολική ηλικία.

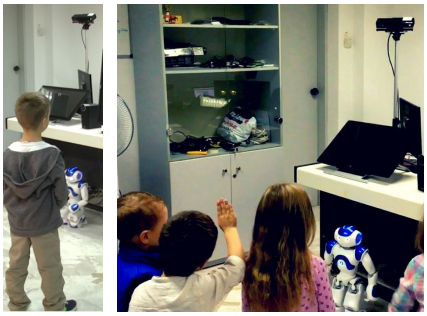
Οι περισσότεροι ρομποτικοί πράκτορες που χρησιμοποιούνται σε τέτοιες μελέτες είναι ημιαυτόνομοι ή τηλεχειριζόμενοι, εφαρμόζοντας την τεχνική Wizard-of-Oz [Belpaeme; 2020]. Πιο πρόσφατα, όμως, λόγω της ραγδαίας προόδου στις τεχνικές μηχανικής μάθησης και στα νευρωνικά δίκτυα, όλο και περισσότερα κοινωνικά ρομπότ έχουν ενσωματώσει έξυπνα συστήματα αντίληψης που μπορούν να αλληλεπιδράσουν με τους ανθρώπους με πιο φυσικό τρόπο [Chalvatzaki et al.; 2020]. Έτσι, η χρήση τους από μη ειδικούς, π.χ. εκπαιδευτικούς ή θεραπευτές, αναμένεται να αυξηθεί και τομείς όπως η εκπαίδευση να επωφεληθούν από αυτή την πρόοδο.

Συστήματα αλληλεπίδρασης παιδιών-ρομπότ που αναπτύχθηκαν στο πλαίσιο της διδακτορικής διατριβής

Κατά τη διάρκεια αυτής της διδακτορικής διατριβής αναπτύξαμε δύο ολοκληρωμένα συστήματα αλληλεπίδρασης παιδιών και ρομπότ. Το πρώτο σύστημα, το οποίο αναφέρουμε ως **ChildBot**, αποτελεί ένα ολοκληρωμένο ρομποτικό σύστημα ικανό να συμμετέχει και να εκτελεί ένα ευρύ φάσμα εκπαιδευτικών και ψυχαγωγικών εργασιών σε συνεργασία με ένα ή περισσότερα παιδιά. Το σύστημα αυτό διαθέτει πολυτροπικές μονάδες αντίληψης (multimodal perception units) και πολλαπλούς ρομποτικούς πράκτορες (multible robotic agents) που παρακολουθούν το περιβάλλον αλληλεπίδρασης και μπορούν να συντονίσουν περίπλοκα σενάρια αλληλεπίδρασης. Προκειμένου να αξιολογήσουμε την αποτελεσματικότητα του συστήματος και των ενσωματωμένων μονάδων του, πραγματοποιήσαμε πολλαπλά πειράματα με συνολικά 52 παιδιά, ενώ επιπρόσθετα, ζητήσαμε τη γνώμη τους σχετικά με την εμπειρία τους κατά τη διάρκεια των σχεδιασμένων αλληλεπιδράσεων.

Το σύστημα ChildBot αναπτύχθηκε σε συνεργασία με άλλα μέλη του εργαστηρίου CVSP κατά τη διάρκεια του EU project BabyRobot[1] και παρουσιάστηκε στην ολότητά του στην εργασία [Efthymiou et al.; 2022]. Αποτελεί μια ολοκληρωμένη επέκταση ενός συνόλου προκαταρκτικών ερευνών και αντίστοιχων δημοσιεύσεων σχετικά με επιμέρους προβλήματα αντίληψης

[1]Περισσότερες πληροφορίες: http://babyrobot.eu/



Σχήμα 2.1: Παραδείγματα αλληλεπίδρασης παιδιών και ρομπότ με χρήση του συστήματος TeachBot που αναπτύξαμε.

και αλληλεπίδρασης κατά τη χρήση πολλαπλών ρομπότ [Efthymiou et al.; 2018, Hadfield et al.; 2018, Tsiami et al.; 2018a, Tsiami et al.; 2018b]. Αντίστοιχα και το TeachBot ήταν αποτέλεσμα συνεργασίας ενώ οι κύριες δημοσιεύσεις που προέκυψαν από αυτό είναι οι [Efthymiou et al.; 2021a, Efthymiou et al.; 2021b]. Η ατομική μου ερευνητική συνεισφορά εστιάζει κύρια στην ανάπτυξη των συστημάτων αναγνώρισης δράσης, όπως παρουσιάζονται αναλυτικά στο Κεφάλαιο 3, αλλά και στον σχεδιασμό και την υλοποίηση των πειραμάτων.

Το δεύτερο σύστημα, το οποίο αναφέρουμε ως **TeachBot**, είναι ένα ελαφρύ σύστημα αλληλεπίδρασης παιδιού-ρομπότ για εκπαιδευτικά σενάρια (Σχήμα 2.1) και αναπτύχθηκε κατά τη διάρκεια του έργου «Ευφυές σύστημα αλληλεπίδρασης παιδιού-ρομπότ για τον σχεδιασμό και την εκτέλεση εκπαιδευτικών-ψυχαγωγικών σεναρίων με έμφαση στην οπτική πληροφορία»[2]. Στόχος του είναι να παρέχει ένα εύχρηστο σύστημα για τον σχεδιασμό και την υλοποίηση τέτοιων σεναρίων και να προσφέρει απλούστερη πρόσβαση στη χρήση των κοινωνικών ρομπότ από εκπαιδευτικούς με μη εξειδικευμένες τεχνικές γνώσεις σε αυτό το δύσκολο τομέα. Το σύστημά μας ενσωματώνει ισχυρές ρομποτικές μονάδες αντίληψης για την αναγνώριση δράσης και συναισθημάτων του παιδιού, επιτρέποντας στο ρομπότ να εκδηλώνει ενσυναίσθηση και να αντιλαμβάνεται τη δραστηριότητα του παιδιού. Επιπρόσθετα η μονάδα αναγνώρισης συνδυάζεται και με τεχνικές επαυξημένης μάθησης με σκοπό να δώσει τη δυνατότητα στον καθηγητή που φιλοδοξεί να αξιοποιήσει τα ρομπότ σε νέες δράσεις, να προσαρμόσει εύκολα το σύστημα στις νέες ανάγκες του μαθήματος. Το σύστημα αξιολογήθηκε ως προς την αποτελεσματικότητα των συστημάτων αναγνώρισης και των υπολογιστικών του απαιτήσεων σε δεδομένα παιδιών.

Τα παραπάνω συστήματα αναπτύχθηκαν σε συνεργασία με άλλα μέλη του εργαστηρίου και είχαν σαν αποτέλεσμα μια σειρά δημοσιεύσεων. όπως

2.2 Επισκόπηση Αντιληπτικών Συστημάτων στο Πλαίσιο της Αλληλεπίδραση Ανθρώπων και Ρομπότ

Μια γενική ανασκόπηση των μεθόδων αντίληψης που χρησιμοποιούνται για την αλληλεπίδραση ανθρώπου-ρομπότ σε κοινωνικά ρομπότ μέχρι το 2014 παρουσιάζεται στο [Yan et al.; 2014]. Στην εργασία αυτή επισημαίνονται τρία σημαντικά ζητήματα που σχετίζονται με τα συστήματα αντίληψης: α) η ανάγκη για ανάπτυξη συστημάτων αντίληψης σε πραγματικά περιβάλλοντα με πραγματικά δεδομένα, β) η απαίτηση δημιουργίας αποδοτικών αναπαραστάσεων σύμφωνων με το πλαίσιο της εκάστοτε αλληλεπίδρασης, και γ) η απαίτηση να συνδυάζεται

[2]Η έρευνα συγχρηματοδοτήθηκε από την Ελλάδα και την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο-ΕΚΤ) μέσω του Επιχειρησιακού Προγράμματος «Ανάπτυξη Ανθρώπινου Δυναμικού, Εκπαίδευση και Δια Βίου Μάθηση 2014-2020» στο πλαίσιο της Πράξης «Ευφυές σύστημα αλληλεπίδρασης παιδιού-ρομπότ για τον σχεδιασμό και την εκτέλεση εκπαιδευτικών-ψυχαγωγικών σεναρίων με έμφαση στην οπτική πληροφορία (MIS 5049533)»

αποτελεσματικά ένα σύστημα αντίληψης με κατάλληλες αποκρίσεις των ρομπότ προκειμένου να δημιουργήσουν ευχάριστες αλληλεπιδράσεις για τους ανθρώπους. Μια πρόσφατη ανασκόπηση σχετικά με τον τρόπο με τον οποίο τα κοινωνικά ρομπότ αντιλαμβάνονται τους ανθρώπους αλλά και τα «κοινωνικά» χαρακτηριστικά που εκδηλώνουν κατά τη διάρκεια των αλληλεπιδράσεών τους παρουσιάζεται από τους Tapus et al. [Tapus et al.; 2019].

Αναφερόμενοι σε πρωτότυπες εργασίες που υλοποιούν ένα πολύπλευρο σύστημα αντίληψης, οι Zaraki et al. [Zaraki et al.; 2017] προσπάθησαν να αναπτύξουν συστήματα αντίληψης για HRI συνδυάζοντας χαμηλού και υψηλού επιπέδου χαρακτηριστικά για να ανιχνεύσουν μια σειρά από χαρακτηριστικά-συμπεριφορές των ανθρώπων που εμφανίζονται κατά τη διάρκεια ενός πραγματικού σεναρίου. Οι Valipour et al. [Valipour et al.; 2017] πρότειναν ένα νέο παράδειγμα για τη σταδιακή βελτίωση της οπτικής αντίληψης ενός ρομπότ κατά τη διάρκεια μιας εμπειρίας HRI. Σημαντικά έργα που έχουν επίσης επικεντρωθεί στην αντίληψη των ρομπότ σε εφαρμογές CRI περιλαμβάνουν το έργο ALIZ-E [Belpaeme et al.; 2012]. Οι Belpaeme et al. έκαναν μια μακροχρόνια μελέτη της αλληλεπίδρασης των παιδιών με κοινωνικά ρομπότ, τόνισαν τις δυσκολίες που προκύπτουν όταν τα πειράματα αφορούν τον πραγματικό κόσμο, δηλαδή γίνονται σε ένα σχολείο και ένα νοσοκομείο, και ανέπτυξαν ένα πλήρες πλαίσιο για πολυτροπική CRI. Μια άλλη ενδιαφέρουσα εργασία που αξιολογήθηκε σε νοσοκομειακό περιβάλλον είναι η πλατφόρμα NAOTherapist [Pulido et al.; 2017], η οποία επικεντρώθηκε σε συνεδρίες αποκατάστασης των άνω άκρων παιδιών με σωματικές αναπηρίες. Το ρομπότ NAO εκτελούσε αυτόνομα συνεδρίες φυσιοθεραπείας, παρατηρώντας τη στάση του ασθενούς μέσω ενός αισθητήρα Kinect δίνοντας κατάλληλες διορθωτικές οδηγίες εάν χρειαζόταν. Μια παρόμοια πλατφόρμα για τη δημιουργία αυτόνομης αλληλεπίδρασης μεταξύ ενός ρομπότ και παιδιών με αυτισμό κατασκευάστηκε από το έργο INSIDE [Melo et al.; 2019]. Αναπτύχθηκε ένα πολυτροπικό σύστημα αντίληψης για τον προσδιορισμό της θέσης και της δραστηριότητας του παιδιού, την αναγνώρισης της ικανοποίησής του όταν το ρομπότ απαντούσε, και την κατάσταση της τρέχουσας δραστηριότητας.

Άλλα παρόμοια ερευνητικά έργα περιλαμβάνουν το L2TOR [Belpaeme et al.; 2015], όπου ένα ρομπότ NAO αναλαμβάνει το ρόλο ενός δασκάλου δεύτερης γλώσσας που έχει τη δυνατότητα πολυτροπικής αντίληψης, και το έργο EASEL, όπου οι Vouloutsi et al. [Vouloutsi et al.; 2016] παρουσίασαν μια κατανεμημένη και προσαρμόσιμη αρχιτεκτονική ελέγχου του ρομποτικού συστήματος και των τεσσάρων επιπέδων του: somatic, reactive, adaptive, and contextual. Οι Esteban et al. [Esteban et al.; 2017] κατασκεύασαν ένα σύστημα πολλαπλών αισθητήρων για αυτόνομη αλληλεπίδραση ενός ρομπότ NAO με παιδιά που βρίσκονται στο φάσμα του αυτισμού, με σκοπό να αντιλαμβάνονται διάφορα χαρακτηριστικά κατά την αλληλεπίδραση, όπως εκτίμηση βλέμματος, αναγνώριση ενεργειών και παρακολούθηση αντικειμένων. Οι δυνατότητες του συστήματος ήταν επαρκείς για τις εφαρμογές που παρουσιάστηκαν στην εργασία, αλλά περιορισμένες για μια πιο γενική αλληλεπίδραση ενώ το σύστημα δεν διέθετε αξιολόγηση σε πραγματικά σενάριο. Στο [Marinoiu et al.; 2018], οι Marinoiu et al. εισήγαγαν ένα σύστημα αναγνώρισης ενεργειών και συναισθημάτων αξιοποιώντας μεθόδους δισδιάστατης και τρισδιάστατης εκτίμησης πόζας και αξιολόγησαν το συστημά τους σε ένα σύνολο δεδομένων μεγάλης κλίμακας από συνεδρίες θεραπείας παιδιών με αυτισμό με τη βοήθεια ενός ρομπότ. Αξίζει επίσης να αναφέρουμε το έργο ANIMATAS που εστιάζει στην εκπαίδευση ερευνητών για την προώθηση της αλληλεπίδρασης ανθρώπου-μηχανής. Συγκεκριμένα, στο [Wallkötter et al.; 2020], οι Valipour et al. υπογράμμισαν τις διαφορές στην αντίληψη του κοινωνικού ρομπότ κατά τη διάρκεια πειραμάτων εικονικού και πραγματικού κόσμου και τη σημασία της διεξαγωγής πραγματικών πειραμάτων για την αποκάλυψη όλων των παραμέτρων αλληλεπίδρασης.

Εστιάζοντας στις τεχνολογίες αντίληψης για παιδιά, οι Kennedy et al. [Kennedy et al.; 2017], αφού αξιολόγησαν πολλά συστήματα αυτόματης αναγνώρισης ομιλίας, κατέληξαν στο συμπέρασμα ότι η παιδική αναγνώριση ομιλίας απαιτεί μια πολύπλευρη προσέγγιση για να είναι

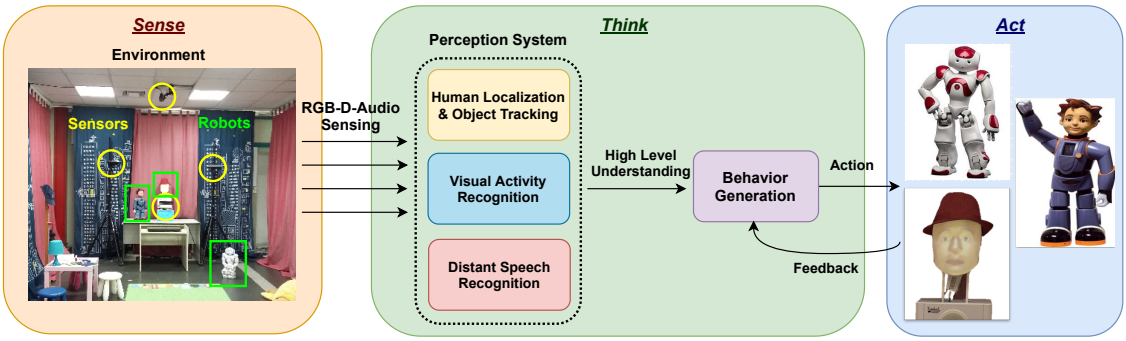
αποτελεσματική και να επιτύχει υψηλότερη απόδοση. Ένα ενδιαφέρον αποτέλεσμα σημειώθηκε από τους Yeung και Alwan [Yeung and Alwan; 2018], όπου διαπιστώθηκε ότι ακόμα και ένας χρόνος διαφοράς στην ηλικία των νηπίων έχει αντίκτυπο στην απόδοση της αυτόματης αναγνώρισης ομιλίας. Όσον αφορά την αναγνώριση δράσης, οι Chiang et al. [Chiang et al.; 2018] ταξινόμησαν οκτώ ανθρώπινες ενέργειες, που πραγματοποιήθηκαν από παιδιά και ενήλικες, χρησιμοποιώντας Histogram of Oriented Gradients (HOG) χαρακτηριστικά από συνδυασμένους χρωματικούς χάρτες βάθους και κίνησης. Επιπλέον, στο [Zhang et al.; 2021], οι Zhang et al. ανέπτυξαν μια μέθοδο που αναγνωρίζει στερεοτυπικές ενέργειες παιδιών με αυτισμό και χρησιμοποίησαν Long-Short Term Memory (LSTM) δίκτυα σε δεδομένα σκελετού. Τέλος, στο [Wu et al.; 2019] οι Wu et al. ενσωμάτωσαν μια μονάδα αναγνώρισης αντικειμένων σε ένα εκπαιδευτικό ρομποτικό σύστημα που παρείχε ενδιαφέρουσες και καινοτόμες υπηρεσίες εκμάθησης δεύτερης γλώσσας για παιδιά προσχολικής ηλικίας στην Κίνα.

2.3 ChildBot: Σύστημα Αντίληψης και Αλληλεπίδρασης Παιδιών με Πολλαπλά Ρομπότ

Μέχρι σήμερα, τα περισσότερα συστήματα κοινωνικών ρομπότ παρουσιάζουν δύο μεγάλες ελλείψεις: (α) *συνήθως ενσωματώνουν μόνο συγκεκριμένες μεθόδους*, αναγκάζοντας τους χρήστες τους να προσαρμοστούν στον τρόπο που το σύστημα αντιλαμβάνεται το περιβάλλον αντί για το αντίθετο, (β) *έχουν αναπτυχθεί και σχεδιαστεί για συγκεκριμένες εφαρμογές και εργασίες*. Ωστόσο, με την επέκταση της χρήσης των κοινωνικών ρομπότ σε διάφορους τομείς εφαρμογών προκύπτει η ανάγκη για ολοκληρωμένα συστήματα ικανά να αξιοποιηθούν σε πολλαπλές εφαρμογές και πραγματικά σενάρια σε απαιτητικά περιβάλλοντα, προσφέροντας φυσική αλληλεπίδραση στους χρήστες τους. Αυτή η αλληλεπίδραση περιλαμβάνει τη δημιουργία έξυπνων προσαρμοστικών ολοκληρωμένων ρομποτικών συστημάτων ικανών για πολλαπλές εργασίες με ένα ευρύ φάσμα αντιληπτικών ικανοτήτων και ενεργειών. Έτσι θα δίνεται η δυνατότητα στους χρήστες του εκάστοτε συστήματος HRI να σχεδιάζουν πολύμορφες διαδραστικές εφαρμογές που μπορούν να διατηρήσουν το ενδιαφέρον και την προσήλωση του χρήστη τους. Αυτό είναι ιδιαίτερα σημαντικό για τα παιδιά χρήστες, στο πλαίσιο της εκπαίδευτικής ψυχαγωγίας (edutainment). Επιπλέον, συστήματα που θα αντιλαμβάνονται πολλές τροπικότητες (modalities) - πολλά κανάλια πληροφορίας - θα επιτρέπουν στους χρήστες τους να δημιουργούν αλληλεπιδράσεις αξιοποιώντας τον τρόπο επικοινωνίας που προτιμούν.

Ένας από τους λόγους που εστιάσαμε στα παραπάνω είναι και το γεγονός ότι τα εμπορικά κοινωνικά ρομπότ έχουν διαφορετικές δυνατότητες. Για παράδειγμα, το ρομπότ NAO [NAO;] μπορεί να κάνει πολλές κινήσεις με το σώμα του, αλλά είναι αδύνατο να εκφράσει συναισθήματα, ο Furhat [Furhat Robotics;] παρουσιάζει μεγάλη ποικιλία εκφράσεων του προσώπου αλλά δεν κινείται, ενώ το ρομπότ Zeno [Robokind. Advanced Social Robots.;] είναι ικανό και να κινηθεί και να εκφράσει συναισθήματα, αλλά δεν είναι εξίσου καλό και στα δύο (η κίνηση του σώματος του Zeno υστερεί σε σύγκριση με το NAO). Επιπλέον, κάθε κοινωνικό ρομπότ έχει διαφορετικούς αισθητήρες, και έτσι περιορίζει τον χρήστη να αξιοποιήσει συγκεκριμένα κανάλια επικοινωνίας.

Με κίνητρο τα παραπάνω, αναπτύξαμε ένα ολοκληρωμένο ρομποτικό σύστημα που μπορεί να χρησιμοποιηθεί για πολλαπλές εφαρμογές ψυχαγωγίας. Για να επιτύχουμε αυτή την ευελιξία, *το ChildBot ενσωματώνει*: (α) *πολλαπλούς αισθητήρες και μονάδες αντίληψης* που επιτρέπουν στον χρήστη να επικοινωνεί με τα ρομπότ μέσω πολλαπλών καναλιών και (β) *πολλαπλά κοινωνικά ρομπότ*, αξιοποιώντας τα δυνατά σημεία του κάθε ρομπότ ώστε να παρακάμψει τις αδυναμίες του. Έτσι, με στόχο την αύξηση της απόδοσης, της ευελιξίας και της ευρωστίας, το ChildBot αποτελείται από πολλαπλά ρομπότ και πολυτροπικές μονάδες αντίληψης σχεδιασμέ-



Σχήμα 2.2: Επισκόπηση του συστήματος ChildBot κατά την αλληλεπίδραση παιδιού-ρομπότ. Ένα δίκτυο αισθητήρων περιβάλλει το περιβάλλον της αλληλεπίδρασης και λαμβάνει την πολυτροπική πληροφορία που προκύπτει κατά την αλληλεπίδραση. Το σύστημα αντίληψης την επεξεργάζεται και εξάγει πληροφορίες υψηλού επιπέδου σχετικά με το πλαίσιο δράσης. Με βάση αυτό, η μονάδα παραγωγής συμπεριφοράς (Behavior Generation) αποφασίζει και ελέγχει τους ρομποτικούς πράκτορες.

νες και προσαρμοσμένες σε παιδιά, επιτρέποντας την αλληλεπίδραση σε έναν σχετικά μεγάλο χώρο για μια ποικιλία ψυχαγωγικών σεναρίων.

Μια επισκόπηση του συστήματος που αναπτύξαμε φαίνεται στο Σχήμα 2.2. Το ChildBot βασίστηκε σε μια αρχιτεκτονική τριών επιπέδων *Sense - Think - Act* [Gat et al.; 1998] και σχεδιάστηκε για χρήση σε έξυπνους χώρους, δηλαδή σε εσωτερικούς χώρους εφοδιασμένους με ένα δίκτυο αισθητήρων. Η χρήση εξωτερικών - ως προς τα ρομπότ - αισθητήρων βοηθά στο να αποφευχθούν προβλήματα που συναντώνται συχνά σε HRI εφαρμογές, όπως οι επικαλύψεις αντικειμένων (occlusions), και επιτρέπει τη συγχώνευση διαφορετικών ροών δεδομένων, βελτιώνοντας την ευρωστία και την απόδοση των μονάδων αντίληψης. Με αυτόν τον τρόπο, επιτυγχάνουμε επίσης την αυτόματη αντίληψη της αλληλεπίδρασης αυτόνομα και χωρίς εξάρτηση από κάποιο ρομπότ παρακάμπτοντας έτσι τους περιορισμούς ανίχνευσης που επιβάλλει η χρήση μεμονωμένων ρομποτικών συστημάτων. Επιπρόσθετα, αυτή η robot-agnostic αρχιτεκτονική, που είναι ανεξάρτητη από την επιλογή ρομπότ, διευκολύνει την εισαγωγή και χρήση νέων ρομπότ από το σύστημα. Το σύστημα συντονίζει μια σύνθετη και συνεχή διαδικασία HRI που περιλαμβάνει τις ενέργειες των παιδιών και τις αποκρίσεις των ρομπότ και αντίστροφα. Πιο συγκεκριμένα, οι αισθητήρες λαμβάνουν την πολυτροπική πληροφορία (Sense) κατά τη διάρκεια των ενεργειών των παιδιών. Στη συνέχεια το σύστημα την επεξεργάζεται και εξάγει πληροφορίες υψηλού επιπέδου σχετικά με το περιεχόμενο των ενεργειών έτσι ώστε να αποφασίσει την κατάλληλη απάντηση/δράση (Think). Τέλος, το σύστημα μεταδίδει το μήνυμα σε έναν ρομποτικό πράκτορα (Act) ο οποίος ενεργεί αναλόγως.

Η συνεισφορά του ChildBot συστήματος έγκειται τόσο στην ποικιλία των επιμέρους μονάδων αντίληψης όσο και στην ενοποίηση τους σε ένα συνολικό σύστημα. Η ενσωμάτωση των μονάδων αντίληψης είναι καίριας σημασίας για την αντιμετώπιση περίπλοκων σεναρίων αλληλεπίδρασης παιδιών-ρομπότ που απαιτούν μέσω πολλών τροπικοτήτων - καναλιών επικοινωνίας. Η επιλογή μας ώστε το σύστημα να έχει αρθρωτή δομή επιτρέπει στο χρήστη να αξιοποιήσει διαφορετικές μονάδες αντίληψης κάθε φορά, σύμφωνα με την αλληλεπίδραση που σχεδιάζει, χωρίς να επηρεάζεται η λειτουργικότητα ολόκληρου του συστήματος. Επιπλέον, η ενσωμάτωση πολλαπλών ρομπότ επιτρέπει την εναλλαγή μεταξύ τους σύμφωνα με τις επιθυμίες του χρήστη, καθώς και την προσθήκη νέων.

Για να δείξουμε την ευελιξία και τις δυνατότητες του ολοκληρωμένου συστήματος και των μεμονωμένων μονάδων αντίληψης του, έχουμε σχεδιάσει πέντε διαφορετικά ψυχαγωγικά σε-

νάρια. Τα σενάρια αυτά είναι ενδεικτικά και έχουν σχεδιαστεί για να εκμεταλλεύονται διαφορετικά στοιχεία του συστήματος, και να αναδεικνύουν τη μεγάλη ποικιλία εφαρμογών που μπορούν να υλοποιηθούν με το ChildBot. Συλλέξαμε δεδομένα από 52 παιδιά κατά την εκτέλεση πειραμάτων - αλληλεπιδράσεων με τα ρομπότ με σκοπό να αξιολογήσουμε αντικειμενικά την απόδοση κάθε μονάδας αντίληψης του συστήματος αλλά και της συνολικής διάδρασης. Επιπλέον, ως μια πρώτη, αδρή υποκειμενική αξιολόγηση της εμπειρίας των μικρών μας χρηστών, ζητήσαμε την άποψη των παιδιών για τις αλληλεπιδράσεις τους με τα ρομπότ.

Συνοψίζοντας, επισημαίνουμε τις **πιο σημαντικές συνεισφορές** του συστήματος *ChildBot*:

- **είναι ένα ολοκληρωμένο σύστημα για HRI** που έχει σχεδιαστεί και υλοποιηθεί με τη αξιοποίηση πολλαπλών ρομποτικών πρακτόρων. Η αρθρωτή του αρχιτεκτονική τριών επιπέδων ενσωματώνει πολλαπλούς αισθητήρες, πολυάριθμες μονάδες αντίληψης και διαφορετικούς ρομποτικούς πράκτορες, καταλήγοντας σε ένα αυτόνομο σύστημα HRI πολλαπλών εφαρμογών.

- **διαθέτει μονάδες αντίληψης για πολυτροπική κατανόηση σκηνής** που έχουν αναπτυχθεί και προσαρμοστεί σε συγκεκριμένες συνθήκες CRI ενσωματώνοντας νέες προσεγγίσεις και εκτενείς μελέτες. Ο οπτικοακουστικός εντοπισμός των ομιλητών, η παρακολούθηση αντικειμένων με 6 βαθμούς ελευθερίας (6-DoF), η οπτική αναγνώριση δράσεων και η αναγνώριση απομακρυσμένης ομιλίας είναι απαραίτητα για την ανάλυση και την παρακολούθηση της ανθρώπινης συμπεριφοράς κατά την εξέλιξη μιας αλληλεπίδρασης. Οι μονάδες αντίληψης αυτού του συστήματος έχουν αναπτυχθεί σύμφωνα με τις state-of-the-art τεχνολογίες ενώ πολλές φορές υπερτερούν αυτών, όπως φαίνεται από την αξιολόγηση των επιμέρους μονάδων.

- **αξιολογείται σε δεδομένα παιδιών που αλληλεπιδρούν αυθόρμητα με τα ρομπότ** και έχουν συλλεχθεί για τον συγκεκριμένο σκοπό. Ορίσαμε και υλοποιήσαμε ενδεικτικά σενάρια χρήσης προκειμένου να αναδείξουμε τη μεγάλη γκάμα εφαρμογών στις οποίες μπορεί να χρησιμοποιηθεί το ChildBot. Τα δεδομένα που συλλέξαμε επέτρεψαν μια εκτενή αντικειμενική αξιολόγηση των δυνατοτήτων του κατά τη διάρκεια πραγματικών σεναρίων, καθώς και μια προκαταρκτική μελέτη εμπειρίας των μικρών χρηστών.

2.3.1 Περιγραφή Συστήματος

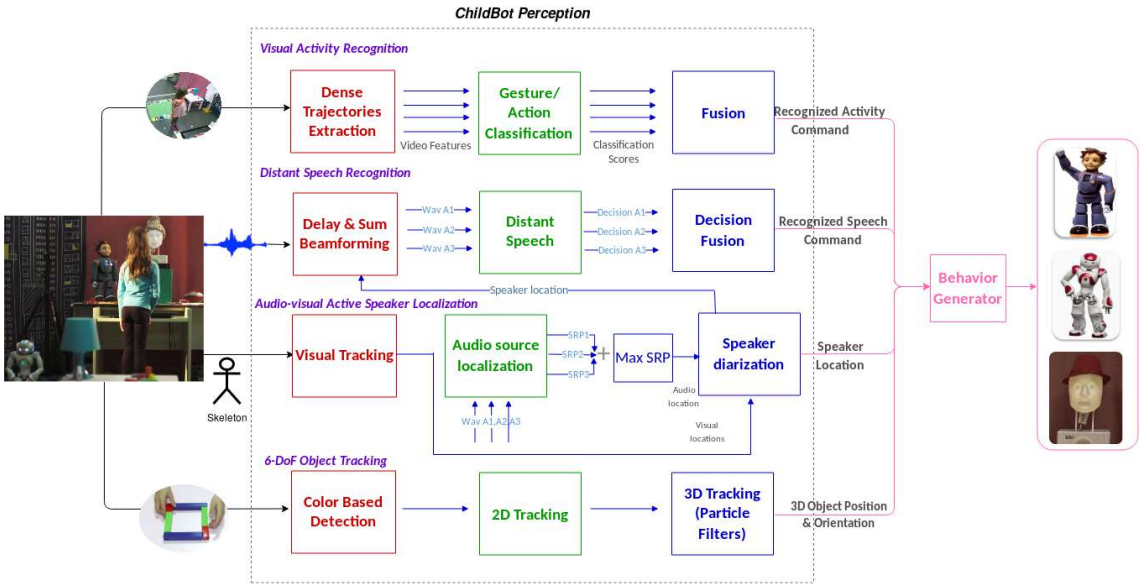
Σε αυτή την ενότητα παρουσιάζουμε το σύστημα αντίληψης που παρέχει ολοκληρωμένη και αποτελεσματική εποπτεία των αλληλεπιδράσεων καθώς και την αρχιτεκτονική του συνολικού συστήματος ChildBot . Αρχικά παρουσιάζεται μια επισκόπηση του συστήματος αντίληψης και των μονάδων του. Στη συνέχεια, περιγράφονται λεπτομερώς η αρχιτεκτονική του συστήματος, οι ενδοεπικοινωνίες και η μονάδα διαχείρισης διαλόγου. Τέλος, στο Κεφάλαιο 3 παρουσιάζονται εκτενώς οι τεχνολογίες που αναπτύξαμε για τη μονάδα αναγνώρισης δράσεων και τα πειραματικά αποτελέσματα.

Μονάδες Αντίληψης

Μια επισκόπηση του συστήματος αντίληψης φαίνεται στο Σχήμα 2.3, όπου παρουσιάζονται οι τρεις κύριες ενότητες:

- Οπτικοακουστικός εντοπισμός του ομιλητή. Εντοπισμός και ιχνηλασία αντικειμένων με 6 βαθμούς ελευθερίας (Audio-Visual Active Speaker Localization and 6-DoF Object Tracking)

- Οπτική αναγνώριση δράσης (Visual Activity Recognition)



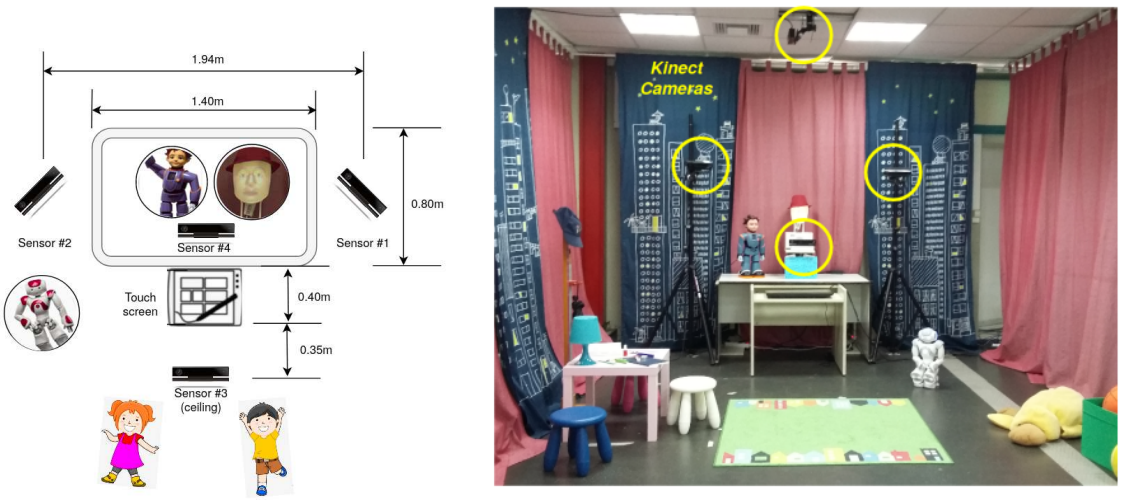
Σχήμα 2.3: Επισκόπηση των μονάδων αντίληψης ChildBot, συμπεριλαμβανομένων των *Audio-Visual Active Speaker Localization και 6-DoF Object Tracking*, *Visual Activity Recognition* και *Distant Speech Recognition*. Το "A" αναφέρεται στη συστοιχία μικροφώνου και το "SRP" στην ισχύ της κατευθυνόμενης απόκρισης. Οι μονάδες χρησιμοποιούνται κατά τη διάρκεια της αλληλεπίδρασης για την παρακολούθηση των πολλαπλών πτυχών της ανθρώπινης συμπεριφοράς και στη συνέχεια οι έξοδοί τους τροφοδοτούν τη μονάδα παραγωγής συμπεριφοράς (Behavior Generator).

- Απομακρυσμένη αναγνώριση ομιλίας (Distant Speech Recognition)

Τέσσερις αισθητήρες Kinect V2 καταγράφουν λεπτομερώς το χώρο και τις δράσεις που λαμβάνουν χώρα σε αυτόν και τροφοδοτούν το σύστημα αντίληψης με την αντίστοιχη ακατέργαστη πληροφορία. Οι αισθητήρες Kinect V2 τοποθετούνται σε διαφορετικές θέσεις και γωνίες θέασης για να καλύπτουν επαρκώς ολόκληρο το περιβάλλον, ώστε να αντιμετωπίζονται προβλήματα επικαλύψεων μεταξύ των αντικειμένων ή/και των ανθρώπων και να προσφέρουν πολλαπλές οπτικές γωνίες. Κάθε αισθητήρας καταγράφει εικόνα RGB, βάθος (depth) και ήχο τεσσάρων καναλιών μέσω της συστοιχίας μικροφώνων κάθε Kinect. Η χωρική διάταξη των αισθητήρων παρουσιάζεται στο Σχήμα 2.4. Στη συνέχεια, παρουσιάζουμε μια επισκόπηση κάθε ενότητας αντίληψης:

Audio-Visual Active Speaker Localization and 6-DoF Object Tracking: Για να επιτραπεί η φυσική αλληλεπίδραση μεταξύ ρομπότ και ανθρώπων, είναι απαραίτητο το ρομπότ να γνωρίζει τη θέσει του ομιλητή ώστε να μπορεί να στραφεί προς τα εκεί, καθώς και να εντοπίζει και να παρακολουθεί σημαντικά αντικείμενα για τη διάδραση.

Αναπτύξαμε μια αποτελεσματική μέθοδο για τον εντοπισμό του ομιλητή σε σκηνές HRI με την αξιοποίηση τόσο της ακουστικής όσο και της οπτικής πληροφορίας. Αρχικά, αξιοποιούνται οι τρισδιάστατοι σκελετοί που παρέχονται από το Kinect V2 για τον εντοπισμό των ατόμων που βρίσκονται στο χώρο. Παράλληλα, γίνεται εντοπισμός της ηχητικής πηγής με χρήση του αλγορίθμου Steered-Response Power Phase Transform (SRP-PHAT) ο οποίος δίνει μια εκτίμηση για το που είναι πιο πιθανό να βρίσκεται η πηγή. Τέλος, ο τελικός εντοπισμός του ομιλητή πραγματοποιείται επιλέγοντας τη θέση του ατόμου που βρίσκεται πλησιέστερα στη θέση της



Σχήμα 2.4: Διάταξη των αισθητήρων στο χώρο διάδρασης.

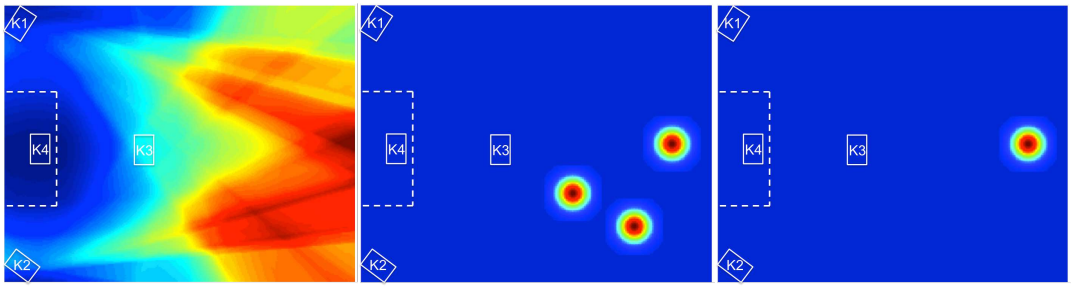
πηγής του ήχου. Ένα παράδειγμα για την υλοποίηση της συγκεκριμένης μεθόδου δίνεται στο Σχήμα 2.5.

Επιπλέον, στο σύστημα έχει ενσωματωθεί μια μονάδα για την αναγνώριση αντικειμένων που μπορεί να ανιχνεύσει πολλά αντικείμενα-παιχνίδια με βάση τα χρώματα και το μέγεθός τους και να παρακολουθήσει την τρισδιάστατη θέση και τον προσανατολισμό του. Η συγκεκριμένη μονάδα αξιοποιεί την πληροφορία του βίντεο τόσο από τα τρία κανάλια χρώματος, όσο και από το κανάλι του βάθους. Η μονάδα 6-DoF Object Tracking αποτελείται από δύο στάδια: το πρώτο εντοπίζει τη θέση του αντικειμένου με τη βοήθεια του χρώματός του, ενώ το δεύτερο χρησιμοποιεί ένα particle φίλτρο που εφαρμόζεται στα δεδομένα βάθους ώστε να τελειοποιήσει το αποτέλεσμα της εξόδου του πρώτου σταδίου και να συμπεράνει την περιστροφή του αντικειμένου.

Visual Activity Recognition: Έχει αναπτυχθεί μια μονάδα για να αναγνωρίζει χειρονομίες που συνοδεύουν την καθημερινή επικοινωνία του ανθρώπου αλλά και γενικότερες κινήσεις του σώματος που αποδίδουν συγκεκριμένα νοήματα. Η μονάδα αναγνώρισης δράσεων αξιοποιεί τις πολλαπλές όψεις που μπορεί να μας δώσει το δίκτυο αισθητήρων και αναγνωρίζει με επιτυχία τη δραστηριότητα του παιδιού ενώ περιφέρεται στο δωμάτιο και αλληλεπιδρά με τα ρομπότ και τα αντικείμενα. Η μονάδα αυτή έχει δύο ενότητες, η μια στοχεύει στην αναγνώριση χειρονομιών που μεταδίδουν ένα εννοιολογικό μήνυμα κατά τη διάρκεια της αλληλεπίδρασης, όπως το να χαιρετά το παιδί ή να κάνει ένα καταφατικό νεύμα. Από την άλλη, η ενότητα αναγνώρισης δράσης στοχεύει σε γενικότερες κινήσεις του σώματος του παιδιού που σχηματίζουν πολύπλοκα νοήματα, όπως οι κινήσεις παντομίμας.

Distant Speech Recognition: Αναπτύξαμε ένα σύστημα απομακρυσμένης αναγνώρισης ομιλίας (DSR) στα ελληνικά για να αλληλεπιδρά το παιδί μέσω της ομιλίας. Καθώς το να φορούν τα παιδιά πάνω τους μικρόφωνα (close-talk microphones) δεν είναι βολικό για αυτά περιορίζοντας τις κινήσεις τους, εκμεταλλευόμαστε τις πολλαπλές συστοιχίες μικροφώνων που βρίσκονται γύρω από το δωμάτιο για εγγραφή του ήχου, και έτσι τα παιδιά μπορούν να κινούνται ελεύθερα και να επικοινωνούν με τα ρομπότ χωρίς περιορισμούς. Για να έχουμε μεγαλύτερη επιτυχία στην αναγνώριση της ομιλίας και να εκμεταλλευτούμε τις κατανεμημένες συστοιχίες μικροφώνων, πειραματιστήκαμε με προσαρμοστικές τεχνικές (adaptation) και τεχνικές σύμμειξης της πληροφορίας (fusion).

Η υψηλού επιπέδου κατανόηση που επιτυγχάνεται από τις μονάδες αντίληψης τροφοδοτείται στο **Dialog Manager**, μαζί με επιπλέον είσοδο από μια οθόνη αφής, η οποία χρησιμο-



(α΄) (β΄) (γ΄)

Σχήμα 2.5: Παράδειγμα οπτικοακουστικού εντοπισμού του ομιλητή. Με κόκκινο αναπαριστώνται οι θέσεις με τη μεγαλύτερη πιθανότητα να βρίσκεται ο ομιλητής. α) Εντοπισμός ηχητικής πηγής, β) Οπτικός εντοπισμός συμμετεχόντων, γ) Οπτικοακουστικός εντοπισμός του ομιλητή.

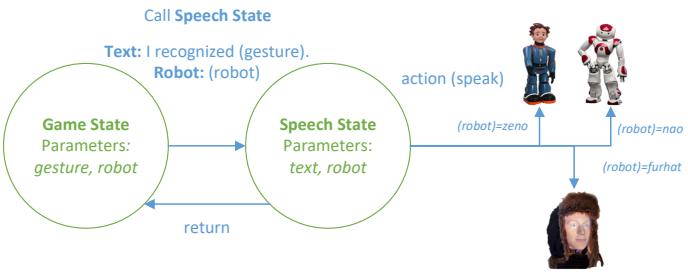
ποιείται ως πρόσθετο μέσο επικοινωνίας κατά τη διάρκεια της αλληλεπίδρασης. Ανάλογα με το γεγονός εισόδου, η μονάδα παραγωγής συμπεριφοράς (Behavioral Generator) αποφασίζει στη συνέχεια για την ενέργεια του συστήματος πολλαπλών ρομπότ και προωθεί την απόφασή της στους μηχανισμούς ενεργοποίησης (actuators). Αυτοί με τη σειρά τους ανταποκρίνονται δίνοντας πληροφορίες πίσω στο σύστημα.

Αρχιτεκτονική

Οι μονάδες αντίληψης είναι ενσωματωμένες στο πλήρες σύστημα αντίληψης με βάση την ακόλουθη αρχιτεκτονική: Το σύστημα λειτουργεί σε τέσσερις κατανεμημένες διασυνδεδεμένες μηχανές, οι τρεις εκ των οποίων τρέχουν σε λειτουργικό σύστημα Linux και ROS (Robotic Operation System) και η τελευταία σε Windows. Καθένα από τα τρία μηχανήματα Linux συνδέεται με έναν αισθητήρα Kinect V2 που παρέχει ακατέργαστα δεδομένα (δηλαδή χρώμα, βάθος και ήχο). Το μηχάνημα των Windows είναι επίσης συνδεδεμένο με έναν αισθητήρα Kinect V2 και μέσω του λογισμικού Microsoft SDK Kinect V2 API παρέχει πρόσθετες πληροφορίες δισδιάστατου σκελετού. Μια οθόνη αφής συνδέεται επίσης με το μηχάνημα των Windows και στέλνει πληροφορίες-συμβάντα σχετικά με τις επιλογές των παιδιών στη μονάδα διαλόγου. Η κύρια επεξεργασία δεδομένων των μονάδων αντίληψης πραγματοποιείται σε καθεμία από τις τρεις μηχανές Linux, ενώ η συγχώνευση των πολλαπλών βίντεο-όψεων γίνεται σε μία από τις μηχανές Linux. Η ροή δεδομένων και η επικοινωνία μεταξύ των μονάδων του συστήματος γίνονται μέσω συμβάντων που μεταδίδονται μέσω του broker TCP/IP, που εκτελείται στο μηχάνημα των Windows και παρέχεται από το λογισμικό IrisTK [Skantze and Al Moubayed; 2012]. Η εργασία που εκτελεί ο broker είναι να μεταφέρει μηνύματα/συμβάντα από το ένα module στο άλλο. Σύμφωνα με τη δομή του IrisTK, χωρίζουμε τα γεγονότα σε τρεις κατηγορίες:

- *Συμβάντα Αίσθησης - Sense*: περιλαμβάνουν πληροφορίες σχετικά με το τι λαμβάνουν οι αισθητήρες του συστήματος
- *Συμβάντα Δράσης - Action*: δίνουν εντολή για μια δράση (δηλ. ένα ρομπότ) να κάνει κάτι
- *Συμβάντα Ενημέρωσης - Monitor*: είναι συμβάντα παρασκηνίου που περιέχουν πληροφορίες ανατροφοδότησης σχετικά με τις ενέργειες του συστήματος (π.χ. όταν ένα ρομπότ έχει σταματήσει να μιλάει)

Ομοίως, η αρχιτεκτονική του συστήματος σχεδιάστηκε με βάση την αρχή Sense - Think - Act [Gat et al.; 1998], όπως φαίνεται στο Σχήμα 2.2. Η ρύθμιση πολλών αισθητήρων του συστήματος αντιπροσωπεύει το τμήμα Sense, ενώ οι μονάδες αντίληψης ταξινομούνται στην αρχή



Σχήμα 2.6: Η «ενέργεια εκφώνησης» χρησιμοποιείται για να πει το ρομπότ ποια χειρονομία αναγνώρισε κατά τη διάρκεια της αλληλεπίδρασης.

Think. Τέλος, τα πολλαπλά ρομπότ ανήκουν στο τμήμα Act της αρχιτεκτονικής. Αυτή η αρχιτεκτονική τριών επιπέδων δίνει μια σπονδυλωτή μορφή στο σύστημα και επιτρέπει την αυτονομία των επιμέρους τμημάτων του συστήματος (modularity), επειδή τα διαφορετικά επίπεδα μπορούν να αντικατασταθούν/τροποποιηθούν χωρίς να επηρεαστούν τα υπόλοιπα.

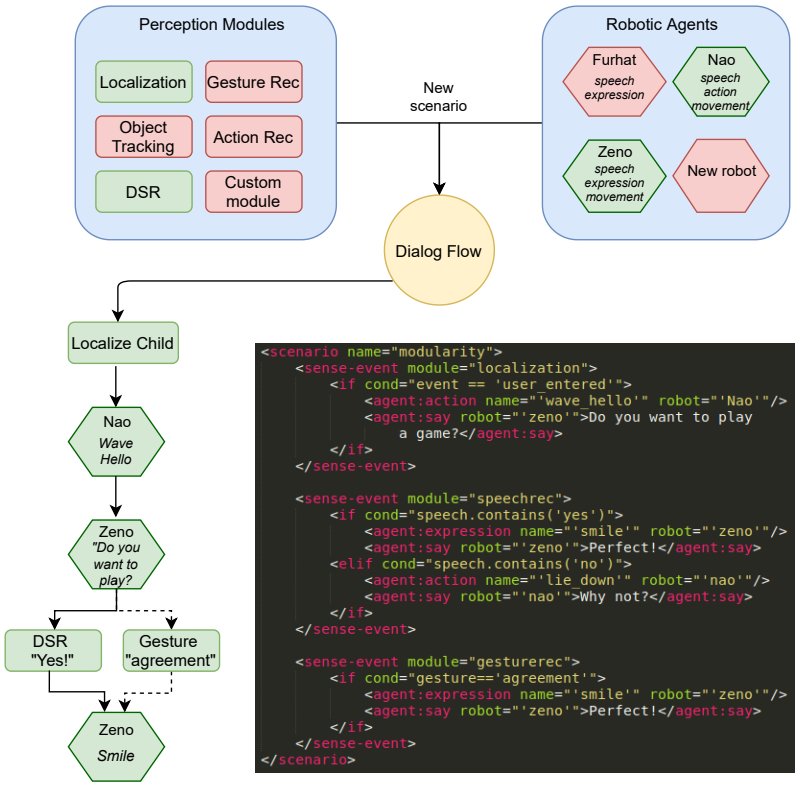
Ο *broker μαζί με την ενότητα διαχείρισης διαλόγου*, που θα περιγραφεί στη συνέχεια, λειτουργεί ως κεντρική μονάδα που λαμβάνει συμβάντα από όλες τις επιμέρους μονάδες του συστήματος και τα διανέμει ανάλογα με το περιεχόμενο, στις κατάλληλες μονάδες. Έτσι είναι δυνατό οι μονάδες να αφαιρούνται ή να προστίθενται εύκολα στην αρχιτεκτονική αφού οριστούν απλώς τα σύνολα γεγονότων που η λειτουργική μονάδα πρέπει να δέχεται - αντιλαμβάνεται ή να στέλνει πίσω στον broker.

Η *μονάδα παραγωγής συμπεριφοράς* (Behavior Generator) είναι η κεντρική μονάδα του συστήματος και μοντελοποιεί τη ροή αλληλεπίδρασης μεταξύ του χρήστη και του συστήματος. Η αλληλεπίδραση μοντελοποιείται χρησιμοποιώντας καταστάσεις Harel [Harel; 1987], δηλαδή καταστάσεις που μπορούν να δομηθούν ιεραρχικά, να εκτελεστούν υπό όρους και να περιέχουν παραμέτρους που αλλάζουν τη ροή και τις μεταβάσεις. Επιπλέον, οι καταστάσεις μπορούν να καλούνται ως συναρτήσεις, πράγμα που σημαίνει ότι η ροή της εκτέλεσης θα συνεχιστεί στην επόμενη κατάσταση κλήσης αφού ολοκληρωθεί η πρώτη.

Επίσης, έχουμε συμπεριλάβει «καταστάσεις δράσης», δηλαδή καταστάσεις που λειτουργούν ως μεσολαβητές μεταξύ της ροής διαλόγου και των ρομπότ, για το σχεδιασμό και τη δημιουργία του γραφήματος καταστάσεων που μοντελοποιεί το διάλογο. Αυτές οι καταστάσεις δράσης περιέχουν τις πληροφορίες που απαιτούνται για να δοθεί εντολή στα ρομπότ του συστήματος να εκτελέσουν μια ενέργεια ενώ περιλαμβάνουν τα ρομπότ ως μια πρόσθετη παράμετρο κατάστασης. Έτσι, η ροή διαλόγου αποσυνδέεται από λεπτομέρειες σχετικές με το κάθε ρομπότ και αποφεύγεται ο πολλαπλός ορισμός ίδιων καταστάσεων για τα διαφορετικά ρομπότ. Αυτή η ιδιότητα του συστήματος επιτρέπει ακόμα την εύκολη ένταξη νέων ρομπότ στη ροή διαλόγου προσθέτοντας μόνο συγκεκριμένες λεπτομέρειες, σχετικές με το ρομπότ, στις καταστάσεις ενεργειών. Οι ενέργειες-δράσεις του ρομπότ δημιουργούνται στο κάθενα ξεχωριστά. Ένα παράδειγμα φαίνεται στο Σχήμα 2.6 όπου μια κατάσταση στη ροή διαλόγου καλεί την «ενέργεια εκφώνησης» και δίνει ως παράμετρο την επιλογή του ρομπότ που θα την εκτελέσει, δηλαδή ποιο ρομπότ θα μιλήσει στο παιδί.

Από τα τρία ρομπότ που χρησιμοποιεί το σύστημα πολλαπλών ρομπότ μας, το ρομπότ Furhat είναι ήδη ενσωματωμένο στο πλαίσιο IrisTK. Για τα ρομπότ NAO και Zeno, αναπτύξαμε ενδιάμεσα API (Application Programming Interface)που χρησιμοποιούμε για την επικοινωνία μεταξύ των ρομπότ και του διαλόγου.

Αρθρωτή δομή και νέα σενάρια: Όπως είδαμε αναλυτικά, η αρθρωτή αρχιτεκτονική του συστήματος, δίνει τη δυνατότητα εναλλαγής/προσθήκης/αφαίρεσης των μονάδων αντίληψης



Σχήμα 2.7: Παρουσίαση της σπονδυλωτής δομής του συστήματος μέσω ενός απλού σεναρίου χρησιμοποιώντας δύο ρομπότ και δύο μονάδες αντίληψης

και των ρομπότ. Ως αποτέλεσμα, το ChildBot μπορεί να θεωρηθεί τόσο ως ένα ολοκληρωμένο σύστημα, αλλά και ως μια συνάθροιση διαφορετικών ρομπότ και μονάδων, τα οποία μπορούν να ενεργοποιηθούν/απενεργοποιηθούν ανάλογα με την επιθυμητή εφαρμογή.

Ένα παράδειγμα αυτής της αρθρωτής δομής φαίνεται στο Σχήμα 2.7, όπου εμφανίζεται ένα απλό σενάριο που χρησιμοποιεί μόνο δύο ρομπότ (Nao και Zeno) και δύο μονάδες αντίληψης (DSR και εντοπισμός). Όπως φαίνεται, λόγω της αρχιτεκτονικής του συστήματος, εκτός από το Dialog Manager, όλες οι άλλες μονάδες (δηλαδή μονάδες αντίληψης και ρομπότ) μπορούν να ενεργοποιηθούν/απενεργοποιηθούν χωρίς να επηρεαστεί η λειτουργικότητα του συστήματος. Για παράδειγμα, στο σενάριο που εμφανίζεται, αφού το ρομπότ Zeno ρωτήσει το παιδί: «Θέλεις να παίξεις;», το σενάριο μπορεί να περιλαμβάνει είτε τη μονάδα DSR είτε τη μονάδα αναγνώρισης χειρονομίας (ή μπορεί να είναι και τα δύο) για να αναγνωρίσει την απάντηση του παιδιού.

Επίσης είναι σημαντικό να σημειώσουμε πως οι ελάχιστες απαιτήσεις για τη λειτουργία του συστήματος περιλαμβάνουν ένα μηχάνημα που τρέχει το ROS, το Dialog Manager που βασίζεται στο IrisTK, και ένα Broker. Όλα αυτά είναι προγράμματα ανοιχτού κώδικα. Επιπλέον, ενώ έχουμε χρησιμοποιήσει αισθητήρες Kinect V2 (οι οποίοι απαιτούν τη σύνδεση κάθε αισθητήρα με έναν υπολογιστή για real-time καταγραφή και επεξεργασία του βίντεο Full HD με 30fps), το σύστημα μπορεί να χρησιμοποιήσει οποιονδήποτε αισθητήρα RGB και μικρόφωνο που υποστηρίζεται από το ROS.

2.3.2 Ενδεικτικά Σενάρια Χρήσης

Ένα σύνολο σεναρίων έχει σχεδιαστεί για να τονίσει τις δυνατότητες του συστήματος κατά τη διάρκεια μιας διασκεδαστικής και εκπαιδευτικής πολυτροπικής αλληλεπίδρασης μεταξύ παιδιών και ρομπότ. Όπως εξηγείται εκτενώς, το ολοκληρωμένο σύστημά μας αντιλαμβάνεται διάφορα γεγονότα που συμβαίνουν κατά τη διάρκεια της αλληλεπίδρασης, όπως η ομιλία και οι κινήσεις των παιδιών, οι θέσεις τους στο δωμάτιο και ο εντοπισμός αντικειμένων. Κάθε σενάριο εστιάζει σε διαφορετικές τεχνολογίες και τις συνδυάζει κατάλληλα για να δημιουργήσει μια ομαλή αλληλεπίδραση. Τα παιδιά καλούνται να ολοκληρώσουν τις παρακάτω εργασίες-παιχνίδια: α) «Δείξε μου τη χειρονομία», β) «Δείχνω τα συναισθήματα μου», γ) «Παντομίμα», δ) «Φτιάχνω σχήματα», ε) «Φτιάχνω μια φάρμα».

Στο πρώτο σενάριο, **«Δείξε μου τη χειρονομία»**, το παιδί αλληλεπιδρά με το ρομπότ μέσω χειρονομιών και ομιλίας. Το ρομπότ ζητά από το παιδί να κάνει μια χειρονομία και προσπαθεί να το αναγνωρίσει. Στη συνέχεια ζητά από το παιδί λεκτική επιβεβαίωση της αναγνώρισης. Τα παιδιά καλούνται να κάνουν μια δράση που να: α) δείχνει συμφωνία, β) καλεί το ρομπότ να πλησιάσει, γ) ζητά από το ρομπότ να καθίσει, δ) δείχνει ένα αντικείμενο στο δωμάτιο, ε) ζητά από τον ρομπότ να σταματήσει αυτό που κάνει, και στ) σχεδιάζει έναν κύκλο στον αέρα. Εκτός από την πρώτη δράση που συνήθως εκτελείται με ένα νεύμα, τα υπόλοιπα συνήθως αποδίδονται με κινήσεις των χεριών. Τα παραπάνω νοήματα χρησιμοποιούνται καθημερινά στην επικοινωνία μεταξύ των ανθρώπων και αποδίδονται συνήθως με κινήσεις είτε μόνα τους είτε σε συνδυασμό με λεκτικές εντολές και συνεπώς είναι ιδιαίτερα χρήσιμες να ενταχθούν σε μια αλληλεπίδραση μεταξύ ανθρώπων και ρομπότ που θέλουμε να γίνεται με φυσικότητα. Έτσι στο σενάριό μας, τα παιδιά επιτρέπεται να κάνουν χειρονομίες αυθόρμητα, όπως θα έκαναν όταν αλληλεπιδρούν με έναν άλλο άνθρωπο.

Το παιχνίδι **«Δείχνω τα συναισθήματα μου»** παρακινεί τα παιδιά να αποκαλύψουν τα συναισθήματά τους χρησιμοποιώντας τόσο το πρόσωπο όσο και το σώμα τους κατά τη διάρκεια μιας διασκεδαστικής αλληλεπίδρασης με το ρομπότ. Σε αυτό το παιχνίδι, το παιδί επιλέγει μία από τις κάρτες που απεικονίζονται στην οθόνη αφής και εκφράζει το συναίσθημα που έχει επιλέξει. Τα συναισθήματα που περιλαμβάνονται σε αυτό το παιχνίδι είναι η χαρά, η λύπη, ο φόβος, ο θυμός, η έκπληξη και η αηδία. Μετά την αντίδραση του παιδιού, το ρομπότ εκφράζει το ίδιο συναίσθημα χρησιμοποιώντας το πρόσωπό του.

Η **«Παντομίμα»** είναι ένα δημοφιλές παιχνίδι, κατά τη διάρκεια του οποίου, ένα άτομο μιμείται τη κίνηση και ένα άλλο άτομο προσπαθεί να καταλάβει τι μιμείται. Το παιδί μπορεί να χρησιμοποιήσει ολόκληρο το σώμα για να μιμηθεί μια δραστηριότητα και να αλληλεπιδράσει εκτενώς με το ρομπότ. Το ρομπότ και το παιδί ανταλλάσσουν επανειλημμένα του ρόλους τους, άλλοτε μιμείται ο ένας και μαντεύει ο άλλος και αντίστροφα. Οι δώδεκα δραστηριότητες που χρησιμοποιούνται σε αυτό το παιχνίδι είναι οι εξής: α) καθάρισμα παραθύρου, β) οδήγηση λεωφορείου, γ) χτύπημα με σφυρί, δ) κολύμπι, ε) γυμναστική, στ) χορός , ζ) διάβασμα ενός βιβλίου, η) σκάψιμο, θ) παίξιμο κιθάρας, ι) σκούπισμα, ια) βάψιμο τοίχου και ιβ) σιδέρωμα πουκάμισου.

Για το **«Φτιάχνω σχήματα»**, ένα ή περισσότερα παιδιά καλούνται να συναρμολογήσουν κάποια σχήματα υπό την επίβλεψη του ρομπότ. Στο παιχνίδι αυτό χρησιμοποιούνται έξι τρισδιάστατα τυπωμένα τουβλάκια διαφορετικού μήκους για τη συνδεσή τους και τη δημιουργία ορθογωνίων και τετραγώνων σχημάτων. Τα τουβλάκια τοποθετούνται σε ένα τραπέζι μπροστά από το παιδί, με το ρομπότ να στέκεται κοντά. Το παιδί είναι υπεύθυνο για το χειρισμό των αντικειμένων της συναρμολόγησης ενώ το ρομπότ παρέχει της οδηγίες σύμφωνα και με την εξέλιξη της συναρμολόγησης. Εάν το παιδί ολοκληρώσει σωστά μια σύνδεση ανάμεσα στα τουβλάκια, το ρομπότ συγχαίρει το παιδί και δίνει την επόμενη οδηγία. Ωστόσο, εάν το παιδί κάνει λάθος, το ρομπότ θα προσπαθήσει να το αναγνωρίσει και να αντιδράσει ανάλογα ώστε το παιδί να διορθώσει το λάθος του. Εκτός από τις προφορικές οδηγίες, για λόγους σαφήνειας, το ρομπότ

	Αναγνώριση Φωνής	Εντοπισμός & Ιχνηλάτηση	Εντοπισμός Ομιλητή	Αναγνώριση Δράσης	
				Χειρονομίες	Κινήσεις
Δείξε μου τη χειρονομία	✓		✓	✓	
Παντομίμα	✓		✓		✓
Φτιάχνω σχήματα		✓	✓		
Φτιάχνω μια φάρμα	✓		✓	✓	
Δείχνω τα συναισθήματα μου					

Πίνακας 2.1: Τα προτεινόμενα σενάρια σε συνδυασμό με τις τεχνολογίες που μελετήσαμε και αναπτύξαμε κατά τη δημιουργία του ChildBot.

κοιτάζει και δείχνει τα τουβλάκια στα οποία αναφέρεται.

Το παιχνίδι **«Φτιάχνω μια φάρμα»** είναι ένα παιχνίδι πολλών συμμετεχόντων που περιλαμβάνει δύο ρόλους που μπορούν να παιχτούν εξίσου από το ρομπότ και το παιδί ή τα παιδιά. Στόχο έχει την ψυχαγωγία, την εκπαίδευση και τη δημιουργία μιας φυσικής αλληλεπίδρασης μεταξύ όλων των συμμετεχόντων. Το παιχνίδι περιλαμβάνει δύο διαφορετικούς ρόλους, ο ένας επιλέγει ένα ζώο και ο άλλο πρέπει να το μαντέψει. Αρχικά ο παίκτης που επιλέγει το ζώο αποκαλύπτει σιγά σιγά κάποια χαρακτηριστικά του και ο άλλος παίκτης προσπαθεί να μαντέψει το ζώο που έχει επιλεγεί. Η αλληλεπίδραση εξελίσσεται ως εξής: Στην αρχή, το ρομπότ επιλέγει ένα τυχαίο ζώο και οι παίκτες μαντεύουν εκ περιτροπής ποιο είναι το ζώο. Σε περίπτωση λανθασμένης εικασίας, το ρομπότ αποκαλύπτει ένα χαρακτηριστικό του ζώου (χρώμα, αριθμός ποδιών, κατηγορία π.χ. θηλαστικά, ερπετά). Σε περίπτωση σωστής αναγνώρισης του ζώου, το ρομπότ ζητά από τα παιδιά να τοποθετήσουν σωστά το ζώο σε μια φάρμα, με καθορισμένες και διακριτές περιοχές, που εμφανίζεται σε μια οθόνη αφής μπροστά τους. Στο δεύτερο γύρο, οι ρόλοι αντιστρέφονται: τα παιδιά συζητούν και επιλέγουν ένα ζώο και αποκαλύπτουν ένα χαρακτηριστικό. Στη συνέχεια, το ρομπότ προσπαθεί να μαντέψει το ζώο που διάλεξαν. Εάν το ρομπότ μαντέψει σωστά, τα παιδιά καλούνται ξανά να τοποθετήσουν το ζώο μέσα στο αγρόκτημα, διαφορετικά αποκαλύπτουν περισσότερα χαρακτηριστικά του ζώου, ένα κάθε φορά. Το παιχνίδι συνεχίζεται εναλλάσσοντας τον ρόλο του μάντη μεταξύ των παιδιών και του ρομπότ σε κάθε γύρο. Το παιχνίδι περιλαμβάνει συνολικά 19 ζώα και τα χαρακτηριστικά τους ανήκουν σε πέντε διαφορετικές κατηγορίες: χρώμα, μέγεθος, είδος, αριθμός ποδιών και ένα χαρακτηριστικό γνώρισμα ή φράση, π.χ. για τον σκύλο: «είναι ο καλύτερος φίλος του ανθρώπου».

Τα προαναφερθέντα σενάρια στοχεύουν στη δημιουργία ενός κατάλληλου πλαισίου για πολυτροπική επικοινωνία μεταξύ παιδιών και ρομπότ, όπως συμβαίνει μεταξύ των ανθρώπων. Με αυτόν τον τρόπο, τα σενάρια αναδεικνύουν τις δυνατότητες του συστήματος και δίνουν μερικά παραδείγματα για το πως μπορεί να χρησιμοποιηθεί το ChildBot ενώ αξιοποιούνται και για την αξιολόγηση του συστήματος. Παρόλο που κάθε παιχνίδι-σενάριο εστιάζει σε μία από τις τεχνολογίες αντίληψης συστήματος, χρησιμοποιούνται παράλληλα περισσότερες από μία μονάδες αντίληψης. Στους Πίνακες 2.1 και 2.2, οι μονάδες που χρησιμοποιούνται συνοψίζονται μαζί με την καταλληλότητα των ρομπότ για συμμετοχή σε κάθε σενάριο.

Τα σενάρια, εφόσον προορίζονται για χρήση από παιδιά, έχουν σχεδιαστεί με την καθοδήγηση ψυχολόγων ύστερα από πιλοτική μελέτη που πραγματοποιήσαμε με οκτώ παιδιά στον ειδικά διαμορφωμένο χώρο του εργαστηρίου μας.

2.3.3 Βάση Δεδομένων

Τα πραγματικά δεδομένα που λαμβάνονται κατά την αλληλεπίδραση ανθρώπων - ρομπότ αποδεικνύονται ιδιαίτερα σημαντικά κατά την ανάπτυξη ενός συστήματος, από την αρχική εκπαίδευση έως το τελικό στάδιο της αξιολόγησης. Τέτοια δεδομένα συμβάλλουν στην προ-

	Ρομπότ			Μονάδα Παραγωγής Συμπεριφοράς	Οθόνη Αφής
	NAO	Furhat	Zeno		
Δείξε μου τη χειρονομία	✓	✓	✓	✓	
Παντομίμα	✓			✓	✓
Φτιάχνω σχήματα	✓		✓	✓	
Φτιάχνω μια φάρμα	✓	✓	✓	✓	✓
Δείχνω τα συναισθήματα μου		✓	✓	✓	✓

Πίνακας 2.2: Τα προτεινόμενα σενάρια σε συνδυασμό με τη δυνατότητα συμμετοχής κάθε ρομποτικού πράκτορα, την αξιοποίηση της οθόνης αφής και της μονάδας παραγωγής συμπεριφοράς σε αυτά.

(α΄) Kinect#1 (β΄) Kinect#2 (γ΄) Kinect#3 (δ΄) Kinect#4

Σχήμα 2.8: Τα τέσσερα διαφορετικά σενάρια-παιχνίδια που έλαβαν χώρα στο εργαστήριό μας. Το καθένα παρουσιάζεται από διαφορετική οπτική-κάμερα. α) Δείξε μου τη χειρονομία, β) Παντομίμα, γ) Φτιάχνω μια φάρμα, δ) Δείχνω τα συναισθήματά μου.

σαρμογή του συστήματος σε πραγματικές συνθήκες χρήσης και στην αυθόρμητη συμπεριφορά των χρηστών. Έτσι, έχει πραγματοποιηθεί **εκτεταμένη συλλογή δεδομένων** με τη συμμετοχή μιας ομάδας 52 παιδιών, ηλικίας από έξι έως έντεκα ετών, σε μια ειδικά διαμορφωμένη αίθουσα και σε μια σχολική τάξη.

Τα περισσότερα από τα δεδομένα έχουν συλλεχθεί σε ένα δωμάτιο που μοιάζει με παιδικό δωμάτιο και όπου έχουν τοποθετηθεί τα ρομπότ και οι αισθητήρες, όπως παρουσιάζεται στην Εικόνα 2.4. Εκεί, η συλλογή δεδομένων πραγματοποιήθηκε σε δύο φάσεις. Στην πρώτη, τα δεδομένα των παιδιών έχουν καταγραφεί ,με ελεγχόμενο τρόπο, κατά την εκτέλεση συγκεκριμένων ενεργειών και την εκφώνηση ορισμένων φράσεων που αναμένεται να προκύψουν κατά τη διάρκεια της αλληλεπίδρασης μεταξύ αυτών και των ρομπότ. Αυτά τα δεδομένα θα αναφέρονται παρακάτω ως ***δεδομένα ανάπτυξης*** *(development data)* καθώς έχουν χρησιμοποιηθεί για την ανάπτυξη του συστήματος. Στη δεύτερη φάση, τα δεδομένα συλλέχθηκαν κατά τη διάρκεια της πειραματικής διαδικασίας όπου τα παιδιά αλληλεπιδρούν με ρομπότ χωρίς διακοπή ή παρέμβαση άλλων ανθρώπων. Αυτά τα δεδομένα θα αναφέρονται ως ***δεδομένα αλληλεπίδρασης*** *(use-case related data).* Και οι δύο τύποι δεδομένων είναι εξίσου σημαντικοί για το CRI, καθώς ο πρώτος είναι απαραίτητος για την εκπαίδευση των μονάδων αντίληψης σε δεδομένα που σχετίζονται με σενάρια χρήσης, ενώ ο δεύτερος είναι απαραίτητος για την αξιολόγηση του συστήματος κατά τη διάρκεια του CRI. Ο Πίνακας 2.3 παρουσιάζει τα πιο σημαντικά γεγονότα που καταγράφηκαν κατά τη διάρκεια των δύο φάσεων και τον συνολικό αριθμό των εμφανίσεων τους.

Οι πληροφορίες που συλλέξαμε κατά τη συλλογή δεδομένων περιλαμβάνουν *βίντεο* Full HD (1920×1080) RGB και βάθους (512×424) και από τις τέσσερις κάμερες Kinect, που τρέχουν στα 30 fps, καθώς και ακατέργαστο *ήχο* από τη συστοιχία μικροφώνου που είναι ενσωματωμένη σε κάθε αισθητήρα Kinect. Εκμεταλλευόμενοι το Kinect v2 API έχουμε επίσης καταγράψει από

τον αισθητήρα που βρίσκεται πάνω από την οθόνη αφής τα ακόλουθα: (α) *Εκτίμηση των σκελετών* των ατόμων τόσο σε 2D όσο και σε 3D συντεταγμένες. (β) *Bounding boxes για το πρόσωπο, landmarks του προσώπου και ένα τρισδιάστατο πλέγμα* αυτού.

Για τα **δεδομένα ανάπτυξης**, 28 παιδιά συμμετείχαν εκτελώντας επτά χειρονομίες και δώδεκα παντομίμες και προφέροντας 40 φράσεις από ένα λεξιλόγιο 120 φράσεων. Αυτή η φάση είναι κρίσιμη για την ανάπτυξη των μοντέλων αντίληψης και την προσαρμογή τους στα παιδιά, καθώς επικεντρώνονται στην ομιλία, τις χειρονομίες και τις ενέργειες που σχετίζονται με τις περιπτώσεις χρήσης. Συγκεκριμένα, τα παιδιά είναι πιο αυθόρμητα και εκφραστικά από τους ενήλικες και ο λόγος τους είναι συνήθως σύντομος και χαμηλόφωνος. Έτσι, για να ελέγξουμε την απόδοση των μονάδων αντίληψης ChildBot, είναι απαραίτητο να έχουμε μια πληθώρα παιδικών δραστηριοτήτων και εκφράσεων. Επιπλέον, έχουν συλλεχθεί δεδομένα ενηλίκων για να ενισχύσουν τα δεδομένα των σεναρίων, να βοηθήσουν στη διερεύνηση για το αν είναι αναγκαία η συλλογή δεδομένων παιδιών για τη βελτίωση της απόδοσης στα μοντέλα αντίληψης για CRI εφαρμογές.

Όσον αφορά τα δεδομένα σχετικά με τα **δεδομένα αλληλεπίδρασης**, 31 παιδιά με μέσο όρο ηλικίας 8,6 ετών, 10 κορίτσια και 21 αγόρια, επιλέχθηκαν τυχαία από ένα σύνολο εθελοντών που γνώρισαν την ομάδα μας σε μια εκδήλωση. Τα παιδιά ήταν ηλικίας από έξι έως έντεκα χρονών, και όλα μιλούσαν ελληνικά και γνώριζαν γραφή και ανάγνωση. Κάθε παιδί συνοδευόμενο από τους γονείς του μπήκε στην ειδικά διαμορφωμένη αίθουσα και γνωρίστηκε με τα ρομπότ. Το παιδί είχε χρόνο να εξοικειωθεί με το χώρο και τα ρομπότ ενώ ο ερευνητής του εξηγούσε τη δομή της διαδικασίας και τα παιχνίδια που θα έπαιζε. Στη συνέχεια, οι γονείς και οι ερευνητές βγήκαν από τον χώρο αλληλεπίδρασης και το παιδί έπαιξε ατομικά παιχνίδια με τα ρομπότ. Μετά την ολοκλήρωση της ατομικής αλληλεπίδρασης, ένα δεύτερο παιδί (που είχε ολοκληρώσει την ίδια αλληλεπίδραση προηγουμένως) μπαίνει στον χώρο και τα δύο από κοινού πλέον συμμετέχουν στο παιχνίδι «Φτιάξε μια φάρμα». Σε περιπτώσεις που δεν υπήρχε διαθέσιμο δεύτερο παιδί, ένας ενήλικας έπαιρνε τη θέση του, τα δεδομένα του οποίου, όπως είναι λογικό, δεν συμπεριλαμβάνονται στην αξιολόγηση του συστήματος. Τέλος, μετά την ολοκλήρωση της διαδικασίας, τα παιδιά κλήθηκαν να συμπληρώσουν ένα ερωτηματολόγιο που περιλάμβανε τη γνώμη τους για την εμπειρία που είχαν. Το ερωτηματολόγιο περιγράφεται και αναλύεται στην Ενότητα 2.3.4.

Η Επιτροπή Ηθικής και Δεοντολογίας του Ερευνητικού Κέντρου Αθηνά ενέκρινε την παραπάνω διαδικασία και το έντυπο συγκατάθεσης που εστάλη μέσω email στους γονείς πριν από τα πειράματα. Επιπλέον, όλα τα πειράματα έχουν πάρει την έγκριση έμπειρων παιδοψυχολόγων.

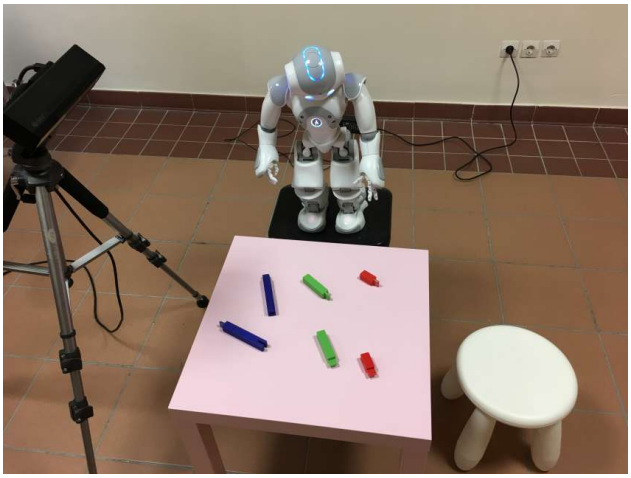
Τα *δεδομένα που σχετίζονται με το σενάριο «Φτιάχνω Σχήματα»* συλλέχθηκαν σε ένα ελληνικό δημοτικό σχολείο με τη συμμετοχή 21 μαθητών, 9-10 ετών. Έξι μαθητές συμμετείχαν μόνοι τους και οι υπόλοιποι οργανώθηκαν σε ομάδες των πέντε, σύμφωνα με τις συμβουλές των δασκάλων, προκειμένου να ενισχυθούν οι συνεργατικές τους δεξιότητες. Για αυτό το πείραμα, μόνο μια κάμερα Kinect και ένας ρομποτικός πράκτορας (NAO robot) έχουν επιλεγεί ως μια ελαφριά έκδοση του συστήματος για να φιλοξενήσει την εκπαιδευτική διαδικασία. Μια τέτοια έκδοση μπορεί εύκολα να εγκατασταθεί σε μια τυπική τάξη και να βοηθήσει τον δάσκαλο να δώσει ένα ζωντανό μάθημα μέσω μιας εμπειρίας CRI (Εικόνα 2.9).

2.3.4 Αξιολόγηση Αλληλεπιδράσεων: Μελέτη Εμπειρίας Χρήστη

Στη συγκεκριμένη παράγραφο παρουσιάζονται τόσο τα στατιστικά στοιχεία από την ανάλυση των πειραμάτων των παιδιών, όσο και μια υποκειμενική τους αξιολόγηση για την εμπειρία που είχαν. Ερευνητικά, ένα ερωτηματολόγιο που δίνεται σε παιδιά μικρής ηλικίας δεν θεωρείται τόσο αξιόπιστο όσο ένα ερωτηματολόγιο που δίνεται σε ενήλικες. Συχνά παρατηρείται το φαινόμενο τα παιδιά να απαντάνε πολύ θετικά ώστε να ευχαριστήσουν τους ενήλικες που θέτουν

Δεδομένα BabyRobot	Τύπος Συμβάντος	# Συμβάντων
Δεδομένα Ανάπτυξης	Φράσεις	977
	Χειρονομίες	196
	Παντομίμες	336
Δεδομένα Αλληλεπίδρασης	Φράσεις	641
	Χειρονομίες	143
	Παντομίμες	109

Πίνακας 2.3: Στατιστικά στοιχεία των σημαντικότερων παιδικών δραστηριοτήτων κατά τη συλλογή δεδομένων.



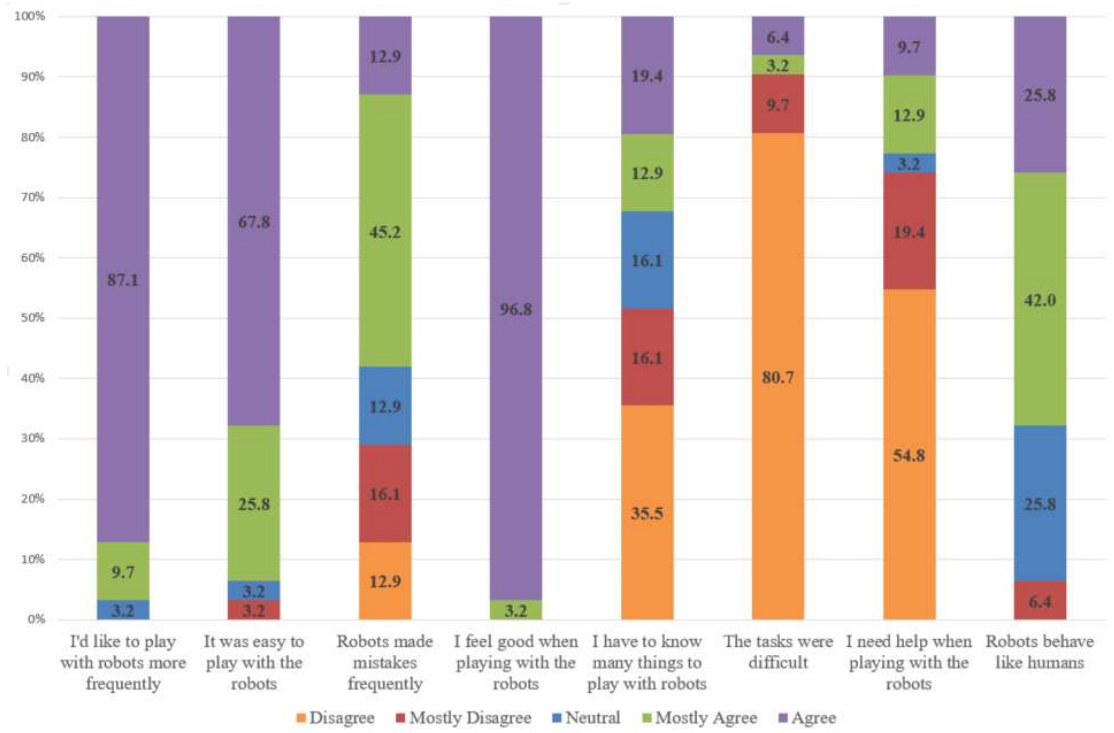
Σχήμα 2.9: Η διάταξη του παιχνιδιού «Φτιάχνω Σχήματα». Τα πειράματα πραγματοποιήθηκαν σε ένα ελληνικό δημοτικό σχολείο.

τις ερωτήσεις ή αντίστροφα να επιλέγουν πολύ αρνητικές απαντήσεις για να αστειευτούν. Παρ' όλα αυτά εμείς κρίναμε πως μελετώντας τις υποκειμενικές τους αξιολογήσεις θα μπορούσαμε να δούμε κάποιες πτυχές για το πως βιώνουν μια αλληλεπίδραση με τα ρομπότ.

Στατιστικά στοιχεία

Όσον αφορά τα **ατομικά παιχνίδια-σενάρια**, όπου κάθε ένα από τα 31 παιδιά συμμετείχε μόνο του, όλα ήταν σε θέση να *ολοκληρώσουν τα παιχνίδια με επιτυχία*. Η μέση διάρκεια ολοκλήρωσης, συμπεριλαμβανομένης της εισαγωγής από τα ρομπότ, ήταν 9 λεπτά, με τυπική απόκλιση 2 λεπτών.

Στο 32% των περιπτώσεων, απαιτήθηκε ανθρώπινη (λεκτική) παρέμβαση έως και δύο φορές κατά τη διάρκεια της πειραματικής ροής, όταν τα παιδιά μπερδεύτηκαν ή είχαν ερωτήσεις σχετικά με τη διαδικασία. Για παράδειγμα, μερικά παιδιά ζήτησαν επιβεβαίωση σχετικά με το τι πρέπει να κάνουν ή χρειάστηκαν μια προτροπή για να ενεργήσουν. Τέτοιες πιθανές αποκλίσεις από το σχεδιασμένο σενάριο ξεπεράστηκαν μέσω του dialog manager όπου όταν αναγνωρίστηκαν τέτοιες περιπτώσεις (π.χ. εάν το παιδί είναι σιωπηλό για ένα εύλογο χρονικό διάστημα) στάλθηκε ένα συμβάν σε κάποιο ρομπότ ώστε να προτρέψει το παιδί να κάνει κάτι ή να του ζητήσει να επαναλάβει την ομιλία/δραστηριότητά του. Σε περιπτώσεις που τα παιδιά αναμενόταν να πουν κάτι ή δεν αναγνωρίστηκε η ομιλία τους, τα ρομπότ ζητούσαν επαναλήψεις έως και δύο συνεχόμενες φορές, ενώ στην περίπτωση της δράσης ενός παιδιού, τα ρομπότ ζητούσαν



Σχήμα 2.10: Υποκειμενική αξιολόγηση των παιδιών για την εμπειρία τους με το ChildBot. Μετά από κάθε ολοκληρωμένη αλληλεπίδραση, τα παιδιά κλήθηκαν να συμπληρώσουν ένα ερωτηματολόγιο με τις εμφανιζόμενες ερωτήσεις σε κλίμακα τύπου Likert από το 1 έως το 5 όπως φαίνεται.

επανάληψη μόνο μία φορά.

Για το **συλλογικό παιχνίδι** «Φτιάχνω μια φάρμα», που παίζεται από δύο παιδιά, παρατηρήθηκε ότι τα μικρότερα παιδιά αντιμετώπιζαν δυσκολίες με τους κανόνες του παιχνιδιού, παρόλο που τα παιδιά του δημοτικού σχολείου είναι εξοικειωμένα με τα ζώα μιας φάρμας. Ως αποτέλεσμα, παιδιά ηλικίας έξι και επτά χρόνων έπαιξαν το παιχνίδι ακολουθώντας τις οδηγίες που προσφέρονταν από έναν ενήλικα. Τα υπόλοιπα παιδιά έπαιξαν το παιχνίδι χωρίς καμία καθοδήγηση. Η μέση διάρκεια του παιχνιδιού ήταν 8 λεπτά. Συνολικά, τα παιδιά ανέλαβαν το ρόλο του μάντη για 24 γύρους και βρήκαν τη σωστή απάντηση χρησιμοποιώντας 2,4 εικασίες κατά μέσο όρο και 4 εικασίες το μέγιστο. Το ρομπότ ανέλαβε τον ρόλο του μάντη για 22 γύρους και βρήκε τη σωστή απάντηση σε 2,2 εικασίες κατά μέσο όρο, με μέγιστο τις 6. Τα παιδιά δεν αναγνώρισαν το ζώο που είχε επιλεγεί από το ρομπότ στο 4% των περιπτώσεων, ενώ το ρομπότ στο 32%. Σε γενικές γραμμές, τα παιδιά κατάφεραν να μαντέψουν το ζώο πιο εύκολα από ότι το ρομπότ, αφού το ρομπότ ήταν προγραμματισμένο να αποκαλύπτει τα χαρακτηριστικά των ζώων από το πιο γενικό στο πιο ειδικό.

Αξιολόγηση ερωτηματολογίου

Όσον αφορά την **υποκειμενική αξιολόγηση της εμπειρίας**, ζητήθηκε από τα παιδιά να συμπληρώσουν ένα ερωτηματολόγιο που περιέχει τις υποκειμενικές δηλώσεις που φαίνονται στο Σχήμα 2.10. Κάθε δήλωση συνοδευόταν από μια κλίμακα τύπου Likert 5 βαθμών με την ένδειξη «διαφωνώ» έως «συμφωνώ», χρησιμοποιώντας χαμογελαστά πρόσωπα [Hall et al.; 2016].

Συμπεριλάβαμε επίσης δύο ερωτήσεις πολλαπλής επιλογής που ζητούσαν από τα παιδιά να αιτιολογήσουν ποια περίπτωση χρήσης τους άρεσε περισσότερο και γιατί, και ποια ικανότητα

αντίληψης των ρομπότ τα κάνει δημοφιλή στα παιδιά. Σε γενικές γραμμές, η πλειονότητα των παιδιών (12/31) δήλωσε ότι η αγαπημένη τους περίπτωση χρήσης ήταν η «Παντομίμα», λόγω των κινήσεων του ρομπότ. Όπως μπορούμε να δούμε στο Σχήμα 2.10, τα περισσότερα παιδιά (27/31) δήλωσαν ότι τους αρέσει να παίζουν με τα ρομπότ, ενώ 22 απολάμβαναν το παιχνίδι επειδή τα ρομπότ κατανοούσαν τόσο τις κινήσεις όσο και την ομιλία τους. Πολλά από τα παιδιά (21/31) βρήκαν επίσης την αλληλεπίδραση και τα παιχνίδια εύκολα και δήλωσαν πως δεν χρειάστηκαν βοήθεια (19/31). Επιπλέον, τα παιδιά έτειναν να συμφωνούν (20 θετικές απαντήσεις από τις 31) ότι τα ρομπότ συμπεριφέρονται σαν τους ανθρώπους. Επίσης, αναλύοντας τις απαντήσεις στο ερωτηματολόγιο, παρατηρήσαμε ότι τα μεγαλύτερα παιδιά δήλωσαν ότι δεν χρειάζονταν προηγούμενες γνώσεις για να παίξουν με τα ρομπότ, σε σύγκριση με τα μικρότερα παιδιά που δήλωσαν ότι χρειάζονταν.

Ομοίως, στο σενάριο όπου τα παιδιά φτιάχνουν σχήματα και πραγματοποιήθηκε στο χώρο του δημοτικού σχολείου, ζητήθηκε από 21 παιδιά ηλικίας 9-10 που συμμετείχαν, να εκφράσουν τη γνώμη τους για την αλληλεπίδραση. Οι ερωτήσεις παρουσιάζονται στον Πίνακα 2.4 και οι διαθέσιμες απαντήσεις ήταν μια κλίμακα Likert 3 βαθμών (Διαφωνώ - Ουδέτερο - Συμφωνώ). Ο Πίνακας παρουσιάζει επίσης τα αποτελέσματα του ερωτηματολογίου αφού χαρτογραφήθηκαν σε κλίμακα 0-2, με το 0 να είναι το πιο αρνητικό. Οι απαντήσεις τους δείχνουν ότι τα παιδιά ήταν ευχαριστημένα με την αλληλεπίδραση (1,81 Mean Opinion Score (MOS) για το αν θα ήθελαν να παίξουν ξανά με το ρομπότ και την άνεση της αλληλεπίδρασης). Ωστόσο, είναι σαφές ότι η επίβλεψη του ρομπότ για την εργασία συναρμολόγησης έχει περιθώρια βελτιώσεων, καθώς αν και οι οδηγίες του ρομπότ ήταν πολύ σαφείς (1,95 MOS), τα παιδιά φάνηκαν ουδέτερα στο αν το ρομπότ ήταν χρήσιμο στη διαδικασία (1,10 MOS) ή έκανε πολλά λάθη (0,95 MOS).

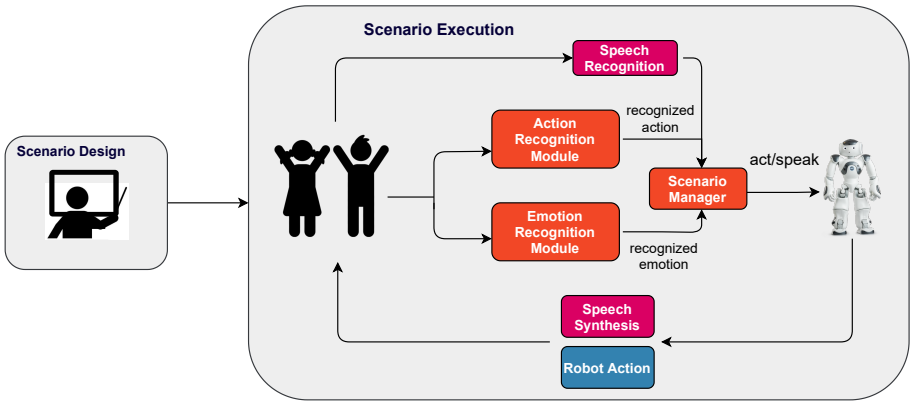
Ερώτηση	Mean Opinion Score
Ένιωθες άνετα να παίζεις με το ρομπότ;	1.81
Θα ήθελες να ξαναπαίξεις με το ρομπότ;	1.81
Σε βοηθούσε όσο παίζατε;	1.10
Το ρομπότ έκανε πολλά λάθη;	0.95
Ήταν οι οδηγίες του ξεκάθαρες;	1.95

Πίνακας 2.4: Οι ερωτήσεις και τα αποτελέσματα του ερωτηματολογίου που δόθηκε στα παιδιά μετά το παιχνίδι «Φτιάχνω σχήματα». Οι διαθέσιμες απαντήσεις ήταν μια κλίμακα Likert 3 βαθμών (Διαφωνώ - Είμαι ουδέτερος - Συμφωνώ) και αντιστοιχήθηκαν σε μία κλίμακα 0-2 απ' όπου προέκυψε το Mean Opinion Score, δηλαδή η μέση βαθμολογία των απαντήσεων.

Γενικά, η *αξιολόγηση της εμπειρίας των χρηστών* κατά τη διάρκεια της αλληλεπίδρασης με το ρομποτικό σύστημα πολλαπλών ρομπότ, πολλαπλών εργασιών και πολλαπλών αισθητήρων *παρείχε ενθαρρυντικά αποτελέσματα.* Διαπιστώθηκε ότι το σύστημα είναι τεχνικά ικανό να φιλοξενήσει μια ολοκληρωμένη εμπειρία CRI, με κάποια παρέμβαση ενηλίκου που απαιτείται σε ορισμένες περιπτώσεις και κυρίως για την ομαδική εργασία. Φυσικά, υπάρχουν περιθώρια βελτίωσης μιας και πολλά παιδιά δήλωσαν ότι τα ρομπότ έκαναν συχνά λάθη (18/31).

2.4 TeachBot: Ευφυές Σύστημα Αλληλεπίδρασης Παιδιού - Ρομπότ για την Σχεδίαση και την Εκτέλεση Εκπαιδευτικών - Ψυχαγωγικών Σεναρίων με έμφαση στην Οπτική Πληροφορία

Το σύστημα που παρουσιάζουμε στην ενότητα αυτή αποτελεί ένα ευφυές σύστημα αλληλεπίδρασης παιδιού-ρομπότ για την σχεδίαση και την εκτέλεση εκπαιδευτικών σεναρίων με έμ-



Σχήμα 2.11: Η δομή του ευφυούς συστήματος αλληλεπίδρασης παιδιών-ρομπότ TeachBot. Με πορτοκαλί χρώμα παρουσιάζονται οι μονάδες που μελετήσαμε και αναπτύξαμε.

φαση στην οπτική πληροφορία. Το σύστημα αυτό, που θα ονομάζουμε **TeachBot**, αξιοποιεί και επεκτείνει ερευνητικά τις υπάρχουσες τεχνολογίες στην αναγνώριση δράσεων και κινήσεων, αλλά και στην αποκωδικοποίηση της συναισθηματικής κατάστασης του παιδιού μέσω επεξεργασίας οπτικής πληροφορίας, έτσι ώστε να λειτουργούν εύρωστα σε πραγματικές συνθήκες εκπαίδευσης. Οι τεχνολογίες αυτές, σε συνδυασμό με υπάρχουσα τεχνολογία αναγνώρισης και σύνθεσης φωνής ενσωματώνονται σε ένα συνολικό σύστημα σχεδίασης εκπαιδευτικών σεναρίων, έτσι ώστε το τελικό αποτέλεσμα να είναι ένα ολοκληρωμένο και εύχρηστο εργαλείο που θα είναι αρωγός στην εκπαιδευτική διαδικασία.

Η αλληλεπίδραση των παιδιών με το ρομπότ γίνεται ευχάριστα και φυσικά καθώς το ρομπότ έχει τη δυνατότητα να αντιλαμβάνεται πολλαπλούς διαύλους επικοινωνίας: την ομιλία, την κίνηση αλλά και τη συναισθηματική κατάσταση του παιδιού. Παράλληλα το σύστημα θα παρέχει στον εκπαιδευτικό ένα εργαλείο για να προσαρμόζει τα υπάρχοντα σενάρια, αλλά και να σχεδιάζει καινούργια, ανάλογα με τους εκπαιδευτικούς σκοπούς που θέλει να εξυπηρετήσει, σε αντίθεση με τωρινά συστήματα ψυχαγωγικής εκπαίδευσης που απαιτούν από τον εκπαιδευτικό να προσαρμοστεί στα σενάρια και στους περιορισμούς τους. Θα πρέπει να τονιστεί ότι το σύστημα δεν έχει ως στόχο την αντικατάσταση του εκπαιδευτικού, αλλά την παροχή ενός συμπληρωματικού εργαλείου, το οποίο θα τον βοηθήσει και θα τον αφήσει να χρησιμοποιήσει τη δημιουργικότητά του κατά την διδασκαλία. Έτσι, ο στόχος του συνολικού συστήματος θα είναι να μπορεί να χρησιμοποιηθεί σε σχολεία, τα οποία αποτελούν ελεύθερα και όχι εργαστηριακά - περιορισμένα περιβάλλοντα.

Ακολουθώντας τις παραπάνω κατευθύνσεις, κατά την ανάπτυξη του συστήματός μας, αντιμετωπίζουμε πολλά ερευνητικά ερωτήματα.

- Πώς μπορούν να αξιοποιηθούν οι τεράστιες ευκαιρίες που προσφέρουν οι βαθιές αρχιτεκτονικές στο οπτικό πεδίο του υπολογιστή προκειμένου να δημιουργηθεί ένα σύστημα οπτικής αντίληψης για ρομποτικές εφαρμογές ειδικά για παιδιά;

- Είναι δυνατόν να αξιοποιήσουμε αυτές τις αρχιτεκτονικές σε βάθος για να δημιουργήσουμε ένα σύστημα CRI με ικανότητες αναγνώρισης δράσεων και συναισθημάτων και μια κατάλληλη αντιστάθμιση μεταξύ της υπολογιστικής αποτελεσματικότητας και της απόδοσης;

- Μπορεί η επαυξητική μάθηση (Incremental Learning - IL) να χρησιμοποιηθεί για να επιτρέψει στο σύστημα αντίληψης να αναγνωρίζει νέες ενέργειες χωρίς να ξεχνά τις παλαιότερες, και ποια κατηγορία μεθόδων IL αποδίδει καλύτερα;

Από όσο γνωρίζουμε, δεν υπάρχει κάποια προηγούμενη εργασία που να εξετάζει την επαυξητική μάθηση για αναγνώριση ενεργειών σε αλληλεπίδραση παιδιών και ρομπότ και να έχει να αντιμετωπιστεί το γεγονός ότι οι νέες κλάσεις πρέπει να αναγνωρίζονται από το σύστημα όταν προστίθενται νέα σενάρια ψυχαγωγίας. Εκτός από αυτή τη βασική καινοτομία της δουλειάς μας, εξετάζουμε επίσης αρκετές παραμέτρους του συστήματος, προκειμένου να εξισορροπήσουμε την απόδοση και την υπολογιστική αποτελεσματικότητα, και προτείνουμε ένα συνδυασμένο σύστημα οπτικής αντίληψης για αναγνώριση δράσης και συναισθημάτων στο πλαίσιο του CRI.

Σε σχέση με το ChildBot, που παρουσιάστηκε στην προηγούμενη ενότητα, το TeachBot είναι ένα πιο ελαφρύ, φορητό σύστημα με τεχνολογίες που επιτρέπουν την εύκολη προσαρμογή του σε νέα εκπαιδευτικά σενάρια και απαιτούν λίγες γνώσεις από τον χειριστή για την προσθήκη και εκτέλεση αυτών. Εν ολίγοις, οι κύριες συνεισφορές αυτής της εργασίας είναι οι εξής:

- Ενσωματώνει ισχυρές αρχιτεκτονικές που αναπτύξαμε και βασίζονται σε βαθιά νευρωνικά δίκτυα για την αντίληψη των ενεργειών των παιδιών και την αποκωδικοποίηση της συναισθηματικής τους κατάστασης μέσω οπτικών πληροφοριών, οι οποίες αξιολογούνται σε δύο βάσεις δεδομένων παιδιών, και αποδεικνύεται ότι ξεπερνούν τις state-of-the-art μεθόδους και έχουν χαμηλό υπολογιστικό κόστος.
- Επεκτείνει την αρχιτεκτονική αναγνώρισης δράσεων με μια μέθοδο επαυξητικής μάθησης που επιτρέπει την εύκολη προσθήκη νέων δράσεων. Με αυτή την επέκταση, το προτεινόμενο σύστημα αντίληψης δίνει τη δυνατότητα σε έναν μη τεχνικά εξειδικευμένο χρήστη να επεκτείνει και να προσαρμόσει τις δράσεις αναγνώρισης με βάση τις ανάγκες που εξυπηρετεί η αλληλεπίδραση.
- Ενσωματώνει υπάρχουσες τεχνολογίες αναγνώρισης και σύνθεσης ομιλίας για να διευκολύνει περαιτέρω την ευχάριστη και φυσική αλληλεπίδραση μεταξύ παιδιών και ρομπότ και μέσω του τρόπου ομιλίας.
- Παρέχει μια νέα και φιλική προς το χρήστη γραφική διεπαφή χρήστη, σχεδιασμένη με γνώμονα τους εκπαιδευτικούς, επιτρέποντάς τους να προσαρμόσουν τα υπάρχοντα σενάρια CRI ή να δημιουργήσουν νέα σύμφωνα με τις απαιτήσεις του προγράμματος σπουδών.

Μια επισκόπηση του προτεινόμενου συστήματος απεικονίζεται στο Σχήμα 2.11.

2.4.1 Περιγραφή Συστήματος

Το προτεινόμενο σύστημα, εκτός από το υπολογιστικό σύστημα, αξιοποιεί μόλις ένα ρομπότ NAO και μια κάμερα με μικρόφωνα, π.χ. ένας αισθητήρας Kinect, ώστε να είναι φορητό και ελαφρύ και να μην έχει μεγάλες υπολογιστικές απαιτήσεις. Στο TeachBot διατηρείται η αρθρωτή αρχιτεκτονική που υπήρχε στο ChildBot, καθώς κρίθηκε αποδοτική ως προς τη δυνατότητα ελέγχου των μονάδων αντίληψης και των ρομπότ. Ως προς τις λειτουργίες του, το TeachBot περιλαμβάνει τρεις κύριες ενότητες που αναπτύχθηκαν ύστερα από μελέτη και δύο υπάρχουσες τεχνολογίες που παρέχουν κάποιες απαραίτητες λειτουργίες.

Οι δύο πρώτες βασικές ενότητες του συστήματός μας αφορούν τις βασικές δυνατότητες αντίληψης του ρομπότ: 1) Η μονάδα αναγνώρισης δράσεων, η οποία, όπως υποδηλώνει το όνομα, έχει την ευθύνη να αναγνωρίζει τις δράσεις του παιδιού και 2) την μονάδα αναγνώρισης συναισθηματικής κατάστασης, η οποία αποκωδικοποιεί τα συναισθήματα του παιδιού. Η τρίτη μονάδα είναι ο διαχειριστής σεναρίων που διευκολύνει τον εκπαιδευτικό να σχεδιάσει νέα σενάρια ψυχαγωγίας χρησιμοποιώντας διαγράμματα ροής. Παράλληλα, χρησιμοποιούνται επίσης ενότητες αναγνώρισης και σύνθεσης ομιλίας, βασισμένες σε υπάρχουσες λύσεις.

Μονάδες Αντίληψης

Στην παρούσα ενότητα περιγράφουμε συνοπτικά τις μονάδες αντίληψης του TeachBot. Οι μέθοδοι και τα πειραματικά αποτελέσματα της μονάδας αναγνώρισης δράσεων που αναπτύχθηκαν στο πλαίσιο της παρούσας διατριβής περιγράφονται αναλυτικά στο Κεφάλαιο 3.

Μονάδα αναγνώρισης δράσεων

Η μονάδα αυτή βασίζεται στην τεχνολογία του Temporal Segment Network (TSN) [Wang et al.; 2016a], που αρχικά εισήχθη για αναγνώριση ενεργειών μεγάλης κλίμακας. Ακολουθώντας την παραπάνω λογική, λαμβάνονται τυχαία δείγματα K διαφορετικών τμημάτων από το βίντεο εισόδου, καθένα από τα οποία αποτελείται από N διαδοχικά καρέ. Αυτή η τυχαία δειγματοληψία βοηθά στη γενίκευση και μειώνει το υπολογιστικό κόστος και τις περιττές πληροφορίες που υπάρχουν σε διαδοχικά καρέ βίντεο. Έτσι η μονάδα μπορεί να είναι αποδοτική χωρίς μεγάλες υπολογιστικές απαιτήσεις. Για μεγαλύτερη αξιοπιστία της μονάδας, χρησιμοποιούνται δύο διαφορετικές ροές, μία χωρική που λαμβάνει ως είσοδο RGB καρέ βίντεο και μία χρονική που λαμβάνει ως είσοδο την οπτική ροή που προέρχεται από το βίντεο.

Λαμβάνοντας υπόψη τη συνεχή ανάγκη εισαγωγής νέων τάξεων δράσης κατά τη διάρκεια νέων σεναρίων ψυχαγωγίας, η μονάδα αναγνώρισης ενεργειών επεκτείνεται με τη χρήση μεθόδων επαυξητικής μάθησης. Αναλυτικά οι μέθοδοι και οι αρχιτεκτονικές της μονάδας αναγνώρισης δράσεων αναπτύσσονται στο Κεφάλαιο 3.

Μονάδα αναγνώρισης συναισθηματικής κατάστασης

Για συνέπεια και ευκολία, η μονάδα αναγνώρισης συναισθημάτων χρησιμοποιεί παρόμοια αρχιτεκτονική με αυτήν της αναγνώρισης ενεργειών. Τα TSN έχουν ήδη αποδειχθεί ότι επιτυγχάνουν αποτελέσματα αιχμής στην αναγνώριση συναισθημάτων βίντεο [Filntisis et al.; 2020a]. Κάτω από αυτό το πλαίσιο, οι στατικές πληροφορίες αξιοποιούνται σε διαφορετικά πλαίσια για την αναγνώριση της έκφρασης ενός παιδιού και συνδυάζονται με τη δυναμική κίνησης του χρησιμοποιώντας την οπτική ροή ως είσοδο.

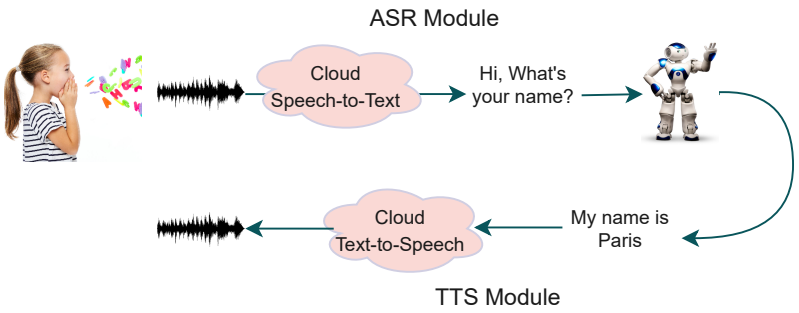
Μονάδες ομιλίας

Για να δημιουργήσουμε συνθήκες που θα θυμίζουν φυσική και ανεμπόδιστη αλληλεπίδραση, το ρομπότ πρέπει ακόμα να κατανοήσει την ομιλία του παιδιού καθώς και να έχει την ικανότητα να μιλήσει στο παιδί. Για το σκοπό αυτό χρησιμοποιούνται υπάρχουσες λύσεις μετατροπής κειμένου σε ομιλία (TTS) [Google Text-to-Speech; 2021] και αυτόματης αναγνώρισης ομιλίας (ASR) [Google Speech-to-Text; 2021] που βασίζονται σε τεχνολογίες νέφους (cloud-based). Η ενοποίηση των δύο μονάδων στο κύριο σύστημα φαίνεται στο Σχ. 2.12. Κατά τη διάρκεια της αλληλεπίδρασης, η ομιλία που καταγράφεται από τα μικρόφωνα μεταδίδεται συνεχώς σε μια υπηρεσία cloud, η οποία επιστρέφει τα αποτελέσματα αναγνώρισης. Στη συνέχεια, το κείμενο τροφοδοτείται στον διαχειριστή σεναρίων (που περιγράφεται στην Ενότητα 2.4.1), ο οποίος αποφασίζει εάν το ρομπότ πρέπει να απαντήσει κάτι πίσω. Το κείμενο που πρόκειται να συντεθεί στη συνέχεια αποστέλλεται σε μια υπηρεσία cloud TTS, η οποία συνθέτει την ομιλία.

Μονάδα Διαχείρισης Σεναρίου και Ενοποίηση Συστήματος

Μοντελοποίηση σεναρίου

Για τη διαχείριση της ροής του διαλόγου, υιοθετούμε το παράδειγμα "Sense, Think, Act" που έχουμε αναφέρει και στο ChildBot. Σύμφωνα με αυτό το παράδειγμα, το ρομποτικό σύστημα



Σχήμα 2.12: Οι μονάδες αναγνώρισης και σύνθεσης ομιλίας του συστήματος.

χρησιμοποιεί πρώτα τις αντιληπτικές του ικανότητες (*Sense*) και στη συνέχεια αποφασίζει για την επόμενη πορεία δράσης (*Think*) και τελικά εκτελεί την επιλεγμένη ενέργεια (*Act*). Στο προτεινόμενο πλαίσιο, μοντελοποιούμε το παράδειγμα χρησιμοποιώντας γεγονότα [Skantze and Al Moubayed; 2012]. Αυτά χωρίζονται σε δύο κατηγορίες: συμβάντα *Action*, που δίνουν εντολή στο ρομπότ να κάνει κάτι, και συμβάντα *Sense*, που ενεργοποιούνται όταν οι μονάδες αντίληψης αντιλαμβάνονται κάτι (δηλαδή, μια συγκεκριμένη ενέργεια ή συναίσθημα) Η ροή της αλληλεπίδρασης μοντελοποιείται χρησιμοποιώντας Harel statecharts [Harel; 1987]. Κάθε κατάσταση στο γράφημα έχει κρυφές παραμέτρους που ελέγχουν τη ροή, μαζί με τα συμβάντα που λαμβάνονται.

Σχεδίαση προσαρμοσμένου σεναρίου

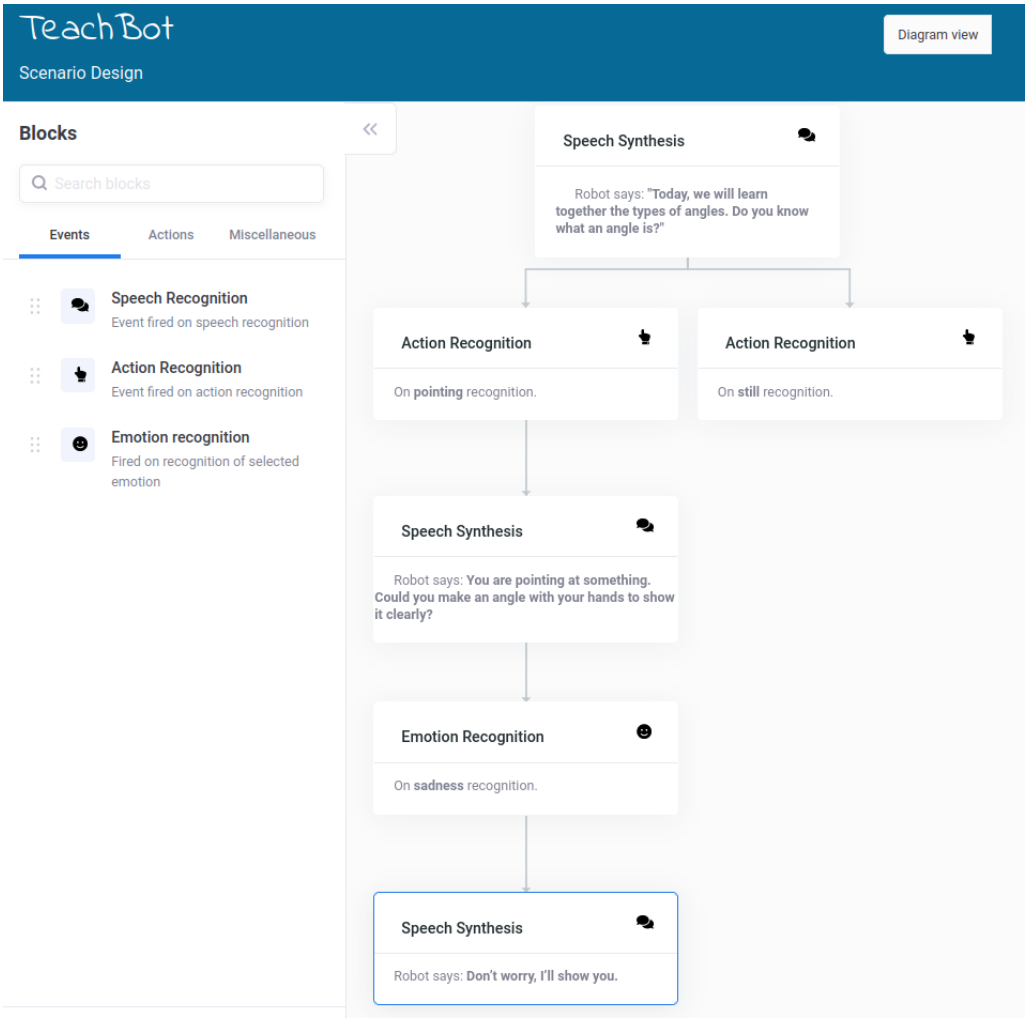
Μια σημαντική καινοτομία του παρόντος συστήματος εκπαιδευτικών σεναρίων είναι η μελέτη που έγινε ώστε να προσφέρει στους εκπαιδευτικούς τη δυνατότητα να σχεδιάζουν τα δικά τους σενάρια, χρησιμοποιώντας ένα ευέλικτο, φιλικό προς το χρήστη και αισθητικά ελκυστικό γραφικό περιβάλλον με μεταφορά και απόθεση. Μέσω αυτής της διεπαφής, ο δάσκαλος μπορεί να δημιουργήσει τα σενάρια που επιθυμεί ακολουθώντας τις προαναφερθείσες αρχές της Μοντελοποίησης Σεναρίων. Στη συνέχεια, το σενάριο μεταγλωττίζεται σε ένα διάγραμμα κατάστασης Harel που μοντελοποιεί τη ροή και την αναπτύσσει στο ρομπότ για να την εκτελέσει. Ένα παράδειγμα δημιουργίας γραφικών σεναρίων μπορεί να δει κανείς στο Σχήμα 2.13.

Ενοποίηση του συστήματος

Οι μονάδες συστήματος επικοινωνούν μέσω ενός broker, που υλοποιείται χρησιμοποιώντας το πρωτόκολλο TCP/IP. Πιο συγκεκριμένα, κατά τη διάρκεια της αλληλεπίδρασης, οι μονάδες αντίληψης στέλνουν συμβάντα *Sense* στον broker, ο οποίος με τη σειρά του μεταδίδει τα συμβάντα στον διαχειριστή σεναρίου. Επιπλέον, όταν ο διαχειριστής σεναρίου απαιτεί από το ρομπότ να κάνει κάτι, στέλνει ένα συμβάν *Act* στην αντίστοιχη ενότητα: τη μονάδα σύνθεσης ομιλίας ή το ίδιο το ρομπότ στην περίπτωση μιας ενέργειας. Όλες οι μονάδες αντίληψης και η διαχείριση σεναρίων αναπτύσσονται σε μια μηχανή Linux με κάρτα γραφικών RTX 2080.

2.4.2 Ενδεικτικό Σενάριο Χρήσης

Για τη βοήθεια του εκπαιδευτικού στην υλοποίηση σεναρίων καθώς και την καλύτερη αξιοποίηση του συστήματός μας κρίνεται απαραίτητο να συμπεριληφθούν βασικά/πρότυπα σενάρια που θα μπορούσαν να χρησιμοποιηθούν. Παρακάτω προτείνεται αναλυτικά ένα πρότυπο σενάριο επισημαίνοντας παράλληλα την αλληλουχία των δράσεων που κάνει το παιδί και ανα-



Σχήμα 2.13: Το γραφικό περιβάλλον χρήστη για τη δημιουργία και προσαρμογή νέων σεναρίων.

γνωρίζεται από το ρομπότ καθώς και τις δράσεις του ρομπότ.

ρομπότ: Σήμερα θα μάθουμε μαζί τα είδη των γωνιών! Αλήθεια, ξέρεις τι είναι μια γωνία; [*Δράση Ομιλίας*]

παιδί: [*Διστάζει και στη συνέχεια δείχνει προς μια γωνία του δωματίου*]

ρομπότ: Αναγνωρίζει την δεικτική χειρονομία [*Αναγνώριση Δράσης*] Βλέπω πως κάτι μου δείχνεις. Μπορείς όμως να μου φτιάξεις μια γωνία χρησιμοποιώντας τα χέρια σου για να καταλάβω καλύτερα; [*Δράση Ομιλίας*]

παιδί: Δεν ξέρω. [*Στενοχωρημένο*].

ρομπότ: [*Αναγνώριση Ομιλίας και Συναισθήματος*] Δεν πειράζει. Δεν είναι δύσκολο! Θα σου δείξω εγώ. [*Δράση Ομιλίας*] Σχηματίζει μια γωνία με τα χέρια του [*Δράση Κίνησης*] Τώρα θα μου κάνεις και εσύ μία; [*Δράση Ομιλίας*]

παιδί: "Ναι!" και σχηματίζει μια γωνία με τα χέρια του

ρομπότ: [*Αναγνώριση Ομιλίας και Δράσης*] Πολύ ωραία. Τώρα λοιπόν θα ήθελα να μου μάθετε τα είδη των γωνιών. [*Δράση Ομιλίας*]

δάσκαλος: Τα παιδιά δεν τα ξέρουν ακόμα. Θα τα μάθουν σήμερα και θα στα μάθουν και εσένα

Σχήμα 2.14: Παράδειγμα εικόνων από τη βάση δράσεων του BabyRobot (πρώτη γραμμή) και από την EmoReact βάση (δεύτερη γραμμή).

ρομπότ: [*Αναγνώριση Ομιλίας*] Εντάξει. Θα περιμένω λοιπόν. [*Δράση Ομιλίας*] Στο τέλος του μαθήματος τα παιδιά θα σχηματίσουν μπροστά στο ρομπότ διάφορα είδη γωνιών, είτε στατικά είτε δυναμικά (π.χ. σχηματίζουν τα παιδιά οξείες γωνίες δείχνοντας όλο το εύρος). Η καταγραφή αυτή θα χρησιμοποιηθεί για την Ενότητα Εκμάθησης του Συστήματος. Στο επόμενο μάθημα το ρομπότ καλεί ένα ένα τα παιδιά να του δείξουν μία γωνία και αυτό πρέπει να το αναγνωρίσει.

ρομπότ: Σχημάτισε μια γωνία για να βρω το είδος της" [*Δράση Ομιλίας*]
παιδί: Σχηματίζει με τα χέρια του μια ορθή γωνία.
ρομπότ: Αναγνωρίζει τη δράση. [*Αναγνώριση Δράσης*] Είναι μία ορθή γωνία γιατί τα χέρια σου είναι κάθετα μεταξύ τους [*Δράση Ομιλίας*]
παιδί: Σωστά.

Έτσι αντίστοιχα διαμορφώνονται και άλλα πρότυπα σενάρια με σκοπό να ενσωματωθούν στην τελική πλατφόρμα. Επίσης, το ρομπότ θα είναι στη διάθεση των μαθητών ώστε ανεξάρτητα από το ποια αλληλεπίδραση-σενάριο μαθαίνουν εκείνη την περίοδο, να μπορούν να αξιοποιούν άλλες διαδράσεις που είχαν παλιότερα με το ρομπότ. Έτσι ουσιαστικά θα μπορούν να παίζουν κάνοντας επανάληψη στις γνώσεις τους, είτε προκαλώντας το ρομπότ να αναγνωρίσει τι κάνουν εκείνη τη στιγμή είτε απαντώντας τις ερωτήσεις-προκλήσεις του ρομπότ.

2.5 Συμπεράσματα Κεφαλαίου

Στο Κεφάλαιο αυτό αρχικά μελετήσαμε την υπάρχουσα βιβλιογραφία σχετικά με τα συστήματα που εξυπηρετούν αλληλεπιδράσεις παιδιών και ρομπότ εντοπίζοντας παράλληλα και τις ανάγκες που προκύπτουν όταν χρησιμοποιούνται εκτός ερευνητικών χώρων, όπως στις σχολικές αίθουσες. Στη συνέχεια περιγράψαμε τα αυτόματα συστήματα αντίληψης που έχουν αναπτυχθεί στο πλαίσιο μεγάλων ερευνητικών έργων, όπως το έργο BabyRobot κατά τη διάρκεια του οποίου αναπτύξαμε το ChildBot, και άλλων εργασιών και συνδυάζουν την αναγνώριση πολλών τροπικοτήτων κατά τη διάρκεια αλληλεπιδράσεων παιδιών και ρομπότ.

Ακόμα παρουσιάσαμε τα δύο συστήματα αλληλεπίδρασης παιδιών και ρομπότ που αναπτύξαμε, το ChildBot και το TeachBot. Πιο συγκεκριμένα αναφερθήκαμε στις μονάδες αναγνώρισης που περιλαμβάνουν, στις αρχιτεκτονικές τους, στα πιθανά σενάρια χρήσης καθώς και

στα πειράματα που πραγματοποιήσαμε για να αξιολογήσουμε τη χρήση τους. Το ChildBot είδαμε πως αποτελεί ένα ολοκληρωμένο σύστημα για HRI που διαθέτει μονάδες αντίληψης για πολυτροπική κατανόηση σκηνής και αξιοποιεί πολλαπλά ρομπότ και αισθητήρες. Το σύστημα αξιολογήθηκε ως προς την εμπειρία των μικρών χρηστών και έδειξε ενθαρρυντικά αποτελέσματα, ενώ στο επόμενο Κεφάλαιο παρουσιάζεται αναλυτικά η αξιολόγησή του ως προς τη μονάδα αναγνώρισης δράσης.

Από την άλλη, το σύστημα TeachBot που αναπτύξαμε είναι ένα ελαφρύ σύστημα που αποσκοπεί να διευκολύνει την ενσωμάτωση των αλληλεπιδράσεων παιδιών και ρομπότ σε αίθουσες εκτός εργαστηρίων. Για το σκοπό αυτό αξιοποιεί μόνο έναν αισθητήρα και ένα ρομπότ και ενσωματώνει ισχυρές μονάδες αντίληψης δράσεων και συναισθημάτων και υπάρχουσες τεχνολογίες αναγνώρισης και σύνθεσης ομιλίας. Σημαντική συνεισφορά του TeachBot αποτελεί η διευρυμένη εκδοχή του συστήματος αναγνώρισης δράσης με μια μέθοδο επαυξητική μάθησης, πράγμα που του επιτρέπει να μαθαίνει νέες δράσεις χωρίς να ξεχνά τις παλιές. Στο επόμενο Κεφάλαιο περιγράφεται αναλυτικά η πειραματική ανάλυση των μεθόδων.

3

Αυτόματη Αναγνώριση Δράσεων σε Αλληλεπιδράσεις Παιδιών-Ρομπότ: Μέθοδοι & Πειράματα

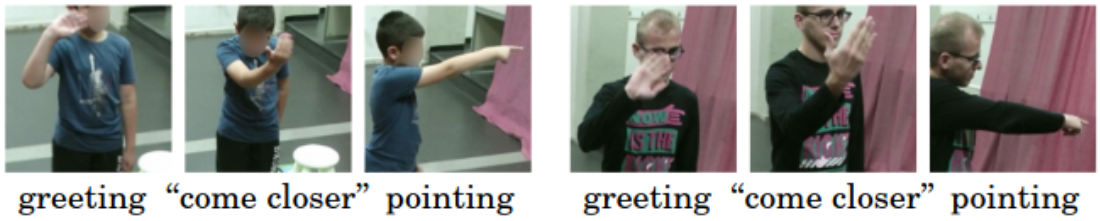
Στο κεφάλαιο αυτό, παρουσιάζουμε τις προτεινόμενες μεθόδους και τις πειραματικές μελέτες που πραγματοποιήσαμε σχετικά με την αυτόματη αναγνώριση δράσεων για την αξιοποίηση σε αλληλεπιδράσεις παιδιών με ρομποτικούς πράκτορες. Ο κύριος στόχος μας ήταν να δημιουργήσουμε ένα ισχυρό σύστημα αντίληψης για την αντιμετώπιση διαφορετικών εργασιών, όπως οι γενικές κινήσεις του σώματος που εκτελούνται από παιδιά καθώς και οι χειρονομίες. Τα δεδομένα παιδιών είναι εξαιρετικά δυσεύρετα κ ο όγκος τους μικρός και συνεπώς τα συστήματα που αναπτύξαμε έπρεπε να λαμβάνουν υπόψιν αυτό τον παράγοντα και παράλληλα να επιτυγχάνουν καλή ακρίβεια αναγνώρισης για να είναι αξιοποιήσιμα σε αλληλεπιδράσεις παιδιών και ρομπότ.

Οι παρακάτω μελέτες για την ανάπτυξη συστημάτων αυτόματης αναγνώρισης δράσεων έγιναν στα πλαίσια των συστημάτων αλληλεπίδρασης παιδιών και ρομπότ *ChildBot* και *TeachBot*. Συνολικά θα μπορούσαμε να συνοψίσουμε τα εξής σημαντικά σημεία για το σύστημα αυτόματης αναγνώρισης δράσεων που αναπτύξαμε:

- έχει εκπαιδευτεί και αξιολογηθεί για δυο διαφορετικά σύνολα δράσεων δημιουργώντας δύο διαφορετικές εκδόσεις του συστήματος που λειτουργούν ανεξάρτητα. Η μία έκδοση αφορά την αναγνώριση χειρονομιών και η άλλη αφορά πιο γενικευμένες κινήσεις του σώματος, όπως οι κινήσεις παντομίμας.
- έχει γίνει μελέτη για την αποτελεσματικότητα της μονάδας όταν χρησιμοποιούνται δεδομένων ενηλίκων έναντι δεδομένων παιδιών.
- μελετήθηκε η ευρωστία του συστήματος από διαφορετικές θέσης καταγραφής καθώς και η αποτελεσματικότητα της σύμμειξης της πληροφορίας που καταγράφεται σύγχρονα από πολλαπλούς αισθητήρες.
- επεκτάθηκε σε ένα σύστημα επαυξητικής μάθησης με στόχο την ευκολότερη αξιοποίησή της σε μη εργαστηριακά περιβάλλοντα.

3.1 Πειραματικά δεδομένα

Στην παρούσα ενότητα παρουσιάζουμε τα δεδομένα που έχουν χρησιμοποιηθεί στο Κεφάλαιο 3 για την εκπαίδευση και την αξιολόγηση του συστήματος αναγνώρισης δράσεων. Τα



Σχήμα 3.1: Παραδείγματα των δεδομένων χειρονομιών από ένα παιδί (αριστερά) και έναν ενήλικα (δεξιά)

δεδομένα ως προς το περιεχόμενο μπορούν να χωριστούν σε δύο κατηγορίες: α) δεδομένα χειρονομιών, β) δεδομένα παντομίμας, ενώ ως προς την ηλικία των συμμετεχόντων που κάνουν τις κινήσεις διακρίνονται σε: α) δεδομένα παιδιών, β) δεδομένα ενηλίκων.

Δεδομένα χειρονομιών: Τα δεδομένα αυτά αποτελούνται από 8 χειρονομίες που κύρια έχουν επικοινωνιακό χαρακτήρα και συχνά συνοδεύουν το λόγο μας. Αυτές είναι: α) δείχνω ένα αντικείμενο (pointing), β) καλώ κάποιον να έρθει κοντά (come closer), γ) κάνω νόημα σε κάποιον να κάτσει (sit down), δ) κάνω νόημα σε κάποιον να σταματήσει (stop), ε) σχεδιάζω έναν κύκλο στον αέρα (cicle), στ) χαιρετάω (greeding), ζ)κάνω ένα καταφατικό νεύμα (agreement), η) κάνω μία τυχαία χειρονομία χωρίς συγκεκριμένο περιεχόμενο (background gesture). Η περιγραφή αυτών των κινήσεων δινόταν περιγραφικά όπως και παραπάνω και το κάθε άτομο επέλεγε να κατά το δοκούν για το πως θα τις εκτελέσει. Στο Σχήμα 3.1 παρουσιάζονται παραδείγματα - στιγμιότυπα από την εκτέλεση τριών χειρονομιών από ένα παιδί και έναν ενήλικα.

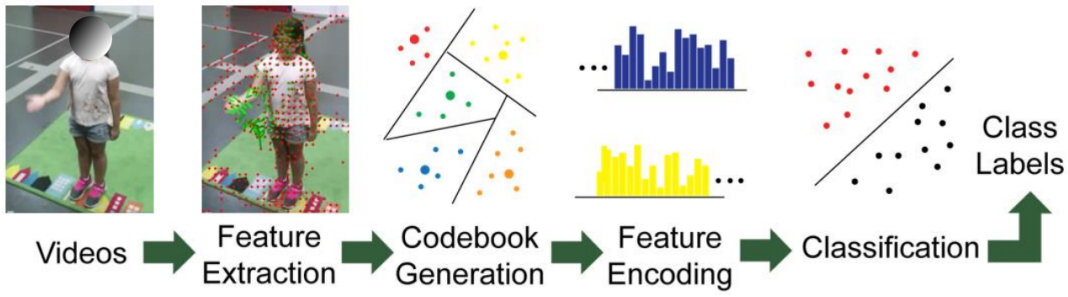
Δεδομένα κινήσεων παντομίμας: Τα δεδομένα αυτά αποτελούνται από 13 γενικές κινήσεις του σώματος που εντάσσονται σε ένα παιχνίδι θα παντομίμας και αναπαριστούν μια εργασία από τις εξής: α) κολυμπώ, β) οδηγώ ένα λεωφορείο, γ) χτυπάω με ένα σφυρί, δ) καθαρίζω ένα παράθυρο, ε) κάνω γυμναστική, στ) χορεύω , ζ) διαβάζω ένα βιβλίο, η) σκάβω, θ) παίζω κιθάρα, ι) σκουπίζω, ια) βάφω έναν τοίχο, ιβ) σιδερώνω, ιγ) εκτελώ μία τυχαία κίνηση χωρίς κάποιο συγκεκριμένο σκοπό. Ζητήθηκε από τους συμμετέχοντες να αποδώσουν τα παραπάνω νοήματα με όποιες κινήσεις επιθυμούν.

Τα δεδομένα που συλλέξαμε για τις παραπάνω δράσεις διακρίνονται σε δύο κατηγορίες:

1. δεδομένα ανάπτυξης (development data) όπου οι συμμετέχοντες (20 ενήλικες και 28 παιδιά) εκτελούν τις κινήσεις αφού τους ζητηθεί. Τα δεδομένα αυτά χρησιμοποιούνται τόσο για την εκπαίδευση όσο και για τον έλεγχο της αξιοπιστίας του συστήματος.
2. δεδομένα αλληλεπίδρασης (use-case related data) είναι τα δεδομένα που έχουν καταγραφεί από 31 παιδιά κατά τη διάρκεια μιας συνεχούς και ελεύθερης αλληλεπίδρασής τους

	Δεδομένα Ανάπτυξης		Δεδομένα Αλληλεπίδρασης
Πλήθος	28 Παιδιά	20 Ενήλικες	31 Παιδιά
Χειρονομίες	196	160	143
Παντομίμες	336	260	109

Πίνακας 3.1: Στατιστικά στοιχεία των δεδομένων δράσεων που χρησιμοποιούνται κατά την εκπαίδευση και αξιολόγηση του συστήματος αναγνώρισης δράσεων.



Σχήμα 3.2: Δομή του συστήματος αναγνώρισης δράσεων με αξιοποίηση χαρακτηριστικών πυκνών τροχιών.

με τρεις ρομποτικούς πράκτορες. Τα δεδομένα αυτά χρησιμοποιούνται μόνο για την αξιολόγηση του συστήματος.

Στον Πίνακα 3.1 συνοψίζεται το πλήθος των δεδομένων που αναφέραμε και χρησιμοποιούνται στη συνέχεια του κεφαλαίου.

3.2 Συστήματα αναγνώρισης μονής όψης

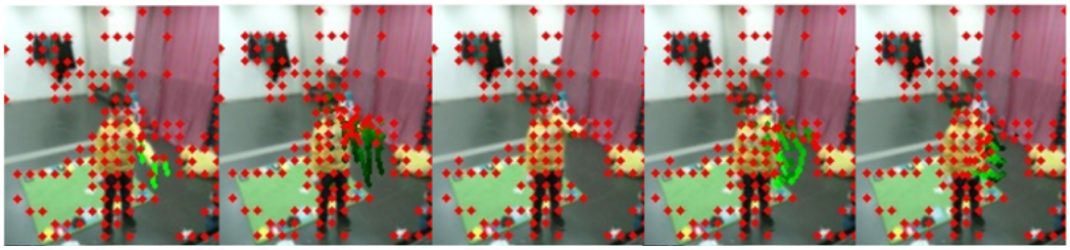
Προκειμένου να σχεδιάσουμε ένα αποτελεσματικό σύστημα αναγνώρισης ενεργειών, διερευνήσαμε τις πιθανές επιλογές σχετικά με τις οπτικές αναπαραστάσεις μιας ροής βίντεο. Κατά την ανάπτυξη των συστημάτων μονής όψης (δηλαδή συστημάτων που αξιοποιούν την οπτική πληροφορία από μια μόνο κάμερα) για την αναγνώριση δράσεων παιδιών εξετάσαμε και υλοποιήσαμε πολλές τεχνικές που βασίζονται τόσο σε κλασικές μεθόδους της όρασης υπολογιστών όσο και σε πιο σύγχρονες που αξιοποιούν νευρωνικά δίκτυα. Πιο συγκεκριμένα, λόγω της φύσης των δεδομένων (τα δεδομένα δράσεων παιδιών είναι δυσεύρετα και ο όγκος τους μικρός) αρχικά μελετήσαμε τεχνικές, όπως οι οι πυκνές τροχιές, που εξάγουν και αξιοποιούν handcrafted χαρακτηριστικά. Στη συνέχεια, μελετήσαμε την αναγνώριση δράσης με χρήση χαρακτηριστικών από ένα τρισδιάστατο συνελικτικό δίκτυο και τέλος αξιοποιήσαμε νευρωνικά δίκτυα που βασίζονται στην χρονική δειγματοληψία. Παρακάτω εξηγείται λεπτομερώς η ανάπτυξη και η λειτουργία των συστημάτων αναγνώρισης δράσης κατηγοριοποιημένα με βάση την κύρια τεχνολογία τους.

3.2.1 Μέθοδοι και Αρχιτεκτονικές

Χαρακτηριστικά πυκνών τροχιών (Dense trajectories features)

Για την ανάπτυξη του συστήματος αναγνώρισης δράσεων με κλασικές προσεγγίσεις αξιοποιήσαμε χαρακτηριστικά πυκνών τροχιών (Dense Trajectories- DT) σε συνδυασμό με τις δημοφιλείς μεθόδους κωδικοποίησης Bag-of-Visual-Words (BoVW) και Vector of Locally Aggregated Descriptors (VLAD). Τα βήματα της μεθόδου που αναλύουμε παρακάτω, συνοψίζονται στο Σχήμα 3.2

Η μέθοδος Πυκνών Τροχιών [Wang et al.; 2011] είναι ιδιαίτερα δημοφιλής λόγω της πολύ καλής απόδοσής της σε απαιτητικά σύνολα δεδομένων. Η κύρια ιδέα της στηρίζεται στην πυκνή δειγματοληψία σημείων ενδιαφέροντος από κάθε στιγμιότυπο (frame) ενός βίντεο και στην παρακολούθηση-ιχνηλάτησή τους στο χρόνο με χρήση οπτικής ροής. Η παρακολούθηση (tracking) πραγματοποιείται σε πολλαπλές χωρικές κλίμακες και οι τροχιές που προκύπτουν έχουν σταθερό μήκος L.



Σχήμα 3.3: Πυκνές τροχιές που προκύπτουν όταν το παιδί κάνει μια χειρονομία.

Έτσι, ο αλγόριθμος για κάθε σημείο ενδιαφέροντος με δείκτη n καταλήγει σε μια τροχιά $\mathbf{x}_n = (P_1, P_2, \ldots, P_L)$, που είναι μια ακολουθία σημείων P_1,..,P_L των επόμενων στιγμιοτύπων, ξεκινώντας από το στιγμιότυπο $t_b(\mathbf{x}_n)$ και τελειώνοντας στο $t_e(\mathbf{x}_n)$. Μετά την εξαγωγή τροχιάς, διαφορετικά οπτικά χαρακτηριστικά υπολογίζονται εντός των χωροχρονικών όγκων κάθε τροχιάς (Εικόνα 3.3). Τα χαρακτηριστικά που χρησιμοποιούμε είναι τα εξής:

- τον περιγραφητή τροχιάς (Trajectory) [Wang et al.; 2011] που κωδικοποιεί το σχήμα των τροχιών. Ο περιγραφητής τροχιάς είναι η ακολουθία των διανυσμάτων μετατόπισης μεταξύ διαδοχικών σημείων τροχιάς, $\Delta P_l = P_{l+1} - P_l$, κανονικοποιημένη με το άθροισμα του μέτρου των διανυσμάτων μετατόπισης κατά μήκος της τροχιά.

- τα ιστογράμματα προσανατολισμένης κλίσης (Histograms of Oriented Gradients - HOG) [Laptev et al.; 2008]. Έχουν τη δυνατότητα να μοντελοποιήσουν την τοπική στατική εμφάνιση από την κατεύθυνση και το μέτρο της κλίσης της φωτεινότητας της εικόνας.

- τα ιστογράμματα οπτικής ροής (Histograms of Optical Flow - HOF) [Laptev et al.; 2008]. Καταγράφουν την πληροφορία που περιλαμβάνει η κίνηση χρησιμοποιώντας την κατεύθυνση και το μέτρο της οπτικής ροής.

- ιστογράμματα συνόρων κίνησης (Motion Boundary Histograms - MBH) [Wang et al.; 2011] και στους δύο άξονες (MBHx, MBHy). Υπολογίζονται με χρήση της κλίσης της οπτικής ροής στον αντίστοιχο άξονα, κάθετο ή παράλληλο σε αυτή, ενώ το MBH αποτελεί τη συνένωσή τους και γι' αυτό είναι πιο εύρωστοι στην κίνηση της κάμερας.

Τα ιστογραφικά χαρακτηριστικά περιγράφουν τοπικά το σχήμα, την εμφάνιση και την κίνηση κατά μήκος κάθε τροχιάς.

Με τη χρήση ενός ανιχνευτή προσώπου θα μπορούσε να βελτιωθεί η αναπαράσταση της δράσης.Για το σκοπό αυτό χρησιμοποιούμε έναν απλό ανιχνευτή ατόμου που βασίζεται σε περιγραφητές HOG [Dalal and Triggs; 2005]. Έτσι, ο χρόνος εξαγωγής χαρακτηριστικών μπορεί να μειωθεί καθώς οι πυκνές τροχιές (DT) υπολογίζονται σε μια μικρή περιοχή της Full HD εικόνας που καταγράφεται.

Στη συνέχεια, τα χαρακτηριστικά που εξάγονται κωδικοποιούνται. Για κάθε περιγραφητή ξεχωριστά δημιουργούνται οπτικά λεξικά κατά τη φάση της εκπαίδευσης από ένα υποσύνολο τυχαία επιλεγμένων χαρακτηριστικών με χρήση του αλγορίθμου K-means. Το κεντροειδές κάθε συστάδας (cluster) αποτελεί μια οπτική λέξη και κάθε τροχιά ανατίθεται στην πλησιέστερη από τις K οπτικές λέξεις του λεξικού, βάση της Ευκλείδειας απόστασης.

Κατά τη χρήση της μεθόδου κωδικοποίησης BoVW λαμβάνουμε μια αραιή αναπαράσταση του βίντεο K-διαστάσεων, η οποία πρακτικά είναι το ιστόγραμμα των συχνοτήτων εμφάνισης των οπτικών λέξεων στο χωροχρονικό όγκο του βίντεο. Έτσι, τα βίντεο ταξινομούνται με βάση την BoVW αναπαράστασή τους, χρησιμοποιώντας μη γραμμικές μηχανές διανυσμάτων υποστήριξης (Support Vector Machines - SVM) με πυρήνα χ^2 [Wang et al.; 2009]. Επιπλέον, διαφορετικοί περιγραφητές συνδυάζονται, υπολογίζοντας τις αποστάσεις μεταξύ των αντίστοιχων

ιστογραμμάτων BoVW τους ως εξής:

$$K\left(\mathbf{h}_i, \mathbf{h}_j\right) = \exp\left(-\sum_m \frac{1}{A_m} L\left(\mathbf{h}_i^m, \mathbf{h}_j^m\right)\right), \tag{3.1}$$

όπου $\mathbf{h}_i^m$ υποδηλώνει την αναπαράσταση BoVW του m-ιοστού περιγραφητή του i-οστού βίντεο. Το A_m είναι η μέση τιμή των χ^2 αποστάσεων που ορίζεται ως

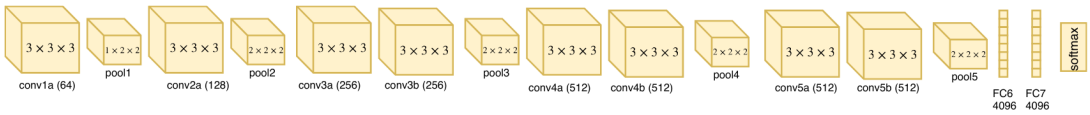
$$L(\mathbf{h}_i^m, \mathbf{h}_j^m) = -\frac{(\mathbf{h}_i^m - \mathbf{h}_j^m)^2}{2(\mathbf{h}_i^m + \mathbf{h}_j^m)} \tag{3.2}$$

μεταξύ όλων των ζευγών των δειγμάτων εκπαίδευσης. Δεδομένου ότι αντιμετωπίζουμε προβλήματα ταξινόμησης πολλών τάξεων, ακολουθούμε την προσέγγιση ενός εναντίον όλων (one-against-all) και επιλέγουμε την τάξη με την υψηλότερη βαθμολογία.

Μια άλλη παραλλαγή της παραπάνω μεθοδολογίας χρησιμοποιεί τα προαναφερθέντα χαρακτηριστικά πυκνών τροχιών (DT) συνδυάζοντάς τα με την κωδικοποίηση VLAD [Jégou et al.; 2010]. Η VLAD κωδικοποίηση περιέχει επιπρόσθετη πληροφορία έναντι της μεθόδου K-means που υπολογίζει μόνο τα κεντροειδή, αφού υπολογίζει επίσης το μέσο διάνυσμα διαφορών ανάμεσα στα μέλη της κλάσης και στα κεντροειδή. Στην προσέγγιση μας κάθε τροχιά ανατίθεται στην πλησιέστερη συστάδα ενός λεξικού μεγέθους $K = 256$. Για κάθε μια από τις K συστάδες, υπολογίζονται οι διαφορές (δηλαδή οι αποκλίσεις) μεταξύ των χαρακτηριστικών της οπτικής λέξης και των χαρακτηριστικών που της έχουν ανατεθεί. Στη συνέχεια, τα κωδικοποιημένα χαρακτηριστικά που προκύπτουν από το VLAD ταξινομούνται χρησιμοποιώντας γραμμικά SVM. Τέλος, κάθε βίντεο κατατάσσεται στην κατηγορία με την υψηλότερη βαθμολογία, όπως συνέβη και με την κωδικοποίηση BoVW.

Χαρακτηριστικά εξαγόμενα από τρισδιάστατα συνελικτικά δίκτυα (C3D-based features)

Η δεύτερη εκδοχή του συστήματος αναγνώρισης δράσεων μονής όψης περιλαμβάνει εξαγωγή χαρακτηριστικών από ένα τρισδιάστατο συνελικτικό νευρωνικό δίκτυο (CNN) [Tran et al.; 2015, Tran et al.; 2017].



Σχήμα 3.4: Η τρισδιάστατη συνελικτική αρχιτεκτονική που αναπτύξαμε για την εξαγωγή τρισδιάστατων χαρακτηριστικών [Tran et al.; 2015]. Ο αριθμός στην παρένθεση για κάθε συνελικτικό μπλοκ υποδηλώνει τον αριθμό των φίλτρων ενώ ο αριθμός μέσα στο μπλοκ υποδηλώνει το μέγεθος του πυρήνα συνέλιξης. Τα πλήρως συνδεδεμένα επίπεδα (Fully Connected - FC) αποτελούνται και τα δύο από 4096 νευρώνες.

Σε ένα 3D CNN, η είσοδος του δικτύου αποτελείται από έναν τρισδιάστατο όγκο εικόνωνστιγμιοτύπων του βίντεο, ενώ στην έξοδο λαμβάνουμε τις πιθανότητες κατηγοριοποίησης του συγκεκριμένου βίντεο σε κάθε κλάση. Αυτή η end-to-end δομή χρησιμοποιείται για την εκπαίδευση του δικτύου. Στη συνέχεια, *χρησιμοποιούμε το δίκτυο για εξαγωγή χαρακτηριστικών.* Μεταξύ της εισόδου και της εξόδου του δικτύου παρεμβάλλονται: συνελικτικά επίπεδα όπου έχουμε συνέλιξη της εισόδου με τρισδιάστατους πυρήνες, επίπεδα χωρικής υποδειγματοληψίας (pooling layers), και πλήρως συνδεδεμένα επίπεδα (fully connected layers) που αντιστοιχούν στα τελικά χαρακτηριστικά που χρησιμοποιούνται για ταξινόμηση στο τελικό επίπεδο. *Τα*

CNN-χαρακτηριστικά εξάγονται από αυτά τα ενδιάμεσα στρώματα και στη συνέχεια τροφοδοτούνται σε μια μηχανή διανυσματικής στήριξης (Support Vector Machines - SVM) για την τελική ταξινόμηση έναντι της αξιοποίησης των πιθανοτήτων που προκύπτουν από την έξοδο. Στην συγκεκριμένη υλοποίηση χρησιμοποιήσαμε την αρχιτεκτονική δικτύου που παρουσιάστηκε στο [Tran et al.; 2015] και φαίνεται στο Σχήμα 3.4.

Συνήθως, τα χαρακτηριστικά που χρησιμοποιούνται στην ταξινόμηση εξάγονται από τα τελικά συνδεδεμένα στρώματα (FC6 και FC7). Ωστόσο, έχει προταθεί επίσης η εξαγωγή χαρακτηριστικών από το τελικό pooling επίπεδο ή από το τελικό επίπεδο συνέλιξης προκειμένου να αξιοποιηθούν οι χωρικές πληροφορίες που περιλαμβάνονται σε αυτά τα επίπεδα και χάνονται στη συνέχεια, στα πλήρως συνδεδεμένα επίπεδα. Στις εργασίες [Peng and Schmid; 2015,Xu et al.; 2015], CNN περιγραφητές εξάγονται από τα ενδιάμεσα επίπεδα και περιέχουν αυτές τις χωρικές πληροφορίες.

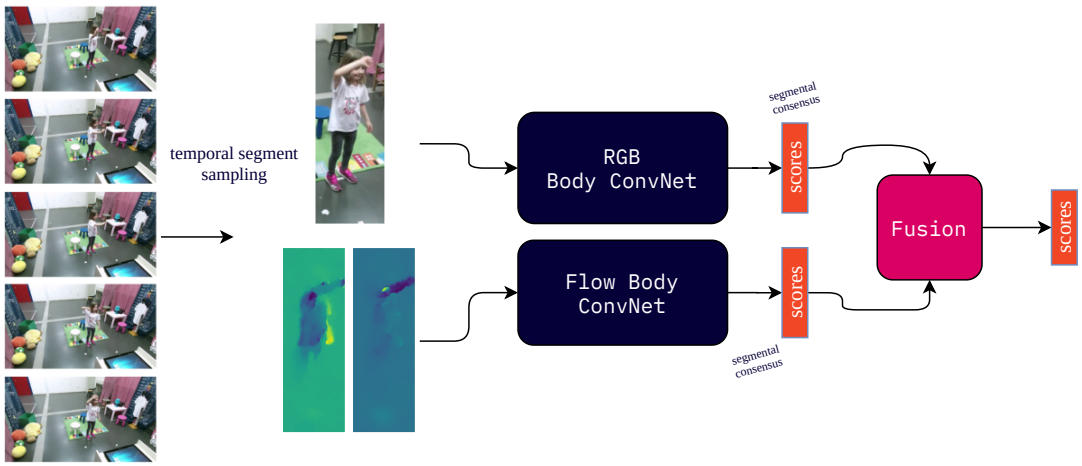
Το μειονέκτημα της χρήσης ενός δικτύου C3D είναι ο μεγάλος όγκος δεδομένων που απαιτείται για την αποφυγή υπερβολικής προσαρμογής (overfitting). Τα δίκτυα συνήθως εκπαιδεύονται σε βάσεις με μεγάλο όγκο δεδομένων όπως η βάση δεδομένων ActivityNet [Heilbron et al.; 2015] που περιέχει 15.410 βίντεο από 200 κλάσεις για εκπαίδευση, το Sports1M [Karpathy et al.; 2014] περιλαμβάνει πάνω από 1 εκατομμύριο βίντεο από 487 κλάσεις και το UCF101 [Soomro et al.; 2012] περιέχει 13.320 βίντεο από 101 κλάσεις. Προκειμένου να αποφευχθεί η υπερβολική προσαρμογή, στην περίπτωσή μας χρησιμοποιούμε τεχνικές μεταφοράς μάθησης (transfer leanring) εξιοποιώντας ένα προεκπαιδευμένο μοντέλο στη βάση Sports1M, το οποίο στη συνέχεια προσαρμόζουμε για να ταξινομήσουμε τις δράσεις σύμφωνα με τη βάση δεδομένων μας. Επιπλέον, δεδομένου ότι έχουμε περιορισμένα δεδομένα, χωρίζουμε κάθε βίντεο σε αποσπάσματα μήκους 16 στιγμιοτύπων με επικάλυψη 15 στιγμιοτύπων και τα χρησιμοποιούμε για την προσαρμογή του δικτύου ακολουθώντας μια προσέγγιση leave-one-out ως προς τα υποκείμενα που εκτελούν τη δράση και learning rate 10^{-4}. Κατά την εξαγωγή χαρακτηριστικών, για κάθε απόσπασμα 16 στιγμιοτύπων εξάγουμε χαρακτηριστικά από τα επίπεδα FC6, FC7, pool5 και conv5b και υπολογίζουμε τον μέσο όρο σε κάθε απόσπασμα προκειμένου να λάβουμε έναν περιγραφητή για ολόκληρο το βίντεο της δράσης.

Συνελικτικά Δίκτυα Συνδυασμένα με Χρονική Δειγματοληψία

Η συγκεκριμένη δομή του συστήματος *βασίζεται στη λογική του Δικτύου Χρονικής Δειγματοληψίας (Temporal Segment Network - TSN)* [Wang et al.; 2016a], που αρχικά εισήχθη για αναγνώριση δράσεων μεγάλης κλίμακας. Σύμφωνα με τη συγκεκριμένη δουλειά, λαμβάνονται τυχαία δείγματα K διαφορετικών τμημάτων από το βίντεο εισόδου, καθένα από τα οποία αποτελείται από N διαδοχικά καρέ. *Αυτή η τυχαία δειγματοληψία βοηθά στη γενίκευση και μειώνει το υπολογιστικό κόστος και τις περιττές πληροφορίες που υπάρχουν σε διαδοχικά καρέ βίντεο.*

Η παρούσα αρχιτεκτονική του δικτύου που αναπτύξαμε για τη μονάδα αναγνώρισης δράσεων φαίνεται στο Σχήμα 3.5. *Συγκεκριμένα, χρησιμοποιούνται δύο διαφορετικές ροές, μία χωρική που λαμβάνει ως είσοδο την RGB πληροφορία του βίντεο και μία χρονική που λαμβάνει ως είσοδο την οπτική ροή που εξάγεται από το βίντεο.* Προκειμένου τα δίκτυα να εστιάσουν στο παιδί και τις ενέργειές του, αρχικά γίνεται εκτίμηση της πόζας του παιδιού και στη συνέχεια περικόπτεται η περιοχή γύρω από το παιδί κατά τη διάρκεια της χρονικής δειγματοληψίας με βάση τον σκελετό του παιδιού που ανιχνεύθηκε με χρήση της τεχνολογίας OpenPose [Cao et al.; 2017]. Η ίδια διαδικασία εφαρμόζεται και στην οπτική ροή του βίντεο.

Αφού λάβουμε τα σκορ για την ταξινόμηση του κάθε τμήματος (segment) του βίντεο, μια συνάρτηση τμηματικής συμφωνίας, λ.χ. μέση τιμή των σκορ, συνυπολογίζει τα σκορ των επιμέρους τμημάτων και εξάγει μια πρόβλεψη-σκορ για κάθε μια από τις ροές του δικτύου. Τέλος, οι *προκύπτουσες προβλέψεις από τα δύο επιμέρους δίκτυα συνδυάζονται υπολογίζοντας τον σταθμισμένο*



Σχήμα 3.5: Η δομή του συστήματος αναγνώρισης δράσεων που βασίστηκε στα δίκτυα χρονικής δειγματοληψίας (Temporal Segment Networks - TSN).

μέσο όρο τους ώστε να δώσουν την τελική ταξινόμηση της δράσης.

3.2.2 Πειραματικά Αποτελέσματα

Αξιολόγηση σχημάτων εκπαίδευσης πυκνών τροχιών

Αρχικά αξιολογήσαμε την ανάπτυξη συστημάτων αναγνώρισης χειρονομιών και δράσεων που βασίζονται σε κλασικές τεχνικές όρασης υπολογιστών, δηλαδή σε χαρακτηριστικά πυκνών τροχιών όπως παρουσιάσαμε αναλυτικά στην υποενότητα 3.2.1. Συνολικά μελετήσαμε τέσσερα είδη χαρακτηριστικών (Trajectories, HOG, HOF, MBH) καθώς και το συνδυασμό τους για την οπτική πληροφορία που συλλέξαμε από κάθε κάμερα και σε συνδυασμό με τις κωδικοποιήσεις BoW και VLAD. Επίσης, τα μοντέλα αναγνώρισης έχουν εκπαιδευτεί ξεχωριστά για κάθε αισθητήρα στα δεδομένα ανάπτυξης και έχουν αξιολογηθεί ακολουθώντας cross-validation προσέγγιση σε διαφορετικούς και μη επικαλυπτόμενους φακέλους όπου ο έλεγχος γινόταν στα δεδομένα ενός μόνο παιδιού που δεν περιεχόταν στα δεδομένα εκπαίδευσης (leave-one-out). *Η ακρίβεια αναγνώρισης των μοντέλων έχει υπολογιστεί τόσο στα δεδομένα ανάπτυξης (development data) όσο και στα δεδομένα που σχετίζονται με τα δεδομένα της συνεχόμενης αλληλεπίδρασης των παιδιών με τα ρομπότ (use-case related data).* Στους Πίνακες 3.2, 3.3 παρουσιάζονται τα αποτελέσματα των συγκρίσεων.

Ως προς το σύστημα αναγνώρισης των χειρονομιών ο Πίνακας 3.2 παρουσιάζει τα αποτελέσματα μέσης ακρίβειας (%) για τις 7 χειρονομίες και μία τυχαία - άσκοπη χειρονομία (background). Παρατηρούμε πως οι περιγραφητές HOF και MBH που σχετίζονται με την αναπαράσταση των κινήσεων, είναι πιο αποτελεσματικοί στην αναγνώρισης απ' ότι π.χ. οι HOG περιγραφητές που είναι στατικοί, αφού οι χειρονομίες που εκτελούν τα παιδιά περιλαμβάνουν την κίνηση ολόκληρου το χεριού και όχι μόνο μια χειρομορφή. Επίσης, *σημαντική παρατήρηση αποτελεί το γεγονός ότι ο συνδυασμός και των τεσσάρων περιγραφητών, που στον πίνακα αναγράφεται ως Comb., δίνει μια σημαντική αύξηση στην ακρίβεια αναγνώρισης.* Έτσι τα αποτελέσματα είναι πολύ ικανοποιητικά αφού κυμαίνονται στο διάστημα 74%-81% για τα δεδομένα ανάπτυξης του συστήματος. Ακόμη παρατηρούμε πως ανάμεσα στους τέσσερις αισθητήρες καταγραφής, η αναγνώριση των χειρονομιών από το Kinect#1 (δεξιά πλάγια όψη) υπερτερεί έναντι των άλλων ανεξάρτητα από τη μέθοδο αναπαράστασης. Αυτό πιθανά οφείλεται στο ότι η πλειοψηφία των παιδιών ήταν δεξιόχειρες και συνεπώς η κάμερα αυτή κατέγραφε πιο καθαρά την κίνηση του δεξιού χεριού

Δεδομένα Ανάπτυξης								
	Kinect #1		Kinect #2		Kinect #3		Kinect #4	
Features	BoW	VLAD	BoW	VLAD	BoW	VLAD	BoW	VLAD
Traj.	68.75	70.83	66.90	64.58	65.74	65.74	**68.52**	62.96
HOG	40.74	33.33	33.33	31.25	29.40	30.79	41.20	31.94
HOF	70.83	72.69	70.37	71.76	**69.21**	66.67	63.43	53.70
MBH	**76.85**	75.93	67.82	**73.38**	68.29	68.75	65.28	57.41
Comb.	77.78	**80.79**	73.84	**78.24**	73.61	**73.84**	**75.00**	70.83
Δεδομένα Αλληλεπίδρασης								
	Kinect #1		Kinect #2		Kinect #3		Kinect #4	
Features	BoW	VLAD	BoW	VLAD	BoW	VLAD	BoW	VLAD
Traj.	45.76	50.55	42.12	45.19	45.41	58.85	45.56	43.84
HOG	24.13	28.32	29.70	22.26	17.09	19.13	41.25	27.79
HOF	56.92	**66.93**	54.49	65.93	57.10	63.01	51.97	37.14
MBH	62.70	63.00	56.47	**65.95**	60.15	**68.04**	54.25	**56.33**
Comb.	57.96	**70.77**	54.08	**67.87**	67.03	**71.73**	59.16	**60.54**

Πίνακας 3.2: Μέση ακρίβεια ταξινόμησης (%) για τις 8 χειρονομίες που εκτέλεσαν τα παιδιά. Αποτελέσματα για τα πέντε διαφορετικά είδη χαρακτηριστικών περιγραφητών και δύο κωδικοποιήσεων για το σύστημα αναγνώρισης χειρονομιών μονής όψης.

τους.

Αναφορικά με την αναγνώριση του συστήματος στα δεδομένα αλληλεπίδρασης, η ακρίβεια του συστήματος παρατηρείται μικρότερη απ' ότι στα δεδομένα ανάπτυξης καθώς τα βέλτιστα αποτελέσματα κυμαίνονται από 61% έως 72%, δηλαδή μείωση της τάξης του 10%. *Αυτό αποκαλύπτει τη δυσκολία που αντιμετώπισαν τα παιδιά όταν προσπάθησαν να κάνουν τις χειρονομίες αυθόρμητα κατά την αλληλεπίδραση,* π.χ. δίσταζαν να ξεκινήσουν να κάνουν μια κίνηση ή και κάποιες φορές την άλλαζαν κατά τη διάρκεια, καθώς και τη δυσκολία της αναγνώρισης που προκύπτει κατά τη διάρκεια μιας ανεμπόδιστης αλληλεπίδρασης παιδιών και ρομπότ. Στα δεδομένα αλληλεπίδρασης είναι ενδιαφέρον να σημειώσουμε πως η μεγαλύτερη ακρίβεια αναγνώρισης επιτυγχάνεται από την κάμερα Kinect#3 που βρίσκεται στην οροφή με ποσοστό σχεδόν ίσο με αυτό των δεδομένων ανάπτυξης. Καθώς μια από τις σημαντικότερες διαφορές μεταξύ των δύο συνόλων δεδομένων είναι πως στα δεδομένα αλληλεπίδρασης τα παιδιά μπορούν να βρεθούν όπου επιθυμούν στο χώρο, η αναπαράσταση από την κατακόρυφη κάμερα μας δίνει ένα πιο γενικευμένο μοντέλο - ικανό να αποσβέσει την επίδραση τέτοιων χωρικών αλλαγών. Ως προς τους περιγραφητές, οι MBH και HOF συνεχίζουν να υπερτερούν ενώ ο συνδυασμός των τεσσάρων περιγραφητών βελτιώνει αρκετά τα αποτελέσματα.

Συνολικά, ανάμεσα στις δύο μεθοδολογίες κωδικοποίησης BoW και VLAD, η δεύτερη κωδικοποιεί πιο αποτελεσματικά την οπτική πληροφορία αφού περιλαμβάνει πλούσια πληροφορία σχετικά με την κατανομή των οπτικών λέξεων στο λεξικό. Τέλος, παρατηρούσαμε πως το μοντέλο μας έκανε συχνά λάθος στην κατηγοριοποίηση του καταφατικού νεύματος των παιδιών ταξινομώντας το στην κλάση της τυχαίας κίνησης, κάτι που θεωρείται εύλογο καθώς είναι μια απαλή κίνηση που δεν περιλαμβάνει κίνηση του χεριού και άρα θα μπορούσε η κίνηση του κεφαλιού να θεωρηθεί μια άσκοπη κίνηση.

Για να επαληθεύσουμε την καταλληλότητα του προτεινόμενου συστήματος αναγνώρισης παιδικών δράσεων σε *πιο απαιτητικές εργασίες, αξιολογούμε το σύστημα αναγνώρισης μας σε παντομίμες.* Για την αναγνώριση των δράσεων αυτών, η δυσκολία δημιουργίας κατάλληλων ανα-

Σχήμα 3.6: Παράδειγμα των εξαγόμενων πυκνών τροχιών από διαφορετικές οπτικές γωνίες αισθητήρων ενώ το παιδί εκτελεί την παντομίμα κολύμβησης.

παραστάσεων είναι μεγαλύτερη καθώς όχι μόνο καλείται να κατηγοριοποιήσει *μεγαλύτερο πλήθος κινήσεων από αυτό του συστήματος αναγνώρισης χειρονομιών αλλά και κινήσεις-παντομίμες που έχουν μεγαλύτερη ελευθερία απόδοσης με διαφορετικούς τρόπους.* Στο Σχήμα 3.6 παρουσιάζεται ένα παράδειγμα των εξαγόμενων πυκνών τροχιών από τις τέσσερις διαφορετικές οπτικές γωνίες που καταγράφουν οι αισθητήρες. Στον Πίνακα 3.3 παρουσιάζονται τα αποτελέσματα μέσης ακρίβειας (%) της εκπαίδευσης του συστήματος αναγνώρισης για τις 12 παντομίμες και για μια πρόσθετη τυχαία-άσκοπη κίνηση (background) που εκτέλεσαν τα παιδιά στα δεδομένα ανάπτυξης.

Ξεκινώντας από την αξιολόγηση στα δεδομένα ανάπτυξης, παρατηρούμε πως το ποσοστό αναγνώρισης των τριών αισθητήρων (Kinect #1, #2, #4) κυμαίνεται από 69% έως 76%, ένα ποσοστό που θα μπορούσαμε να χαρακτηρίσουμε ως ικανοποιητικό, όμως ο αισθητήρας κάτοψης (Kinect#3) δεν μπορεί να μας δώσει μια αποτελεσματική αναπαράσταση για τις κινήσεις παντομίμας. *Απ' ότι γίνεται αντιληπτό (και επιβεβαιώνεται και στα δεδομένα αλληλεπίδρασης) η κατακόρυφη οπτική δεν μπορεί να καταγράψει το εύρος και την ποικιλία των κινήσεων που εμφανίζεται σε ένα παιχνίδι παντομίμας.* Επίσης, παρατηρούμε πως ο συνδυασμός των χαρακτηριστικών πυκνών τροχιών αποδίδει ελαφρώς καλύτερα απ' ότι μεμονωμένα τα MBH που διακρίνονται ως η βέλτιστη επιλογή για το συγκεκριμένο σύστημα αναγνώρισης τόσο για τις κωδικοποιήσεις VLAD όσο και τις BoW. Τέλος, ανάμεσα στις δύο κωδικοποιήσεις, λαμβάνουμε τα καλύτερα αποτελέσματα για την κωδικοποίηση VLAD. *Επιπλέον, το διάνυσμα VLAD βελτιώνει περαιτέρω την απόδοση,* καθώς κωδικοποιεί πλούσιες πληροφορίες σχετικά με την κατανομή των οπτικών λέξεων.

Σχετικά με τα δεδομένα της αδιάκοπης αλληλεπίδρασης παιδιών και ρομπότ κατά τη διάρκεια του παιχνιδιού της παντομίμας, όπως είναι αναμενόμενο η *ακρίβεια της αναγνώρισης μειώνεται ακόμα περισσότερο* και κυμαίνεται από το 51% έως το 59% για τις τρεις κάμερες, ενώ η κάμερα οροφής αποτυγχάνει στην αναγνώριση των δράσεων. Τα καλύτερα αποτελέσματα λαμβάνονται από την κωδικοποίηση της οπτικής πληροφορίας μέσω του αισθητήρα Kinect #1 που μας έδινε και πολύ καλά αποτελέσματα και στα δεδομένα ανάπτυξης. Η VLAD κωδικοποίηση υπερτερεί συνολικά στην κωδικοποίηση του συνδυασμού όλων των χαρακτηριστικών πυκνών τροχιών, ενώ στην περίπτωση χρήσης MBH χαρακτηριστικών για την Kinect #1 μας δίνει την καλύτερη κωδικοποίηση η μέθοδος BoW.

Συνολικά παρατηρήσαμε πως ανάμεσα στα χαρακτηριστικά πυκνών τροχιών που χρησιμοποιήσαμε, τα MBH χαρακτηριστικά αποδίδουν καλύτερα στην κωδικοποίηση των δράσεων, ενώ ο συνδυασμός και των τεσσάρων ειδών χαρακτηριστικών μας αποφέρει καλύτερα αποτελέσματα αναγνώρισης από ότι τα μεμονωμένα χαρακτηριστικά. Από την πλευρά της κωδικοποίησης, φαίνεται πως η μέθοδος VLAD υπερτερεί συνολικά έναντι της BoW καθώς περιλαμβάνει πλουσιότερη πληροφορία σχετικά με την κατανομή των οπτικών λέξεων στο οπτικό λεξικό. Τέλος, είδαμε πως δεν μπορούμε να θεωρήσουμε πως η θέση κάποια συγκεκριμένης κάμερας αποδίδει το ίδιο ικανά και στις δύο περιπτώσεις αναγνώρισης και στα δύο σύνολα δεδομένων των παιδιών. Γενικά μπορούμε να πούμε πως η θέση της Kinect #1 (πλαϊνή μπροστινή όψη) δίνει καλά αποτελέσματα, χωρίς όμως να είναι πάντα τα βέλτιστα, πιθανά επειδή η πλειοψηφία των

Δεδομένα Ανάπτυξης								
	Kinect #1		Kinect #2		Kinect #3		Kinect #4	
Features	BoW	VLAD	BoW	VLAD	BoW	VLAD	BoW	VLAD
Traj.	63.08	60.31	48.62	48.62	45.45	46.46	49.73	55.08
HOG	36.69	36.69	32.00	38.15	27.69	34.46	28.30	50.15
HOF	68.31	69.85	56.31	63.08	48.62	50.46	53.85	63.69
MBH	70.77	**72.92**	60.92	**68.62**	**61.85**	60.00	55.22	**72.92**
Comb.	73.85	**74.15**	63.38	**69.23**	**60.00**	58.46	61.45	**76.31**
Δεδομένα Αλληλεπίδρασης								
	Kinect #1		Kinect #2		Kinect #3		Kinect #4	
Features	BoW	VLAD	BoW	VLAD	BoW	VLAD	BoW	VLAD
Traj.	47.89	50.84	38.49	43.75	25.52	22.61	44.14	44.64
HOG	30.98	23.10	23.95	27.51	16.53	19.73	21.76	34.14
HOF	46.34	51.00	46.19	**51.78**	25.50	26.13	47.70	44.67
MBH	**61.42**	56.86	46.28	45.82	**31.59**	29.44	45.57	**56.11**
comb.	52.59	**59.16**	46.74	**51.51**	36.62	**38.22**	48.16	**55.11**

Πίνακας 3.3: Μέση ακρίβεια ταξινόμησης (%) για τις 13 παντομίμες που εκτελέστηκαν από τα παιδιά. Αποτελέσματα για τα πέντε διαφορετικά είδη χαρακτηριστικών περιγραφητών και δύο κωδικοποιήσεων για το σύστημα αναγνώρισης δράσεων μονής όψης.

συμμετεχόντων είναι δεξιόχειρες και η θέση της κάμερας αυτής βοηθά στην καταγραφή των κινήσεων του δεξιού χεριού χωρίς επικάλυψη π.χ. από άλλα μέλη του σώματος.

Αξιολόγηση σχημάτων εκπαίδευσης ως προς τις ηλικιακές ομάδες

Κύριο ζήτημα στην εκπαίδευση συστημάτων αναγνώρισης παιδικών δράσεων, όπως έχει αναφερθεί παραπάνω, αποτελεί η έλλειψη μεγάλου πλήθους δεδομένων παιδιών που εκτελεί αυτές τις δράσεις. Για το λόγο αυτό, διερευνήσαμε την αναγκαιότητα ή μη συλλογής δεδομένων παιδιών για την εκπαίδευση και των έλεγχων των συστημάτων που αναπτύσσουμε. Έτσι πραγματοποιήσαμε πειράματα για να διαπιστώσουμε εάν υπάρχει κάποια σημαντική διαφορά στην αποτελεσματικότητα των συστημάτων αναγνώρισης που σχετίζεται με την ηλικιακή διαφορά των ατόμων που εκτελούν τις δράσεις. Για να ελέγξουμε την υπόθεση αυτή δημιουργήσαμε τρία σύνολα δεδομένων σύμφωνα με την ηλικία των συμμετεχόντων που εκτελούν τις δράσεις και τις χειρονομίες. Το πρώτο σύνολο περιέχει αποκλειστικά δράσεις και χειρονομίες παιδιών, το δεύτερο σύνολο ενηλίκων και το τρίτο σύνολο περιέχει δεδομένα ατόμων όλων των ηλικιών. Ο έλεγχος πραγματοποιήθηκε τόσο για παιδιά όσο και για δεδομένα ενηλίκων ξεχωριστά, εξαιρώντας κάθε φορά το άτομο ελέγχου από το σύνολο εκπαίδευσης και εφαρμόζοντας τον έλεγχο σε διαφορετικούς μη επικαλυπτόμενους φακέλους (leave-one-out cross-validation).

Στον Πίνακα 3.4 παρουσιάζονται τα αποτελέσματα της αξιολόγησης των συστημάτων αναγνώρισης μονής όψης χειρονομιών και δράσεων για τους διάφορους συνδυασμούς ηλικιακών ομάδων για καθέναν από τους τρεις αισθητήρες καθώς και ο μέσος όρος αυτών. Στα συγκεκριμένα πειράματα έχουμε κάνει χρήση MBH χαρακτηριστικών με BoW κωδικοποίηση. Θεωρούμε πως η διερεύνηση μιας μόνο επιλογής μεθόδων είναι αρκετή ώστε να μας βοηθήσει να αντιληφθούμε τις διαφορές και δυνατότητες που υπάρχουν αξιοποιώντας σύνολα δεδομένων διαφορετικών ηλικιών για την εκπαίδευση και τον έλεγχο των συστημάτων αναγνώρισης που αναπτύσσουμε.

Ως προς τις χειρονομίες, παρατηρούμε πως όταν θέλουμε να αναγνωρίσουμε χειρονομίες παιδιών

		Ομάδα Εκπαίδευσης					
		Αναγνώριση Χειρονομιών			Αναγνώριση Κινήσεων		
Ομάδα Ελέγχου		Ενήλικες	Παιδιά	Μείξη	Ενήλικες	Παιδιά	Μείξη
Ενήλικες	Kinect #1	84.79	60.21	**87.81**	**79.67**	67.58	78.02
	Kinect #2	89.27	53.13	**92.19**	**83.52**	62.09	79.12
	Kinect #3	**85.42**	55.63	82.08	71.98	59.34	**78.02**
	Average	86.49	56.32	**87.36**	**78.39**	63.00	**78.39**
Παιδιά	Kinect #1	60.42	76.85	**77.31**	53.85	**73.85**	73.67
	Kinect #2	46.99	67.82	**68.75**	47.63	63.38	**64.20**
	Kinect #3	42.36	68.29	**70.83**	38.18	**60.00**	59.76
	Average	49.92	70.99	**72.30**	46.55	65.74	**65.88**

Πίνακας 3.4: Αξιολόγηση των συστημάτων αναγνώρισης χειρονομιών και κινήσεων παντομίμας ως προς τις ηλικιακές ομάδες ελέγχου και εκπαίδευσης

και ενηλίκων το βέλτιστο είναι να εκπαιδεύσουμε το σύστημά μας σε μεικτά δεδομένα όλων των ηλικιών. Για την περίπτωση των ενηλίκων, βλέπουμε πως η εκπαίδευση των μοντέλων μας στην ίδια ηλικιακή ομάδα δίνει ικανοποιητικά αποτελέσματα αναγνώρισης (άνω του 80%) και συγκεκριμένα για την κατακόρυφη κάμερα οροφής (Kinect #3) έχουμε το βέλτιστο αποτέλεσμα. Αντίστοιχα, στην περίπτωση των παιδιών παρατηρούμε πως ανάμεσα στις δύο διακριτές ομάδες, για την εκπαίδευση μοντέλων παιδιών είναι απαραίτητα τα δεδομένα παιδιών, ενώ βλέπουμε πως και σ' αυτή την περίπτωση η ενσωμάτωση δεδομένων ενηλίκων αυξάνει λίγο την ακρίβεια της αναγνώρισης. *Συνολικά αντιλαμβανόμαστε πως για κάθε ηλικιακή ομάδα είναι απαραίτητη η χρήση δεδομένων εκπαίδευσης από την ίδια ηλικιακή ομάδα ενώ η ενίσχυση των δεδομένων με δεδομένα άλλης ηλικιακής ομάδας προσφέρει μια μικρή βελτίωση του μοντέλου.* Η βελτίωση αυτή θα μπορούσε να οφείλεται είτε στο διπλασιασμό των δεδομένων εκπαίδευσης (η μεικτή ομάδα αποτελείται από το άθροισμα των δεδομένων των δύο άλλων ομάδων) είτε και στη διαφορά του τρόπου εκτέλεσης των χειρονομιών και άρα στη βοήθεια γενίκευσης του μοντέλου.

Ως προς τις δράσεις, παρατηρούμε πως στις δύο από τις τρεις κάμερες η βέλτιστη αναγνώριση επιτυγχάνεται όταν η εκπαίδευση των μοντέλων γίνεται στην ίδια ηλικιακή ομάδα που θέλουμε να αναγνωρίσουμε ενώ για μια κάμερα κάθε φορά, η μείξη των δεδομένων δίνει καλύτερα αποτελέσματα. Η απόκλιση της επιτυχίας αναγνώρισης μεταξύ των συστημάτων που εκπαιδεύονται σε διαφορετικό σύνολο απ' ότι ελέγχονται - μεταξύ των ξένων ομάδων - είναι αρκετά μεγάλη, της τάξης του 10%-20%. Συνολικά, υπολογίζοντας τον μέσο όρο ακρίβειας αναγνώρισης των τριών καμερών παρατηρούμε πως η μείξη των δεδομένων πρακτικά μας δίνει το ίδιο ποσοστό αναγνώρισης με την επιλογή εκπαίδευσης και ελέγχου στην ίδια ηλικιακή ομάδα. Και στην περίπτωση των δράσεων λοιπόν παρατηρούμε την τεράστια σημαντικότητα της συλλογής δεδομένων παιδιών για την αυτόματη αναγνώρισή τους.

Συμπερασματικά, η εκπαίδευση των μοντέλων αυτόματης αναγνώρισης χειρονομιών και δράσεων είναι θεμιτό και αναγκαίο να γίνεται σε δεδομένα παιδιών όταν τα συστήματα αυτά αξιοποιούνται σε αλληλεπιδράσεις παιδιών και ρομπότ. Στη συνέχεια λοιπόν θα επικεντρωθούμε μόνο στην εκπαίδευση συστημάτων σε δεδομένα παιδιών. Σε επόμενη ενότητα που θα μελετήσουμε τα συστήματα αναγνώρισης πολλαπλών όψεων θα επανεξετάσουμε αυτή μας την επιλογή.

Αξιολόγηση σχημάτων εκπαίδευσης βαθιάς μάθησης

Στην παρούσα υποενότητα εξετάζουμε τη χρήση τεχνικών βαθιάς μάθησης για την αντιμετώπιση του προβλήματος αναγνώρισης χειρονομιών και δράσεων και συγκεκριμένα τη χρήση

C3D Layers	C3D Network		
	Kinect #1	Kinect #2	Kinect #3
conv5b	54.46	52.00	40.92
pool5	57.23	54.15	42.46
FC6	**59.38**	**54.46**	42.77
FC7	57.85	52.92	42.15
Comb.	56.92	**54.46**	**44.31**
end-to-end	58.03	52.05	41.87

Πίνακας 3.5: Αξιολόγηση του συστήματος αναγνώρισης δράσεων για παιδιά μονής όψης με χρήση χαρακτηριστικών συνελικτικών δικτύων και end-to-end χρήση δικτύου.

συνελικτικών δικτύων. Η αρχική μας προσέγγιση υλοποιήθηκε με χρήση τρισδιάστατων συνελικτικών δικτύων ενώ στη συνέχεια μελετήσαμε το συνδυασμό απλού συνελικτικού δικτύου με τη λογική των δικτύων χρονικής δειγματοληψίας.

C3D: Τρισδιάστατα Συνελικτικά Δίκτυα

Ο Πίνακας 3.5 παρουσιάζει τα αποτελέσματα μέσης ακρίβειας (%) για τις 12 δράσεις παντομίμας και την τυχαία-άσκοπη δράση που εκτελέστηκαν από παιδιά, εξαιρώντας κάθε φορά τα δεδομένα του παιδιού-ελέγχου από το σύνολο εκπαίδευσης και εφαρμόζοντας τον έλεγχο σε διαφορετικούς μη επικαλυπτόμενους φακέλους (leave-one-out cross-validation). Όπως εξηγήσαμε αναλυτικά στην υποενότητα 3.2.1, χρησιμοποιούμε το τρισδιάστατο συνελικτικό δίκτυο για να πάρουμε CNN-χαρακτηριστικά που εξάγονται από τα ενδιάμεσα επίπεδα του δικτύου (conv5b, pool5), τα τελικά συνδεδεμένα επίπεδα (FC6, FC7) και ένα συνδυασμό αυτών. Στη συνέχεια τα τροφοδοτούμε σε ένα SVM για να λάβουμε την τελική ταξινόμηση έναντι της αξιοποίησης των πιθανοτήτων που προκύπτουν από την έξοδο. Επιπρόσθετα συγκρίνουμε και την ακρίβεια αναγνώρισης του δικτύου εάν το αξιοποιήσουμε ως έχει, σε μια end-to-end λογική. Το σύστημα αναγνώρισης αξιολογείται στα δεδομένα ανάπτυξης.

Όσον αφορά την απόδοση των χαρακτηριστικών του CNN βλέπουμε ότι έχουν χαμηλή απόδοση, ιδιαίτερα αν τα συγκρίνουμε με τα χαρακτηριστικά πυκνών τροχιών όπου η βέλτιστη ακρίβεια αναγνώρισης ανά αισθητήρα είναι $+15\%$ *μεγαλύτερη.* Ο κύριος λόγος που η ακρίβεια αναγνώρισής τους είναι πολύ χαμηλή είναι ότι τα συγκεκριμένα δίκτυα απαιτούν πολύ μεγάλο όγκο δεδομένων για την εκπαίδευσή τους και συνεπώς η χρήση προεκπαιδευμένων δικτύων είναι αναγκαία. Στην περίπτωση μας λοιπόν χρησιμοποιήσαμε ένα προεκπαιδευμένο δίκτυο στη βάση Sports1M, ύστερα από κατάλληλη προσαρμογή (fine-tuning), το οποίο είναι μεν κατάλληλο για ανάπτυξη συστημάτων αναγνώρισης δράσεων όμως δεν είναι αρκετά σχετικό με τις δράσεις που περιλαμβάνονται στο δικό μας σύνολο δεδομένων και σε αλληλεπιδράσεις παιδιών και ρομπότ. Επιπλέον, η εκπαίδευση από άκρο σε άκρο ενός δικτύου 3D CNN απαιτεί τεράστιο όγκο δεδομένων παιδιών, κάτι που δεν είναι ρεαλιστικό σενάριο όπως έχουμε ήδη εξηγήσει. Ωστόσο, βλέπουμε ότι το καλύτερο αποτέλεσμα στις περισσότερες περιπτώσεις επιτυγχάνεται χρησιμοποιώντας τα χαρακτηριστικά που εξάγονται από το επίπεδο FC6.

Συμπερασματικά, παρατηρούμε πως η χρήση των τρισδιάστατα συνελικτικά δίκτυα δεν προτείνεται για την ανάπτυξη συστημάτων αναγνώρισης ενεργειών σε αλληλεπιδράσεις παιδιών και ρομπότ.

Πλήθος Δειγμάτων	Ακρίβεια Αναγνώρισης (%)	Χρόνος/Εποχή Εκπαίδευσης (s)	Χρόνος/Εποχή Αξιολόγησης (s)
RGB			
1	36.74	5.2	0.4
3	40.95	6.0	0.8
5	47.43	8.8	1.0
10	49.56	14.6	1.4
Οπτική Ροή			
1	58.75	5.4	0.6
3	71.77	10.3	1.2
5	75.96	16.3	1.8
10	**76.82**	31.3	3.2

Πίνακας 3.6: Μέση ακρίβεια αναγνώρισης δράσεων και απαιτούμενος χρόνος ανά περίοδο εκπαίδευσης και αξιολόγησης για διαφορετικό πλήθος τμημάτων δειγματοληψίας για τα δεδομένα ανάπτυξης από την κάμερα Kinect #1.

Συνελικτικά Δίκτυα Συνδυασμένα με Χρονική Δειγματοληψία

Στη συνέχεια μελετάμε πως *η χρήση χρονικής δειγματοληψίας στα βίντεο δράσεων και ο συνδυασμός της με απλά συνελικτικά δίκτυα*, όπως περιγράφεται στην υποενότητα 3.2.1, *μπορεί να βοηθήσει στο πρόβλημα της αναγνώρισης των παιδικών δράσεων*. Αρχικά διερευνούμε την επίδραση της χρονικής δειγματοληψίας στην απόδοση του συστήματος και στη συνέχεια την επίδραση της προεκπαίδευσης στις δυο ροές πληροφορίας RGB και Flow καθώς και στο συνδυασμό τους.

Πλήθος τμημάτων δειγματοληψίας: Στην πρώτη μας μελέτη, διερευνούμε την επίδραση του αριθμού των τμημάτων δειγματοληψίας για την RGB και την οπτική ροή (Πίνακας 3.6), ως συνάρτηση τόσο της τελικής απόδοσης όσο και της υπολογιστικής πολυπλοκότητας του συστήματος. Στο σύστημα αναγνώρισης δράσεων που αναπτύσσουμε, χρησιμοποιούμε μια αρχιτεκτονική Batch Normalization Inception (BNInception) [Szegedy et al.; 2016] για κάθε μια από τις ροές RGB και Optical Flow με προεκπαίδευση στη βάση δεδομένων αναγνώρισης δράσεων Kinetics [Carreira and Zisserman; 2017]. Εκπαιδεύουμε κάθε δίκτυο 60 εποχές με τη μέθοδο στοχαστικής καθόδου κλίσης (Stochastic Gradient Descent - SGD) με συνάρτηση κόστους cross-entropy loss ενώ επιλέγουμε πέντε τμήματα δειγματοληψίας μήκους 1 στιγμιοτύπου (frames) για την RGB πληροφορία και μήκους 5 στιγμιοτύπων για την οπτική ροή. Τα δίκτυά αυτά εκπαιδεύτηκαν και αξιολογήθηκαν στα βίντεο δράσεων-παντομίμας που έχουν καταγραφεί από την κάμερα Kinect#1 και περιλαμβάνονται στα δεδομένα ανάπτυξης με τη μέθοδο leave-one-out cross-validation.

Στον Πίνακα 3.6, παρατηρούμε ότι η απόδοση του συστήματος αυξάνεται καθώς αυξάνουμε τον αριθμό των τμημάτων που χρησιμοποιούνται στο πλαίσιο TSN και για τις δύο ροές. *Μάλιστα για την οπτική ροή παρατηρούμε πως μόλις με 10 δειγματοληπτημένα τμήματα μπορούμε να λάβουμε τη μεγαλύτερη ακρίβεια αναγνώρισης που έχουμε πετύχει στη συγκεκριμένη εργασία, 76.82% έναντι 74.15% που είχαμε πετύχει με τα χαρακτηριστικά πυκνών τροχιών* (Πίνακας 3.3). Ανάμεσα στις δύο ροές πληροφορίας, όπως είναι εύκολο να αντιληφθούμε, η RGB πληροφορία μας δίνει φτωχά αποτελέσματα μιας και παρέχει στατική πληροφορία ενώ η οπτική ροή περιλαμβάνει μεταβολές στην κίνηση του υποκειμένου και άρα είναι πιο κατάλληλη για την κωδικοποίηση της πληροφορίας σε ένα σύστημα αναγνώρισης δράσεων. Σχετικά με τη λογική της χρονικής

Μοντέλο Προεκπαίδευσης	Ακρίβεια Αναγνώρισης(%)
RGB-Kinetics	47.43
RGB-ImageNet	42.46
Flow-Kinetics	75.96
Flow-ImageNet	65.26
RGB-Kinetics + Flow-Kinetics	**76.55**
RGB-ImageNet + Flow-ImageNet	64.37

Πίνακας 3.7: Μέση ακρίβεια αναγνώρισης δράσεων για τα διαφορετικά σύνολα προεκπαίδευσης και κανάλια πληροφορίας για δειγματοληψία 5 δειγμάτων, στα δεδομένα ανάπτυξης από την κάμερα Kinect #1 για τις 13 κλάσεις των κινήσεων παντομίμας.

δειγματοληψίας παρατηρούμε πως μόλις με 3 τμήματα-δείγματα ενός βίντεο-δράσης που θα λέγαμε αδρά ότι αντιστοιχεί σε 3 * 5 ακολουθιακά στιγμιότυπα, έχουμε πολύ καλή ταξινόμηση μέσω της οπτικής ροής και άρα για κάθε παντομίμα-δράση που μπορεί να διαρκεί μερικά δευτερόλεπτα, θα λέγαμε πως περιέχονται μικρότερα μοτίβα χαρακτηριστικά της κάθε κίνησης.

Ωστόσο, εάν συνυπολογίσουμε στην ακρίβεια αναγνώρισης και το υπολογιστικό κόστος των προσεγγισεών μας, παρατηρούμε πως μετά από ένα όριο, το αυξημένο υπολογιστικό κόστος *δεν αντιστοιχεί σε μια αντίστοιχη αύξηση της ακρίβειας του συστήματος.* Συγκεκριμένα, στη RGB κανάλι, αύξηση περίπου 65% στο χρόνο εκπαίδευσης και 40% στο χρόνο αξιολόγησης επιφέρει μόλις 2% αύξηση στην ακρίβεια ταξινόμησης, ενώ στην οπτική ροή αύξηση περίπου 90% στο χρόνο εκπαίδευσης και 80% στο χρόνο αξιολόγησης επιφέρει λιγότερο από 1% αύξηση στην ακρίβεια ταξινόμησης της δράσης. Αντιλαμβανόμαστε λοιπόν πως σε περιπτώσεις συστημάτων όπως το TeachBot (βλ. υποενότητα 2.4) όπου θέλουμε μια απλή και εύχρηστη εκδοχή ενός συστήματος αναγνώρισης είναι ιδιαίτερο σημαντικό να λάβουμε υπόψιν και το υπολογιστικό κόστος της αναγνώρισης. *Έτσι, για τα επόμενα πειράματα κρατάμε ως βέλτιστη επιλογή τη δειγματοληψία 5 τμημάτων τόσο για το κανάλι του χρώματος όσο και για το κανάλι της οπτική ροής.*

Προεκπαίδευση: Στην παρούσα μελέτη εξετάζουμε δύο διαφορετικές αρχιτεκτονικές προεκπαίδευσης: α) στο σύνολο δεδομένων αναγνώρισης δράσεων Kinetics [Carreira and Zisserman; 2017], όπως παρουσιάστηκε και στην προηγούμενη παράγραφο, και β) στη βάση δεδομένων ImageNet [Deng et al.; 2009]. Η εκπαίδευση και η αξιολόγηση των μοντέλων μας γίνεται όμοια με την προηγούμενη παράγραφο ενώ τα αποτελέσματα παρουσιάζονται στον Πίνακα 3.7. Παρατηρούμε λοιπόν πως χρησιμοποιώντας προεκπαιδευμένα μοντέλα στη βάση δεδομένων Kinetics επιτυγχάνεται σημαντικά υψηλότερη ακρίβεια απ' ότι με τα προεκπαιδευμένα μοντέλα στην ImageNet, τόσο για το RGB όσο και για το Flow. Στον ίδιο πίνακα, παρουσιάζουμε επίσης το τελικό αποτέλεσμα της σύμμειξης των δύο ροών υπολογίζοντας το σταθμισμένο μέσο όρο τους. Πειραματικά υπολογίσαμε πως μια καλή αναλογία των δύο ροών επιτυγχάνεται με βάρος 0,8 στην οπτική ροή και 0,2 στο RGB.

Συμπερασματικά, από όλα τα παραπάνω πειραματικά αποτελέσματα, μπορούμε να καταλήξουμε στο συμπέρασμα πως *αξιοποιώντας τη λογική των Δικτύων Χρονικής Δειγματολήψίας σε συνδυασμό με την αποτελεσματικότητα των απλών συνελικτικών δικτύων και την προστειθέμενη αξία που δινει η προεκπαίδευση σε βάσεις αναγνώρισης δράσεων όπως η Kinetics, μπορούμε να αναπτύξουμε ένα αποτελεσματικό και ελαφρύ δίκτυο αναγνώρισης παιδικών δράσεων για χρήση σε πλούσιες αλληλεπιδράσεις παιδιών και ρομπότ.* Στον Πίνακα 3.8 συνοψίζονται τα καλύτερα αποτελέσματα για τις διάφορες εκδοχές του συστήματος αναγνώρισης παιδικών δράσεων σύμφωνα με τις κύριες μεθοδολογίες που αξιοποιήθηκαν.

Μοντέλο Εκπαίδευσης	Ακρίβεια Αναγνώρισης(%)
RGB-Kinetics + Flow-Kinetics (3.2.2)	**76.55**
Dense Trajectories Combined (3.2.2)	74.15
C3D (3.2.2)	59.38

Πίνακας 3.8: Συνοπτικός πίνακας μέσης ακρίβεια αναγνώρισης παιδικών δράσεων για τις διάφορες μεθοδολογίες. Τα αποτελέσματα αφορούν τον έλεγχο στα δεδομένα ανάπτυξης που λαμβάνονται από τον αισθητήρα Kinect#1.

3.3 Συστήματα αναγνώρισης πολλαπλών όψεων

Στην προηγούμενη ενότητα παρατηρήσαμε πως η ακρίβεια στην ταξινόμηση της οπτικής πληροφορίας και τελικά της αναγνώρισης των παιδικών ενεργειών που καταγράφηκαν από τους αισθητήρες μεταβάλλεται ανάλογα με την θέση του κάθε αισθητήρα. Όπως μπορεί να γίνει εύκολα αντιληπτό η αναγνώριση των δράσεων ενός ατόμου επηρεάζεται εύκολα από το σημείο θέασης ή καταγραφής, αν αναφερόμαστε σε άνθρωπο ή μηχανή αντίστοιχα. Έτσι, *θέλοντας να αναπτύξουμε ένα σύστημα που θα αντιλαμβάνεται με ακρίβεια τις αυθόρμητες ενέργειες των παιδιών χωρίς όμως αυτά να περιορίζονται στο χώρο κατά τη διάρκεια της αλληλεπίδρασης, μελετήσαμε τη σύμμειξη της πληροφορίας από τις πολλαπλές κάμερες που έχουμε εγκαταστήσει στο χώρο.* Άλλωστε αυτός ήταν ο κύριος λόγος της τοποθέτησης πολλών αισθητήρων, να δημιουργήσουμε ένα έξυπνο δωμάτιο στο οποίο τα παιδιά θα μπορούν να αλληλεπιδράσουν με τα ρομπότ ελεύθερα. Στην παρούσα ενότητα μελετάμε και επεκτείνουμε κάποιες από τις προσεγγίσεις που είδαμε κατά την σχεδίαση συστημάτων μονής όψης. Αναπτύσσουμε συστήματα αναγνώρισης δράσεων πολλαπλών όψεων με σύμμειξη της οπτικής πληροφορίας που αναπαρίσταται με χρήση μεθόδων πυκνών τροχιών (βλ. υποενότητα 3.2.1) και τρισδιάστατων νευρωνικών δικτύων C3D (βλ. υποενότητα 3.2.1).

3.3.1 Μέθοδοι Σύμμειξης και Αρχιτεκτονικές

Σύμμειξη της οπτικής πληροφορίας με χρήση χαρακτηριστικών πυκνών τροχιών και τρισδιάστατων συνελικτικών δικτύων

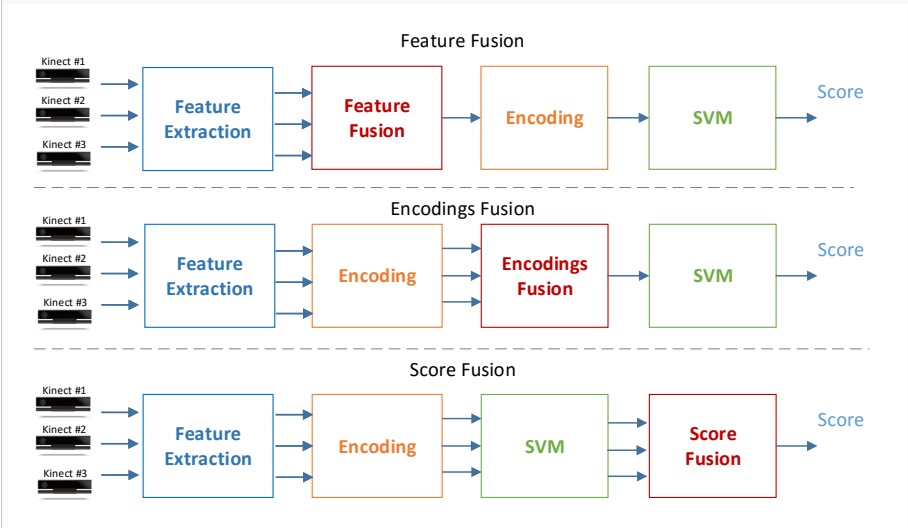
Στην παρούσα ενότητα διερευνούμε διαφορετικές προσεγγίσεις για τη σύμμειξη της οπτικής πληροφορίας που λαμβάνεται από τους πολλαπλούς αισθητήρες σε διάφορα στάδια του συστήματος:

1. σύμμειξη των εξαγόμενων χαρακτηριστικών (feature fusion),
2. σύμμειξη της πληροφορίας μετά την κωδικοποίηση (encodings fusion) των χαρακτηριστικών,
3. σύμμειξη των βαθμολογιών που προκύπτουν κατά την ταξινόμηση των βίντεο (score fusion).

Τροποποιούμε τα γενικά πλαίσια των BoVW και VLAD προκειμένου να τα αξιοποιήσουμε κατάλληλα στην αναγνώριση δράσεων από πολλαπλές όψεις. Στο Σχήμα 3.7 βλέπουμε τις προσεγγίσεις που αναπτύξαμε για τη σύμμειξη πληροφορίας από πολλαπλές όψεις σε διαφορετικά στάδια του συστήματος αναγνώρισης.

Σύμμειξη χαρακτηριστικών

Στη μέθοδο αυτή, η σύμμειξη της οπτικής πληροφορίας γίνεται σε ένα αρχικό στάδιο στο οποίο έχουμε μόνο χαμηλού επιπέδου περιγραφητές με χαρακτηριστικά D-διαστάσεων $\mathbf{x}_m^i \in$



Σχήμα 3.7: Σύμμειξη της πληροφορίας από πολλαπλές όψεις σε διαφορετικά στάδια της αναγνώρισης δράσης: 1) σύμμειξη χαρακτηριστικών πυκνών τροχιών (feature fusion), 2) σύμμειξη πληροφορίας μετά την κωδικοποίηση (encodings fusion), 3) σύμμειξη των βαθμολογιών ταξινόμησης (score fusion).

$\mathbb{R}^D$, δηλαδή έχουμε $m = 1, \ldots, M_i$ τοπικούς περιγραφητές πυκνής τροχιάς από κάθε έναν από τους $i = 1, \ldots, S$ αισθητήρες. Παρ' ότι οι S αισθητήρες καταγράφουν ακριβώς τις ίδιες δράσεις, ο αριθμός σημείων-δειγμάτων που λαμβάνονται για να δημιουργηθούν οι τροχιές σε κάθε βίντεο δεν είναι σταθερός καθώς εξαρτάται από την οπτική ροή. Συνεπώς, δεν μπορούμε να κάνουμε απλά μια συνένωση (concatenation) των περιγραφητών και να δημιουργήσουμε έναν νέο περιγραφητή διάνυσμα $\tilde{\mathbf{x}}_m$ μεγέθους $D \cdot S$. Έτσι, τροποποιούμε τον τρόπο δημιουργίας του οπτικού λεξικού, της κωδικοποίησης των δεδομένων, η οποία βασίζεται στον αλγόριθμο K-means, προκειμένου να αντιμετωπίσουμε τα δεδομένα από τους πολλαπλούς αισθητήρες. Πιο συγκεκριμένα, δεδομένου ενός συνόλου χαρακτηριστικών-περιγραφητών $\mathbf{x}_m^i$, ο στόχος μας είναι να διαχωρίσουμε το σύνολο των χαρακτηριστικών σε K συστάδες (clusters) $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_K]$, όπου $\mathbf{c}_k \in \mathbb{R}^D$ είναι το κεντροειδές της k-ιοστής συστάδας. Τα κεντροειδή $\mathbf{c}_k$ είναι κοινά για την συσταδοποίηση των χαρακτηριστικών όλων των αισθητήρων. Αξιοποιώντας τον συμβολισμό του [Peng et al.; 2016], εάν ένας περιγραφητής $\mathbf{x}_m^i$ αντιστοιχίζεται στη συστάδα k τότε η βοηθητική τιμή - δείκτης είναι $r_{m,i,k} = 1$ ενώ $r_{m,i,\ell} = 0$ όταν $\ell \neq k$. Ο υπολογισμός των κεντροειδών $\mathbf{c}_k$ γίνεται ελαχιστοποιώντας το συναρτησιακό:

$$\min_{\mathbf{c}_k, r_{m,i,k}} \sum_{k=1}^{K} \sum_{i=1}^{S} \sum_{m=1}^{M_i} r_{m,i,k} \|\mathbf{x}_m^i - \mathbf{c}_k\|_2^2. \tag{3.3}$$

Στη συνέχεια, τροποποιούμε την διαδικασία κωδικοποίησης για τις μεθόδους BoVW και VLAD για να μπορέσουμε να τις εφαρμόσουμε στα δεδομένα πολλαπλών όψεων. Για ένα σύνολο περιγραφητών-χαρακτηριστικών $\mathbf{X}^i = [\mathbf{x}_1^i, \ldots, \mathbf{x}_{N_j}^i]$, που εξάγονται N τροχές από το j-ιοστό βίντεο και καταγράφονται από τον i-οστό αισθητήρα, η κωδικοποίηση της τροχιάς $\mathbf{x}_{n_j}^i$ σύμφωνα με την BoVW μέθοδο και το οπτικό λεξικό $\mathbf{C}$ δίνεται ως εξής:

$$\mathbf{s}_{n_j}^i(k) = 1, \text{ if } k = \underset{\ell}{\operatorname{argmin}} \|\mathbf{x}_{n_j}^i - \mathbf{c}_\ell\|_2^2, \text{ s.t. } \|\mathbf{s}_{n_j}^i\|_0 = 1. \tag{3.4}$$

Στην περίπτωση της μεθόδου VLAD, όπου κρατάμε τα στατιστικά πρώτης τάξης, η κωδι-

(α΄) Παντομίμα παιδιού για την εργασία σκάβω (β΄) Παντομίμα παιδιού για την εργασία κολυμπάω

Σχήμα 3.8: Δύο παραδείγματα δράσεων από την βάση δεδομένων πολλαπλών όψεων.

κοποίηση του $\mathbf{x}_{n_j}^i$ δίνεται ως εξής:

$$\mathbf{s}_{n_j}^i(k) = [\mathbf{0}, \ldots, \mathbf{x}_{n_j}^i - \mathbf{c}_k, \ldots, \mathbf{0}], \quad k = \underset{\ell}{\operatorname{argmin}} \|\mathbf{x}_{n_j}^i - \mathbf{c}_\ell\|_2^2. \tag{3.5}$$

Η ολική αναπαράσταση $\mathbf{h}$ του βίντεο από τις πολλαπλές όψεις χρησιμοποιώντας ένα αθροιστικό pooling σχήμα δίνεται:

$$\mathbf{h} = \sum_{i=1}^{S} \sum_{n_j=1}^{N_j} \mathbf{s}_{n_j}^i. \tag{3.6}$$

Τελικά, για την BoVW μέθοδο εφαρμόζουμε $L2$ κανονικοποίηση [Perronnin et al.; 2010] ενώ για την VLAD ακολουθούμε μια εσωτερική κανονικοποίηση σύμφωνα με τη στρατηγική που προτείνεται στο [Arandjelovic and Zisserman; 2013].

Σύμμειξη της πληροφορίας μετά την κωδικοποίηση

Σε αυτή την προσέγγιση έχουμε ένα διαφορετικό ολικό διάνυσμα $\mathbf{h}^i$ για κάθε αισθητήρα i. Αυτή η αναπαράσταση θα μπορούσε να είναι είτε μια κωδικοποίηση των χαρακτηριστικών πυκνής τροχιάς χρησιμοποιώντας ένα διαφορετικό οπτικό βιβλίο $\mathbf{C}^i$ για κάθε αισθητήρα ή ένα διάνυσμα χαρακτηριστικών που λαμβάνεται από διαφορετικό δίκτυο C3D. Για τις κωδικοποιήσεις BoVW εφαρμόζουμε τη σύμμειξη πολλαπλών όψεων προσθέτοντας τους πυρήνες χ^2:

$$K(\mathbf{h}_j, \mathbf{h}_q) = \sum_{i=1}^{S} \sum_{m=1}^{N_c} \exp\left(-\frac{1}{A_c} L\left(\mathbf{h}_j^{m,i}, \mathbf{h}_q^{m,i}\right)\right), \tag{3.7}$$

όπου $\mathbf{h}_j^{m,i}$ συμβολίζει την BoVW αναπαράστασης του m-ιοστού περιγραφητή του j-οστού βίντεο που καταγράφηκε από τον αισθητήρα i, και A_m είναι η μέση τιμή των χ^2 αποστάσεων $L(\mathbf{h}_j^{m,i}, \mathbf{h}_q^{m,i})$ ανάμεσα σε όλα τα ζεύγη των δειγμάτων εκπαίδευσης από έναν συγκεκριμένο αισθητήρα i. Στην περίπτωση της VLAD κωδικοποίησης καθώς και όταν αξιοποιούμε τα C3D εφαρμόζουμε μια απλή ένωση των διανυσμάτων (concatenation) που αντιστοιχούν στους διάφορους αισθητήρες: $\mathbf{h} = [\mathbf{h}^1, \ldots, \mathbf{h}^S]$.

Σύμμειξη των βαθμολογιών ταξινόμησης

Για ένα δεδομένο αισθητήρα i εκπαιδεύουμε ένα διαφορετικό SVM για όλες τις κλάσεις που χρησιμοποιούνται και λαμβάνουμε τις πιθανότητες $\mathbf{P}^i$ όπως περιγράφεται στο [Chang and Lin; 2011]. Στη συνέχεια εφαρμόζουμε μια κανονικοποίηση softmax στις πιθανότητες SVM

Δεδομένα Ανάπτυξης						
	Features		Encodings		Scores	
Features	BoW	VLAD	BoW	VLAD	BoW	VLAD
Traj.	75.00	76.39	76.85	79.63	77.31	78.24
HOG	39.81	40.28	41.67	39.35	41.20	37.04
HOF	71.76	74.07	77.78	81.48	75.93	81.94
MBH	76.39	76.85	81.02	81.48	82.87	**83.80**
Comb.	81.48	82.87	82.87	83.80	82.87	**85.19**
Δεδομένα Αλληλεπίδρασης						
	Features		Encodings		Scores	
Features	BoW	VLAD	BoW	VLAD	BoW	VLAD
Traj.	54.16	58.90	51.88	58.39	49.00	65.37
HOG	37.84	35.59	31.79	27.76	34.48	31.78
HOF	54.56	71.61	58.01	74.73	63.26	**74.83**
MBH	65.32	72.70	67.72	72.52	66.73	72.72
Comb.	61.51	69.85	63.38	**73.95**	64.82	73.35

Πίνακας 3.9: Μέση ακρίβεια αναγνώρισης χειρονομιών (%) για τις 8 χειρονομίες που εκτέλεσαν τα παιδιά. Αποτελέσματα για τα πέντε διαφορετικά είδη χαρακτηριστικών περιγραφητών και δύο κωδικοποιήσεων για το σύστημα αναγνώρισης χειρονομιών πολλαπλών όψεων.

κάθε αισθητήρα. Εναλλακτικά, στην περίπτωση που χρησιμοποιούμε ένα δίκτυο C3D, αυτές οι πιθανότητες θα μπορούσαν να ληφθούν από το τελευταίο στρώμα softmax. Για τη σύμμειξη των διαφορετικών πιθανοτήτων εξόδου αισθητήρα εφαρμόζουμε απλώς μια μέση σύμμειξη: $\mathbf{P} = \frac{1}{S}\sum_{i=1}^{S}\mathbf{P}^i$. Τέλος, επιλέγουμε την κατηγορία με την υψηλότερη συγχωνευμένη βαθμολογία, ακολουθώντας την ίδια προσέγγιση όπως και στην περίπτωση μεμονωμένου αισθητήρα.

3.3.2 Πειραματικά αποτελέσματα

Αξιολόγηση σχημάτων εκπαίδευσης σύμμειξης πυκνών τροχιών

Αρχικά αξιολογούμε το σύστημα αναγνώρισης ενεργειών σε δύο εφαρμογές, στην ακρίβεια αναγνώρισης των παιδικών χειρονομιών και των κινήσεων παντομίμας για τη σύμμειξη της πληροφορίας των τεσσάρων αισθητήρων με τη μέθοδο των πυκνών τροχιών. Στους Πίνακες 3.9 και 3.10 παρουσιάζονται τα αποτελέσματά των πειραμάτων για τα δύο σύνολα αναγνώρισης.

Συνολικά *παρατηρείται ότι η ακρίβεια της αναγνώρισης οπτικής δραστηριότητας είναι χαμηλότερη στα δεδομένα αλληλεπίδρασης,* καθώς τα παιδιά ενεργούν πιο αυθόρμητα ενώ κινούνται στο δωμάτιο και αλληλεπιδρούν ελεύθερα με τα ρομπότ, όπως είδαμε και στην προηγούμενη ενότητα στην αναγνώριση μονής όψης. Επιπλέον, δεδομένου ότι η διακύμανση των οπτικών πληροφοριών στην εργασία της παντομίμας είναι μεγαλύτερη από ό,τι στην εργασία με χειρονομίες, *η πρώιμη συγχώνευση των χαρακτηριστικών αποδίδει καλύτερα για την παντομίμα ενώ η συγχώνευση σε επίπεδο τελικών σκορ ταξινόμησης είναι ικανοποιητική για την αναγνώριση των χειρονομιών.*

Πιο συγκεκριμένα, το 3.9 παρουσιάζει αποτελέσματα μέσης ακρίβειας (%) για τις επτά κινήσεις και ένα μοντέλο τυχαίας κίνησης. *Τα αποτελέσματα υποδεικνύουν ότι το καλύτερο μοντέλο πολλαπλών όψεων υπερτερεί του καλύτερου μοντέλου μίας όψης κατά περίπου 7%, υπογραμμίζοντας την ανάγκη για ένα σύστημα πολλαπλών όψεων για πιο ελεύθερο CRI.* Τα δεδομένα ανάπτυξης δείχνουν ότι ο συνδυασμός διαφορετικών τύπων χαρακτηριστικών αποδίδει καλύτερα από ότι τα

Δεδομένα Ανάπτυξης						
	Features		Encodings		Scores	
Features	BoW	VLAD	BoW	VLAD	BoW	VLAD
Traj.	62.15	60.00	66.15	66.77	64.31	69.50
HOG	48.62	50.15	49.85	54.15	44.31	58.00
HOF	66.77	67.08	68.00	69.23	68.62	75.50
MBH	76.00	76.69	76.92	76.92	74.46	76.50
Comb.	75.08	76.92	77.23	77.85	75.08	**79.00**
Δεδομένα Αλληλεπίδρασης						
	Features		Encodings		Scores	
Features	BoW	VLAD	BoW	VLAD	BoW	VLAD
Traj.	58.26	52.30	50.79	54.87	47.45	56.99
HOG	37.41	31.49	26.95	36.26	29.14	31.44
HOF	63.08	61.02	49.87	56.17	52.59	57.99
MBH	**70.25**	67.97	57.70	59.04	62.18	62.49
Comb.	63.52	**69.37**	60.75	61.55	55.00	64.90

Πίνακας 3.10: Μέση ακρίβεια αναγνώρισης κινήσεων παντομίμας (%) για τις 13 παντομίμες που εκτέλεσαν τα παιδιά. Αποτελέσματα για τα πέντε διαφορετικά είδη χαρακτηριστικών περιγραφητών και δύο κωδικοποιήσεων για το σύστημα αναγνώρισης κινήσεων πολλαπλών όψεων.

μεμονωμένα χαρακτηριστικά. *Η καλύτερη ακρίβεια αναγνώρισης 85,2% παρατηρείται για τη σύμμειξη της πληροφορίας στο τελικό βήμα της διαδικασίας με τις κωδικοποιήσεις VLAD και τον συνδυασμό χαρακτηριστικών. Όσον αφορά τα δεδομένα της αλληλεπίδρασης,* η ακρίβεια είναι γύρω στο 12% χαμηλότερη στις περιπτώσεις κωδικοποίησης με VLAD (στην κωδικοποίηση με BoW είναι ακόμα μεγαλύτερη) από ότι στα δεδομένα ανάπτυξης, γεγονός που αποκαλύπτει τη δυσκολία που αντιμετώπισαν τα παιδιά όταν προσπαθούσαν να εκτελέσουν τη χειρονομία αυθόρμητα. *Η μεγαλύτερη απόδοση αναγνώρισης είναι ελαφρώς καλύτερη για τη σύμμειξη των κωδικοποιήσεων από τη σύμμειξη των σκορ και πλησιάζει το 74%.*

Για να επαληθεύσουμε την καταλληλότητα του προτεινόμενου συστήματος αναγνώρισης δράσεων σε πιο απαιτητικές εργασίες, αξιολογούμε το σύστημα στις παντομίμες. Ο πίνακας 3.10 παρουσιάζει τα αποτελέσματα μέσης ακρίβειας (%) για τις 12 παντομίμες και την τυχαία κίνηση. *Η σύμμειξη των πληροφοριών πολλαπλών όψεων βελτιώνει την απόδοση της αναγνώρισης,* όπως παρατηρήθηκε και στην περίπτωση χειρονομιών, κατά ένα μικρότερο όμως ποσοστό, περίπου 3%. Η υψηλότερη ακρίβεια για τον έλεγχο σε δεδομένα ανάπτυξης παρουσιάζεται με τις κωδικοποιήσεις VLAD στη σύμμειξη των σκορ, καθώς οι οπτικές πληροφορίες σε αυτά τα δεδομένα είναι πιο συνεπείς από τα δεδομένα που σχετίζονται με την περίπτωση χρήσης, π.χ. παρόμοια χρονική διάρκεια μιας παντομίμας ή παρόμοιες παιδικές τοποθεσίες στην αίθουσα.Επιπλέον, *η σύμμειξη χαρακτηριστικών, δηλαδή η συγχώνευση των χαρακτηριστικών σε ένα πρώιμο στάδιο της διαδικασίας έχει ως αποτέλεσμα την καλύτερη απόδοση του συστήματος ανεξάρτητα από τη μέθοδο κωδικοποίησης.* Η ακρίβεια αναγνώρισης είναι μεγαλύτερη κατά 9% από την καλύτερη επίδοση του συστήματος μονής όψης.

Αξιολόγηση σχημάτων εκπαίδευσης σύμμειξης συνελικτικών δικτύων

Στον Πίνακα 3.11 παρουσιάζουμε τα αποτελέσματα αξιολόγησης των μεθόδων σύμμειξης πολλαπλών όψεων για τα C3D χαρακτηριστικά. *Μπορούμε να δούμε ότι τα σχήματα σύμμειξης*

C3D Feats.	Σύμμειξη Πληροφορίας		
	Feature	Encodings	Score
conv5b	58.77	61.23	62.46
pool5	60.31	61.23	63.08
FC6	60.31	63.08	62.46
FC7	63.08	63.08	62.15
Comb.	60.31	61.23	**63.69**
end-to-end	-	-	61.72

Πίνακας 3.11: Μέση ακρίβεια αναγνώρισης κινήσεων παντομίμας (%) για τις 13 παντομίμες που εκτέλεσαν τα παιδιά. Αποτελέσματα της σύμμειξης της πληροφορίας στα διάφορα επίπεδα για τα διάφορα είδη C3D χαρακτηριστικών για το σύστημα αναγνώρισης κινήσεων πολλαπλών όψεων.

βελτιώνουν την απόδοση της αντίστοιχης μεθόδου μονής όψης σε όλες τις περιπτώσεις. Η σύμμειξη των σκορ έχουν την καλύτερη απόδοση στην περίπτωση που χρησιμοποιούμε τα χαρακτηριστικά C3D και μάλιστα έχουμε αύξηση περίπου 9% στην ακρίβεια αναγνώρισης όταν συνδυάζουμε όλα τα χαρακτηριστικά. Επίσης παρατηρείται αυξημένη απόδοση του συστήματος με χρήση των χαρακτηριστικών του επιπέδου FC7 ανεξάρτητα από τον τρόπο σύμμειξης της πληροφορίας των καμερών.

Γενικά, συγκρίνοντας τη σύμμειξη της πληροφορίας για τις μεθόδους πυκνών τροχιών και τις μεθόδους που απορρέουν από τη χρήση τρισδιάστατων συνελικτικών δικτύων μπορούμε να πούμε ξεκάθαρα πως οι πρώτες υπερέχουν για την αναγνώριση δράσεων σε αλληλεπιδράσεις παιδιών και ρομπότ. Αυτό μπορεί να εξηγηθεί αν λάβουμε ότι οι πυκνές τροχιές καταγράφουν χωροχρονικές τοπικές πληροφορίες, οι οποίες κωδικοποιούνται περαιτέρω για να σχηματίσουν μια καθολική αναπαράσταση, ενώ το C3D καταγράφει μια καθολική αναπαράσταση ενός κλιπ 16 στιγμιοτύπων που στη συνέχεια υπολογίζεται κατά μέσο όρο σε όλο το βίντεο.

3.4 Συστήματα επαυξητικής μάθησης

Η επαυξητική ή σταδιακή μάθηση (Incremental Learning - IL) είναι ένα παράδειγμα μηχανικής μάθησης όπου ένα μοντέλο εκπαιδεύεται σε νέα δεδομένα σταδιακά, με την πάροδο του χρόνου, αντί να εκπαιδεύεται το μοντέλο από την αρχή κάθε φορά που γίνονται διαθέσιμα νέα δεδομένα. Αυτή η προσέγγιση επιτρέπει στο μοντέλο να μαθαίνει διαρκώς νέα δεδομένα.

Μερικές γνωστές κατηγορίες [Goodfellow et al.; 2016, Shin et al.; 2017, Liu et al.; 2022] που αποτελούν μεθόδους επαυξητικής μάθησης είναι:

1. το online learning όπου το μοντέλο μαθαίνει από δεδομένα που φτάνουν σε συνεχή ροή,

2. το transfer learning, όπου ένα μοντέλο εκπαιδεύεται εκ των προτέρων σε ένα μεγάλο σύνολο δεδομένων και, στη συνέχεια, ρυθμίζεται με ακρίβεια σε ένα μικρότερο, σχετικό σύνολο δεδομένων ώστε να αξιοποιήσει τη γνώση που αποκτάται στο μεγαλύτερο σύνολο δεδομένων, επιταχύνοντας τη διαδικασία εκμάθησης και βελτιώνοντας την απόδοση στο μικρότερο σύνολο δεδομένων,

3. το lifelong learning, όπου ένα μοντέλο μαθαίνει μια σειρά εργασιών με την πάροδο του χρόνου, με στόχο τη διατήρηση της γνώσης που αποκτήθηκε από προηγούμενες εργασίες, ενώ προσαρμόζεται σε νέες.

Η σταδιακή μάθηση αν και έχει μια λογική μάθησης παρόμοια με την ανθρώπινη, και άρα ακούγεται ιδανική, στην πράξη εμπεριέχει πολλές δυσκολίες στην υλοποίησή της. Οι πιο σημαντικοί παράγοντες που πρέπει να ληφθούν υπ' όψιν κατά την ανάπτυξη ενός συστήματος επαυξητικής μάθησης είναι:

- Catastrophic forgetting (καταστροφική λήθη): αναφέρεται στο φαινόμενο όπου το μοντέλο ξεχνά τις γνώσεις που έχει μάθει προηγουμένως όταν εκπαιδεύεται σε νέα δεδομένα. Αυτό συμβαίνει καθώς το μοντέλο ενημερώνει τις παραμέτρους του στα νέα δεδομένα, προκαλώντας την αντικατάσταση των παλιών παραμέτρων που είχαν μάθει σε προηγούμενα δεδομένα. Η καταστροφική λήθη αποτελεί το σημαντικότερο πρόβλημα που πρέπει να αντιμετωπιστεί.

- Scalability (δυνατότητα επέκτασης): η σταδιακή μάθηση μπορεί να απαιτεί σημαντικούς υπολογιστικούς πόρους κάτι που μπορεί να επιφέρει μεγαλύτερη δυσκολία κατά την εκμάθηση μεγάλων συνόλων δεδομένων. Αυτό μπορεί να δυσκολέψει την εκπαίδευση και την ανάπτυξη μοντέλων σε σενάρια πραγματικού κόσμου.

- Επιλογή μοντέλου: απαιτείται προσεκτική επιλογή της αρχιτεκτονικής του μοντέλου και της εκπαιδευτικής προσέγγισης για να εξισορροπηθούν οι αντισταθμίσεις μεταξύ της πολυπλοκότητας, της απόδοσης και των περιορισμών μνήμης του μοντέλου.

Η υπέρβαση αυτών των δυσκολιών απαιτεί προσεκτική εξέταση της προσέγγισης της σταδιακής μάθησης και βαθιά κατανόηση των υποκείμενων αλγορίθμων και τεχνικών. Στη συνέχεια θα μελετήσουμε μερικούς αλγορίθμους επαυξητικής μάθησης ενώ θα τους συγκρίνουμε πειραματικά για να καταλήξουμε σε έναν αποδοτικό τρόπο προσέγγισης της επαυξητικής μάθησης σε ένα σύστημα αλληλεπίδρασης παιδιού-ρομπότ.

3.4.1 Επισκόπηση μεθόδων επαυξητικής μάθησης

Δεδομένου ότι δεν υπάρχουν σχετικές εργασίες σχετικά με την επαυξητική μάθηση για CRI, θα δώσουμε μια γενική επισκόπηση του πεδίου IL και των πιο δημοφιλών μεθόδων του, αρκετές από τις οποίες χρησιμοποιούνται και αξιολογούνται παρακάτω.

Μια ενδιαφέρουσα κατηγοριοποίηση των μεθόδων νευρωνικών δικτύων για συνεχή δια βίου μάθηση (continual lifelong learning) παρουσιάζεται στο [Parisi et al.; 2019] σύμφωνα με τον τρόπο με τον οποίο περιορίζουν την καταστροφική λήθη. Εννοιολογικά, αυτές οι προσεγγίσεις μπορούν να χωριστούν σε (α) μεθόδους κανονικοποίησης (regularization methods) που επιβάλλουν περιορισμούς στην ενημέρωση των τιμών των βαρών του νευρωνικού δικτύου (π.χ. [Kirkpatrick et al.; 2017, Li and Hoiem; 2017, Aljundi et al.; 2018]), (β) δυναμικές αρχιτεκτονικές δηλαδή αρχιτεκτονικές που αλλάζουν τις βασικές τους ιδιότητες όπως τον αριθμό των νευρώνων που χρησιμοποιούνται (π.χ. [Rusu et al.; 2016]) και (γ) σε άλλα συμπληρωματικά συστήματα εκμάθησης όπως οι μέθοδοι replayed memory (π.χ. [Rebuffi et al.; 2017], [Castro et al.; 2018], [Shin et al.; 2017], [Maracani et al.; 2021]).

Στο End-to-End Incremental Learning (EEIL) [Castro et al.; 2018], ένας συνδυασμός δεδομένων μνήμης (ονομάζεται επίσης experience replay) και distillation knowledge - η μέθοδος αυτή αρχικά προτάθηκε για μεταφορά μάθησης μεταξύ διαφορετικών δικτύων - χρησιμοποιήθηκε για την σταδιακή προσθήκη νέων εικόνων και νέων κατηγοριών σε ένα δίκτυο ταξινόμησης εικόνων. Στο Incremental Classifier and Representation Learning (iCaRL) [Rebuffi et al.; 2017] προτάθηκε ένα παρόμοιο σύστημα για IL για χρήση αποσυνδέοντας την αναπαράσταση των δεδομένων και του ταξινομητή. Από την άλλη πλευρά, το [Shin et al.; 2017] πρότεινε τη χρήση Generative Adversarial Networks (GAN) για να γίνει μίμηση των δεδομένων που είχε δει

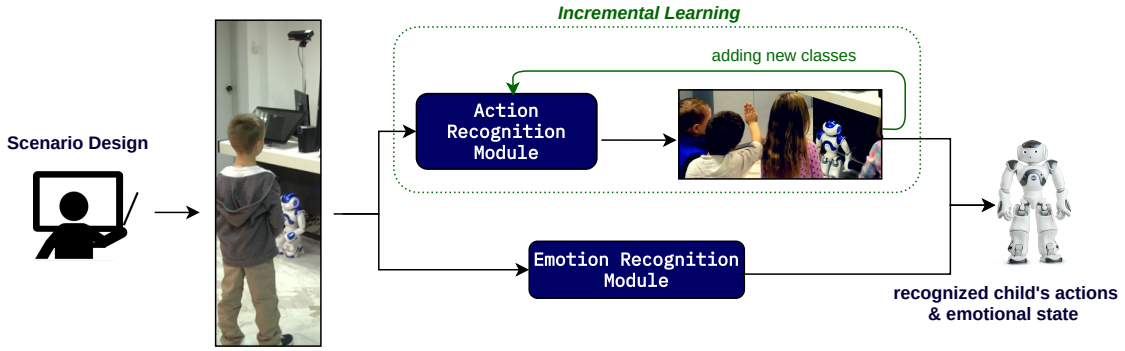
το μοντέλο στο παρελθόν, ενώ η εργασία [Van de Ven et al.; 2020] πρότεινε μια επανάληψη-αναπαραγωγή των εσωτερικών αναπαραστάσεων του μοντέλου εμπνευσμένη από τις λειτουργίες του εγκεφάλου. Στο [Belouadah and Popescu; 2019], οι συγγραφείς πρότειναν το δίκτυο Incremental Learning with Dual Memory (IL2M), το οποίο διορθώνει τις προβλέψεις χρησιμοποιώντας μια διπλή μνήμη η οποία βασίζεται στα αποθηκευμένα στατιστικά των προβλέψεων των κλάσεων που ήδη έχει δει το μοντέλο. Η μέθοδος Memory Aware Synapses (MAS) [Aljundi et al.; 2018] βασίζεται στον διαρκή υπολογισμό της σημαντικότητας των παραμέτρων του νευρωνικού δικτύου. Τέλος, η μέθοδος Learning without Forgetting (LwF) [Li and Hoiem; 2017] χρησιμοποίησε distillation knowledge και αξιοποίησε την έξοδο του δικτύου από παλιά δεδομένα σε νέα δεδομένα ώστε να αποφύγει να ξεχάσει να αναγνωρίζει παλιά δεδομένα.

Σχετικά με τη συνεχή μάθηση στο HRI, οι Churamani et al. [Churamani et al.; 2020] συζήτησαν τη σημασία του για τη δημιουργία ρομπότ που προσαρμόζουν τα ``συναισθήματά'' τους και πως μπορεί να αξιοποιηθεί για να μάθει να αντιλαμβάνεται και να συμπεριφέρεται λαμβάνοντας υπ' όψιν τις συνθήκες. Στο [Dehghan et al.; 2019], ένας ταξινομητής CNN που ανιχνεύει αντικείμενα εμπλουτίστηκε με δυνατότητες επαυξητικής μάθησης ώστε να έχει τη δυνατότητα να προστίθενται νέες κατηγορίες αντικειμένων για ταξινόμηση, ενώ στο [Zhang et al.; 2016], προτάθηκε η επαυξητική μάθηση με προσαρμογή μέσω της αλληλεπίδρασης των κοινωνικών ρομπότ με τους ανθρώπους. Το online incremental classification resonance network που παρουσιάζεται στο [Park and Kim; 2020] εμπλουτίζει το σύστημα αναγνώρισης προσώπων του ρομπότ Mybot αυξάνοντας σταδιακά τον αριθμό των προσώπων που μπορεί να αναγνωρίσει. Οι Tuyen et al. [Viet Tuyen et al.; 2018] χρησιμοποίησαν επίσης ένα μοντέλο σταδιακής μάθησης, το οποίο εντόπισε τα πολιτισμικά χαρακτηριστικά των ανθρώπων με τα οποία αλληλεπιδρούσε.Τέλος, οι Lesort et al. στο [Lesort et al.; 2020] συνοψίζουν αρκετές πραγματικές περιπτώσεις χρήσης συνεχούς μάθησης για ρομποτικές εφαρμογές, τους λόγους για την ανάπτυξη της σταδιακής μάθησης και τις προκλήσεις που αντιμετωπίζουν τέτοιες εργασίες.

3.4.2 Μέθοδος Επαυξητικής Μάθησης

Όταν ένα βαθύ νευρωνικό δίκτυο επανεκπαιδεύεται με νέες κλάσεις, υποφέρει από ***καταστροφική λήθη****: η γνώση για τις προηγούμενες κατηγορίες τείνει να ξεχαστεί και η απόδοσή του δικτύου να μειωθεί δραματικά.* Μια επιφανειακή προσέγγιση για να αντιμετωπιστεί θα εξέταζε την επανεκπαίδευση του δικτύου με ένα ολόκληρο σύνολο δεδομένων που περιέχει όλες τις νέες και παλιές κλάσεις προς ταξινόμηση. Ωστόσο, είναι εύκολο να καταλάβει κανείς ότι αυτό πολλές φορές γίνεται υπολογιστικά απαγορευτικό, ειδικά όταν χρειάζεται να προστίθενται συνεχώς νέες κλάσεις στο δίκτυο. Για να επιτρέψουμε στη μονάδα αναγνώρισης δράσεων του συστήματός μας να έχει σταδιακές δυνατότητες εκμάθησης, την συνδυάζουμε με ένα πλαίσιο επαυξητικής μάθησης (IL). Για να γίνει αυτό, επεκτείνουμε τη μέθοδο iCaRL [Rebuffi et al.; 2017], επιτρέποντάς της να εφαρμοστεί σε βίντεο σύμφωνα με το πλαίσιο TSN που υλοποιήσαμε παραπάνω. Στο Σχήμα 3.9 παρουσιάζεται η προτεινόμενη μέθοδος στο πλαίσιο του συστήματος TeachBot.

Η μέθοδος iCaRL είναι μια state-of-the-art τεχνική που συνδυάζει τα πλεονεκτήματα τόσο της αναπαράστασης χαρακτηριστικών όσο και της ταξινόμησης με χρήση νευρωνικών δικτύων και αποτελείται από δύο στάδια: ένα στάδιο εκπαίδευσης και ένα στάδιο ταξινόμησης. Κατά τη διάρκεια του σταδίου εκπαίδευσης, η μέθοδος μαθαίνει ένα νέο σύνολο κλάσεων και χρησιμοποιεί τα παλιά δεδομένα για να σχηματίσει υποδείγματα, τα οποία είναι μικρά σύνολα αντιπροσωπευτικών παραδειγμάτων των παλιών κλάσεων. Τα υποδείγματα χρησιμοποιούνται για τη διατήρηση της παλιάς γνώσης και τη μείωση των επιπτώσεων της καταστροφικής λήθης. Η μέθοδος iCaRL ενημερώνει τα βάρη του δικτύου ταξινόμησης με βάση τα νέα δεδομένα εκπαίδευσης και τα υποδείγματα. Με αυτόν τον τρόπο, το μοντέλο μπορεί να συνεχίσει να μαθαίνει



Σχήμα 3.9: Το σύστημα οπτικής αντίληψης του TeachBot με την προτεινόμενη μέθοδο επαυξητικής μάθησης για την προσθήκη νέων κλάσεων δράσης κατά την αλληλεπίδραση παιδιών και ρομπότ.

διατηρώντας παράλληλα τη γνώση που είχε μάθει προηγουμένως.

Τα σύνολα υποδειγμάτων κατασκευάζονται επιλέγοντας ένα υποσύνολο των παλαιών δεδομένων που είναι αντιπροσωπευτικό των παλαιών κλάσεων. Ο αλγόριθμος επιλέγει ένα σύνολο υποδειγμάτων που αντιπροσωπεύουν καλύτερα τις παλιές κατηγορίες και επίσης ελαχιστοποιούν την απόσταση μεταξύ των υποδειγμάτων και των δεδομένων εκπαίδευσης.

Κατά το στάδιο της ταξινόμησης, η μέθοδος iCaRL χρησιμοποιεί έναν ταξινομητή πλησιέστερου μέσου όρου των υποδειγμάτων για την ταξινόμηση των νέων παραδειγμάτων. Ο ταξινομητής υπολογίζει την απόσταση μεταξύ ενός νέου παραδείγματος και του μέσου όρου κάθε υποδειγματικού συνόλου και εκχωρεί το παράδειγμα στην κλάση με τη μικρότερη απόσταση. Ο ταξινομητής αυτός αποτελεί έναν αποτελεσματικό και αποδοτικό τρόπος ταξινόμησης νέων δεδομένων, καθώς δεν απαιτεί την αποθήκευση όλων των παλαιών δεδομένων, αλλά μόνο των συνόλων παραδειγμάτων.

Η μέθοδος iCaRL *συνδυάζει μια μέθοδο δυναμικής στάθμισης για να εξισορροπήσει τη σημασία των νέων και των παλαιών δεδομένων* κατά τη διάρκεια της εκπαιδευτικής διαδικασίας. Η μέθοδος στάθμισης δίνει μεγαλύτερη σημασία στα παλιά δεδομένα κατά τα πρώτα στάδια της εκπαιδευτικής διαδικασίας, μετατοπίζοντας σταδιακά προς τα νέα δεδομένα καθώς προχωρά η εκπαίδευση.

Ένα από τα βασικά πλεονεκτήματα της συγκεκριμένης μεθόδου είναι η ικανότητά της να χειρίζεται αποτελεσματικά μεγάλο αριθμό τάξεων. Αυτό επιτυγχάνεται καθώς χρησιμοποιεί έναν ταξινομητή δύο επιπέδων, όπου το πρώτο επίπεδο αποτελείται από έναν δυαδικό ταξινομητή που καθορίζει εάν ένα νέο παράδειγμα ανήκει σε μια νέα κλάση ή σε μια παλιά κλάση. Το δεύτερο επίπεδο χρησιμοποιεί τον ταξινομητή για να καθορίσει την κλάση του νέου παραδείγματος.

Πιο συγκεκριμένα, στη δική μας εκδοχή της iCaRL μεθόδου, πολλά υποδείγματα από τις κατηγορίες που ήδη έχει δει το σύστημα - και καλούνται exemplars - διατηρούνται και δημιουργούν ένα σύνολο που λέγεται *memory budget* B. Στη βιβλιογραφία [Masana et al.; 2020], περιγράφονται δύο μέθοδοι για τον ορισμό του B. Η πρώτη μέθοδος αφορά ένα σταθερό μέγεθος του budget ανεξάρτητο της μεταβολής του πλήθους των κλάσεων, ενώ στη δεύτερη προσέγγιση ο αριθμός των υποδειγμάτων ανά κλάση είναι σταθερός οπότε παρατηρείται μια γραμμική αύξηση του memory budget ανάλογη του αριθμού των κλάσεων στο σύστημα. Σε αυτή τη δεύτερη περίπτωση, αν ορίσουμε ως E τον αριθμό των υποδειγμάτων/κλάση και C τον συνολικό αριθμό των κλάσεων που έχει δει το σύστημα, το memory budget υπολογίζεται $B = C \cdot E$. Στη δική μας IL υλοποίηση, ακολουθούμε τη δεύτερη προσέγγιση καθώς στη συγκεκριμένη περίπτωση δεν αναμένεται το πλήθος των κλάσεων να γίνει πολύ μεγάλο. Η γραμμική αύξηση του προϋπολογισμού γίνεται εμπόδιο μόνο σε εφαρμογές μεγάλης κλίμακας.

Όταν λαμβάνει χώρα μια νέα φάση του IL για τη μονάδα αναγνώρισης δράσεων (δηλαδή, όταν χρειάζεται να εισαχθούν σε αυτήν μία ή περισσότερες νέες κλάσεις), τα δείγματα των νέων κλάσεων συνδυάζονται με το υποδείγματα της μνήμης για να σχηματίσει το συνδυασμένο σύνολο δεδομένων, το οποίο στη συνέχεια χρησιμοποιείται για την επανεκπαίδευση της μονάδας αναγνώρισης ενεργειών. Κατά τη διάρκεια της εκπαίδευσης, το δίκτυο μαθαίνει να ελαχιστοποιεί τόσο το cross-entropy loss όσο και το distillation loss για τις παλιές κλάσεις [Rebuffi et al.; 2017]. Για την επιλογή των υποδειγμάτων κάθε κλάσης γίνεται τυχαία δειγματοληψία. Διάφορες εργασίες έχουν δείξει [Castro et al.; 2018, Masana et al.; 2020] ότι η τυχαία δειγματοληψία επιτυγχάνει συγκρίσιμα αποτελέσματα με άλλες προσεγγίσεις όπως το herding [Welling; 2009] που χρησιμοποιεί την απόσταση των δειγμάτων κλάσης από το μέσο υπόδειγμα.

Όπως αναφέραμε η επαυξητική μέθοδος επεκτείνει τη μέθοδο μονής όψης που παρουσιάσαμε στο 3.2.2, δηλαδή τη μέθοδο που προέκυψε από το συνδυασμό συνελικτικών δικτύων και χρονικής δειγματοληψίας. Σε αυτή την υλοποίηση, το κανάλι που μας ενδιαφέρει (RGB ή οπτική ροή) του αρχικού βίντεο V χωρίζεται σε K διαφορετικά τμήματα (segments) $\{S_1, S_2, ..., S_K\}$ ίσης διάρκειας, από τα οποία στη συνέχεια δειγματοληπτούνται ισάριθμα αποσπάσματα του βίντεο T_k μήκους N διαδοχικών frames. Στη συνέχεια, εφαρμόζεται ένα CNN σε κάθε απόσπασμα, το οποίο αναπαρίσταται ως $F(T_k; \boldsymbol{W_{cnn}})$, και παράγεται ένα διάνυσμα χαρακτηριστικών $L_k(V)$ για κάθε ένα από αυτά.

Σε αντίθεση με άλλες προσεγγίσεις επαυξητικής μάθησης, η ταξινόμηση νέων δειγμάτων στο iCaRL (και η δική μας μέθοδος που βασίζεται σε αυτό) χρησιμοποιεί έναν representation-based ταξινομητή. Για κάθε κλάση το σύστημα διατηρεί στη μνήμη ένα πρωτότυπο διάνυσμα M_i, το οποίο είναι ο μέσος όρος όλων των διανυσμάτων των χαρακτηριστικών των υποδειγμάτων της κλάσης, $M_i = \frac{1}{E}\sum_{j=1}^{E} L_j(V)$. Συνδυάζοντάς τα παραπάνω με το πλαίσιο TSN το διάνυσμα χαρακτηριστικών του υποδείγματος-exemplar $L(V)$ υπολογίζεται ως εξής:

$$L(V) = G(L_k(V)) = G(F(T_k; \boldsymbol{W_{cnn}})|_{k\in K}). \tag{3.8}$$

Έτσι, το κάθε νέο βίντεο V_n αποδίδεται στην κλάση που ελαχιστοποιεί την απόσταση από το διάνυσμα χαρακτηριστικών του υποδείγματος της $\text{argmin}_{i=1..C} \|L(V) - M_i\|$.

3.4.3 Πειραματικά Αποτελέσματα

Για να αξιολογήσουμε τη μέθοδο IL, δημιουργούμε ένα αυξημένο και πιο απαιτητικό σύνολο δεδομένων. *Συγκεκριμένα δημιουργήσαμε ένα νέο σύνολο δεδομένων που αποτελείται και από τα δεδομένα χειρονομιών όσο και από τα δεδομένα κινήσεων παντομίμας όπως καταγράφονται από την ίδια κάμερα (Kinect #1) για τα δεδομένα των παιδιών.* Μετά τη συγχώνευση, το **επαυξημένο σύνολο δεδομένων περιλαμβάνει πλέον συνολικά 20 κλάσεις**. Για να επιταχύνουμε τη διαδικασία εκπαίδευσης και αξιολόγησης, δημιουργούμε ένα διαχωρισμό εκπαίδευσης/αξιολόγησης με 20 παιδιά στο σετ εκπαίδευσης και 5 παιδιά στο σετ ελέγχου (αντί για τη leave-one-child-out cross-validation μέθοδο που χρησιμοποιούσαμε στα προηγούμενα πειράματα και θα αύξανε πολύ τον αριθμό των μοντέλων που απαιτούνται για την εκπαίδευση).

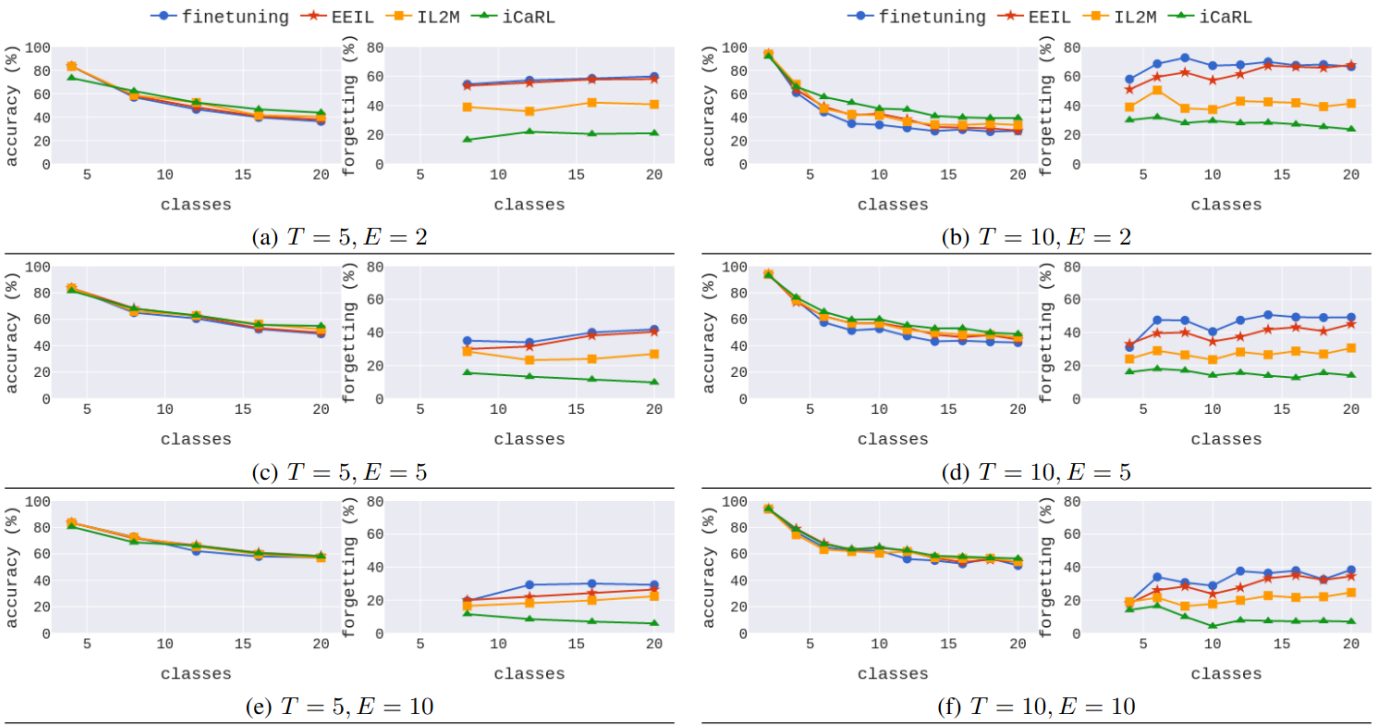
Στην επαυξητική εκδοχή του συστήματος αναγνώρισης δράσεων ακολουθούμε την ίδια αρχιτεκτονική σύστημα που αναπτύξαμε στο 3.2.1. Χρησιμοποιούμε δηλαδή μια αρχιτεκτονική BNInception με προεκπαίδευση στη βάση δεδομένων αναγνώρισης δράσεων Kinetics [Carreira and Zisserman; 2017]. Εκπαιδεύουμε κάθε δίκτυο 60 εποχές σε κάθε φάση της επαυξητικής μάθησης με τη μέθοδο στοχαστικής καθόδου κλίσης (Stochastic Gradient Descent - SGD) με συνάρτηση κόστους cross-entropy loss και μείωση του ρυθμού μάθησης στις 20 και 40 εποχές. Επαναλαμβάνουμε όλα τα πειράματα 10 φορές για να υπολογίσουμε το μέσο όρο για να πάρουμε την τελική εκτίμηση για την απόδοση του συστήματος για τα διάφορα υποσύνολα που

προκύπτουν προσθέτοντας νέες κλάσεις στις κλάσεις που ήδη έχει δει το σύστημα. Η σειρά των κλάσεων σε κάθε εκτέλεση επιλέγεται τυχαία.

Όπως είδαμε κατά την αξιολόγηση στον Πίνακα 3.7, *η ροή πληροφορίας RGB προσφέρει ελάχιστη αύξηση στην ακρίβεια*, αλλά απαιτεί σημαντικό χρόνο για την εκπαίδευση. Ως αποτέλεσμα, *επιλέγουμε να καταργήσουμε τη λειτουργία RGB στο σύστημα αναγνώρισης τελικής δράσης και να χρησιμοποιήσουμε μόνο την οπτική ροή για τα πειράματα IL.*

Μελέτη πλήθους δειγμάτων ανά κλάση Η πρώτη μας μελέτη αξιολόγησης συγκρίνει τη μέθοδο iCaRL για TSN με άλλες μεθόδους που χρησιμοποιούν experience replay, δηλαδή επανάληψη των δειγμάτων που έχει δει το σύστημα, τροποποιημένες για το πλαίσιο TSN: EEIL [Castro et al.; 2018], IL2M [Belouadah and Popescu; 2019] , και απλό fine-tuning (δηλαδή, απλή εκπαίδευση στο νέο σύνολο δεδομένων). Τα συγκριτικά αποτελέσματα για διάφορους αριθμούς παραδειγμάτων ανά κλάση ($E = 2, 5$ και 10) και δύο διαφορετικούς αριθμούς συνολικών σταδίων επαυξητικής μάθησης ($T = 5, 10$) φαίνονται στο Σχήμα 3.10.

Ανάλογα με τον συνολικό αριθμό των φάσεων/σταδίων, προστίθεται διαφορετικό πλήθος κλάσεων ανά φάση (ο συνολικός αριθμός κλάσεων 20 διαιρεμένος με τον αριθμό των φάσεων). Για 5 συνολικά στάδια, προστίθενται 4 νέες κλάσεις ανά στάδιο εκμάθησης. Στα διαγράμματα ακρίβειας (accuracy) ο άξονας x ξεκινά από $x = 4$ ενώ στα διαγράμματα που παρουσιάζεται η καταστροφική λήθη (forgetting) ο άξονας x ξεκινά όπως είναι λογικό μια φάση αργότερα στα $x = 8$. Ομοίως, για $T = 10$, ο αριθμός νέων κλάσεων που προστίθενται ανά φάση είναι 2, ο άξονας x στα διαγράμματα ακρίβειας ξεκινά από $x = 2$ και ο άξονας x για τη λήθη ξεκινά από $x = 4$.



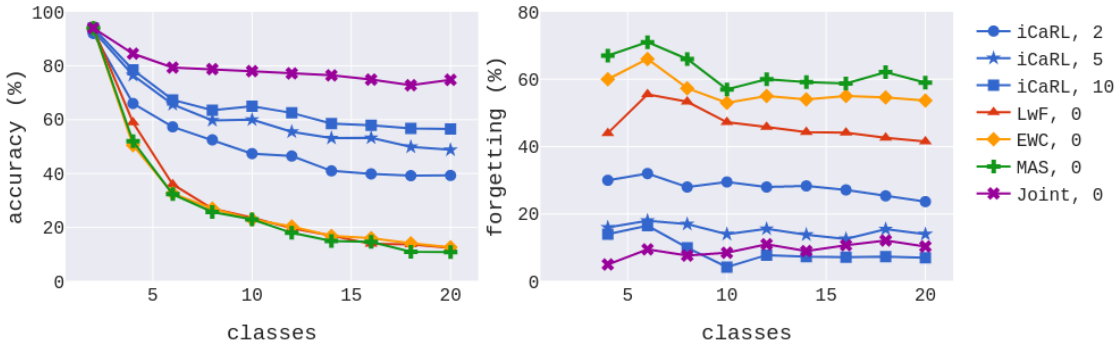
(a) $T = 5, E = 2$ (b) $T = 10, E = 2$

(c) $T = 5, E = 5$ (d) $T = 10, E = 5$

(e) $T = 5, E = 10$ (f) $T = 10, E = 10$

Σχήμα 3.10: Σύγκριση του προτεινόμενου εκτεταμένου iCaRL για βίντεο έναντι εναλλακτικών αλγορίθμων που χρησιμοποιούν experience replay με διαφορετικό αριθμό υποδειγμάτων/κλάση (E). Η αριστερή στήλη δείχνει αποτελέσματα για συνολικά $T = 5$ στάδια IL και η δεξιά στήλη για $T = 10$ στάδια.

Μπορούμε να παρατηρήσουμε ότι το εκτεταμένο iCaRL για TSN εμφανίζει το μικρότερο ποσοστό λήθης σε όλες τις διαφορετικές ρυθμίσεις και την υψηλότερη ακρίβεια στις περισσότερες περιπτώσεις. Είναι ενδιαφέρον ότι καθώς αυξάνουμε τον αριθμό των δειγμάτων, όλες οι μέθοδοι IL παρουσιάζουν ανταγωνιστικά αποτελέσματα όσον αφορά την ακρίβεια. Ωστόσο, το EEIL και το fine-tuning παρουσιάζουν σημαντικό πρόβλημα καταστροφικής λήθης. Αυτό σημαίνει ότι *το iCaRL (και το IL2M σε κάποιο βαθμό) σε αυτές τις περιπτώσεις έχουν μια εγγενή αντιστάθμιση, προσπαθώντας να εξισορροπήσουν την απόδοση στις κλάσεις που έγινε η εκμάθησή τους πρόσφατα με τις παλιότερες κλάσεις.* Από την άλλη πλευρά, το EEIL και το fine-tuning αγνοούν σε μεγάλο βαθμό την παλιά γνώση και επιτυγχάνουν υψηλές επιδόσεις μόνο σε νέες κατηγορίες. Πιστεύουμε ότι το iCaRL για TSN, αν και χρησιμοποιεί τυχαία τμήματα βίντεο για τη δημιουργία της αναπαράστασης βίντεο, αντλεί τη δύναμή του από τον ταξινομητή που βασίζεται στην εκμάθηση αναπαράστασης (representation-based classifier).



Σχήμα 3.11: Αξιολόγηση του εκτεταμένου μοντέλου iCaRL για βίντεο έναντι των μεθόδων κανονικοποίησης ($T = 10$).

Μελέτη απόδοσης έναντι μεθόδων κανονικοποίησης Στο Σχήμα 3.11, συγκρίνουμε επίσης την εκτεταμένη μέθοδο iCaRL για TSN με μεθόδους που βασίζονται σε κανονικοποίηση και δεν χρησιμοποιούν experience replay. Αυτές είναι Learning without Forgetting (LwF) [Li and Hoiem; 2017], Elastic Weight Consolida-tion(EWC) [Kirkpatrick et al.; 2017] και Memory Aware Synapsis (MAS) [Aljundi et al.; 2018]. Στο ίδιο σχήμα, δείχνουμε επίσης το αποτέλεσμα της εκπαίδευσης κάθε φάσης με το πλήρες σύνολο δεδομένων (δηλαδή, όταν δεν πραγματοποιείται επαυξητική μάθηση - αναφέρεται ως "Joint"). *Το iCaRL με τα παραδείγματα $E = 5$ και 10 παρουσιάζει συγκρίσιμο ποσοστό λήθης με το "Joint", το οποίο ενισχύει περαιτέρω την άποψή μας σχετικά με την υπεροχή του ταξινομητή που βασίζεται στο representation learning.* Οι άλλες μέθοδοι υποφέρουν σημαντικά από καταστροφική λήθη και καθώς ο αριθμός των τάξεων αυξάνεται, παρουσιάζουν κακή απόδοση.

Μελέτη χρόνου εκπαίδευσης και απόδοση Τέλος, ο Πίνακας 3.12 παρουσιάζει τα συγκεντρωτικά αποτελέσματα των μεθόδων που εξετάσαμε, συμπεριλαμβανομένης της μέσης ακρίβειας και της λήθης σε όλες τις φάσεις, καθώς και του μέσου χρόνου που απαιτείται για την εκπαίδευση μιας φάσης. *Μπορούμε να δούμε ότι το iCaRL επιτυγχάνει την υψηλότερη μέση ακρίβεια και τη λιγότερη λήθη, ενώ έχει παρόμοια υπολογιστικό κόστος σε σύγκριση με τις άλλες μεθόδους που χρησιμοποιούν υποδείγματα.* Από την άλλη πλευρά, ενώ οι μέθοδοι κανονικοποίησης είναι υπολογιστικά αποδοτικές, έχουν κακή απόδοση. *Τέλος, όταν συγκρίνουμε τη μέθοδο "Joint" (δηλαδή, χωρίς IL) με το iCaRL με παραδείγματα 10, παρατηρούμε μια σχετική μείωση του χρόνου ανά φάση*

Method	# Exemplars	Accuracy(%)	Forgetting(%)	Time(s)
EEIL [Castro et al.; 2018]	2	45.06	62.09	443
	5	58.44	39.47	600
	10	65.05	28.73	845
Fine-tuning	2	41.16	67.34	289
	5	54.98	45.77	365
	10	63.18	32.78	503
Proposed Method	2	52.11	28.00	305
(iCaRL [Rebuffi et al.; 2017])	5	61.55	15.17	387
	10	**66.06**	**9.04**	533
IL2M [Belouadah and Popescu; 2019]	2	46.37	41.41	292
	5	58.92	27.09	370
	10	64.03	20.59	507
EWC [Kirkpatrick et al.; 2017]	0	30.72	56.51	252
LwF [Li and Hoiem; 2017]	0	31.62	46.50	253
MAS [Aljundi et al.; 2018]	0	29.66	62.22	252
Joint (Πλήρης Εκπαίδευση χωρίς IL)	–	**79.06**	**9.32**	873

Πίνακας 3.12: Μέση ακρίβεια (accuracy), καταστροφική λήθη (forgetting) και χρόνος που απαιτείται για μία φάση της επαυξητικής μάθησης για το εκτεταμένο iCaRL για TSN και άλλες IL μεθόδους ($T = 10$). Η υλοποίηση των παραπάνω μεθόδων για τη σύγκριση τους έγινε από εμάς και τα πειράματα πραγματοποιήθηκαν στο σύνολο που προέκυψε από την ένωση των συνόλων των χειρονομιών και των δράσεων παντομίμας των παιδιών.

39%, *αλλά μόνο μια σχετική μείωση της ακρίβειας* 16% *και παρόμοια ποσοστά λήθης, επιβεβαιώνοντας την απόφασή μας για τη δημιουργία ενός πλαισίου επαυξητικής μάθησης για την αναγνώριση δράσεων.*

3.4.4 Συμπεράσματα

Συνοψίζουμε τα πιο σημαντικά ευρήματά μας για την state-of-the-art εκδοχή του συστήματος αναγνώρισης δράσεων που βασίζεται σε TSN λογική και συνδυάζεται με την iCarl μέθοδο επαυξητικής μάθησης και επισημαίνουμε παράγοντες που επηρεάζουν την απόδοσή του.

- Η επιλογή της βάσης δεδομένων στην οποία είναι προεκπαιδευμένα τα μοντέλα οπτικής αντίληψης επηρεάζει σε μεγάλο βαθμό την ακρίβεια αναγνώρισης. Η αποτελεσματικότητα του συστήματος ενισχύεται σημαντικά με τη χρήση προεκπαιδευμένων μοντέλων σε σύνολα δεδομένων που σχετίζονται άμεσα με τις επιθυμητές μας εργασίες αναγνώρισης. Πιο συγκεκριμένα, για την αναγνώριση ενεργειών, η προεκπαίδευση στο σύνολο δεδομένων Kinetics (που περιλαμβάνει ανθρώπινες ενέργειες) έχει ως αποτέλεσμα καλύτερη απόδοση αναγνώρισης ενεργειών σε σύγκριση με την προεκπαίδευση στο σύνολο δεδομένων ImageNet (για την αναγνώριση αντικειμένων). Ομοίως, για την εργασία αναγνώρισης συναισθημάτων, η προεκπαίδευση στο σύνολο δεδομένων εκφράσεων του προσώπου AffectNet ενισχύει σημαντικά την απόδοση του συστήματος, σε σύγκριση με την προεκπαίδευση ImageNet.

- Ο αριθμός των τμημάτων του δείγματος για το TSN συσχετίζεται σε μεγάλο βαθμό με

την εργασία αναγνώρισης (π.χ. αναγνώριση δράσης ή αναγνώριση συναισθήματος). Για την αναγνώριση δράσεων, σημειώνουμε ότι υπάρχει σημαντική (αλλά φθίνουσα) βελτίωση στην ακρίβεια αυξάνοντας τον αριθμό των τμημάτων. Η μέση διάρκεια των βίντεο στο σύνολο δεδομένων δράσεων είναι 4,23 δευτερόλεπτα. Πράγματι, περισσότερα δείγματα τμημάτων ενός βίντεο δράσης συνεπάγονται μια πιο ολοκληρωμένη κατανόηση της παρουσιαζόμενης ενέργειας, καθώς συνήθως μια ενέργεια αποτελείται από πολλές διαφορετικές κινήσεις. Παρ' όλα αυτά η χρήση της λογικής της χρονικής δειγματοληψίας έδειξε πως δεν είναι απαραίτητη η διατήρηση του συνόλου του εκάστοτε βίντεο.

- Όσον αφορά τις ροές πληροφοριών, πειραματιστήκαμε με μια χωρική ροή που λαμβάνει ως είσοδο καρέ βίντεο RGB και μια χρονική που παίρνει την οπτική ροή που προέρχεται από το βίντεο ως είσοδο. Τα πειραματικά αποτελέσματα καταδεικνύουν ότι η κύρια ροή πληροφοριών για την αναγνώριση δράσεων είναι η χρονική, με τη χωρική ροή να προσφέρει μια μικρή μόνο ώθηση απόδοσης.

- Όσον αφορά τη σταδιακή μάθηση, συγκρίνουμε διάφορες μεθόδους που χρησιμοποιούν experience replay (όπως το προτεινόμενο εκτεταμένο iCaRL για TSN) και άλλες που επιβάλλουν περιορισμούς στην ενημέρωση των δικτύων. Η καταστροφική λήθη είναι μεγαλύτερη στις μεθόδους κανονικοποίησης, ενώ οι μέθοδοι memory-replay τείνουν να αποδίδουν καλύτερα. Το προτεινόμενο εκτεταμένο iCaRL για TSN πέτυχε το χαμηλότερο ποσοστό λήθης σε όλες τις περιπτώσεις που μελετήσαμε και την υψηλότερη ακρίβεια στις περισσότερες από αυτές. Επιπλέον, πραγματοποιήσαμε επίσης μελέτες σχετικά με το μέγεθος της μνήμης και τον αντίκτυπό της στην ακρίβεια, τη λήθη και τον χρόνο εκπαίδευσης, αποδεικνύοντας την αποτελεσματικότητα του προτεινόμενου συστήματος. Οι μέθοδοι που χρησιμοποιούν δυναμικές αρχιτεκτονικές δεν έχουν διερευνηθεί ακόμη, καθώς θεωρήθηκαν πιο απαιτητικές υπολογιστικά.

3.5 Συμπεράσματα Κεφαλαίου

Στο Κεφάλαιο αυτό παρουσιάσαμε μια πληθώρα μεθόδων, κλασικών αλλά και πιο σύγχρονων, για την αντιμετώπιση της αναγνώρισης παιδικών δράσεων από έναν ή πολλαπλούς αισθητήρες. Ο πειραματισμός μας ήταν εκτενής και οδήγησε στην εξαγωγή πολλών και χρήσιμων συμπερασμάτων. Ιδιαίτερα σημαντική ήταν και η συνεισφορά μας στο θέμα της επαυξητικής μάθησης αφού, απ' όσο γνωρίζουμε, πρώτη φορά χρησιμοποιούνται και συγκρίνονται εκτενώς μέθοδοι επαυξητικής μάθησης σε δεδομένα αλληλεπιδράσεων παιδιών και ρομπότ.

Πιο συγκεκριμένα, παρατηρήσαμε πως ο συνδυασμός τεσσάρων ειδών χαρακτηριστικών καθώς και η χρήση VLAD κωδικοποίησης υπερτερεί συνολικά έναντι των υπολοίπων handcrafted μεθόδων που μελετήθηκαν για το πρόβλημα της αναγνώρισης δράσεων μονής όψης, τόσο στα δεδομένα ανάπτυξης όσο και στα δεδομένα αλληλεπίδρασης. Ως προς τις μεθόδους βαθιών νευρωνικών δικτύων, η χρήση συνελικτικών δικτύων σε συνδυασμό με χρονική δειγματοληψία των βίντεο από δύο κανάλια πληροφορίας έφερε πολύ καλά αποτελέσματα βελτιώνοντας αυτά των κλασικών μεθόδων από μια όψη και διατηρώντας χαμηλά το υπολογιστικό κόστος του προτεινόμενου συστήματος.

Ως προς την αναγκαιότητα συλλογής δεδομένων δράσεων από παιδιά παρατηρήσαμε πως είναι επιβεβλημένη και αναντικατάστατη. Η επιπρόσθετη συλλογή και χρήση δεδομένων ενηλίκων μπορεί να προσφέρει στη γενίκευση των μοντέλων, κυρίως στις χειρονομίες που παρουσιάζουν μικρότερη μεταβλητότητα μεταξύ παιδιών και ενηλίκων. Στο κομμάτι της σύμμειξης της πληροφορίας από πολλούς οπτικούς αισθητήρες, παρατηρήσαμε στα δεδομένα ανάπτυξης η σύμμειξη σε επίπεδο βαθμολογιών των επιμέρους βίντεο βοηθά στην αναγνώριση των δρά-

σεων ενώ στα δεδομένα ανάπτυξης βοηθά η σύμμειξη να γίνεται σε κάποιο στάδιο πριν από την εξαγωγή των βαθμολογιών.

Τέλος, η επέκταση του συστήματος αναγνώρισης δράσεων με χρήση επαυξητικής μεθόδου έδωσε ενθαρρυντικά αποτελέσματα για τη χρήση του σε αλληλεπιδράσεις παιδιών και ρομπότ καθώς έδειξε χαμηλό ποσοστό λήθης και υψηλή ακρίβεια αναγνώρισης.

Μέρος ΙΙ

Αναγνώριση Δράσεων Με Χρήση Βιοσημάτων

4

Αντίληψη Δράσεων σε Εφαρμογές Ηλεκτρονικής Υγείας

4.1 Επισκόπηση Αντιληπτικών Συστημάτων σε Εφαρμογές Ηλεκτρονικής Υγείας

Τα συστήματα αυτόματης αντίληψης δράσεων που σχετίζονται με την υγεία, αφορούν τεχνολογικά εργαλεία που έχουν σχεδιαστεί για την παρακολούθηση ποσοτικών δεικτών ικανών να περιγράφουν επιμέρους στοιχεία της υγείας των ανθρώπων χωρίς όμως να απαιτείται κάποια συνεισφορά από τους ίδιους. Τα συστήματα αυτά συνήθως συναντώνται ως φορητές (portable) ή φορετές (wearable) συσκευές αλλά και ως εφαρμογές σε κινητές συσκευές (mobile) και συχνά αξιοποιούνται στο πλαίσιο της ηλεκτρονικής παρακολούθησης της υγείας των ανθρώπων στην καθημερινότητά τους. Πιο συγκεκριμένα, *ο τομέας της ηλεκτρονικής υγείας (e-health)*, όπως ορίζεται από τον Παγκόσμιο Οργανισμό Υγείας [WHO;], *αφορά στην αποδοτική και ασφαλή χρήση των τεχνολογιών πληροφορίας και επικοινωνιών για την υποστήριξη της υγείας, συμπεριλαμβανομένης της υγειονομικής περίθαλψης, της παρακολούθησης, της αγωγής και της έρευνας.* Η χρήση των κινητών ασύρματων τεχνολογιών στη δημόσια υγεία, που συχνά αναφέρεται ως *mobile health (m-health)*, είναι αναπόσπαστο μέρος της ηλεκτρονικής υγείας και είναι ιδιαίτερα σημαντική λόγω της *ευκολίας χρήσης, της ευρείας εμβέλειας και αποδοχής της από τους χρήστες.*

Έτσι, *συνδυάζοντας αισθητήρες, ασύρματες τεχνολογίες, προηγμένους αλγόριθμους και μοντέλα μηχανικής μάθησης αναπτύσσονται εύρωστα συστήματα που συλλέγουν και αναλύουν δεδομένα σχετικά με τη σωματική δραστηριότητα του ατόμου,* την ποιότητα του ύπνου του και άλλους πολυάριθμους δείκτες. Οι πληροφορίες που προκύπτουν μπορούν να χρησιμοποιηθούν για τη διάγνωση ασθενειών, την παρακολούθηση χρόνιων παθήσεων και την παροχή εξατομικευμένων συστάσεων για αλλαγές στον τρόπο ζωής. Καθώς ο τομέας της υγειονομικής περίθαλψης ενσωματώνει ολοένα και περισσότερο την ψηφιακή καινοτομία, τα συστήματα αυτόματης αντίληψης που σχετίζονται με την υγεία έχουν τη δυνατότητα να βοηθήσουν στην πρόληψη ασθενειών, να αποτελέσουν πολύτιμα ιατρικά εργαλεία και να βελτιώσουν την ποιότητα ζωής των ανθρώπων.

Τα προηγούμενα χρόνια, η πανδημία COVID-19 έφερε σημαντική αύξηση στη χρήση των συστημάτων ηλεκτρονικής υγείας παγκοσμίως. Μια μελέτη των Wang et al. [Wang et al.; 2022] που διεξήχθη στην επαρχία Χουμπέι της Κίνας, το αρχικό επίκεντρο της πανδημίας, αποκαλύπτει τον τρόπο με τον οποίο αυτή η κρίση έχει αναδιαμορφώσει τις πρακτικές υγειονομικής περίθαλψης. Εξωτερικοί παράγοντες, όπως ο αυξημένος κίνδυνος μόλυνσης σε εξωτερικούς χώρους, οι διακοπές στις μετακινήσεις και οι ασφυκτικά γεμάτες εγκαταστάσεις υγειονομι-

κής περίθαλψης που έδιναν προτεραιότητα στους ασθενείς με COVID-19, οδήγησαν πολλούς ασθενείς στο να στραφούν σε ψηφιακές λύσεις υγείας. Οι ψηφιακές πλατφόρμες λειτούργησαν αποτελεσματικά ως υποκατάστατο της παραδοσιακής ιατρικής περίθαλψης, συγκεντρώνοντας υψηλή ικανοποίηση μεταξύ των ασθενών αλλά και την προθυμία τους να συνεχίσουν να τις χρησιμοποιούν μετά την πανδημία. Αντίστοιχα η εργασία των Asadzadeh et al. [Asadzadeh and Kalankesh; 2021] υπογραμμίζει τον καθοριστικό ρόλο που έχουν διαδραματίσει οι εφαρμογές m-health στην επιτάχυνση της διάγνωσης και του προληπτικού ελέγχου της COVID-19 παρέχοντας κρίσιμη ιατρική καθοδήγηση, ενώ στην εργασία [Abbaspur-Behbahani et al.; 2022], μελετήθηκε η χρήση τέτοιων εφαρμογών από τους ηλικιωμένους και διαπιστώθηκε πως η χρήση τους επιτάχυνε την παροχή υπηρεσιών υγείας και μείωσε τον κίνδυνο νοσηρότητας και θνησιμότητα κατά τη διάρκεια αυτής της άνευ προηγουμένου παγκόσμιας κρίσης υγείας.

Επιπρόσθετα, η εξέλιξη των τεχνολογιών των ασύρματων δικτύων, π.χ. δίκτυα 5G, έφεραν μεγάλη αλλαγή στις m-health εφαρμογές προσφέροντας μεγαλύτερες ταχύτητες δεδομένων, μειωμένη καθυστέρηση και βελτιωμένη συνδεσιμότητα [Devi et al.; 2023]. Ένα βήμα πιο πέρα, τα δίκτυα 6G θα επιτρέψουν την απομακρυσμένη παρακολούθηση ασθενών σε πραγματικό χρόνο υποστηρίζοντας περισσότερες δυνατότητες και βελτιώνοντας τόσο την ποιότητα υπηρεσιών όσο και της εμπειρίας του χρήστη [Nasralla et al.; 2023]. Οι παραπάνω εξελίξεις σε συνδυασμό με την πρόοδο και την αξιοποίηση συστημάτων τεχνητής νοημοσύνης θα διασφαλίσει την ταχεία ανάλυση δεδομένων και θα παρέχουν στο ιατρικό προσωπικό καινοτόμα και ισχυρά εργαλεία για να επιτελέσει το έργο του.

Ένα από τα σημαντικότερα οφέλη των αυτόματων συστημάτων αντίληψης με εφαρμογές στην υγεία είναι η ικανότητά τους να παρέχουν εξατομικευμένες συστάσεις για αλλαγές στον τρόπο ζωής. Αναλύοντας δεδομένα που σχετίζονται με τη σωματική δραστηριότητα, τις συνήθειες ύπνου και τη διατροφή ενός ατόμου, αυτά τα συστήματα μπορούν να προτείνουν αλλαγές που είναι προσαρμοσμένες στις ανάγκες και τους στόχους του κάθε ατόμου [Böhm et al.; 2019]. Ένα ακόμα πλεονέκτημά των αυτόματων συστημάτων αντίληψης που σχετίζονται με την υγεία είναι η δυνατότητα να επεξεργάζονται και να αξιοποιούν πολλαπλά δεδομένα που καταγράφονται μέσα στο χρόνο με αποτέλεσμα να μπορούν να παρακολουθούν χρόνιες παθήσεις. Για παράδειγμα, μια φορητή συσκευή που παρακολουθεί τις επιληπτικές κρίσεις μπορεί να παρέχει χρήσιμες πληροφορίες για αυτές, π.χ.για τη διάρκειά τους, και να ειδοποιεί τους οικείους όταν ο ασθενής χρειάζεται βοήθεια [Beniczky et al.; 2021]. Ομοίως, ένας αισθητήρας που παρακολουθεί τα επίπεδα γλυκόζης μπορεί να βοηθήσει τους διαβητικούς να διαχειριστούν την κατάστασή τους πιο αποτελεσματικά [Domingo-Lopez et al.; 2022].

Συνολικά, τα συστήματα αυτόματης αντίληψης δράσεων που σχετίζονται με την υγεία έχουν ένα τεράστιο εύρος εφαρμογών ενώ αξιοποιούν μεγάλη γκάμα αισθητήρων. Συνήθως χρησιμοποιούνται φορητοί αισθητήρες που συλλέγουν πολυτροπικά δεδομένα, π.χ. επιταχυνσιόμετρα, γυροσκόπια, μετρητές καρδιακών παλμών, και έχουν ως στόχο τη μέτρηση της φυσικής δραστηριότητας του χρήστη [Patel et al.; 2012, Boletsis et al.; 2015, Chaspari; 2022]. Η εύκολη χρήση αυτών των αισθητήρων και το μικρό τους κόστος έχει προκαλέσει αύξηση του ενδιαφέροντος για αξιοποίηση τους στον τομέα της ψυχικής υγείας. Όλο και περισσότερα στοιχεία δείχνουν ότι συμπεριφορικοί και βιομετρικοί δείκτες θα μπορούσαν να εισαχθούν στην κλινική ψυχιατρική [Aung et al.; 2017], αξιοποιώντας τους στην παρακολούθηση της κατάθλιψης [De Choudhury et al.; 2013, Canzian and Musolesi; 2015, Saeb et al.; 2015], της διπολικής διαταραχής [Gravenhorst et al.; 2015], ή της σχιζοφρένειας [Wang et al.; 2016b].

4.2 Επισκόπηση Αντιληπτικών Συστημάτων σε Εφαρμογές Ψυχικής Υγείας

Η καθημερινή δραστηριότητα του ανθρώπου μπορεί να αποτελέσει έναν δείκτη τόσο για τη σωματική όσο και για την ψυχική υγεία του. Έτσι παρατηρείται πως για τα άτομα με ψυχικές διαταραχές, οι σωματικές δραστηριότητες συνήθως μειώνονται ή διακόπτονται κατά τη διάρκεια ή ακόμη και πριν από μια υποτροπή της ασθένειάς τους. Σήμερα, *υπάρχουν αυξανόμενες επιστημονικές ενδείξεις ότι η παρακολούθηση της δραστηριότητας θα μπορούσε να αποτελέσει έναν σταθερό δείκτη προκειμένου να ταξινομηθούν διάφοροι τύποι επεισοδίων σε τέτοιους ασθενείς* [Maxhuni et al.; 2016], κάτι που συνήθως πραγματοποιείται από το ιατρικό προσωπικό με χρήση εξειδικευμένων ερωτηματολογίων [Bauer et al.; 2008]. Ωστόσο, η χρήση έξυπνων συστημάτων αντίληψης μπορεί να βοηθήσει στον πιο έγκαιρο εντοπισμών των αλλαγών στην καθημερινή δραστηριότητα. Κάποιες εργασίες [Bauer et al.; 2008, Blum and Magill; 2008] έχουν επίσης προτείνει τη χρήση του ψηφιακού φαινοτύπου για ακριβή και συνεχή παρακολούθηση του ασθενούς με σκοπό την κατανόηση του τρόπου με τον οποίο η καθημερινή ρουτίνα επηρεάζει τη συμπτωματολογία [Henson et al.; 2020].

Υπάρχουν εργασίες που προσπάθησαν να αντιμετωπίσουν αυτό το πρόβλημα προσφέροντας πολλά στοιχεία για τη χρήση τέτοιων αισθητηριακών δεδομένων. Στην εργασία [Maxhuni et al.; 2016] οι συγγραφείς χρησιμοποιούν επιταχυνσιόμετρο και πληροφορίες ήχου από έξυπνα κινητά για την ταξινόμηση της κατάστασης πέντε ασθενών με διπολική διαταραχή. Συνέλεξαν πληροφορίες σχετικά με τη δραστηριότητά των ασθενών σε μια χρονική περίοδο 12 εβδομάδων και έδειξαν ότι οι σημαντικές μεταβολές στη διάθεση ή οι υποτροπές μπορούν να προβλεφθούν με μεγάλη σιγουριά. Στην εργασία [Chapman et al.; 2017] χρησιμοποιήθηκαν στατιστικές μέθοδοι για τον χαρακτηρισμό προτύπων δραστηριότητας που καταγράφονται από το επιταχυνσιόμετρο σε διάστημα μιας εβδομάδας από 99 ενήλικες με διάφορες ψυχικές ασθένειες, υποστηρίζοντας ότι τα πρότυπα δραστηριότητας ποικίλλουν μεταξύ διαφορετικών διαταραχών. Επίσης, στην εργασία [Cai et al.; 2018] αποδείχτηκε πως η γραμμική επιτάχυνση και η γωνιακή ταχύτητα του καρπού είναι ιδιαιτέρως χρήσιμες μετρήσεις για την ανίχνευση του τρόμου σε ασθενείς με Πάρκινσον, και πιθανά σε ασθενείς με ψυχικές διαταραχές μιας και ο τρόμος είναι ανάμεσα στα συνήθη συμπτώματα. Τέλος, στην εργασία [Barnett et al.; 2018] οι συγγραφείς έδειξαν ότι αλλαγές στην κινητικότητα και την κοινωνική συμπεριφορά, που μετρώνται μέσω smartphone, θα μπορούσαν να εντοπίσουν στατιστικά σημαντικές ανωμαλίες στη συμπεριφορά των ασθενών κατά τις ημέρες πριν από την υποτροπή.

Οι προηγούμενες εργασίες χρησιμοποίησαν κυρίως smartphones [Reyes-Ortiz et al.; 2014] και επικεντρώθηκαν σε λειτουργίες που σχετίζονται με την κοινωνική ζωή του χρήστη όπως μηνύματα κειμένου, διάρκεια κλήσης και διάρκεια ύπνου [Barnett et al.; 2018, Adler et al.; 2020, Ben-Zeev et al.; 2017] ενώ η διάρκεια καταγραφής των δεδομένων τους διαρκούσε από μερικές ώρες έως μερικές εβδομάδες [Barnett et al.; 2018, Cella et al.; 2018, Valenza et al.; 2014], με ορισμένες εξαιρέσεις [Adler et al.; 2020]. Σε σύγκριση με τα smartphone, τα wearables είναι πιο διακριτικά, ελαφριά και μπορούν να χρησιμοποιηθούν για την παρακολούθηση των καθημερινών δραστηριοτήτων [Mukhopadhyay; 2014]. Επίσης έχει αποδειχθεί ότι τα άτομα με ψυχικές διαταραχές είναι άνετα και πρόθυμα να τα εντάξουν στην καθημερινότητά τους, κάτι που υποστηρίζει το γεγονός ότι χρησιμοποιώντας έξυπνα ρολόγια θα μπορούσαμε να έχουμε πρόσβαση σε δεδομένα που καταγράφουν διαρκώς τη δραστηριότητα των χρηστών μέσα στη μέρα του και να συλλεχθούν εύκολα με χαμηλό κόστος [Robotham et al.; 2016].

4.3 Το Ερευνητικό Έργο e-Prevention: Στόχοι, Συνιστώσες και Σύστημα

Τα ερευνητικά μέρη αυτής της διατριβής που αφορούν την εφαρμογή της αντίληψης δράσεων στον τομέα της ηλεκτρονικής υγείας και συγκεκριμένα στην αναγνώριση ψυχωτικών υποτροπών πραγματοποιήθηκαν στο πλαίσιο του ερευνητικού έργου *e-Prevention*[1]. Το έργο αυτό αποτέλεσε μια μακράς διάρκειας διεπιστημονική μελέτη (άνω των τριών ετών) με στόχο την ανάπτυξη καινοτόμων και προηγμένων ηλεκτρονικών υπηρεσιών ιατρικής παρακολούθησης και υποστήριξης για τη διευκόλυνση, την αποτελεσματική παρακολούθηση και την πρόληψη των υποτροπών σε ασθενείς με ψυχικές διαταραχές. Κατά τη διάρκεια του έργου αναπτύχθηκε ένα καινοτόμο ολοκληρωμένο σύστημα, το οποίο προσφέρει τις ακόλουθες τεχνολογίες:

- μακροπρόθεσμη συνεχή παρακολούθηση και καταγραφή βιομετρικών δεικτών και δεικτών συμπεριφοράς μέσω ενός μη παρεμβατικού και ευκολοφόρετου αισθητήρα, συγκεκριμένα ενός έξυπνου ρολογιού (smartwatch),

- καταγραφή σύντομων οπτικοακουστικών βίντεο του ασθενούς κατά τις συνέντευξεις που πραγματοποιούνταν από κλινικό γιατρό μέσω μιας φορητής κινητής συσκευής (tablet) εγκατεστημένης στο σπίτι του ασθενούς, με στόχο τη συλλογή κοινωνικών χαρακτηριστικών, συμπεριλαμβανομένων των εκφράσεων του λόγου και του προσώπου,

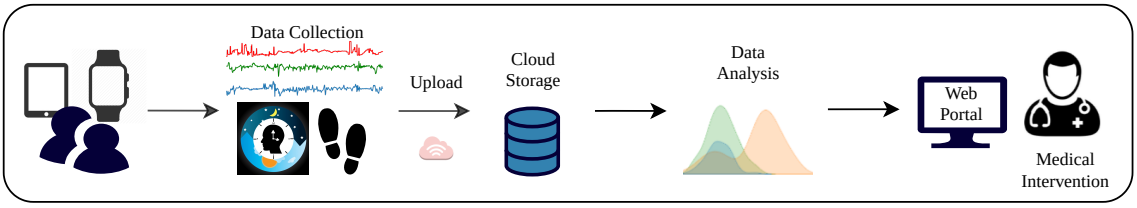
- αυτόματη και συστηματική αποθήκευση αυτών των δεδομένων σε έναν διακομιστή cloud [Maglogiannis et al.; 2020].

Μέσω αυτών, σκοπός του έργου είναι η ανάπτυξη καινοτόμων συνδυασμών ιατρικής και απομακρυσμένης ηλεκτρονικής υποστήριξης που θα διευκολύνουν την αποτελεσματική παρακολούθηση της πορείας - θεραπείας και την πρόληψη υποτροπών ασθενών με διπολική διαταραχή και σχιζοφρένεια. Ο υπεύθυνος ιατρός θα μπορεί να λαμβάνει μέσω του συστήματος πληροφορίες για την καθημερινή κατάσταση του ασθενούς, παρακολουθώντας τα δεδομένα από τους ενεργούς αισθητήρες βιοσημάτων φυσιολογίας, από τα ολιγόλεπτα αραιά βίντεο, καθώς και από τα αποτελέσματα των συστημάτων τεχνητής νοημοσύνης που θα εντοπίζουν αξιοσημείωτα γεγονότα, βασισμένα στην επεξεργασία δεδομένων μεγάλης κλίμακας.

Στο πλαίσιο του έργου, μελετήθηκαν και αναπτύχθηκαν ειδικοί αλγόριθμοι για την μηχανική κατανόηση των συλλεγμένων πληροφοριών (δηλαδή τις συνεχείς μετρήσεις βιομετρικών δεικτών μέσω φορητών αισθητήρων και τις οπτικοακουστικές εκφράσεις προσώπου μέσω ολιγόλεπτων αραιών βίντεο), έτσι ώστε να καθοδηγείται καλύτερα ο θεράπων ιατρός στην έγκαιρη λήψη αποφάσεων. Συνολικά όλες οι ερευνητικές προσπάθειες του έργου αποσκοπούν στη δημιουργία ενός ολοκληρωμένου συστήματος που θα μπορεί να βελτιώσει σημαντικά την εξωνοσοκομειακή περίθαλψη ασθενών με ψυχωτικές διαταραχές, να μειώσει το κόστος θεραπείας και να βελτιώσει την ποιότητα ζωής των ασθενών και των οικείων τους. Η επιτυχημένη εφαρμογή του συστήματος για τους ασθενείς με ψυχωτικές διαταραχές θα αποτελέσει πρότυπο για την ευρεία χρήση του και περαιτέρω ανάπτυξή για χρήση σε άλλες ψυχικές και νευρολογικές διαταραχές, προσφέροντας έτσι ένα ισχυρό εργαλείο στους επαγγελματίες λειτουργούς της ψυχικής υγείας.

Όπως λοιπόν αναφέρθηκε, η εφαρμογή για την οποία μελετάμε την ανάπτυξη έξυπνων υπολογιστικών συστημάτων αφορά τις ψυχωτικές διαταραχές. *Η ψύχωση είναι ένα φάσμα καταστάσεων που προκαλούνται από διάφορους αιτιολογικούς μηχανισμούς που επηρεάζουν το Κεντρικό Νευρικό Σύστημα (ΚΝΣ), με αποτέλεσμα να παρουσιάζονται κοινά συμπτώματα στα άτομα που τις*

[1] http://eprevention.gr



Σχήμα 4.1: Η δομή του συστήματος e-Prevention.

εμφανίζουν [Os and Kapur; 2009]. Τα τελευταία 60 χρόνια έχουν διεξαχθεί διάφορες μελέτες σχετικών ψυχιατρικών καταστάσεων στη νευροβιολογία και τη νευροφυσιολογία, ωστόσο οι αιτίες τους παραμένουν ακόμη άγνωστες. Έτσι, ακόμα δεν έχουν ανακαλυφθεί αποτελεσματικοί βιοδείκτες είτε για τη διάγνωση είτε για την πρόβλεψη της πορείας της ψυχωτικής συμπτωματολογίας. Για το λόγο αυτό, ο εντοπισμός και η χρήση τέτοιων δεικτών αποτελεί έναν από τους πιο εξέχοντες τομείς μελέτης στην ψυχιατρική για έγκαιρη διάγνωση και πρόληψη ψυχωτικών υποτροπών [Wiersma et al.; 1995, Koutsouleris et al.; 2011, McGorry et al.; 2014]. Στην πραγματικότητα, η έγκαιρη αναγνώριση των επιδεινούμενων συμπτωμάτων στα αρχικά στάδια της ψυχωτικής διαδικασίας και η έγκαιρη πρόληψη των υποτροπών έχει βρεθεί ότι συμβάλλει σημαντικά στην καλύτερη έκβαση της διαταραχής [Bertelsen et al.; 2008, Norman and Malla; 1995, Hegelstad et al.; 2012] και στην πρόληψη των καταστροφικών επιπτώσεων που συχνά έχουν οι υποτροπές στη ζωή των ασθενών [Insel; 2007].

Δεδομένου ότι η ψύχωση εξελίσσεται συνεχώς και η υποτροπή είναι μια βιολογική διαδικασία που αναπτύσσεται με την πάροδο του χρόνου [McCandless-Glimcher et al.; 1986, Gaebel et al.; 1993, Wiersma et al.; 1998], θα ήταν λογικό να προβλεφθούν διακυμάνσεις στη συμπεριφορά τέτοιων σχετικών βιοδεικτών, και πιθανώς να προηγείται της εμφάνισης ή/και επιδείνωσης τέτοιων ψυχικών διαταραχών. Μερικά από τα τυπικά πρώιμα προειδοποιητικά συμπτώματα ψυχιατρικών καταστάσεων περιλαμβάνουν ακαμψία, τρόμο, απότομες κινήσεις των χεριών, άτυπες κινήσεις ή στάσεις και απόσυρση από υπαίθριες δραστηριότητες, μεταξύ άλλων [Maxhuni et al.; 2016, Schizophrenia;]. *Έτσι, με βάση τα παραπάνω στο έργο e-Prevention μελετήσαμε την ανάπτυξη ενός έξυπνου σύστηματος που θα μπορούσε να μετρά διαρκώς την ανθρώπινη συμπεριφορά, παθητικά και μη παρεμβατικά, για να ανιχνεύσει αυτές τις αλλαγές και να αποτρέψει τις ψυχωτικές υποτροπές πριν εκφραστούν πλήρως τα συμπτώματα.*

Στην παρούσα διατριβή ασχολούμαστε κύρια με τη μελέτη των βιοσημάτων που συλλέγονται μέσω ενός έξυπνου ρολογιού από χρήστες που φορούν το ρολόι καθημερινά. Κύριος στόχος είναι η μελέτη και ο εντοπισμός των αλλαγών που μπορεί να προκύπτουν στα βιοσήματα ασθενών με ψυχωτικές διαταραχές κατά τη διάρκεια ή πριν την έναρξη υποτροπών της νόσου τους. Η συλλογή και η μελέτη των δεδομένων έγινε στο πλαίσιο του ερευνητικού έργου e-Prevention[2] καθώς μια τέτοια προσπάθεια απαιτεί μια διεπιστημονική συνέργεια και καθοδήγηση από εξειδικευμένη ιατρική ομάδα. Για την επίτευξη αυτού του στόχου πραγματοποιήθηκαν πολλές αναλύσεις των πολυτροπικών δεδομένων που συλλέχθηκαν [Zlatintsi et al.; 2022] καθώς και των ιατρικών κλιμάκων αξιολόγησης που αξιολογούσαν την φάση της ασθένειας των συμμετεχόντων [Kalisperakis et al.; 2023] και έγιναν προσπάθειες ανάπτυξης υπολογιστικών μοντέλων εντοπισμού των υποτροπών [Filntisis et al.; 2020b, Efthymiou et al.; 2023, Garoufis et al.; 2022, Fekas et al.; 2023].

Όπως ήδη αναφέραμε το έργο e-Prevention είναι αποτέλεσμα μιας μεγάλης ομαδικής δουλειάς πολλών χρόνων. Η δική μου ερευνητική συνεισφορά αποτυπώνεται κύρια στη μελέτη της εξαγωγής των βιοδεικτών, της δημιουργίας και ταυτοποίησης ψηφιακών φαινοτύπων και της αξιοποίησής τους για τον εντοπισμό των ψυχωτικών υποτροπών όπως παρουσιάζεται στα

[2]Περισσότερες πληροφορίες: https://eprevention.gr/

Δημογραφικά Στοιχεία	Σύνολο Ελέγχου	Σύνολο Ασθενών
Άντρες / Γυναίκες	12/11	26/12
Ηλικία (χρόνια)	27.8 ± 3.9	30.55 ± 7.28
Εκπαίδευση (χρόνια)	16.9 ± 1.8	13.36 ± 2.18
Χρόνος Ασθένειας (χρόνια)	-	7.34 ± 6.41

Πίνακας 4.1: Δημογραφικά στοιχεία των συνόλων ελέγχου και ασθενών, υγιών και ασθενών εθελοντών αντίστοιχα, κατά την έναρξη της συμμετοχής τους.

επόμενα Κεφάλαια αλλά και στις δημοσιεύσεις [Efthymiou et al.; 2023, Filntisis et al.; 2020b, Efthymiou et al.;]. Στην επόμενη ενότητα 4.3.1 παρουσιάζουμε τη βάση δεδομένων που συλλέχθηκε κατά τη διάρκεια του έργου, ενώ στην ενότητα 4.3.2 κάνουμε μια εκτενή αναφορά στις εργασίες που αξιοποίησαν τα βιοσήματα της βάσης για τη διεξαγωγή πειραμάτων στην κατεύθυνση της ανίχνευσης υποτροπών.

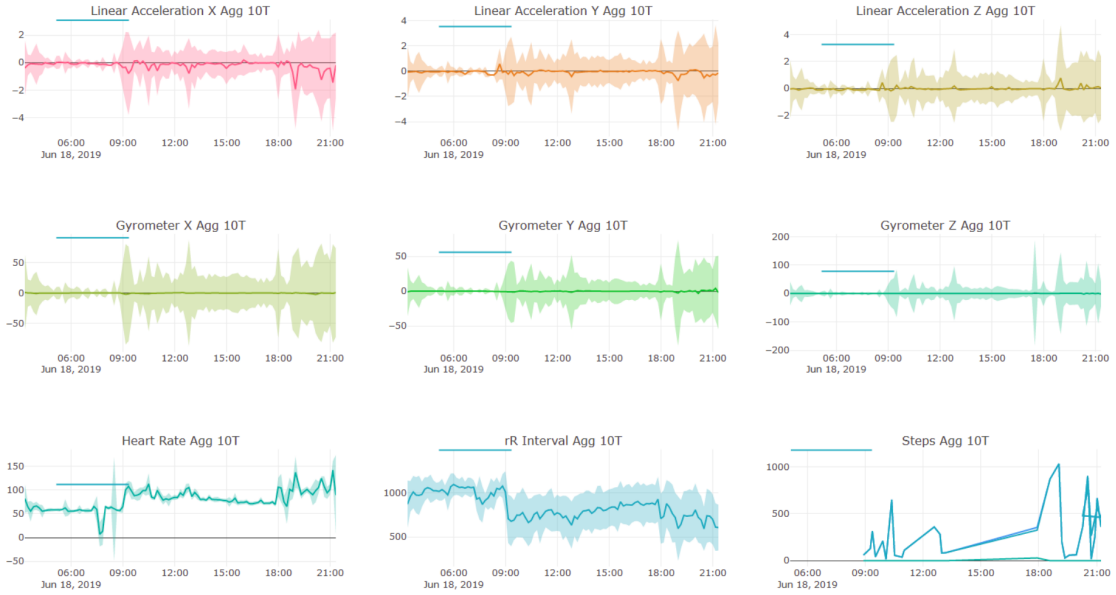
4.3.1 Η βάση δεδομένων e-Prevention

Κατά τη διάρκεια των τριών και πλέον ετών υλοποίησης του έργου e-Prevention δημιουργήθηκε μια από τις μεγαλύτερες βάσεις δεδομένων με πολυτροπικά δεδομένα από 60 συμμετέχοντες με συνολικά περισσότερες από 20.000 ημέρες καταγραφής της ανθρώπινης δραστηριότητας. Στην παρούσα ενότητα περιγράφεται συνοπτικά η βάση e-Prevention.

Στην αρχική φάση του έργου, 23 άτομα χωρίς ψυχωτικές διαταραχές συμμετείχαν για τρεις μήνες για να αποτελέσουν την ομάδα ελέγχου, ενώ στην επόμενη φάση συμμετείχαν συνολικά 38 ασθενείς ηλικίας 19-43 ετών. 23 εθελοντές ασθενείς συμπλήρωσαν το διάστημα των 24 μηνών που οριζόταν αρχικά ως το μέγιστο διάστημα συμμετοχής, 15 ασθενείς αποχώρησαν από τη μελέτη για λόγους που δεν σχετίζονταν με αυτήν, σε διαφορετικά χρονικά διαστήματα. Ο Πίνακας 4.1 εμφανίζει πληροφορίες σχετικά με τα δημογραφικά στοιχεία τόσο των υγιών όσο και των ασθενών εθελοντών κατά την έναρξη της συμμετοχής τους στο e-Prevention.

Η επιλογή των συμμετεχόντων στο έργο e-Prevention έγινε από το Ερευνητικό Πανεπιστημιακό Ινστιτούτο Ψυχικής Υγείας, Νευροεπιστημών και Ιατρικής Ακριβείας «Κώστας Στεφανής» (ΕΠΨΥ) στην Αθήνα, ενώ η επιστημονική ομάδα του ινστιτούτου είχε και τη συνολική επίβλεψη της διαδικασίας. Οι συμμετέχοντες στο έργο e-Prevention έδωσαν γραπτώς τη συγκατάθεση τους για την αξιοποίηση των ανωνυμοποιημένων δεδομένων τους για ερευνητική χρήση σύμφωνα με τις διατάξεις του Γενικού Κανονισμού (ΕΕ) 2016/679. Το πρωτόκολλο της διαδικασίας εγκρίθηκε από την Επιτροπή Δεοντολογίας του Ιδρύματος.

Πριν από την έναρξη της συμμετοχής των εθελοντών, οι κλινικοί γιατροί συναντήθηκαν με τους συμμετέχοντες για να πραγματοποιήσουν αξιολόγηση των συμπτωμάτων και της γενικής λειτουργίας τους. Συγκεκριμένα, κάθε εθελοντής υποβλήθηκε σε αρχική ατομική αξιολόγηση, διάρκειας περίπου 180 λεπτών, κατά την οποία συλλέχθηκαν δημογραφικά στοιχεία (ηλικία, φύλο, έτη εκπαίδευσης, επάγγελμα, οικογενειακή κατάσταση, τόπος γέννησης και διαμονής), ιστορικό σωματικής και ψυχικής υγείας, στοιχεία για τις πιθανές περιγεννητικές επιπλοκές και την κατάχρηση ουσιών. Στη συνέντευξη αυτή όλοι οι συμμετέχοντες έκαναν επίσης μια νευροψυχολογική αξιολόγηση από εκπαιδευμένο νευροψυχολόγο για να διασφαλιστεί ότι δεν υπάρχει καμία νευρολογική διαταραχή. Επιπλέον, για την ομάδα ελέγχου, διασφαλίστηκε ότι δεν υπάρχει ιστορικό ψυχικών διαταραχών ή κατάχρησης ουσιών. Επιπρόσθετα, στην περίπτωση των ασθενών, καταγράφηκε το οικογενειακό ιστορικό ψυχικής νόσου, ο χρόνος που είχε παρέλθει από την εμφάνιση των πρώτων συμπτωμάτων της ψυχικής νόσου, ο βαθμός συμμόρφωσης με τη θεραπεία τους τελευταίους 6 μήνες καθώς και η φαρμακευτική αγωγή που λαμβάνουν.



Σχήμα 4.2: Αναπαράσταση των σημάτων που καταγράφονται από το έξυπνο ρολόι για μια ημέρα ενός ατόμου. Στην πρώτη γραμμή απεικονίζεται η γραμμική επιτάχυνση (άξονες x,y,z) και στη δεύτερη η γωνιακή επιτάχυνση (άξονες x,y,z). Στην τελευταία γραμμή απεικονίζονται, από τα αριστερά προς τα δεξιά, ο καρδιακός ρυθμός τα RR-intervals (χρονικά διαστήματα ανάμεσα στους χτύπους της καρδιάς) και ο αριθμός των βημάτων. Σε κάθε διάγραμμα (εκτός των βημάτων), η κύρια γραμμή υπολογίζεται από το μέσο όρο όλων των μετρούμενων τιμών κατά τη διάρκεια κάθε δεκαλέπτου, ενώ με πιο αχνό χρωματισμό φαίνεται και η διακύμανση της υπολογιζόμενης τιμής. Η μπλε οριζόντια γραμμή σε κάθε διάγραμμα, δίνει το χρονικό διάστημα που ο χρήστης του ρολογιού κοιμόταν.

Κατά την έναρξη της συμμετοχής τους, οι ασθενείς λάμβαναν τη θεραπεία τους και οι περισσότεροι ήταν σε περίοδο ύφεσης της νόσου. Αποκλείστηκαν από τη συμμετοχή στο πρόγραμμα οι ασθενείς που είχαν οποιοδήποτε από τα ακόλουθα: (α) προβλήματα ακοής, όρασης ή κινητικής ανεπάρκειας, (β) επίπεδο ανάγνωσης κάτω της έκτης τάξης ή (γ) αδυναμία παροχής ενημερωμένης συγκατάθεσης.

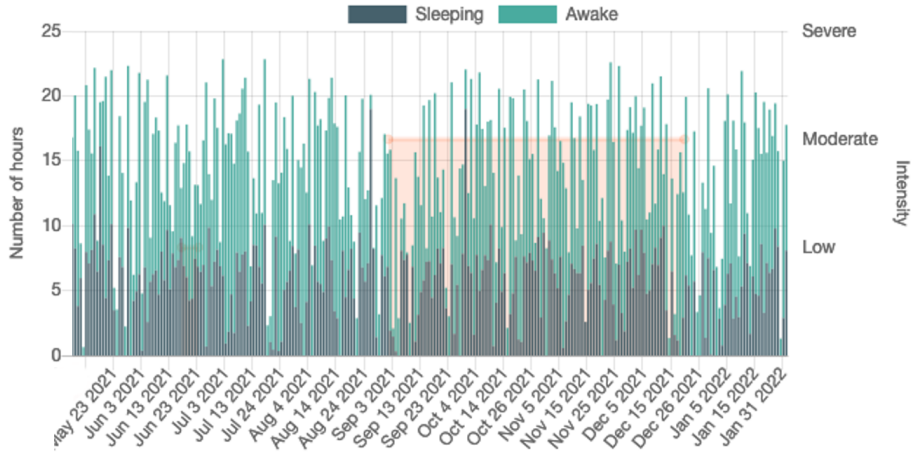
Κατά τη συμμετοχή των εθελοντών στο έργο έλαβαν ένα έξυπνο ρολόι Samsung Gear S3 που παρακολουθούσε συνεχώς τη γραμμική και γωνιακή επιτάχυνση, τον καρδιακό ρυθμό, τη μεταβλητότητα του καρδιακού παλμού, το πρόγραμμα ύπνου τους και τον αριθμό των βημάτων τους. Στο Σχήμα 4.2 παρουσιάζεται μια οπτική αναπαράσταση των σημάτων που καταγράφονται μέσω του ρολογιού. Περισσότερα στοιχεία για το είδος και τη συχνότητα δειγματοληψίας των δεδομένων μέσω των αισθητήρων του ρολογιού δίνονται στον Πίνακα 4.2.

Οι εθελοντές φορούσαν το ρολόι τους διαρκώς και τα δεδομένα που καταγράφονταν μεταφορτώνονταν καθημερινά κατά τη διάρκεια της φόρτισης του. Στη συνέχεια, αποθηκεύονταν σε μια πλατφόρμα που αναπτύχθηκε κατά τη διάρκεια του ερευνητικού έργου και βασίζεται σε cloud τεχνικές [Maglogiannis et al.; 2020]. Επίσης, οι εθελοντές κατά τη συμμετοχή τους έκαναν κάποιες συνεντεύξεις με τους γιατρούς, μέσης διάρκειας 5-10 λεπτών, μέσω της ειδικής διαδικτυακής εφαρμογής που αναπτύχθηκε για το έργο e-Prevention ή μέσω τηλεφώνου, προκειμένου να αξιολογηθεί η φυσική δραστηριότητα τους με χρήση της ελληνικής σύντομης έκδοσης του International Physical Activity Questionnaire (IPAQ-Gr) [Papathanasiou et al.; 2009].

Επιπλέον, η κλινική ομάδα διεξήγαγε εκτιμήσεις της κατάστασης των ασθενών μία φορά κάθε μήνα κατά τις οποίες αξιολογήθηκε κλινικά η πορεία και η θεραπεία τους με χρήση αξιό-

Αισθητήρας	Δεδομένα	Μονάδα Μέτρησης	Συχνότητα Δειγματοληψίας
Επιταχυνσιόμετρο	Γραμμική Επιτάχυνση (3-άξονες)	m/s^2	20Hz
Γυροσκόπιο	Γωνιακή Επιτάχυνση (3-άξονες)	$degrees/s^2$	20Hz
Φωτοπληθυσμογράφος	Μεταβολή Καρδιακού Παλμού RR-Intervals	beats/min seconds	5Hz
Βηματομετρητής	Βήματα & Συνολική Απόσταση	steps/min	Σύνολο βημάτων ανά λεπτό

Πίνακας 4.2: Συλλογή δεδομένων από τους αισθητήρες του έξυπνου ρολογιού. Παρουσιάζεται ο αισθητήρας, το είδος των δεδομένων, η μονάδα μέτρησης και η συχνότητα δειγματοληψίας.

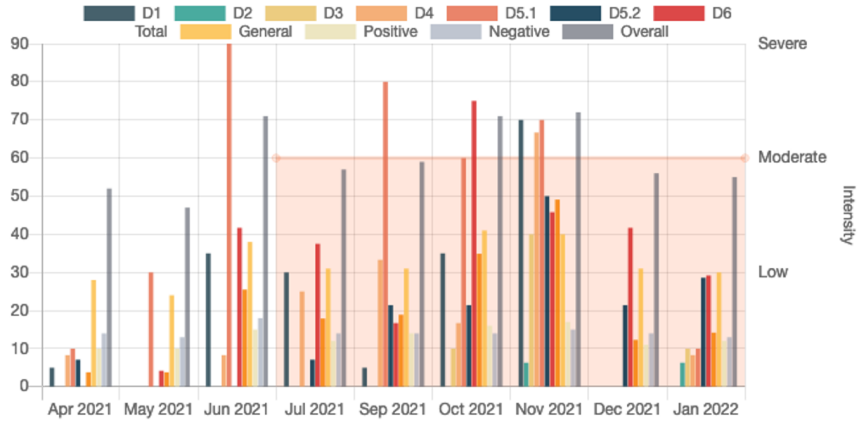


Σχήμα 4.3: Οπτική αναπαράσταση του ημερολογίου ύπνου ενός ασθενούς. Με σκούρο μπλε εμφανίζεται ο αριθμός των ωρών του ύπνου ενώ με ανοιχτό μπλε οι υπόλοιπες ώρες καταγραφής που ο ασθενής ήταν ξύπνιος. Το ροζ παραλληλόγραμμο δείχνει τη διάρκεια μιας μέτριας έντασης ψυχωτικής υποτροπής που βίωσε ο ασθενής.

πιστων κλιμάκων για εκτίμηση της ψυχοπαθολογίας (Positive and Negative Syndrome Scale - PANSS), της λειτουργικότητας (WHO Disability Assessment Schedule 2.0 - WHODAS 2.0), των ανεπιθύμητων ενεργειών των φαρμάκων (Glasgow Antipsychotic Side-effect Scale - GASS), των ακούσιων ανώμαλων κινήσεων (bnormal Involuntary Movement Scale - AIMS) και των εξωπυραμιδικών ανεπιθύμητων ενεργειών (Simpson-Angus Scale - SAS), ενώ καταχωρήθηκε εκ νέου ο δείκτης μάζας σώματος (Body Mass Index - BMI).

Μέσω της διαδικτυακής εφαρμογής που έχει αναπτυχθεί στο πλαίσιο του έργου, η ιατρική ομάδα έχει εύκολη πρόσβαση και επεξεργασία στα προσωπικά δεδομένα των ασθενών ενώ έχει πρόσβαση και σε εργαλεία που μπορούν να δώσουν μια σύνοψη των καταγραφών των αισθητήρων. Στα Σχήματα 4.3 και 4.4 παρουσιάζονται δύο τέτοια παραδείγματα, όπου στο πρώτο παρουσιάζεται το ημερολόγιο ύπνου ενός ασθενή όπως καταγράφηκε από το έξυπνο ρολόι και στο δεύτερο μια οπτική αναπαράσταση των βαθμολογιών του ασθενή ως προς τις κλίμακες PANSS και WHODAS 2, όπως έχουν καταγραφεί από τους κλινικούς ιατρούς. Και στα δύο σχήματα, το αχνό ροζ παραλληλόγραμμο που εμφανίζεται υποδεικνύει την περίοδο μιας ψυχωτικής υποτροπής μέτριας έντασης του εκάστοτε ασθενούς.

Κατά τη διάρκεια συμμετοχής στο πρόγραμμα, κάποιοι ασθενείς παρά τη λήψη της φαρμα-



Σχήμα 4.4: Οπτική αναπαράσταση της κλίμακας ψυχοπαθολογίας PANSS και της κλίμακας λειτουργικότητας WHODAS 2 για έναν ασθενή κατά τη διάρκεια δέκα μηνών συμμετοχής στο έργο. Το ροζ πλαίσιο δείχνει τη διάρκεια μιας μέτριας έντασης ψυχωτικής υποτροπής που βίωσε ο ασθενής. Οι ενδείξεις D1-D6, το Total αναφέρονται στην κλίμακα WHODAS 2, ενώ οι υπόλοιπες στην κλίμακα PANSS.

κευτικής τους αγωγής και της παρακολούθησης από τον θεράποντα ιατρό τους, παρουσίασαν υποτροπές στη νόσο τους. Οι υποτροπές ανιχνεύθηκαν και αξιολογήθηκαν με βάση τη μηνιαία κλινική αξιολόγηση των εθελοντών ασθενών, τις μεταβολές στην ψυχοπαθολογία όπως υπολογίστηκαν με την κλίμακα PANSS, καθώς και από τις πληροφορίες που συλλέχθηκαν από τους θεράποντες ψυχιάτρους, το οικογενειακό περιβάλλον και τους φροντιστές των ασθενών ή το προσωπικό της κλινικής σε περιπτώσεις νοσηλείας.

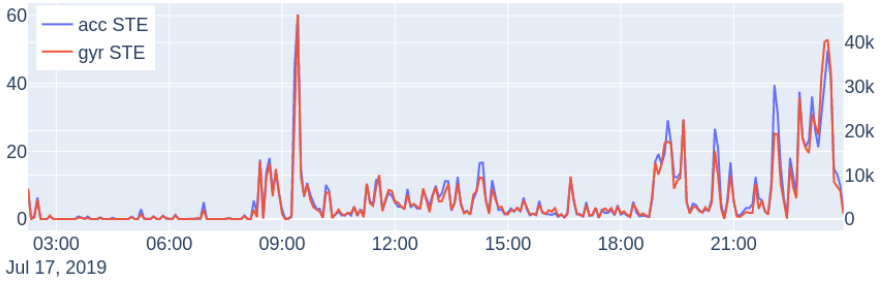
Σχετικά με τις υποτροπές, η ιατρική ομάδα όρισε ως υποτροπή τη σαφή κλινική επιδείνωση των ασθενών με την επανεμφάνιση ψυχωτικών συμπτωμάτων (παραληρητικές ιδέες, ψευδαισθήσεις) ή με την εμφάνιση μανιακών, καταθλιπτικών ή μικτών επεισοδίων με ή χωρίς την παρουσία ψυχωτικών συμπτωμάτων. Το τέλος μιας υποτροπής ορίστηκε ως το χρονικό σημείο εκείνο, όπου τα συμπτώματα του ασθενούς υποχώρησαν σε τέτοιο βαθμό, ώστε ο ασθενής επέστρεψε κλινικά και λειτουργικά σε παρόμοιο επίπεδο με αυτό πριν την υποτροπή.

Όσον αφορά στο είδος των υποτροπών, ως ψυχωτικές υποτροπές ορίστηκαν εκείνες κατά τις οποίες ανιχνεύθηκαν ψυχωτικά συμπτώματα και αντιστοίχως, ως μη ψυχωτικές υποτροπές ορίστηκαν εκείνες κατά τις οποίες δεν ανιχνεύθηκαν ψυχωτικά συμπτώματα. Σύμφωνα με τη βαρύτητά τους οι υποτροπές διακρίθηκαν σε ήπιες, μέτριες και σοβαρές. Έτσι κατά τη διάρκεια της μελέτης, σημειώθηκαν συνολικά 45 υποτροπές σε 21 διαφορετικούς εθελοντές ασθενείς, 27 ψυχωτικές σε 16 διαφορετικούς ασθενείς και 18 μη ψυχωτικές σε 11 διαφορετικούς ασθενείς.

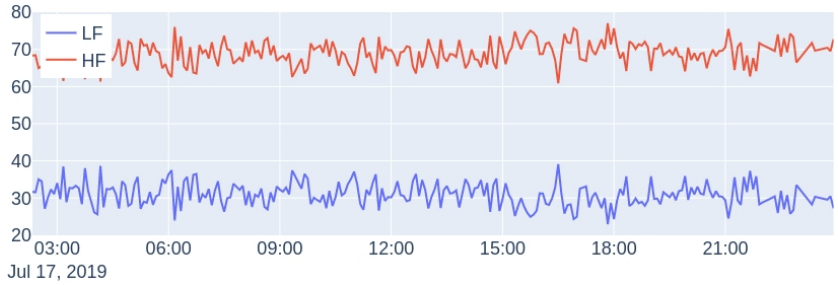
4.3.2 Επισκόπηση ερευνητικών εργασιών στα βιοσήματα της βάσης e-Prevention

Κατά τη διάρκεια των τριών και πλέον χρόνων που διήρκεσε το έργο e-Prevention, οι ερευνητικές ομάδες έχουν δημοσιεύσει πολλές εργασίες σε διαφορετικά υποσύνολα της βάσης. Ακόμα και αν αναφερθούμε μόνο στα δεδομένα που προέκυψαν από τους αισθητήρες του ρολογιού, πάλι θα εντοπίσουμε ερευνητικές εργασία σε διαφορετικά υποσύνολα, καθώς αυτές πραγματοποιήθηκαν ενώ ήταν σε εξέλιξη η συλλογή των δεδομένων.

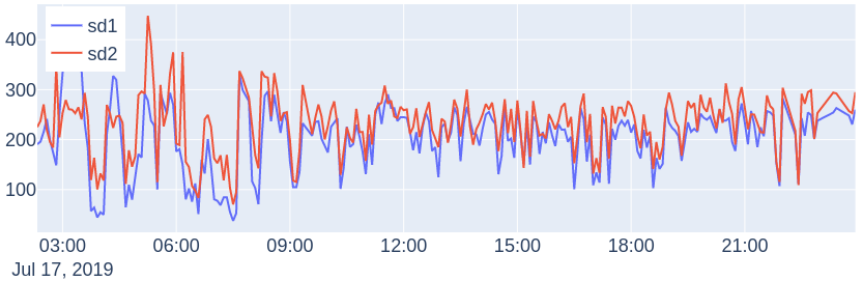
Η πρώτη σημαντική εργασία μας που μελέτησε την εξαγωγή χαρακτηριστικών από τα ακατέργαστα συνεχή δεδομένα των ρολογιών ήταν η [Filntisis et al.; 2020b] όπου παρουσιάστηκε μια διεξοδική στατιστική ανάλυση των χαρακτηριστικών αυτών. Με έμπνευση από τις παραδοσιακές τεχνικές επεξεργασίας σήματος, εξήγαμε απλά αλλά και πιο σύνθετα χαρακτηριστικά



(α΄) Ενεργεία βραχέως χρόνου της ευκλείδειας νόρμας της γραμμικής και της γωνιακής επιτάχυνσης



(β΄) Κανονικοποιημένες δυνάμεις Low and High Frequency της μεταβολής των καρδιακών παλμών



(γ΄) Οι άξονες της Poincare έλλειψης (sd1,sd2) που προκύπτουν από την μεταβολή των καρδιακών παλμών

Σχήμα 4.5: Παραδείγματα χαρακτηριστικών που μπορούν να εξαχθούν από τα δεδομένα της βάσης του e-Prevention για τη διάρκεια μιας ημέρας για έναν χρήστη του έξυπνου ρολογιού.

χρησιμοποιώντας ανάλυση βραχέως χρόνου και τα μελετήσαμε μέσω περιγραφικής στατιστικής προκειμένου να λάβουμε μια αδρή εκτίμηση του τρόπου με τον οποίο διαφοροποιούνται μεταξύ υγιών και ασθενών με ψυχωτικές διαταραχές εθελοντών. Στο Σχήμα 4.5 παρουσιάζονται μερικά παραδείγματα αυτών των χαρακτηριστικών κατά τη διάρκεια μιας ημέρας. Η πειραματική αξιολόγηση μας έδειξε ότι τόσο τα πιο απλά, αλλά και μερικά από τα νέα μη γραμμικά χαρακτηριστικά που εξετάστηκαν είναι ισχυρά στη διάκριση μεταξύ των δύο ομάδων. Αναλυτικότερα παρουσιάζονται τα αποτελέσματα της συγκεκριμένης έρευνας στην Ενότητα 5.1.

Μια ακόμα σημαντική εργασία που αποτέλεσε γνώμονα στον τρόπο αντιμετώπισης του προβλήματος της ανίχνευσης υποτροπών και πιο ειδικά στην επιλογή των χαρακτηριστικών στη διδακτορική διατριβή είναι η εργασία [Kalisperakis et al.; 2023]. Το επίκεντρο αυτής της μελέτης ήταν να ελεγχθεί εάν οι ψηφιακοί φαινότυποι των ασθενών, που αποτελούνται από ένα σύνολο χαρακτηριστικών υπολογιζόμενων ανά πέντε λεπτά καταγραφών του ρολογιού, σχετίζονται με αλλαγές στη θετική και αρνητική ψυχοπαθολογία αυτών των ασθενών με την πάροδο του χρόνου (μετρούμενες με τη χρήση PANSS). Αξιοποιώντας τα δεδομένα 35 ασθενών (20 με

σχιζοφρένεια και 15 με διαταραχές διπολικού φάσματος) για περίοδο έως και 14 μήνες υπολογίστηκαν και μελετήθηκαν τα εξής χαρακτηριστικά: η συνολική κινητική δραστηριότητα μέσω του μέτρου της γραμμικής επιτάχυνσης, ο μέσος καρδιακός ρυθμός, η μεταβλητότητα καρδιακού ρυθμού, ο αριθμός των συνολικών βημάτων ανά ημέρα και αναλογία ύπνου/αφύπνισης. Μετά τη συγκέντρωση δεδομένων φαινοτύπου, ο μηνιαίος μέσος όρος και η διακύμανσή τους συσχετίστηκαν σε κάθε ασθενή με βαθμολογίες ψυχοπαθολογίας (PANSS) που αξιολογήθηκαν μηνιαίως.

Τα αποτελέσματά έδειξαν ότι ο αυξημένος μέσος καρδιακός ρυθμός κατά τη διάρκεια της εγρήγορσης και του ύπνου συσχετίστηκε με αυξήσεις στη θετική ψυχοπαθολογία. Επιπλέον, η μειωμένη μεταβλητότητα καρδιακού ρυθμού και η αύξηση της μηνιαίας διακύμανσής της συσχετίστηκαν με αυξήσεις στην αρνητική ψυχοπαθολογία. Η αυτοαναφερόμενη σωματική δραστηριότητα (οι μετρήσεις της κλίμακας IPAQ) δεν συσχετίστηκε με αλλαγές στην ψυχοπαθολογία. Αυτές οι επιδράσεις ήταν ανεξάρτητες από δημογραφικές και κλινικές μεταβλητές καθώς και από αλλαγές στη δόση των αντιψυχωτικών φαρμάκων. Τα ευρήματά αυτά υποδηλώνουν ότι διακριτοί ψηφιακοί φαινότυποι που προέρχονται παθητικά από ένα έξυπνο ρολόι μπορούν να προβλέψουν διακυμάνσεις στις θετικές και αρνητικές διαστάσεις της ψυχοπαθολογίας ασθενών με ψυχωτικές διαταραχές, με την πάροδο του χρόνου, παρέχοντας βασικά στοιχεία για την πιθανή κλινική χρήση τους και θα μπορούσαν να χρησιμοποιηθούν ως δεδομένα για τη μοντελοποίηση του εντοπισμού και της πρόβλεψης υποτροπής.

Σημαντική προσπάθεια στη διάδοση, την αξιοποίηση της βάσης e-Prevention καθώς και στην σύγκριση των ερευνητικών αποτελεσμάτων αποτέλεσε ο διαγωνισμός **Signal Processing Grand Challenge e-Prevention** που διοργανώθηκε στα πλαίσια του συνεδρίου ICASSP 2023. Ο διαγωνισμός είχε ως στόχο την προώθηση της έρευνας στον τομέα της ηλεκτρονικής υγείας και συγκεκριμένα στις εφαρμογές ψυχικής υγείας, ενθαρρύνοντας την ανάπτυξη ισχυρών και αποτελεσματικών συστημάτων για τον εντοπισμό και την πρόληψη ψυχωτικών υποτροπών. Πιο συγκεκριμένα, οι συμμετέχοντες κλήθηκαν να αξιοποιήσουν τα δεδομένα που συλλέχθηκαν από το έξυπνο ρολόι για δύο ανεξάρτητες εφαρμογές: την ταυτοποίηση του χρήστη του ρολογιού ανάμεσα σε 46 χρήστες και τον εντοπισμό των ψυχωτικών υποτροπών δέκα ασθενών.

Και οι δυο εφαρμογές του διαγωνισμού είχαν ως στόχο την κατανόηση των ψηφιακών φαινοτύπων των χρηστών και την εξαγωγή κατάλληλων συμπερασμάτων για τις μεταβολές αυτών. Έτσι, στη συνέχεια της ενότητας, διατηρούμε αυτό το διαχωρισμό για να ομαδοποιήσουμε τις επόμενες εργασίες σύμφωνα με τον στόχο τους: α) την ταυτοποίηση του χρήστη, β) τον εντοπισμό των ψυχωτικών ή μη-ψυχωτικών υποτροπών.

Ταυτοποίηση ατόμου

Πρώτη προσπάθεια στην προσέγγιση του προβλήματος της ταυτοποίησης των ατόμων μέσω του ψηφιακού αποτυπώματός τους έγινε στην εργασία μας [Retsinas et al.; 2020]. Ο κύριος στόχος μας ήταν να προσδιορίσουμε εάν τα σήματα μικρής διάρκειας (από 10 δευτερόλεπτα έως 10 λεπτά) αποκρύπτουν διακριτά μοτίβα για την πρόβλημα της αναγνώρισης ατόμων, τα οποία θα μπορούσαν να οδηγήσουν στη δημιουργία ενός ενδιαφέροντος προφίλ συμπεριφοράς για κάθε άτομο. Έτσι, αναπτύξαμε ένα βαθύ νευρωνικό δίκτυο, βασισμένο σε μονοδιάστατες συνελίξεις (1D convolution networks - CNN), το οποίο λαμβάνει ακατέργαστα σήματα από τρεις διαφορετικούς αισθητήρες, γραμμική και γωνιακή επιτάχυνση και καρδιακό παλμό, και προβλέπει ποιο είναι το άτομο που φορά το ρολόι. Ένα σημαντικό εμπόδιο για τον στόχο μας ήταν η ύπαρξη μοναδικού θορύβου αισθητήρα, ο οποίος παραπλάνησε το νευρωνικό δίκτυο να ταξινομήσει τον αισθητήρα αντί για το άτομο. Αυτό αντιμετωπίστηκε αποτελεσματικά με την προσθήκη θορύβου στα ακατέργαστα σήματα που καταγράφονται και ο οποίος αποδείχτηκε πως ήταν επαρκής για να καλύψει τον θόρυβο των αισθητήρων. Το προτεινόμενο νευρωνικό

δίκτυο αναφέρει εξαιρετικά αποτελέσματα αναγνώρισης, ειδικά όταν ο χρήστης περπατά, επαληθεύοντας την υπόθεση μας για την ύπαρξη μοναδικών βραχυπρόθεσμων μοτίβων ανά άτομο.

Οι επόμενες εργασίες έγιναν στα πλαίσια του διαγωνισμού e-Prevention Challenge ICASSP 2023 και παρουσιάζονται κατά φθίνουσα σειρά κατάταξης σύμφωνα με τις επιδόσεις των συστημάτων που πρότειναν για την ταυτοποίηση των ατόμων. Οι Wu et al., στην εργασία τους [Wu and Tu; 2023] παρουσιάζουν ένα πολύ αποδοτικό σύστημα αναγνώρισης ατόμων που βασίζεται στην σύμμειξη της πληροφορίας που προκύπτει από τρεις διαφορετικές εκπαιδεύσεις του ίδιου μονοδιάστατου συνελικτικού δικτύου με διαφορετικά τμήματα των σημάτων. Πιο συγκεκριμένα, αφού απέρριψαν τις μη φυσιολογικές τιμές των σημάτων και τα κανονικοποίησαν, τα έτμησαν με τρεις διαφορετικούς τρόπους: α) σε διαστήματα των 30 λεπτών, β) ανάλογα με την κατάσταση ύπνου ή εγρήγορσης του ατόμου, γ) στα τμήματα που δεν είχαν πληροφορία για τους καρδιακούς παλμούς. Έτσι, για κάθε ένα από αυτά τα τμήματα λάμβαναν ένα σκορ ταυτοποίησης του ατόμου και με ένα γραμμικό συνδυασμό αυτών που υλοποιούσε μια ψηφοφορία για την κάθε ημέρα, λάμβαναν την τελική ταυτότητα του χρήστη.

Στο [Calcagno et al.; 2023] οι συγγραφείς αξιοποιούν όλα τα δεδομένα των αισθητήρων και επιπρόσθετα, από το πλήθος των βημάτων και την απόσταση που διένυσε το άτομο εξάγουν τρεις ακολουθίες: τον αριθμό των βημάτων, τις καταναλισκόμενες θερμίδες ανά μονάδα χρόνου, και την ταχύτητα της κίνησής του. Επαυξάνουν τα δεδομένα με μια προσέγγιση κυλιόμενου παραθύρου χρησιμοποιώντας μη επικαλυπτόμενα παράθυρα με διαφορετικά πλάτη (1.5 και 3 ωρών), ενώ κατά τη διάρκεια της εκπαίδευσης του δικτύου αξιοποιούν παρεμβολή πλησιέστερου γείτονα (Nearest Neighbor interpolation) για να αντικαταστήσουν τις μη έγκυρες τιμές της βάσης. Το μοντέλο αναγνώρισης αποτελείται από ένα συνδυασμό ενός μονοδιάστατου CNN και πέντε αρχιτεκτονικές μετασχηματιστών (transformers). Τέλος, για την κωδικοποίηση της πληροφορίας του χρόνου μέσω γίνεται μέσω της χρήσης της τεχνικής Time2Vec [Kazemi et al.; 2019] που έχει ως στόχο είναι να παρέχει μια αναπαράσταση του χρόνου με τη μορφή embeddings.

Η ομάδα των Mohapatra et al. [Mohapatra et al.; 2023] πραγματοποίησε εκ νέου δειγματοληψία όλων των δεδομένων με διάστημα 30 δευτερολέπτων και εξήγαγαν τρεις νέες δειγματοληπτημένες ακολουθίες για τη γραμμική επιτάχυνση, τη γωνιακής επιτάχυνση και την πληροφορία του καρδιακού ρυθμού για διαστήματα μιας ώρας με επικάλυψη 30% από την επόμενη χρονικά ακολουθία. Για την πληροφορία του καρδιακού παλμού οι συγγραφείς δημιούργησαν και ενσωμάτωσαν έναν επιπλέον δείκτη ύπαρξης ή απουσίας της πληροφορίας. Οι τρεις αυτές ακολουθίες τροφοδότησαν τρία ξεχωριστά μονοδιάστατα συνελικτικά δίκτυα των οποίων οι έξοδοι συνενώθηκαν σε ένα διάνυσμα και τροφοδότησαν ένα δίκτυο Ĺong Short-Term Memory (LSTM). Επίσης εξήγαγαν στατικά χαρακτηριστικά, όπως οι καταναλισκόμενες θερμίδες, η απόσταση, η διάρκεια του ύπνου, και η χρονική περίοδος της ημέρας που αφορούν τα δεδομένα (νύχτα, πρωί, μεσημέρι, απόγευμα), και μαζί με τα embeddings που προέκυψαν από το LSTM συνδυάστηκαν και τροφοδότησαν ένα πλήρως συνδεδεμένο επίπεδο (Fully Connected - FC). Η τελική ταυτοποίηση του χρήστη έγινε με majority voting, δηλαδη επιλέγοντας σε ποιο χρήστη ταξινομήθηκαν τα περισσότερα δείγματα της ημέρας.

Εντοπισμός ψυχωτικών υποτροπών

Στην πρώτη εργασία μας που πραγματοποιήθηκε με στόχο τη διερεύνηση των ψυχωτικών υποτροπών [Panagiotou et al.; 2022] παρουσιάστηκαν τέσσερις διαφορετικές αρχιτεκτονικές αυτόματου κωδικοποιητή, που βασίζονται σε Transformers, Fully Connected Neural Networks (FNN), Convolution Neural Networks (CNN) και Gated Recurrent Units (GRU) [Vaswani et al.; 2017, Baldi; 2012], με τα μοντέλα να μελετώνται τόσο στην εξατομικευμένη τους μορφή όσο και στην καθολική, για όλους τους ασθενείς ενιαία. Αξιοποιήθηκαν πεντάλεπτα χαρακτηριστικά

που προέκυψαν από την εργασία [Filntisis et al.; 2020b] για τα δεδομένα δέκα ασθενών, λαμβάνοντας ενθαρρυντικά αποτελέσματα. Οι αρχιτεκτονικές CNN και FNN autoencoders φάνηκε πως αποδίδουν καλύτερα, η πρώτη για τα εξατομικευμένα μοντέλα ενώ η δεύτερη για το καθολικό. Επίσης πραγματοποιήθηκε μια ανάλυση χρησιμοποιώντας τα μοντέλα με τις καλύτερες επιδόσεις, για να εξεταστεί η δυνατότητα του επιπέδου σοβαρότητας της υποτροπής (χαμηλή, μέση, υψηλή) και παρατηρήθηκε πως υπήρξε σταδιακή αύξηση του σφάλματος ανακατασκευής όταν αυξανόταν η σοβαρότητα της υποτροπής. Ακόμα, μελετήθηκε η αλλαγή του μήκους της εισόδου του δικτύου διατηρώντας σταθερή τη χρονική ανάλυση των πέντε λεπτών και φάνηκε πως τα μεγαλύτερα μήκη εισόδου βελτίωσαν τη συνολική προγνωστική ικανότητα του μοντέλου.

Σημαντική εργασία στην πρόβλεψη του κινδύνου εμφάνισης υποτροπών (ψυχωτικών ή μη ψυχωτικών) είναι η εργασία των Fekas et al. [Fekas et al.; 2023] που βασίζεται σε μεθόδους αυτοεπιβλεπόμενης μάθησης (Self Supervised Learning - SSL) και μοντέλα επιβίωσης (Survival Analysis). Συγκεκριμένα, στο πρώτο σκέλος αξιοποιούνται και συγκρίνονται τρεις μέθοδοι contrastive SSL ([Wickstrøm et al.; 2022, Eldele et al.; 2021, Zhang et al.; 2022]) με σκοπό την εκμάθηση αναπαραστάσεων από δεδομένα 14 χρηστών σε ένα πλαίσιο ταυτοποίησης τους. Στη συνέχεια, αυτές οι αναπαραστάσεις χρησιμοποιήθηκαν για την εκπαίδευση τεσσάρων διαφορετικών μοντέλων ανάλυσης επιβίωσης ([Wright et al.; 2017, Geurts et al.; 2006, Fotso; 2018, Katzman et al.; 2018]) που προβλέπουν μια συνάρτηση επιβίωσης ή ρίσκου, δηλαδή την ύπαρξη ή όχι επακόλουθων υποτροπών. Ακόμη, εκτός της συγκριτικής μελέτης των παραπάνω τεχνικών, έγινε σύγκριση του χρονικού παραθύρου των δεδομένων που εξετάστηκαν καθώς και της χρήσης στατικών χαρακτηριστικών που μεταβάλλονται όταν εμφανίζεται μια υποτροπή, όπως το πλήθος των προηγούμενων υποτροπών, το επίπεδο σοβαρότητας της τελευταίας υποτροπής και το είδος της υποτροπής.

Οι επόμενες εργασίες παρουσιάστηκαν στα πλαίσια του διαγωνισμού e-Prevention Challenge ICASSP 2023 και παρουσιάζονται σύμφωνα με τη σειρά κατάταξης στα τελικά αποτελέσματα του διαγωνισμού. Ξεκινώντας με την ομάδα των Calcagno et al. [Calcagno et al.; 2023], οι συγγραφείς επέλεξαν να ακολουθήσουν εξατομικευμένη αντιμετώπιση του προβλήματος για κάθε ασθενή. Τρεις διαφορετικές αρχιτεκτονικές (ένας αυτόματος κωδικοποιητής που βασίζεται στο CNN, ένας αυτόματος κωδικοποιητής για χρονικές σειρές και ένας transformer autoencoder) χρησιμοποιήθηκαν για τη δημιουργία των μοντέλων. Κάθε ένα από αυτά εκπαιδεύτηκαν ξεχωριστά τόσο σε ελαφρώς επεξεργασμένα δεδομένα (σύμφωνα με την προσέγγιση της ομάδας στο πρόβλημα της ταυτοποίησης) όσο και σε συγκεντρωτικά δεδομένα (εξαγωγή χαρακτηριστικών σε aggregated δεδομένα των πέντε λεπτών) χρησιμοποιώντας τα δεδομένα περιόδων ύφεσης της διαταραχής και το Μέσο Τετράγωνο Σφάλμα (MSE) ως συνάρτηση κόστους. Στη συνέχεια έγινε επιλογή του μοντέλου με βάση την μεγαλύτερη ακρίβεια στο σύνολο επικύρωσης. Κατά την αξιολόγηση του μοντέλων υπολογίστηκε το σφάλμα ανακατασκευής και η Συνάρτηση Αθροιστικής Κατανομής (CDF) του σφάλματος ανακατασκευής ανά κανάλι (ανά χαρακτηριστικό) χρησιμοποιήθηκε ως βαθμολογία ανωμαλίας. Η τελική αξιολόγηση της εμφάνισης ή όχι υποτροπής πραγματοποιήθηκε χρησιμοποιώντας τη διάμεση τιμή της βαθμολογίας ανωμαλίας σε όλα τα διαθέσιμα παράθυρα για κάθε ημέρα.

Η ομάδα Emotion [Hamieh et al.; 2023] χρησιμοποίησαν ένα απλό νευρωνικό δίκτυο αυτοκωδικοποιητή (AE) με ένα κρυφό στρώμα και δέκα νευρώνες για την πρόβλεψη υποτροπής σε ασθενείς με ψυχωτικές διαταραχές ενώ πειραματίστηκαν και με άλλες μεθόδους(one-cass Support Vector Machine, Local Outlier Factor, και Elliptical Envelope). Η τακτική που ακολούθησαν για την αντιμετώπιση των μη έγκυρων τιμών των δεδομένων ήταν η αντικατάστασή τους με τους μέσους όρους των τιμών αυτών όπως προκύπτουν για τον εκάστοτε ασθενή. Τα χαρακτηριστικά που χρησιμοποίησαν συμπεριλαμβανομένων των καρδιακών παλμών, της γραμμικής και γωνιακής επιτάχυνσης, του ποσοστού χρόνου ύπνου και του αριθμού βημάτων, υπολογίστηκαν

σε διαστήματα τεσσάρων ωρών και κανονικοποιήθηκαν. Το τελικό τους σύστημα αυτοκωδικοποιητή εκπαιδεύτηκε χρησιμοποιώντας ως συνάρτηση κόστους την απώλεια μέσου τετραγώνου σφάλματος (MSE) ενώ το μέσο σφάλμα ανακατασκευής χρησίμευσε ως βαθμολογία ανωμαλίας για την ανίχνευση υποτροπής.

Τέλος, η ομάδα των Avramidis et al. [Avramidis et al.; 2023] αντιμετώπισε το πρόβλημα διερευνώντας τη χρήση της συμπεριφοράς του ασθενή κατά τη διάρκεια του ύπνου για να εκτιμήσουν τις ημέρες υποτροπής ως ακραίες τιμές σε ένα μη εποπτευόμενο περιβάλλον μηχανικής εκμάθησης. Οι συγγραφείς εξήγαγαν χαρακτηριστικά από τα δεδομένα της δραστηριότητας και των καρδιακών παλμών και αξιολόγησαν διάφορους συνδυασμούς τύπων χαρακτηριστικών και αναλύσεων χρόνου. Διαπίστωσαν πως η χρήση χαρακτηριστικών που περιγράφουν τη συμπεριφορά του ύπνου απέδωσε καλύτερα από τη χρήση των αντίστοιχων χαρακτηριστικών της εγρήγορσης, παρ' ότι τα πρώτα περιγράφουν μικρότερα χρονικά διαστήματα της καθημερινότητας ενός ατόμου. Σχετικά με την υλοποίηση της παραπάνω ιδέας, η ομάδα επέλεξε τη χρήση Isolation Forest [Liu et al.; 2008] όπου τα χαρακτηριστικά επιλέχθηκαν τυχαία και χωρίστηκαν σε ακραίες τιμές. Ο αριθμός των διαχωρισμών που απαιτούνται για την απομόνωση ενός δείγματος χρησίμευσε ως μέτρο κανονικότητας, καθώς οι ανωμαλίες είναι πιθανό να έχουν μικρότερες διαδρομές λόγω τυχαίας κατάτμησης. Αυτό το μέτρο υπολογίζεται κατά μέσο όρο σε ένα δάσος τέτοιων τυχαίων δέντρων για να εκτιμηθεί η απόδοση ανίχνευσης ακραίων τιμών. Στην εργασία τους, η χρήση της δραστηριότητας ύπνου, του αριθμού βημάτων και των χαρακτηριστικών καρδιακών παλμών διερευνήθηκαν επίσης αξιολογώντας διάφορους συνδυασμούς χαρακτηριστικών και αναλύσεων χρόνου.

Τέλος, σε μια πολυτροπική προσέγγιση του προβλήματος, μελετήθηκε η αναγνώριση υποτροπής με παράλληλη χρήση ηχητικών δεδομένων από συνεντεύξεις μεταξύ των ασθενών και των κλινικών ιατρών, και των βιοσημάτων που συλλέχθηκαν κατά τη διάρκεια της ημέρας των συνεντεύξεων και παρουσιάστηκε στο [Zlatintsi et al.; 2022]. Η εργασία μας επεκτείνει την [Garoufis et al.; 2022] όπου διερευνήθηκαν οι δυνατότητες των δισδιάστατων Convolutional Variational Autoencoder στην ανίχνευση των υποτροπών από αυθόρμητη ομιλία. Τα αποτελέσματα είχαν δείξει ότι στην εξατομικευμένη περίπτωση, αυτά τα μοντέλα λειτουργούν ισοδύναμα με τη ντετερμινιστική εκδοχή τους. Στην περίπτωση του καθολικού μοντέλου, πέτυχαν συγκρίσιμες επιδόσεις με τις εξατομικευμένες όταν ακολουθήθηκε ένα πρωτόκολλο κανονικοποίησης ανά ασθενή ενώ χρησιμοποιώντας norm pooling για τη συγκέντρωση των αποτελεσμάτων ανά συνεδρία βελτιώνεται περαιτέρω η απόδοση του συστήματος.

Έτσι, στη μελέτη μας για τη σύμμειξη των δύο τύπων πληροφορίας (ομιλία και βιοσήματα) [Garoufis et al.; 2022] έγινε πειραματισμός με διαφορετικά σχήματα σύμμειξης της πληροφορίας και στο επίπεδο των δύο τροπικοτήτων και στο επίπεδο της συσσώρευσης της πληροφορίας (aggrega- tion). Παρατηρήθηκε πως ο συνδυασμός της ομιλίας και των βιοσημάτων είτε προσθετικά είτε πολλαπλασιαστικά σε επίπεδο late fusion αποδίδει καλύτερα αποτελέσματα από τη χρήση ενός μόνο τύπου, αναδεικνύοντας τις δυνατότητες της σύμμειξης της πληροφορίας από διαφορετικές τροπικότητες για πιο αποτελεσματική ανίχνευση υποτροπών. Η σύμμειξη με προσθετικό τρόπο αποδίδει ελαφρώς καλύτερα από την πολλαπλασιαστική σύμμειξη, ενώ σε επίπεδο aggregation παρατηρήθηκε πως το norm pooling λειτουργεί καλύτερα για τα δεδομένα ήχου σε αντίθεση με τα δεδομένα του ρολογιού όπου είναι βέλτιστη η επιλογή του ημερήσιου μέσου όρου των βαθμολογιών ανά ώρα.

5

Δημιουργία Ψηφιακού Φαινοτύπου - Ταυτοποίηση Ατόμου

Ο φαινότυπος ορίζεται ως το το σύνολο των χαρακτηριστικών ενός οργανισμού που προκύπτει από τον συνδυασμό των γονιδίων του και την επίδραση του περιβάλλοντος του. Η εξέλιξη της τεχνολογίας και η αξιοποίησή της στον τομέα της υγείας έχουν επιτρέψει την επέκταση αυτού του ορισμού ως προς την ψηφιακό κόσμο ορίζοντας πλέον και τον ψηφιακό φαινότυπο ενός ατόμου. Έτσι ο ψηφιακός φαινότυπος αναφέρεται στο σύνολο των χαρακτηριστικών που καταγράφονται μέσω των ψηφιακών συσκευών-εφαρμογών, π.χ.της χρήσης των μέσων κοινωνικής δικτύωσης και των φορητών-έξυπνων συσκευών [Torous et al.; 2016]. Δημιουργείται λοιπόν ένα ψηφιακό αποτύπωμα του ατόμου, ένα σύνολο δηλαδή συμπεριφορών και γνωρισμάτων που είναι ανιχνεύσιμα και μπορούν κατ' επέκταση, στον τομέα της υγείας, να αποκαλύψουν συνήθειες και να βοηθήσουν στην παρακολούθηση ασθενειών. Τέτοια δεδομένα δίνουν πρόσθετη αξία στην κλινική εξέταση, στους εργαστηριακούς δείκτες και στα δεδομένα κλινικής απεικόνισης - δηλαδή στις παραδοσιακές ως τώρα προσεγγίσεις για τον χαρακτηρισμό ενός φαινοτύπου μιας ασθένειας. Με τη συγκέντρωση και την κατάλληλη ανάλυση αυτών των δεδομένων μπορεί να αλλάξει θεμελιωδώς η αντίληψή μας για τις εκδηλώσεις μιας νόσου παρέχοντας μια πιο ολοκληρωμένη και πιο λεπτομερή άποψη του πως ο ασθενής βιώνει την ασθένεια. Μέσω του ψηφιακού φαινοτύπου, η χρήση των ψηφιακών τεχνολογιών μπορεί να επηρεάσει όλη την πορεία της ανθρώπινης νόσου από τη διάγνωση, τη θεραπεία και τη διαχείριση χρόνιων ασθενειών [Jain et al.; 2015].

Ως επέκταση των παραδοσιακών μορφών παρατήρησης μιας νόσου, οι ψηφιακοί φαινότυποι μπορούν να διευρύνουν την ικανότητά μας να εντοπίσουμε και να διαγνώσουμε διαταραχές. Πληροφορίες από ψηφιακά προϊόντα και μέσα κοινωνικής δικτύωσης παρέχουν νέους χώρους για τον εντοπισμό και την παρακολούθηση των συμπτωμάτων της νόσου. Για παράδειγμα, οι ερευνητές έχουν αποδείξει ότι τα δεδομένα αναζήτησης Google μπορούν να χρησιμοποιηθούν για τον προσδιορισμό του αυτοκτονικού ιδεασμού [Gunn III and Lester; 2013]. Έτσι, η εκτίμηση του ψηφιακού φαινοτύπου ενός ατόμου θα μπορούσε να βοηθήσει στην έγκαιρη ανίχνευση της νόσου, στον εντοπισμό των συμπτωμάτων πριν από την παραδοσιακή φαινοτυπική έκφραση - και δυνητικά τη δημιουργία εργαλείων για έγκαιρη παρέμβαση. Ο ρόλος των ψηφιακών φαινοτύπων στη διάγνωση εκτείνεται πέρα από την επιτήρηση και την έγκαιρη ανίχνευση. Οι ψηφιακοί φαινότυποι επαναπροσδιορίζουν την έκφραση της νόσου μέσω της εμπειρίας των ατόμων, γεγονός που διευρύνει την ικανότητά μας να ταξινομούμε και να κατανοούμε τις ασθένειες. Για έναν ασθενή με αϋπνία, τα δεδομένα σχετικά με το χρόνο και τις ώρες του ψηφιακού του αποτυπώματος μπορούν να θεωρηθούν μέρος της έκφρασης της νόσου. Ομοίως, για έναν διπολικό ασθενή του οποίου η μανία εκδηλώνεται σε γρήγορη, αδιάλειπτη ομιλία ή υπεργραφία (hypergraphia), η ασθένειά του θα μπορούσε να χαρακτηριστεί από τη συχνότητα, τη διάρ-

κεια και το περιεχόμενο της συμμετοχής στα μέσα κοινωνικής δικτύωσης. Μέσω αυτών των ποικίλων εφαρμογών, οι ψηφιακοί φαινότυποι μπορούν να βοηθήσουν στην παρατήρηση των πρώιμων εκδηλώσεων της νόσου και να επιτρέψουν στο σύστημα υγειονομικής περίθαλψης να αναπτύξει πιο ευκίνητες, στοχευμένες και γρήγορες παρεμβάσεις. Ως συνεχείς μετρούμενες εκδηλώσεις βιολογικής νόσου, οι ψηφιακοί φαινότυποι μπορούν να αποτελέσουν χρήσιμο εργαλείο στις παραδοσιακές προσεγγίσεις στη θεραπεία και τη διαχείριση της νόσου. Με τον επαναπροσδιορισμό της εκδήλωσης της ασθένειας, παρέχουν νέους τρόπους μέτρησης της νόσου και της θεραπευτικής ανταπόκρισης με τρόπους που έχουν μεγαλύτερη σημασία για τους ασθενείς [Jain et al.; 2015].

Για το λόγο αυτό, **πρώτος σταθμός στην ανάλυσή μας είναι η δημιουργία και μελέτη ψηφιακών φαινοτύπων ικανών να περιγράφουν το προφίλ των χρηστών έξυπνων ρολογιών και η αποκάλυψη μοτίβων των δράσεων της καθημερινότητάς τους. Απώτερος στόχος είναι η μελέτη των διακυμάνσεων, από αυτό το «κανονικό» μοτίβο του κάθε χρήστη και άρα η αποκάλυψη κάποιας απόκλισης που συσχετίζεται με πιθανή υποτροπή της ψυχωτικής διαταραχής του**. Πρόσφατα, όπως είδαμε στο Κεφάλαιο 4, μια σειρά από εργασίες χρησιμοποίησαν με επιτυχία βιομετρικά δεδομένα από φορητές συσκευές προκειμένου να προσδιορίσουν την ταυτότητα του χρήστη, π.χ. [Retsinas et al.; 2020, Maiorana et al.; 2022]. Όπως φαίνεται δηλαδή οι μέθοδοι βαθιάς μάθησης μπορούν να χρησιμοποιηθούν με επιτυχία για τον εντοπισμό διακριτών προτύπων συμπεριφοράς.

Έτσι, στο παρόν κεφάλαιο αρχικά παρουσιάζουμε, για καλύτερη κατανόηση του αναγνώστη, μια στατιστική διερεύνηση των βιοδεικτών που εξήγαμε από τα σήματα που συλλέξαμε μέσω των έξυπνων ρολογιών. Στη συνέχεια, αναπτύσσουμε και αναλύουμε μια αρχιτεκτονική βαθιών νευρωνικών δικτύων για την ταξινόμηση ψηφιακών φαινοτύπων, δηλαδή, για την ταυτοποίηση του εκάστοτε χρήστη χρησιμοποιώντας σήματα κίνησης και καρδιακά σήματα από ένα έξυπνο ρολόι.

5.1 Μελέτη Βιοδεικτών για τη Δημιουργία Φαινοτύπων

Αρχικά στις εργασίες [Filntisis et al.; 2020b, Zlatintsi et al.; 2022] εξαγάγαμε μια πληθώρα γραμμικών και μη γραμμικών χαρακτηριστικών με σκοπό τη μελέτη της σημαντικότητάς τους για τη δημιουργία ψηφιακών φαινοτύπων. Ακόμα, για να εξετάσουμε την ύπαρξη διαφορών ανάμεσα στις δύο ομάδες των συμμετεχόντων στο έργο e-Prevention, τους μάρτυρες (control group) και τους ασθενείς, παρουσιάσαμε μια συγκριτική ανάλυση ως προς τα χαρακτηριστικά αυτά. Παρακάτω παρουσιάζονται τα πιο σημαντικά σημεία που θα βοηθήσουν στην περαιτέρω κατανόηση της διατριβής.

5.1.1 Προεπεξεργασία δεδομένων

Αρχικά, μετά την συλλογή των ακατέργαστων δεδομένων από τις καταγραφές των ρολογιών, κάνουμε μια πρώτη επεξεργασία με σκοπό τον έλεγχο των δεδομένων, την αποθορυβοποίηση των σημάτων καθώς και τη μείωση της διάστασής τους με σκοπό την πιο εύκολη διαχείρισή τους. Στην παράγραφο αυτή αναλύεται η διαδικασία που ακολουθείται για την προεπεξεργασία των δεδομένων.

Στην επεξεργασία των σημάτων αξιοποιήσαμε την ανάλυση σημάτων με χρήση παραθύρων βραχέως χρόνου που αποτελεί μια παραδοσιακή μέθοδο επεξεργασίας σημάτων. Ακολουθώντας αυτή την τεχνική, αλλά αυξάνοντας σε μεγάλο βαθμό τη χρονική κλίμακα, προχωρήσαμε στην ανάλυση των σημάτων σε παράθυρα των πέντε λεπτών τόσο για τα δεδομένα κίνησης όσο και για δεδομένα μεταβολής του καρδιακού παλμού (Heart-Rate Variability). Η επιλογή

	Υγιείς	Ασθενείς
Άντρες/Γυναίκες	12/11	16/8
Προεπεξεργασμένα Δεδομένα		
# Μέρες καταγραφής	84.3 ± 30.9	68.5 ± 41.7
# 5λεπτα κινησιακών δεδομένων (εγρήγορση)	15746 ± 4837	13210 ± 6908
# 5λεπτα HRV (εγρήγορση)	12909 ± 3589	12221 ± 6656
# 5λεπτα κινησιακών δεδομένων (ύπνο)	7670 ± 2606	8865 ± 4767
# 5λεπτα HRV (ύπνο)	6924 ± 2331	8578 ± 4741

Πίνακας 5.1: Δημογραφικές πληροφορίες των υγιών και ασθενών εθελοντών και όγκος δεδομένων μετά την προεπεξεργασία τους και την εξαγωγή πεντάλεπτων χαρακτηριστικών (κινησιακών και καρδιακών δεδομένων) για κάθε ομάδα κατά τη διάρκεια της εγρήγορσης και του ύπνου. Δεν υπήρχαν σημαντικές διαφορές μεταξύ των ποσοτήτων των καταγεγραμμένων δεδομένων μεταξύ των δύο ομάδων.

του συγκεκριμένου χρονικού παραθύρου έγινε καθώς στην εργασία [Retsinas et al.; 2020] παρατηρήσαμε πως πεντάλεπτα διαστήματα σήματος περιέχουν σημαντικές πληροφορίες για τη διάκριση βραχυπρόθεσμων μοτίβων.

Για το τελικό σήμα της μεταβολής του καρδιακού παλμού, λαμβάνουμε την ακολουθία των RR-intervals (των χρονικών διαστημάτων ανάμεσα στους καρδιακούς παλμούς) που προκύπτει από την καταγραφή του φωτοπληθυσμογράφου συχνότητας 5 Hz και απορρίπτουμε τις όμοιες διαδοχικές τιμές. Επίσης όσες τιμές των RR-intervals είναι μεγαλύτερες από 2000 ms και μικρότερες από 300 ms τις αντιμετωπίζουμε ως πιθανούς μη ανιχνευμένους παλμούς και τις αντικαθιστούμε με τις κατάλληλες τιμές εφαρμόζοντας γραμμική παρεμβολή. Λαμβάνουμε αυτή την απόφαση καθώς είναι εξαιρετικά πιθανό οι συγκεκριμένες τιμές να είναι εσφαλμένες καθώς αντιστοιχούν σε λιγότερους από 30 παλμούς/λεπτό ή πάνω από 200 παλμούς/λεπτό αντίστοιχα. Μετά την προεπεξεργασία, εξάγουμε χαρακτηριστικά από τα πρώτα 4.5 λεπτά (90%) της ακολουθίας των RR-intervals απορρίπτοντας οποιεσδήποτε τιμές έπονται ώστε όλες οι ακολουθίες να έχουν σταθερό μήκος.

Στα δεδομένα που συλλέχθηκαν από τους αισθητήρες καταγραφής της γραμμικής και της γωνιακής επιτάχυνσης, πρώτα αφαιρούμε τα διαστήματα για τα οποία περισσότερες από 50 συνεχόμενες τιμές δεν έχουν καταγραφεί. Στη συνέχεια, αντικαθιστούμε τις τιμές αυτές κάνοντας παρεμβολή με χρήση της πλησιέστερης τιμής και εξάγουμε χαρακτηριστικά από τα πρώτα 5940 (99%) δείγματα του κάθε πεντάλεπτου διαστήματος. Επίσης εφαρμόζουμε αποθορυβοποίηση του σήματος (high-frequency wavelet denoising [Vantuch; 2018]) προκειμένου να εξομαλύνουμε τον εγγενή θόρυβο από τους αισθητήρες.

Ο μέσος όρος και η τυπική απόκλιση του πλήθους των διαστημάτων των πεντάλεπτων χαρακτηριστικών που προέκυψαν μετά την προεπεξεργασία και αξιοποιήθηκαν στη συγκεκριμένη έρευνα, για κάθε ομάδα χρηστών, αναφέρονται στον Πίνακα 5.1.

5.1.2 Εξαγωγή Βιοδεικτών

Στη συνέχεια παρουσιάζουμε το πως οι υπολογίζονται οι διάφοροι βιοδείκτες από τα σήματα που συλλέγουμε από τα ρολόγια: **Ενέργεια γραμμικής και γωνιακής επιτάχυνσης**: Εξάγουμε την ενέργεια βραχέως χρόνου (Śhort-Time Energy - STE) της ευκλείδειας νόρμας των σημάτων του επιταχυνσιόμετρου *acc*) και του γυροσκόπιου (*gyr* (οι μετρήσεις του ρολογιού υπολογίζονται σε τρεις άξονες). Χρησιμοποιούμε αυτά τα χαρακτηριστικά ως αντικειμενικό μέτρο της σωματικής δραστηριότητας και της γενικής συμπεριφοράς κίνησης. Για κάθε πεντάλεπτο

διάστημα υπολογίζουμε τη μέση τιμή και την τυπική απόκλιση των μετρήσεων.

Οι επόμενοι δείκτες υπολογίζονται με τη χρήση των ακολουθιών RR-intervals.

Φασματικά χαρακτηριστικά: Η Φασματική Πυκνότητα Ισχύος (Power Spectral Density - PSD) είναι μια κοινή και ισχυρή μέθοδος για την ανάλυση του συχνοτικού περιεχομένου των σημάτων που περιγράφει τη σχετική ενέργεια των διακυμάνσεων του σήματος.Σύμφωνα με ιατρικές μελέτες, το φάσμα του HRV χωρίζεται αποδοτικά σε τέσσερις ζώνες συχνοτήτων: Ultra Low Frequency (ULF $\leq 0,003$ Hz), Very Low Frequency (VLF $0,0033$–$0,04$ Hz), Low Frequency (LF $0,04$– $0,15$ Hz) και High Frequency (HF $0,15$–$0,40$ Hz) [Shaffer and Ginsberg; 2017]. Δεδομένου ότι το HRV είναι, εξ ορισμού, ένα μη ομοιόμορφα δειγματοληπτημένο σήμα, εκτελούμε φασματική ανάλυση χρησιμοποιώντας το περιοδόγραμμα Lomb-Scargle [Scargle; 1982]. Το περιοδόγραμμα Lomb-Scargle είναι μια μέθοδος εκτίμησης φάσματος ισχύος που μπορεί να εφαρμοστεί απευθείας σε σήματα μη ομοιόμορφα δειγματοληπτημένα και ως εκ τούτου είναι κατάλληλο για μετρήσεις HRV. Το περιοδόγραμμα ορίζεται ως εξής:

$$P_{LS}(\Omega) = \frac{1}{2}\left\{\frac{\left[\sum_{n=0}^{N-1} x[n]\cos\left(\Omega\left(t_n-\tau\right)\right)\right]^2}{\sum_{n=0}^{N-1}\cos^2\left(\Omega\left(t_n-\tau\right)\right)} + \frac{\left[\sum_{n=0}^{N-1} x[n]\sin\left(\Omega\left(t_n-\tau\right)\right)\right]^2}{\sum_{n=0}^{N-1}\sin^2\left(\Omega\left(t_n-\tau\right)\right)}\right\}, \quad (5.1)$$

όπου το τ δίνεται ως εξής:

$$\tau = \frac{1}{2\Omega}\tan^{-1}\left(\frac{\sum_{n=0}^{N-1}\sin\left(2\Omega t_n\right)}{\sum_{n=0}^{N-1}\cos\left(2\Omega t_n\right)}\right), \quad (5.2)$$

και Ω είναι η γωνιακή συχνότητα ($rad/second$), t_n ο χρόνος κατά τον οποίο έγινε η δειγματοληψία ($second$) και $x[n]$ η τιμή του σήματος τη στιγμή t_n. Χρησιμοποιώντας το περιοδόγραμμα Lomb-Scargle, εξάγουμε για κάθε διάστημα την κανονικοποιημένη ισχύ σε δύο ζώνες: LF και HF, καθώς και την αναλογία LF/HF.

Στη συνέχεια υπολογίζουμε με χρήση μη γραμμικών μεθόδων τους επόμενους δείκτες, δηλαδή με μεθόδους που αντιμετωπίζουν τις χρονοσειρές RR-intervals ως την έξοδο ενός μη γραμμικού συστήματος. Ένα τυπικό χαρακτηριστικό ενός μη γραμμικού συστήματος είναι η πολυπλοκότητά του.

Sample Entropy: Το πρώτο μέτρο πολυπλοκότητας που εξετάζουμε είναι η εντροπία του δείγματος (SampEn). Η εντροπία δείγματος είναι ένα μέτρο του ρυθμού της πληροφορίας που δημιουργείται από το σύστημα και έχει θεωρηθεί ότι είναι μια βελτιωμένη έκδοση της κατά προσέγγιση εντροπίας [Richman and Moorman; 2000], λόγω της αμερόληπτης φύσης της.

Higuchi Fractal Διάσταση: Υπολογίζουμε την φράκταλ διάσταση Higuchi [Higuchi; 1988], η οποία έχει χρησιμοποιηθεί εκτενώς στη νευροφυσιολογία λόγω της απλότητας και της υπολογιστικής της ταχύτητας [Al-Nuaimi et al.; 2017, Kesić and Spasić; 2016, Khoa et al.; 2012].

Multiscale Fractal Διάσταση: Η Multiscale Fractal Dimension (MFD) είναι ένας αποτελεσματικός αλγόριθμος [Maragos; 1994] που μετρά τη βραχυχρόνια διάσταση φράκταλ, με βάση τη διάσταση Minkowski-Bouligand [Falconer; 2004]. Ο αλγόριθμος αυτός μετρά τη σύντομη χρονική διάσταση φράκταλ χρησιμοποιώντας μη γραμμικά μορφολογικά φίλτρα πολλαπλής κλίμακας που μπορούν να δημιουργήσουν γεωμετρικά καλύμματα γύρω από το γράφημα ενός σήματος, του οποίου η φράκταλ διάσταση D μπορεί να βρεθεί από:

$$D = \lim_{s\to 0}\frac{\log[\text{Area of dilated graph by disks of radius } s/s^2]}{\log(1/s)}.$$

Όπως είναι γνωστό, το D είναι μεταξύ 1 και 2 για μονοδιάστατα σήματα και όσο μεγαλύτερο είναι το D, τόσο μεγαλύτερος είναι ο βαθμός γεωμετρικού κατακερματισμού του σήματος. Στην

πράξη, τα σήματα του πραγματικού κόσμου δεν έχουν την ίδια δομή σε διαφορετικές κλίμακες. Εδώ εκμεταλλευόμαστε το γεγονός ότι η διάσταση φράκταλ σύντομου χρόνου στη μικρότερη διακριτή κλίμακα ($s = 1$) έχει βρεθεί ότι παρέχει κάποια διάκριση μεταξύ διαφόρων συμβάντων. Έτσι, συνοψίσαμε τα βραχέως χρόνου υπολογισμένα MFD profiles παίρνοντας τη μέση τιμή της fd[1] φράκταλ διάστασης για κάθε διάστημα 5 λεπτών HRV.

Μετρήσεις Γραφήματος Poincare: Η γραφική παράσταση Poincare [Brennan et al.; 2001] είναι μια γραφική παράσταση επανάληψης, όπου κάθε δείγμα μιας χρονοσειράς σχεδιάζεται έναντι της προηγούμενης και, στη συνέχεια, προσαρμόζεται μια έλλειψη στο διάγραμμα διασποράς. Το πλάτος της έλλειψης (SD1) είναι ένα μέτρο των βραχυπρόθεσμων διακυμάνσεων του HRV, ενώ το μήκος (SD2) είναι ένα μέτρο των μακροπρόθεσμων διακυμάνσεων του HRV.

Τέλος, για λόγους συνοχής, θα παρουσιάσουμε εδώ και κάποια στατιστικά μέτρα που υπολογίζονται από τα διαστήματα RR και χρησιμοποιούνται στη συνέχεια του κεφαλαίου. Οι τιμές αυτές ποσοτικοποιούν τη μεταβολή των χρονικών διαστημάτων μεταξύ των διαδοχικών καρδιακών παλμών.

Μέση Τιμή και Standard Deviation of RR intervals (): Υπολογίζουμε τη μέση τιμή και την τυπική απόκλιση των RR διαστημάτων. Στις εργασίες [Henry et al.; 2010, Voss et al.; 2006], η μέση τιμή και η μεταβλητότητα του καρδιακού ρυθμού φάνηκε πως είναι ένα επαρκές, αυτόνομο χαρακτηριστικό για τη διάκριση μεταξύ ασθενών και μαρτύρων. Οι περισσότερες μελέτες χρησιμοποιούν κάποια παραλλαγή της τυπικής απόκλισης στις αναλύσεις τους. Παρόλο που, συνήθως, δεν είναι στατιστικά σημαντικό μέτρο λόγω της δικής του υψηλής τυπικής απόκλισης, οι Voss et al. [Voss et al.; 2006] διαπίστωσαν ότι το SDRR ήταν επαρκές μέτρο για τη διάκριση μεταξύ ασθενών που λαμβάνουν φαρμακευτική αγωγή και αυτών χωρίς αγωγή.

Root Mean Squared of Successive RR interval Differences (): Η τετραγωνική ρίζα της μέσης τιμής των τετράγωνων των διαφορών των διαδοχικών παλμών, δηλαδή

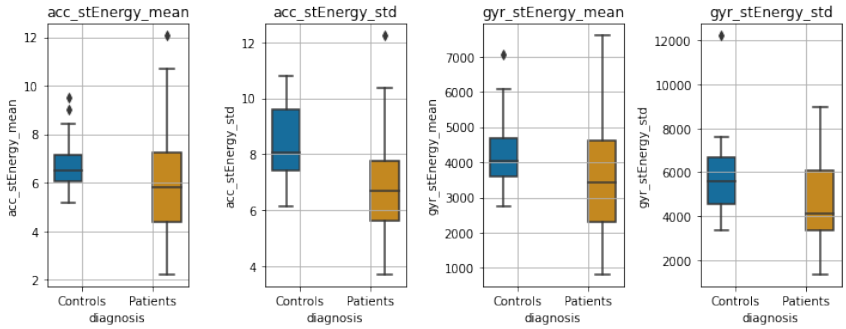
$$\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(RR_i - RR_{i-1})^2}, \tag{5.3}$$

είναι ένα τυπικό χαρακτηριστικό στις περισσότερες αναλύσεις. Παρ' ότι δεν έχει αποδειχθεί επαρκής για τη διάκριση των μαρτύρων και των ασθενών, ωστόσο οι [Voss et al.; 2006] βρήκαν ότι αυτό το μέτρο μπορεί επίσης να διακρίνει τους ασθενείς που δεν λαμβάνουν φαρμακευτική αγωγή.

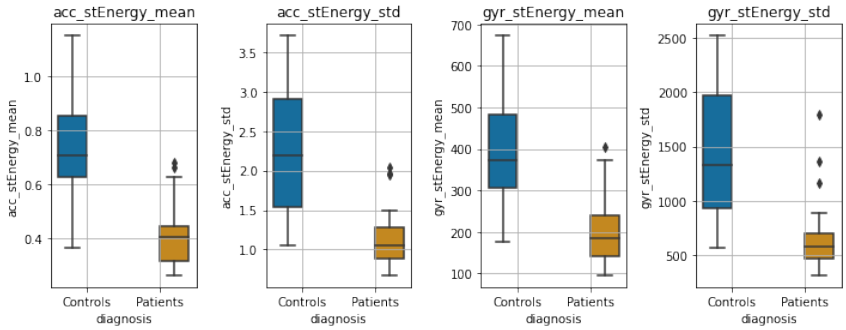
Χρονικός Διαχωρισμός Χαρακτηριστικών

Για την παρούσα μελέτη διακρίνουμε δύο περιόδους στην ημέρα κάθε ατόμου, την περίοδο του ύπνου και την περίοδο εγρήγορσης (περίοδος κατά την οποία είναι ξύπνιος). Έτσι, χρησιμοποιώντας τις πληροφορίες από το πρόγραμμα ύπνου κάθε ατόμου, χωρίζουμε τα διαστήματα σε δύο ομάδες. Στη συνέχεια υπολογίσαμε τη μέση τιμή (mean) και την τυπική απόκλιση (std) σε όλα τα μεμονωμένα διαστήματα, με αποτέλεσμα δύο τιμές για κάθε άτομο και τύπο χαρακτηριστικού. Πραγματοποιήσαμε Student's t-test οι οποίες δεν έδειξαν στατιστικά σημαντικές διαφορές μεταξύ των διαστημάτων που καταγράφηκαν για την κίνηση και την μεταβολή της καρδιακής μεταβολής για κάθε ομάδα και περίοδο ημέρας (δηλαδή, κατά τη διάρκεια του ύπνου και της εγρήγορσης).

Εκτός από τα παραπάνω χαρακτηριστικά, εξάγουμε επίσης για κάθε άτομο τη μέση και τυπική απόκλιση της αναλογίας ύπνου/εγρήγορσης και τον μέσο αριθμό βημάτων ανά ημέρα. Δεδομένου ότι ο αριθμός των καταγεγραμμένων ωρών κάθε μέρα κυμαίνεται, για αυτές τις μετρήσεις χρησιμοποιούμε μόνο ημέρες με τουλάχιστον 20 καταγεγραμμένες ώρες (δεν βρέθηκε σημαντική διαφορά μεταξύ του αριθμού των ημερών για τους ελέγχους και των ασθενών που χρησιμοποιούν τη δοκιμή Mann-Whitney U [Mann and Whitney; 1947]).



(α) Εγρήγορση



(β) Ύπνος

Σχήμα 5.1: Boxplots διαγράμματα για τα χαρακτηριστικά που προκύπτουν από το επιταχυνσιόμετρο και το γυροσκόπιο των μαρτύρων (μπλε) και των ασθενών (κίτρινο) ενώ (α) είναι ξύπνιοι και (β) κοιμούνται. Η μαύρη γραμμή σε κάθε boxplot αντιπροσωπεύει τη διάμεσο και τα έγχρωμα παραλληλόγραμμα εκτείνονται μεταξύ του 1ου και του 3ου τεταρτημορίου του εύρους των τιμών. Οι κατακόρυφες μαύρες γραμμές εκτείνονται έως τη μικρότερη και τη μεγαλύτερη τιμή των χαρακτηριστικών εντός ενός εύρους 1.5*IQR και του 1ου ή 3ου τεταρτημορίου αντίστοιχα. Οι ακραίες τιμές (outliers) εμφανίζονται ως διαμάντια.

Στατιστική Ανάλυση

Στο Σχήμα 5.1 παρουσιάζονται τα boxplots των χαρακτηριστικών που εξάγονται από τα δεδομένα του επιταχυνσιόμετρου και του γυροσκόπιου κατά τη διάρκεια της εγρήγορσης και του ύπνου, ενώ στο Σχήμα 5.2 τα boxplots των χαρακτηριστικών HRV για τις δύο καταστάσεις, αντίστοιχα. Λόγω των διαφορών μεταξύ των κατανομών στα περισσότερα χαρακτηριστικά που παρατηρούνται εύκολα από τα boxplots γραφήματα, ελέγξαμε αν υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ των κατανομών χρησιμοποιώντας μη παραμετρικούς two-tailed Mann-Whitney U ελέγχους [Mann and Whitney; 1947]. Ως μηδενική υπόθεση θεωρήσαμε πως η κατανομή των χαρακτηριστικών μεταξύ των δύο ομάδων (ελέγχου και ασθενών) είναι όμοιες. Λόγω των πολλαπλών ελέγχων, προσαρμόσαμε τα p-values χρησιμοποιώντας τη μέθοδο Benjamini-Hochberg (BH) [Benjamini and Hochberg; 1995]. Ο Πίνακας 5.2 δείχνει τα αποτελέσματα των ελέγχων Mann-Whitney U για όλα τα χαρακτηριστικά.

Κατά τη διάρκεια της εγρήγορσης, τα χαρακτηριστικά που σχετίζονται με τις κινήσεις φαίνεται να παρουσιάζουν μεγαλύτερη μεταβλητότητα στην ομάδα ασθενών σε σύγκριση με την ομάδα ελέγχου, όπως φαίνεται στο Σχήμα 5.1(α). Το ίδιο φαίνεται να ισχύει και για ορισμένα μη γραμμικά χαρακτηριστικά HRV, όπως για παράδειγμα SampEn mean, Higuchi mean και std, SD1 and SD2, MFD mean και ορισμένα από τα χαρακτηριστικά του τομέα συχνότητας, όπως φαίνεται στο Σχήμα 5.2(α). Επιπλέον, ο έλεγχος σημαντικότητας, που παρουσιάζεται

		Εγρήγορση			Ύπνος		
	Features	Controls	Patients	p-value	Controls	Patients	p-value
acc	STE mean	6.517 (1.058)	5.832 (2.902)	0.15	0.708 (0.228)	0.406 (0.129)	< **0.001**
	STE std	**8.065** (**2.174**)	**6.694** (**2.147**)	**0.02**	2.199 (1.375)	1.057 (0.393)	< **0.001**
gyr	STE mean	4045 (1080)	3431 (2313)	0.14	372 (177.542)	185.166 (97.117)	< **0.001**
	STE std	**5572** (**2125**)	**4110** (**2728**)	**0.05**	1324 (1032)	578 (235.108)	< **0.001**
HRV	SampEn mean	**1.446** (**0.217**)	**1.260** (**0.281**)	**0.03**	1.435 (0.180)	1.505 (0.169)	0.14
	SampEn std	**0.407** (**0.052**)	**0.452** (**0.059**)	**0.03**	0.370 (0.063)	0.376 (0.117)	0.61
	Higuchi mean	1.974 (0.010)	1.966 (0.017)	0.07	1.875 (0.067)	1.915 (0.086)	0.18
	Higuchi std	0.040 (0.007)	0.043 (0.016)	0.17	0.089 (0.020)	0.079 (0.024)	0.41
	SD1 mean	214.040 (20.128)	194.322 (33.829)	0.08	78.814 (29.462)	72.437 (16.211)	0.61
	SD1 std	**56.058** (**7.166**)	**63.894** (**9.414**)	**0.02**	60.625 (21.202)	55.604 (14.806)	0.45
	SD2 mean	237.053 (23.944)	219.853 (41.005)	0.14	112.232 (26.737)	104.269 (29.907)	0.32
	SD2 std	**58.511** (**6.954**)	**67.642** (**13.689**)	**0.01**	63.169 (13.386)	61.827 (8.968)	0.94
	LF/HF mean	0.449 (0.001)	0.449 (0.001)	0.48	0.445 (0.002)	0.445 (0.005)	0.93
	LF/HF std	**0.066** (**0.001**)	**0.067** (**0.001**)	**0.02**	**0.066** (**0.001**)	**0.067** (**0.001**)	< **0.001**
	LF mean	30.857 (0.062)	30.858 (0.067)	0.43	30.652 (0.107)	30.644 (0.236)	0.84
	LF std	**3.109** (**0.018**)	**3.130** (**0.031**)	**0.02**	**3.134** (**0.047**)	**3.192** (**0.043**)	< **0.001**
	HF mean	69.143 (0.062)	69.142 (0.067)	0.43	69.348 (0.107)	69.356 (0.236)	0.84
	HF std	**3.109** (**0.018**)	**3.130** (**0.031**)	**0.02**	**3.134** (**0.047**)	**3.192** (**0.043**)	< **0.001**
	MFD mean mean	1.696 (0.035)	1.655 (0.055)	0.05	1.529 (0.046)	1.516 (0.069)	0.26
	MFD mean std	**0.093** (**0.014**)	**0.108** (**0.021**)	**0.01**	0.085 (0.023)	0.086 (0.013)	0.93
walk	Steps mean	**7054** (**2358**)	**3960** (**2928**)	**0.01**	-	-	-
	Steps std	**3513** (**1505**)	**2755** (**756**)	**0.05**	-	-	-
sleep	Ratio mean	-	-	-	**0.579** (**0.107**)	**0.886** (**0.471**)	< **0.001**
	Ratio std	-	-	-	**0.240** (**0.149**)	**0.389** (**0.304**)	**0.01**

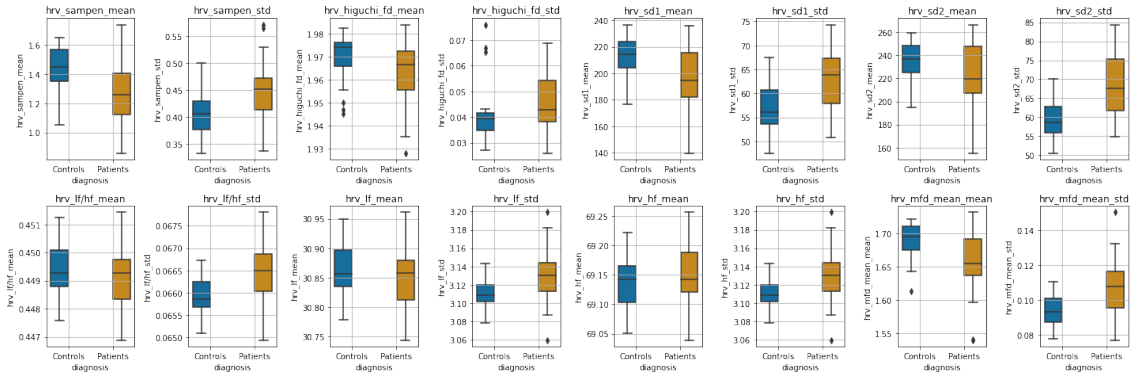
Πίνακας 5.2: Ανάλυση στατιστικών διαφορών χρησιμοποιώντας ελέγχους U Mann-Whitney με BH διόρθωση για κάθε κατάσταση (εγρήγορση, ύπνος). Οι έντονες τιμές υποδηλώνουν στατιστική σημαντικότητα για επίπεδα εμπιστοσύνης 95%. Σε παρένθεση εμφανίζεται για κάθε ομάδα η διάμεσος και το ενδοτεταρτημοριακό εύρος (IQR) για κάθε χαρακτηριστικό.

στον Πίνακα 5.2 έδειξε σημαντικές διαφορές κατανομής στην τυπική απόκλιση *std* της ενέργειας βραχέως-χρόνου των *acc* και *gyr*, η *mean* και *std* του SampEn, η *std* του SD1 και SD2, η *std* του λόγου LF, HF και LF/HF και η *std* του *MFD mean*. Τα υπόλοιπα χαρακτηριστικά απέτυχαν να απορρίψουν τη μηδενική υπόθεση, δηλαδή ότι υπάρχει στατιστικά σημαντική διαφορά μεταξύ των δύο ομάδων του πληθυσμού.

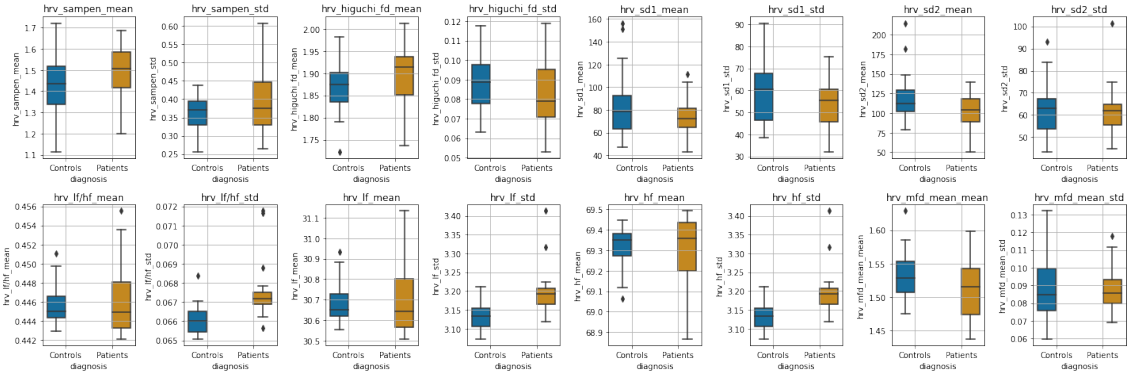
Όμοια, το Σχήμα 5.1(β) παρουσιάζει τις κατανομές χαρακτηριστικών του επιταχυνσιόμετρου και του γυροσκόπιου για κάθε ομάδα κατά τη διάρκεια του ύπνου, ενώ το Σχήμα 5.2(β) δείχνει τις κατανομές των χαρακτηριστικών HRV. Είναι προφανές ότι ειδικά τα σχετιζόμενα με την κίνηση χαρακτηριστικά παρουσιάζουν σημαντική διαφορά, η οποία επαληθεύεται επίσης στα αποτελέσματα του ελέγχου Mann-Whitney U που εμφανίζονται στον Πίνακα 5.2. Ο μέσος όρος της εντροπίας του δείγματος μεταξύ άλλων φαίνεται επίσης να είναι διαφορετικός (μεγάλες διακυμάνσεις), ωστόσο η μηδενική υπόθεση δεν μπορούσε να απορριφθεί, πιθανώς λόγω προσαρμογών της τιμής p για τις πολλαπλές δοκιμές. Από τα υπόλοιπα χαρακτηριστικά, η *std* των LF, HF και η αναλογία τους βρέθηκε να διαφέρουν σημαντικά.

Τέλος, το Σχήμα 5.3 δείχνει τα boxplots των στατιστικών βημάτων ανά ημέρα και την αναλογία ύπνου/εγρήγορσης για τις δύο ομάδες. Παρατηρούμε μια μεγάλη σημαντική διαφορά τόσο μεταξύ των κατανομών των *mean* και *std* του λόγου ύπνου/εγρήγορσης ($p < 0.001$ και $p = 0.01$, αντίστοιχα), καθώς και των *mean* και *std* των συνολικών βημάτων ανά ημέρα ($p = 0.01$ και $p = 0.05$ αντίστοιχα).

Στόχος της στατιστικής ανάλυσης ήταν ο εντοπισμός κοινών δεικτών/χαρακτηριστικών



(α) Εγρήγορση

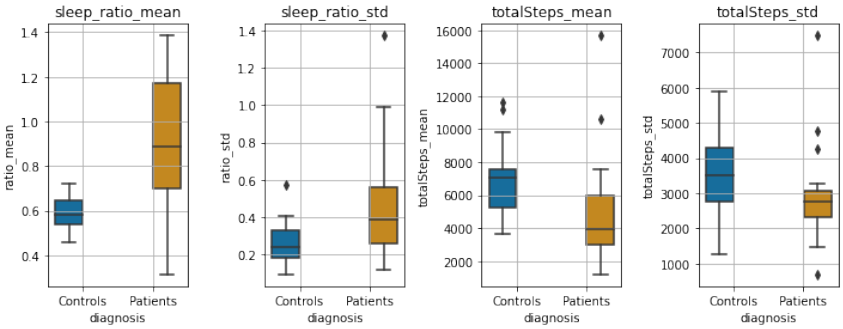


(β) Ύπνος

Σχήμα 5.2: Boxplots διαγράμματα για τα χαρακτηριστικά μεταβλητότητας του καρδιακού ρυθμού των μαρτύρων (μπλε) και των ασθενών (κίτρινο) ενώ (α) είναι ξύπνιοι και (β) κοιμούνται. Η μαύρη γραμμή σε κάθε boxplot αντιπροσωπεύει τη διάμεσο και τα έγχρωμα παραλληλόγραμμα εκτείνονται μεταξύ του 1ου και του 3ου τεταρτημορίου του εύρους των τιμών (δείχνουν το ενδοτεταρτημοριακό εύρος, Inter-Quantile Range - IQR). Οι κατακόρυφες μαύρες γραμμές εκτείνονται έως τη μικρότερη και τη μεγαλύτερη τιμή των χαρακτηριστικών εντός ενός εύρους 1.5*IQR και του 1ου ή 3ου τεταρτημορίου αντίστοιχα. Οι ακραίες τιμές (outliers) εμφανίζονται ως διαμάντια.

που διαφέρουν σημαντικά όταν ένα άτομο έχει ψυχωσική διαταραχή. Τα ευρήματά μας έδειξαν ότι οι ασθενείς τείνουν να συμπεριφέρονται με μεγαλύτερη μεταβλητότητα και παρουσιάζουν μεγάλες ακραίες τιμές (κάποιοι συμπεριφέρονται κοντά στους ελέγχους, ενώ άλλοι μπορεί να παρουσιάζουν ακραίες τιμές). Κατά τη διάρκεια της εγρήγορσης, παρόλο που η μέση ενέργεια δεν διέφερε σε σύγκριση με τους ελέγχους, η τυπική απόκλιση έδειξε σημαντική διαφορά, υποδεικνύοντας ότι οι ασθενείς τείνουν να απεικονίζουν μεγάλες διακυμάνσεις στην κινητική τους συμπεριφορά. Αντίθετα, κατά τη διάρκεια του ύπνου οι ασθενείς παρουσίασαν μικρή μέση τιμή και τυπική απόκλιση της ενέργειας σε κάθε μεσοδιάστημα ύπνου σε σύγκριση με τις αντίστοιχες τιμές του δείγματος ελέγχου. Θα πρέπει ωστόσο να σημειώσουμε ότι οι παρατηρούμενες διαφορές στον ύπνο μεταξύ των δύο ομάδων θα μπορούσαν να αποδοθούν στη φαρμακευτική αγωγή που χορηγείται στους ασθενείς, η οποία πιθανώς προκαλεί διακύμανση στη διάρκεια του ύπνου.

Μερικά από τα μη γραμμικά χαρακτηριστικά που μετρήθηκαν για τα δεδομένα HRV έδειξαν σημαντικές διαφορές στις κατανομές μεταξύ των μαρτύρων και των ασθενών, π.χ. κατά τη διάρκεια της εγρήγορσης, π.χ. η μέση τιμή και η τυπική απόκλιση της εντροπίας του δείγματος (Sample Entropy). Επιπλέον, η τυπική απόκλιση των κανονικοποιημένων ζωνών χαμηλής και



Σχήμα 5.3: Boxplots του λόγου ύπνου/εγρήγορσης (sleep/wake ratio) και των βημάτων ανά ημέρα.

υψηλής συχνότητας του HRV, καθώς και η αναλογία τους, βρέθηκαν να διαφέρουν σημαντικά τόσο κατά τη διάρκεια της εγρήγορσης όσο και του ύπνου. Κατά τη διάρκεια του ύπνου, δεν βρήκαμε άλλες μετρήσεις HRV να διαφέρουν σημαντικά. Τέλος, η αναλογία ύπνου των δύο ομάδων, καθώς και η μέση τιμή και η τυπική απόκλιση του αριθμού των βημάτων ανά ημέρα, παρουσίασαν σημαντική διακύμανση μεταξύ των δύο ομάδων.

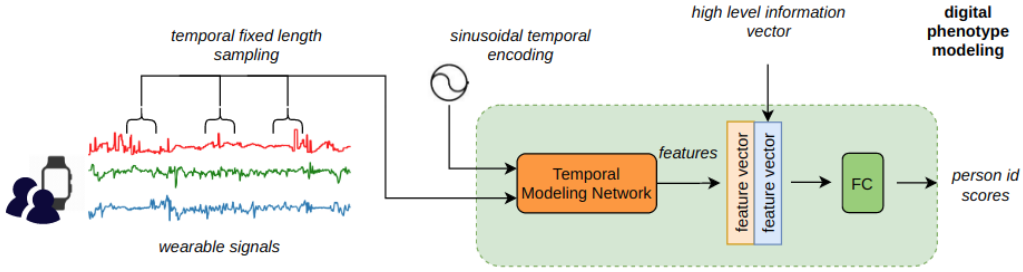
5.2 Αρχιτεκτονική Συστήματος Ταυτοποίησης Χρήστη

Στη συνέχεια του Κεφαλαίου παρουσιάζουμε την πρότασή μας για την ανάπτυξη ενός συστήματος ταυτοποίησης χρήστη μέσω του ψηφιακού του φαινότυπου. Αφού λοιπόν είδαμε μια στατιστική μελέτη των εξαγόμενων χαρακτηριστικών από τα σήματα καταγραφής των έξυπνων ρολογιών, πάμε ένα βήμα πιο πέρα και διερευνούμε πως θα μπορούσαμε να φτιάξουμε μια ισχυρή αναπαράσταση που αναδεικνύει τη διαφορετικότητα του κάθε χρήστη και εν συνεχεία να αναπτύξουμε ένα ισχυρό σύστημα ταξινόμησης-ταυτοποίησης των χρηστών. Ο απώτερος λόγος αυτής της επιλογής είναι η μετέπειτα προσπάθεια εντοπισμού σημαντικών μεταβολών στο φαινότυπο των ασθενών κατά τη διάρκεια μιας ψυχωτικής υποτροπής.

Στο Σχήμα 5.4 σκιαγραφείται η προτεινόμενη μεθοδολογία για την ανάπτυξη ενός συστήματος ταυτοποίησης των ατόμων που φορούν τα έξυπνα ρολόγια. Αρχικά, μοντελοποιούμε τον ψηφιακό φαινότυπο του κάθε ατόμου χρησιμοποιώντας τα κινητικά και καρδιακά σήματα που συλλέγονται από ένα έξυπνο ρολόι, σχεδιάζοντας και εκπαιδεύοντας μια βαθιά αρχιτεκτονική για την ταυτοποίηση των χρηστών. Η εκπαίδευση γίνεται χρησιμοποιώντας δεδομένα από ασθενείς που πάσχουν από ψυχωτικές διαταραχές αλλά βρίσκεται σε ύφεση η ασθένειά τους. Έτσι το εκπαιδευμένο δίκτυο μαθαίνει τα ημερήσια μοτίβα των ασθενών και δοθέντος ενός νέου φαινοτύπου αποκτά την ικανότητα να τον ταξινομήσει στον κατάλληλο ασθενή. Μετά την εκπαίδευση, αξιολογούμε το δίκτυο σε ένα σύνολο δοκιμών που περιλαμβάνει πάλι αποκλειστικά δεδομένα του ασθενούς κατά τη διάρκεια της ύφεσης.

5.2.1 Εκπαίδευση του Συστήματος και Αξιολόγηση

Ως είσοδο στο σύστημα έχουμε ένα πολυδιάστατο σήμα $\mathbf{S}_t^L \in R^N$, το οποίο αντιστοιχεί στα δεδομένα του χρήστη που καταγράφηκαν από το ρολόι κατά τη διάρκεια μιας ημέρας, με L το πλήθος των χρονικών δειγμάτων και N το πλήθος των χαρακτηριστικών που έχουμε εξαγάγει από τα ακατέργαστα δεδομένα. Επειδή το μήκος L του πολυδιάστατου σήματος που προκύπτει μετά την προεπεξεργασία ποικίλλει ανάμεσα σε διαφορετικές ημέρες (βλ. Ενότητα 5.3), κάνουμε αρχικά τυχαία χρονική δειγματοληψία από το προεπεξεργασμένο σήμα. Αυτό οδηγεί



Σχήμα 5.4: Το προτεινόμενο σύστημα για τη δημιουργία ψηφιακών φαινοτύπων και την ταυτοποίηση του χρήστη. Κατά τη διάρκεια της εκπαίδευσης, το μοντέλο μας μαθαίνει τα πρότυπα συμπεριφοράς διαφορετικών χρηστών και στη συνέχεια εξετάζει το ποσοστό σωστής ταξινόμησης σε περιόδους ομαλότητας-ύφεσης της νόσου.

σε ένα πολυδιάστατο σήμα $\mathbf{S}_t^K \in R^N$, όπου το K είναι μια υπερπαράμετρος που συμβολίζει το σταθερό μήκος του δειγματοληπτημένου σήματος. Η δειγματοληψία πραγματοποιείται έτσι ώστε η προκύπτουσα χρονοσειρά να είναι χρονικά συνεκτική, δηλαδή να διατηρείται η χρονική σειρά των δειγμάτων του αρχικού σήματος. Αυτή η τεχνική είναι εμπνευσμένη από τα Temporal Segment Networks (TSN) [Wang et al.; 2016a] και μας επιτρέπει να επεξεργαζόμαστε αποτελεσματικά σήματα διαφορετικού μήκους, τα οποία, όπως αναλύεται στην Ενότητα 5.3, αποτελούν σημαντικό εμπόδιο στη χρήση των δεδομένων που συλλέγονται μέσω φορητών συσκευών. Επίσης, λειτουργεί ως μια μορφή αύξησης δεδομένων βελτιώνοντας την ικανότητα γενίκευσης του εκπαιδευμένου μοντέλου. Επιπλέον, βοηθά στη μοντελοποίηση της δομής των σημάτων κατά τη διάρκεια της ημέρας και επιτρέπει την αποφυγή περιττής πληροφορίας από διαδοχικά δείγματα, συμβάλλοντας έτσι στην αποφυγή υπερβολικής προσαρμογής του μοντέλου μας (overfitting).

Μετά τη χρονική δειγματοληψία, επαυξάνουμε το αρχικό σήμα $\mathbf{S}_t^K$ συνενώνοντας το με δύο χαρακτηριστικά που προκύπτουν από την κυκλική αναπαράσταση της χρονική συνιστώσας του σήματος, παρόμοια με την κωδικοποίηση θέσης [Vaswani et al.; 2017]. Στη συνέχεια, οι ακολουθίες τροφοδοτούνται στο δίκτυο χρονικής μοντελοποίησης (Temporal Modeling Network), το οποίο εξάγει ένα σταθερό διάνυσμα μεγέθους 512. Το διάνυσμα χαρακτηριστικών συνδέεται τώρα με ένα διάνυσμα πρόσθετων χαρακτηριστικών υψηλού επιπέδου: το χρόνο της αντίστοιχης ημέρας που κοιμήθηκε ο χρήστης και την ημέρα της εβδομάδας. Τέλος, αυτό το επαυξημένο διάνυσμα χαρακτηριστικών τροφοδοτείται σε δύο fully connected layers που εξάγουν τα τελικά σκορ της αναγνώρισης του χρήστη. Για την εκπαίδευση το δίκτυο χρησιμοποιεί cross-entropy loss. Μετά την εκπαίδευση του μοντέλου και κατά τη διάρκεια της εκτίμησης των δειγμάτων ελέγχου, εκτελούμε τυχαία δειγματοληψία για κάθε ημέρα - όπως και στην εκπαίδευση - πέντε φορές για την ίδια μέρα και αφού πάρουμε το σκορ για κάθε μία από τις πέντε ακολουθίες, αθροίζουμε τις μη κανονικοποιημένες βαθμολογίες (log score) προκειμένου να λάβουμε την τελική βαθμολογία.

5.2.2 Νευρωνικό Δίκτυο Χρονικής Μοντελοποίησης

Για το δίκτυο χρονικής μοντελοποίησης, σχεδιάζουμε δύο διαφορετικές αρχιτεκτονικές. Η μία αρχιτεκτονική βασίζεται σε συνελικτικά νευρωνικά δίκτυα (CNN) και χρησιμοποιεί συνελικτικούς πυρήνες μιας διάστασης, ενώ η δεύτερη βασίζεται σε επίπεδα μακράς βραχυπρόθεσμης μνήμης (LSTM). Το δίκτυο που βασίζεται στο CNN αποτελείται από ένα πλήρως συνδεδεμένο επίπεδο (fully connected layer) που ακολουθείται από μια συνάρτηση ενεργοποίησης ReLU και

dropout (τεχνική προσωρινής απόρριψης νευρώνων). Στη συνέχεια, ακολουθούν πέντε συνελικτικά μπλοκ, το καθένα αποτελούμενο από ένα συνελικτικό επίπεδο, μια ReLU και dropout. Η έξοδος του τελευταίου συνελικτικού μπλοκ τροφοδοτείται στη συνέχεια σε ένα adaptive average polling επίπεδο προκειμένου να ληφθεί ένα σταθερό διάνυσμα χαρακτηριστικών μεγέθους 512. Τέλος, εισάγουμε το διάνυσμα χαρακτηριστικών σε δύο διαδοχικά fully connected layers (με ReLU ενδιάμεσα) για να λάβουμε τα μη κανονικοποιημένα log scores. Αντίστοιχα, στην αρχιτεκτονική που βασίζεται σε LSTM, τα συνελικτικά μπλοκ αντικαθίστανται με δύο αμφίδρομα επίπεδα LSTM με dropout και παίρνουμε την έξοδο του τελευταίου χρονικού βήματος ως διάνυσμα χαρακτηριστικών. Όπως αναφέραμε στην Ενότητα 5.2.1, η έξοδος διανύσματος σταθερών χαρακτηριστικών από το δίκτυο CNN ή LSTM συνενώνεται στη συνέχεια με ένα διάνυσμα που περιλαμβάνει πληροφορίες υψηλού επιπέδου.

5.2.3 Επαύξηση Δεδομένων

Η εκπαίδευση ενός δικτύου για την ταυτοποίηση ατόμων, ειδικά σε περιορισμένο όγκο δεδομένων όπως στην περίπτωσή μας, μπορεί να οδηγήσει σε φαινόμενα υπερβολικής προσαρμογής (overfitting). Μια ακραία περίπτωση αυτού αναφέρθηκε στο [Retsinas et al.; 2020], όπου μοναδικά μοτίβα σήματος από τους αισθητήρες ενός έξυπνου ρολογιού θα μπορούσαν να οδηγήσουν σε ένα σύστημα που αναγνωρίζει τον αισθητήρα που είναι προσαρμοσμένος στο ρολόι και όχι το πραγματικό άτομο που το φοράει. Παρά το γεγονός ότι μια τέτοια περίπτωση θα προέκυπτε πιο συχνά κατά τη χρήση ακατέργαστων σημάτων, όπως στο [Retsinas et al.; 2020], παρόμοιες περιπτώσεις overfitting μπορούν επίσης να επηρεάσουν τη μελέτη μας. Για το σκοπό αυτό, εισάγουμε ένα στάδιο επαύξησης, το οποίο αποτελείται από τρία διακριτά στάδια. Πιο συγκεκριμένα, εφαρμόζουμε σταδιακά:

- **Προσθήκη θορύβου**: Δειγματοληπτούμε μια κανονική κατανομή και εφαρμόζουμε το θορυβώδες σήμα που λαμβάνουμε, στο αρχικό σήμα με πολλαπλασιαστικό τρόπο, ως εξής: $s_n = s(1 + r_n n)$, όπου s είναι το αρχικό σήμα που αποτελείται από ακολουθίες βιοδεικτών, s_n το επαυξημένο σήμα που προκύπτει από την προσθήκη θορύβου, n είναι το σήμα θορύβου, το οποίο πραγματοποιείται ως τυχαία μεταβλητή δειγματοληψίας από μια κατανομή $\mathcal{N}(0, 1)$ και r_n μια προκαθορισμένη παράμετρος που ελέγχει τη διαφορά κλίμακας (ύστερα από πειραματισμό, στα πειράματα που παρουσιάζονται παρακάτω ορίστηκε ίση με 0.1). Χρησιμοποιήσαμε αυτήν την πολλαπλασιαστική παραλλαγή, αντί για τον τυπικό πρόσθετο Gaussian θόρυβο, προκειμένου να διατηρήσουμε την κλίμακα του αρχικού σήματος και να εισαγάγουμε μόνο θόρυβο «υψηλής συχνότητας» που δεν επηρεάζει σημαντικά το περίγραμμα του σήματος.

- **Εφαρμογή τυχαίας μάσκας**: Χρησιμοποιούμε επίσης μια επαυξητική τεχνική εμπνευσμένη από τη μέθοδο dropout που έχει αποδειχθεί αποτελεσματική σε ένα ευρύ φάσμα εφαρμογών βαθιάς εκμάθησης, όπου εφαρμόζουμε μια τυχαία μάσκα στο σήμα εισόδου. Εφόσον τα δεδομένα εισόδου είναι πολυδιάστατα (αποτελούνται από πολλά χαρακτηριστικά), αυτή η μάσκα είναι ένας 2D αραιός δυαδικός πίνακας που μηδενίζει συγκεκριμένες τιμές εισόδου με μη δομημένο τρόπο, δηλαδή στα πειράματά μας, η πιθανότητα να μηδενιστεί η κάθε τιμή λαμβάνεται ως δείγμα από μια κατανομή Bernoulli με $p = 0.05$. Αυτή η μέθοδος επαύξησης ωθεί το δίκτυο να μάθει χρήσιμες αναπαραστάσεις ακόμη και απουσία συγκεκριμένων χαρακτηριστικών, αυξάνοντας έτσι τις ικανότητες γενίκευσης του δικτύου.

- **Ανάμειξη δειγμάτων**: Τέλος, χρησιμοποιήσαμε επίσης την state-of-the-art τεχνική επαύξησης mixup [Zhang et al.;], όπου συγχωνεύουμε δύο εισόδους x_1 και x_2 παίρνοντας τον

κυρτό συνδυασμό τους: $\lambda x_1 + (1-\lambda)x_2$, με το λ να δειγματοληπτείται από μια κατανομή βήτα, με συντελεστές $a = \beta = 0.2$. Η επαυξημένη έξοδος αναμένεται να είναι ένας κυρτός συνδυασμός των αντίστοιχων εξόδων y_1, y_2, με τον ίδιο παράγοντα ανάμειξης (mixup factor) λ. Η απλότητα αυτής της προσέγγισης, μαζί με τις αξιοσημείωτες ικανότητες κανονικοποίησης και γενίκευσης, την καθιστούν ιδανική προσθήκη στο σύστημά μας.

5.3 Σύνολο Δεδομένων και Επιλογή Χαρακτηριστικών

5.3.1 Σύνολο Δεδομένων

Αναλύοντας τα προεπεξεργασμένα δεδομένα της βάσης του e-Prevention παρατηρήσαμε πως η μέση διάρκεια καταγραφής των δεδομένων του ρολογιού είναι περίπου 14 ώρες την ημέρα, αρκετά χαμηλότερη από τον αναμενόμενο μέσο όρο των 18-20 ωρών την ημέρα, για ασθενείς με καλή συμμόρφωση. Οι πιο συχνοί λόγοι για τους οποίους παρατηρείται αυτό είναι:

- η αναποτελεσματική φωτοπληθυσμογραφία που παρέχεται από το ρολόι, π.χ. λόγω ιδρωμένου καρπού του χρήστη,
- η αστοχία της εφαρμογής εγγραφής των δεδομένων στο ρολόι,
- διάφορα τυπικά σφάλματα στη χρήση, π.χ. παράλειψη έγκαιρης φόρτισης ή παράλειψη στο να φορέσει ο χρήστης το ρολόι,
- η κακή συμμόρφωση ασθενών στις οδηγίες του ιατρικού προσωπικού,
- τα ζητήματα δικτύου για τη σύνδεση ρολογιού-cloud.

Ωστόσο, παρά τις όποιες αδυναμίες της, η βάση του e-Prevention αποτελεί μια μοναδική συλλογή πραγματικών δεδομένων μακροχρόνιας καταγραφής βιοσημάτων ασθενών με ψυχωτικές ασθένειες. Τα δεδομένα της βάσης είναι αξιοσημείωτου μεγέθους και περιλαμβάνουν αντικειμενικές μετρήσεις της καθημερινής ζωής των ασθενών. Κατά συνέπεια, η χρήση μιας τέτοιας βάσης βιοσημάτων για τη μοντελοποίηση των ψηφιακών φαινοτύπων των ασθενών, μας φέρνει ένα βήμα πιο κοντά στην αντιμετώπιση ενός προβλήματος σε πραγματικές συνθήκες που ενσωματώνουν τις αναμενόμενες και μη δυσκολίες της καθημερινής χρήσης ενός έξυπνου ρολογιού.

Στην παρούσα ενότητα, επιλέγουμε να πραγματοποιήσουμε τα πειράματά μας σε ένα υποσύνολο της βάσης του e-Prevention που αποτελείται από δεδομένα ασθενών με περισσότερες από 180 μέρες καταγραφών μετά την προεπεξεργασία τους. Για ευκολία, θα καλούμε το συγκεκριμένο υποσύνολο ως e-Prevention subset A. Πιο συγκεκριμένα, το σύνολο περιέχει όλα τα δεδομένα που καταγράφηκαν κατά τη συμμετοχή 29 ασθενών και εκτείνονται από έξι μήνες έως δυόμισι χρόνια. Σχετικά με την ασθένεια τους, 2 ασθενείς παρουσιάζουν Σχιζοσυναισθηματική διαταραχή, 2 Σχιζοφρενική διαταραχή, 15 Σχιζοφρένεια, 8 Διπολική Διαταραχή I και 2 με Διπολική Διαταραχή II. Τέλος, κατά το διάστημα της μελέτης έντεκα ασθενείς από αυτούς παρουσίασαν 16 ψυχωτικές υποτροπές.

Στον Πίνακα 5.3 παρουσιάζονται τα στατιστικά στοιχεία του e-Prevention subset A, σχετικά με τις μέρες συμμετοχής ανά άτομο, τις μέρες καταγραφής και τις ώρες καταγραφής ανά ημέρα, καθώς και το σύνολο των καταγεγραμμένων ημερών για τους 29 ασθενείς. Όπως φαίνεται στον πίνακα, ενώ η μέση συμμετοχή των ασθενών είναι στις 737 ημέρες, παρατηρείται πως η μέση καταγραφή, μετά την προεπεξεργασία, είναι στις 613 ημέρες, δηλαδή έχουμε κατά μέσο όρος 4 μήνες καταγραφών λιγότερους από το αναμενόμενο. Οι τιμές αυτές μπορούν να δώσουν μια αίσθηση για την επιλογή των έξι μηνών (180 καταγεγραμμένων ημερών) ως ορόσημο για τη

Ημέρες	Μέση τιμή	Τυπική Απόκλιση	Σύνολο
Συμμετοχής στη μελέτη	736.7	168.90	21365
Καταγραφής	613.0	187.71	17777
Υποτροπής	29.3	43.59	849
Ώρες καταγραφής / ημέρα	14.53	1.94	-

Πίνακας 5.3: Στατιστικά στοιχεία του e-Prevention subset A που χρησιμοποιείται στο κεφάλαιο αυτό για πειραματισμό. Η μέση τιμή και η τυπική απόκλιση αναφέρονται στις καταγεγραμμένες μέρες ανά άτομο ενώ το σύνολο αφορά το άθροισμα όλων των ημερών που περιέχει το υποσύνολο.

δημιουργία του υποσυνόλου που θα χρησιμοποιήσουμε. Θέλουμε ο πειραματισμός να γίνει σε ένα σημαντικό όγκο δεδομένων ώστε να μπορούμε να βγάλουμε ασφαλή συμπεράσματα χωρίς να κινδυνεύουμε να πέσουμε σε λάθη λόγω έλλειψης δεδομένων.

5.3.2 Διερεύνηση Αναπαράστασης Ψηφιακού Φαινοτύπου

Πρώτος στόχος μας είναι να βρούμε μια συγκεκριμένη αναπαράσταση του ψηφιακού φαινοτύπου κάθε ασθενούς προκειμένου να τον αναγνωρίσουμε χρησιμοποιώντας τα βιοσήματά του. Σε προηγούμενες μελέτες, εξετάσαμε διάφορα χαρακτηριστικά που εξήχθησαν από τα ακατέργαστα σήματα και πραγματοποιήσαμε εκτεταμένη στατιστική ανάλυση για να αποκαλύψουμε διαφορές στη φυσική δραστηριότητα και στα πρότυπα αυτόνομης λειτουργίας [Zlatintsi et al.; 2022,Filntisis et al.; 2020b]. Επίσης, στο [Kalisperakis et al.; 2023], τα αποτελέσματα δείχνουν ότι οι μετρήσεις του καρδιακού ρυθμού (HRV) και τα RR-intervals είναι οι δείκτες που μπορούμε να λάβουμε από το έξυπνο ρολόι και να υπάρχει συσχέτιση με τις ψυχοπαθολογικές αλλαγές των ασθενών. Λαμβάνοντας υπόψη τα παραπάνω και εμβαθύνοντας στις δυνατότητες αναγνώρισης που προσφέρει η πληθώρα των εξαγόμενων χαρακτηριστικών, παρουσιάζουμε μια διερευνητική ανάλυση για την εύρεση μιας αποτελεσματικής αναπαράστασης για την αναγνώριση-ταυτοποίηση του χρήστη.

Μήκος ακολουθίας δεδομένων: Αρχικά, χωρίζουμε τα δεδομένα ανά ημέρα, καθώς είναι η μικρότερη δυνατή μονάδα που μπορεί να αξιοποιηθεί για τον εντοπισμό μοτίβων συμπεριφοράς, συγκεντρώνοντας τις τιμές των πεντάλεπτων χαρακτηριστικών από τις 00:00 έως τις 23:59. Επομένως, εάν ένας ασθενής φορούσε το ρολόι για ολόκληρη την ημέρα, θα συγκεντρώναμε 288 τιμές ανά χαρακτηριστικό (αφού υπολογίζονται δώδεκα πεντάλεπτες τιμές ανά ώρα).Όπως αναλύθηκε στο 5.3, ένα από τα μεγαλύτερα εμπόδια της μελέτης μας είναι ότι κατά τη συλλογή δεν έχουμε ούτε σταθερό αριθμό ημερήσιων δεδομένων για κάθε χρήστη ούτε κάποιο σταθερό διάστημα εγγραφής. Αυτή η διαφορά θα μπορούσε να έχει αρνητικό αντίκτυπο στην εκπαίδευση του μοντέλου και να εισάγει μεροληψία ως προς την αναγνώριση του ατόμου. Για να είμαστε πιο ακριβείς, εάν π.χ. χρησιμοποιούμε μεγάλες ακολουθίες, θα μπορούσαμε να αφαιρέσουμε χρήστες που φορούν το ρολόι μόνο για λίγες ώρες την ημέρα, ενώ το δίκτυο θα μπορούσε να μάθει να ξεχωρίζει τα άτομα με βάση τις ώρες καταγραφής. Επομένως, προτιμούμε να δημιουργούμε κάθε ακολουθία δεδομένων επιλέγοντας τυχαία διανύσματα διαστήματος 5 λεπτών κατά τη διάρκεια της ημέρας, συνενώνοντας τα και ταξινομώντας διατηρώντας τη χρονική τους σειρά. Έτσι, δημιουργείται μια τυχαία αναπαράσταση της ημέρας. Στη συνέχεια, για να καθορίσουμε την αναπαράσταση της κάθε ημέρας, δοκιμάζουμε διάφορους συνδυασμούς χαρακτηριστικών και κυμαινόμενων μηκών σε συνδυασμό με τις δυο διαφορετικές εκδοχές του συστήματος (με LSTM και CNN).

Χαρακτηριστικά ενδιαφέροντος: Συγκεντρώνουμε πέντε ομάδες χαρακτηριστικών που μπορούν να αποδειχθούν χρήσιμες για τη δημιουργία των φαινοτύπων των χρηστών. Ομαδο-

ποιούμε τα χαρακτηριστικά ως εξής:

α) τα στατιστικά στοιχεία μεταβλητότητας καρδιακού ρυθμού (ο μέσος όρος των διαστημάτων RR, οι ελάχιστες και μέγιστες τιμές τους, καθώς και η τυπική απόκλιση των διαστημάτων RR (SDRR)) που υποδηλώνονται ως *RR_intervals* στο Σχήμα 5.5,

β) τα στατιστικά των καρδιακών παλμών (beats per minute), (υπολογίζονται η μέση τιμή, η ελάχιστη, η μέγιστη και η τυπική απόκλιση του καρδιακού ρυθμού) και δηλώνεται ως *Heart Rate*,

γ) τα χαρακτηριστικά της ομάδας (α), το RMSSD και το πλήθος των καταγεγραμμένων τιμών που χρησιμοποιήθηκαν για τον υπολογισμό των χαρακτηριστικών στο εκάστοτε πεντάλεπτο, και συμβολίζονται ως *RR_intervals + RR stats* ,

δ) τα χαρακτηριστικά της ομάδας (α) και οι κανονικοποιημένες δυνάμεις των υψηλών και χαμηλών συχνοτήτων (LF, HF) που υπολογίζονται από τη φασματική ανάλυση ισχύος Lomb-Scargle των RR-intervals, υποδηλώνονται ως *RR_intervals + Lomb-Scargle feats*, και

ε) τα χαρακτηριστικά της ομάδας (γ) και το μέτρο της γραμμικής επιτάχυνσης. Δεν χρησιμοποιήσαμε το μέτρο της γωνιακής επιτάχυνσης, καθώς είναι συσχετίζεται άμεσα με αυτό της γραμμικής επιτάχυνσης και δεν θα προσφέρει επιπρόσθετη πληροφορία.

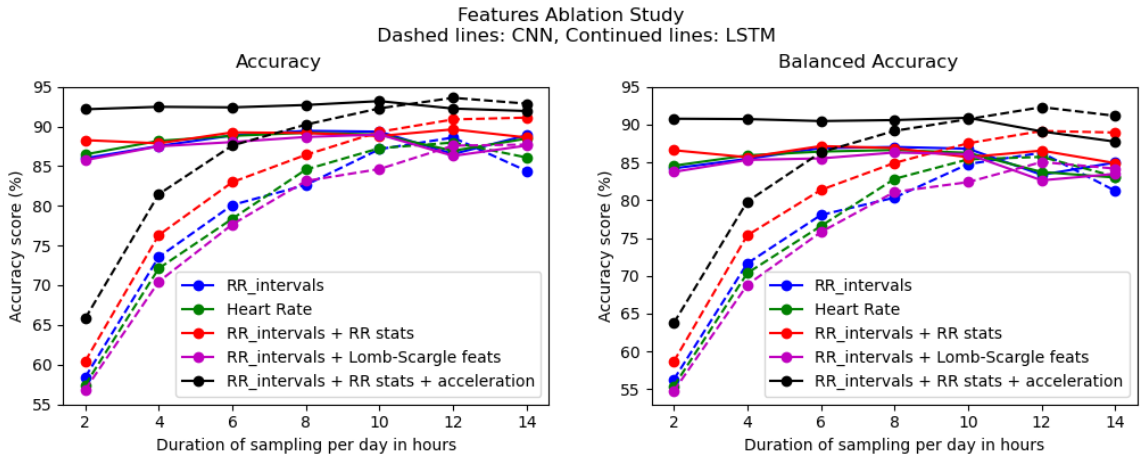
5.4 Πειραματική Ανάλυση

Πειραματική διάταξη (setup): Τα πειράματα αξιολογήθηκαν στο σύνολο δεδομένων e-Prevention subset A που περιγράφεται στην Ενότητα 5.3.1. Η εκπαίδευση και ο έλεγχος πραγματοποιήθηκαν σύμφωνα με ένα five-fold cross-validation σχήμα, χρησιμοποιώντας μόνο τα δεδομένα των περιόδων ύφεσης της ασθένειας των 29 ασθενών. Και τα δύο δίκτυα χρονικής μοντελοποίησης (με βάση το CNN και το LSTM) εκπαιδεύονται με χρήση RAdam optimizer [Liu et al.; 2019a] για 100 εποχές με αρχικό ποσοστό εκμάθησης (learning rate) 0.01, το οποίο πέφτει σε 0.001 στις 75 εποχές.

5.4.1 Πειραματική Διερεύνηση Χαρακτηριστικών

Στην πρώτη ανάλυση, αξιολογούμε την απόδοση των προτεινόμενων δικτύων ταυτοποίησης ατόμων, CNN και LSTM, για τα διαφορετικά σύνολα χαρακτηριστικών που έχουμε εξάγει. Προκειμένου να ληφθεί υπόψη η ανισορροπία του όγκου των δεδομένων μεταξύ των ασθενών, παρουσιάζουμε στο Σχήμα 5.5 αποτελέσματα τόσο σε απλή ακρίβεια (accuracy) στα αριστερά όσο και σε εξισορροπημένη ακρίβεια (balanced accuracy) στα δεξιά για την ταυτοποίηση των 29 ασθενών. Ως εξισορροπημένη ακρίβεια αναφερόμαστε στον αριθμητικό μέσο όρο της ευαισθησίας (sensitivity = true positive rate) και της εξειδίκευσης (specificity = true negative rate) του μοντέλου. Στο ίδιο Σχήμα, διερευνούμε επίσης την επίδραση της μεταβολής του μήκους της ακολουθίας. Ο άξονας x αντιπροσωπεύει τη συνολική διάρκεια δειγματοληψίας ανά ημέρα σε ώρες. Για παράδειγμα, 10 ώρες αντιστοιχούν σε ένα διάνυσμα χαρακτηριστικών μήκους 120 τιμών ανά χαρακτηριστικό (10 ώρες * 12 διαστήματα πέντε λεπτών ανά ώρα).

Σύνολα Χαρακτηριστικών: Όσον αφορά τους τύπους χαρακτηριστικών, τα στατιστικά στοιχεία μεταβλητότητας καρδιακού ρυθμού (RR-intervals) αποδίδουν ελαφρώς καλύτερα από τα στατιστικά στοιχεία καρδιακού παλμού (Heart Rate). Προσθέτοντας το RMSSD και το πλήθος των καταγεγραμμένων RR-intervals, το οποίο μπορεί να λειτουργήσει ως βαθμός βεβαιότητας των εξαγόμενων χαρακτηριστικών, παρατηρούμε μια μικρή αύξηση στη συνολική ακρίβεια, υποδηλώνοντας μια μικρή βελτίωση στην ταξινόμηση ανά ημέρα. Τέλος, όταν προσθέ-



Σχήμα 5.5: Μελέτη μεταβολής της ακρίβειας ταυτοποίησης του συστήματος σε διαφορετικά σύνολα χαρακτηριστικών και μήκη χρονοσειρών. Παρουσιάζεται τόσο η ακρίβεια ταυτοποίησης (accuracy) όσο και η εξισορροπημένη ακρίβεια (balanced accuracy) για τις αρχιτεκτονικές CNN (διακεκομμένες γραμμές) και LSTM (συνεχείς γραμμές).

τουμε και την πληροφορία της γραμμικής επιτάχυνσης, το μοντέλο επιτυγχάνει τα καλύτερα αποτελέσματα κατηγοριοποίησης και στις δυο μετρικές ακρίβειας. Παρατηρούμε ακόμα πως η προσθήκη των συχνοτικών χαρακτηριστικών Lomb-Scargle των RR-intervals δεν είχε ως αποτέλεσμα αυξημένη απόδοση. Τέλος, συμπεραίνουμε επίσης ότι οι διαφορές στην υλοποίηση του δικτύου δεν έχουν ουσιαστικό αντίκτυπο στην υπεροχή ενός συνόλου χαρακτηριστικών έναντι του άλλου, επομένως για τα υπόλοιπα πειράματά μας προχωράμε με το ακόλουθο σύνολο χαρακτηριστικών: *RR_intervals + RR stats + acceleration* .

Μήκος Ακολουθίας Δεδομένων: Από το ίδιο Σχήμα βλέπουμε επίσης ότι το μήκος της ακολουθίας εισόδου επηρεάζει την απόδοση ταυτοποίησης. Διερευνούμε ένα δείγμα ημερήσιων εγγραφών που εκτείνονται από 24 έως 168 χρονικά βήματα (δηλαδή, από 2 έως 14 ώρες). Βλέπουμε ότι για την αρχιτεκτονική του CNN, η τυχαία δειγματοληψία σε μήκος ακολουθίας 144 (=12 ώρες) επιτυγχάνει την υψηλότερη ακρίβεια. Για την εφαρμογή LSTM, ένα μήκος ακολουθίας 120 αποφέρει καλύτερα αποτελέσματα σε σύγκριση με το CNN, ειδικά όσον αφορά την εξισορροπημένη ακρίβεια. Ως αποτέλεσμα, για το υπόλοιπο αυτής της μελέτης, χρησιμοποιούμε ένα πολυδιάστατο διάνυσμα χαρακτηριστικών 120 χρονικών βημάτων για να αναπαραστήσουμε την πληροφορία από την καθημερινότητα των χρηστών. Ένα άλλο σημαντικό συμπέρασμα που μπορούμε να εξάγουμε είναι η διαφορά στον αριθμό των δειγμάτων (μήκος ακολουθίας) που απαιτούνται για την αναγνώριση του χρήστη μεταξύ των δύο αρχιτεκτονικών. Ενώ το CNN απαιτεί πληροφορίες για περισσότερες από οκτώ ώρες για να επιτύχει υψηλά ποσοστά ταξινόμησης, τα δίκτυα LSTM, τα οποία είναι γνωστά για την αποτελεσματικότητά τους με δεδομένα χρονοσειρών, δημιουργούν ισχυρές αναπαραστάσεις για μια ποικιλία μεγεθών δειγμάτων.

5.4.2 Διερεύνηση Αρχιτεκτονικών

Έχοντας διερευνήσει διαφορετικά σύνολα χαρακτηριστικών και τη συμβολή τους στην αναγνώριση φαινοτύπων, προχωράμε τώρα στην πειραματική μελέτη των διαφορετικών αρχιτεκτονικών του δικτύου χρονικής μοντελοποίησης. Εστιάζουμε στη σύγκριση των δύο διαφορετικών χρονικών αρχιτεκτονικών, δηλαδή CNN έναντι LSTM, εκπαιδεύοντας τα δίκτυα με τα δεδομένα των 29 ασθενών σε διαφορετικά σχήματα επαύξησης των δεδομένων όπως προτάθηκαν στο 5.2.3, δηλαδή με προσθήκη θορύβου, εφαρμογή τυχαίας μάσκας και ανάμειξη των

			Χωρίς Επαύξηση	Προσθήκη Θορύβου	+ Εφαρμογή Τυχαίας Μάσκας	+ Mixup Δειγμάτων
CNN	**Base**	Acc.	**92.29**	89.58	89.19	70.36
		Bal.Acc.	**90.72**	88.42	87.58	65.41
	+ temporal encoding	Acc.	**91.56**	88.87	87.83	69.79
		Bal.Acc.	**89.81**	86.78	85.60	64.01
LSTM	**Base**	Acc.	**93.20**	91.32	92.89	91.54
		Bal.Acc.	**90.90**	88.64	90.46	89.18
	+ temporal encoding	Acc.	92.03	90.27	92.31	**92.57**
		Bal.Acc.	88.97	87.08	89.58	**90.06**

Πίνακας 5.4: Σύγκριση των αρχιτεκτονικών CNN και LSTM ως προς την ταυτοποίηση του χρήστη κατά τη διάρκεια των περιόδων ύφεσης της ασθένειας. Η ακρίβεια (Accuracy and Balanced Accuracy) υπολογίζονται με ένα 5-fold cross-validation σχήμα και για τις δύο αρχιτεκτονικές (CNN & LSTM) για τα αρχικά χαρακτηριστικά (base), αξιοποιώντας τη χρονική κωδικοποίηση (+ temporal encoding) και την επαυξημένη εκδοχή τους.

δειγμάτων (mixup). Πιο συγκεκριμένα, προσθέτουμε προοδευτικά κάθε τεχνική επαύξησης των δεδομένων, σχηματίζοντας μια αλυσίδα τεχνικών που εφαρμόζονται σύμφωνα με τη σειρά που αναφέραμε. Εξετάσαμε επίσης τις παραλλαγές της αρχιτεκτονικής, όπου συμπεριλάβαμε τις πληροφορίες χρονικής κωδικοποίησης (βλ. Ενότητα 5.2.1), προκειμένου να κατανοήσουμε εάν η χρονική πληροφορία μιας δράσης είναι σημαντική για την ταυτοποίηση του ατόμου. Τα αποτελέσματα συνοψίζονται στον Πίνακα 5.4, όπου αναφέρονται τόσο η ακρίβεια όσο και η εξισορροπημένη ακρίβεια.

Σύμφωνα με τα παραπάνω αποτελέσματα, τα μοντέλα CNN, όπως και στην περίπτωση της προηγούμενης μελέτης, είναι πολύ πιο ευαίσθητα από την παραλλαγή LSTM. Συγκεκριμένα, η προσθήκη θορύβου στην είσοδο CNN μειώνει σημαντικά την απόδοση του δικτύου. Αντίθετα το δίκτυο LSTM εμφανίζει σταθερή συμπεριφορά σε όλες τις παραλλαγές της επαύξησης. Όσον αφορά τις επιπλέον πληροφορίες χρονικής κωδικοποίησης, παρατηρήσαμε μια μικρή μείωση στις περισσότερες περιπτώσεις. Παρά τη μείωση αυτή, θεωρούμε ότι η επιπρόσθετη χρονική πληροφορία είναι σημαντική για την εργασία της ταυτοποίησης, καθώς μπορεί να μοντελοποιήσει καλύτερα τον κύκλο συμπεριφοράς μέσα σε μια ημέρα και γι' αυτό θα συνεχίσουμε τη διερεύνηση της χρήσης της και σε επόμενα πειράματα.

Είναι σημαντικό να σημειώσουμε πως η ταυτοποίηση του χρήστη είναι μια ενδιάμεση μελέτη με σκοπό την κατανόηση των ψηφιακών φαινοτύπων και την διερεύνηση αρχιτεκτονικών που μπορούν να κατηγοριοποιήσουν κατάλληλα τις δράσεις των ατόμων μέσω της χρήσης των βιοσημάτων και δεν είναι απόλυτα ευθυγραμμισμένη με την τελική εργασία της αναγνώρισης των ψυχωτικών υποτροπών που παραμένει το κύριο πρόβλημα της μελέτης. Έτσι η μικρή μείωση της απόδοσης κατά τη χρήση τέτοιων διαισθητικών χαρακτηριστικών (π.χ. χρονικής πληροφορίας) δεν θεωρείται σημαντικός λόγος για την απόρριψή τους. Συγκεκριμένα, το μοντέλο LSTM, εξοπλισμένο με τη χρονική κωδικοποίηση και το πλήρες σύνολο των επαυξητικών μεθόδων επιτυγχάνει παρόμοια απόδοση με το σύστημα κορυφαίας απόδοσης του βασικού LSTM χωρίς θόρυβο. Συνεπώς, η χρήση των τεχνικών αυτών βοηθάνε στη γενίκευση του μοντέλου μας και στην αποφυγή του overfitting. Καταλήγοντας, παρατηρήσαμε πως οι LSTM αρχιτεκτονικές είναι ιδιαιτέρως σταθερές ως προς την ακρίβεια ταυτοποίησης του χρήστη, με καλά αποτελέσματα και συνεπώς αποτελεί μια καλή επιλογή για αξιοποίηση σε ένα σύστημα αναγνώρισης ψυχωτικών υποτροπών.

5.5 Συμπεράσματα Κεφαλαίου

Κύριος στόχος του κεφαλαίου αποτέλεσε η δημιουργία ισχυρών αναπαραστάσεων των σημάτων που καταγράφονται από έξυπνα ρολόγια με σκοπό την περιγραφή των καθημερινών και μοναδικών μοτίβων των ανθρώπων που τα φορούν. Για το σκοπό αυτό αρχικά παρουσιάσαμε μια στατιστική ανάλυση των εξαγόμενων βιοδεικτών και στη συνέχεια αναπτύξαμε ένα σύστημα ταυτοποίησης των ατόμων μέσω του ψηφιακού τους φαινότυπου.

Πιο συγκεκριμένα, τα ευρήματά μας έδειξαν ότι οι ασθενείς τείνουν να συμπεριφέρονται με μεγαλύτερη μεταβλητότητα και παρουσιάζουν μεγάλες ακραίες τιμές. Κατά τη διάρκεια της εγρήγορσης, παρόλο που η μέση ενέργεια δεν διέφερε σε σύγκριση με τους μάρτυρες, η τυπική απόκλιση έδειξε σημαντική διαφορά, υποδεικνύοντας ότι οι ασθενείς τείνουν να απεικονίζουν μεγάλες διακυμάνσεις στην κινητική τους συμπεριφορά. Αντίθετα, κατά τη διάρκεια του ύπνου οι ασθενείς παρουσίασαν μια μικρή μέση και τυπική απόκλιση της ενέργειας σε κάθε μεσοδιάστημα ύπνου σε σύγκριση με τους μάρτυρες. Αντίστοιχα παρατηρήθηκε πως ορισμένα από τα μη γραμμικά χαρακτηριστικά που μετρήθηκαν για τα δεδομένα της μεταβολής του καρδιακού ρυθμού έδειξαν στατιστικά σημαντικές διαφορές στις κατανομές μεταξύ των μαρτύρων και των ασθενών κατά τη διάρκεια του ύπνου και της εγρήγορσης, ενώ η φασματική ανάλυση δεν αποκάλυψε σημαντικές διαφορές.

Επιπρόσθετα, κατά τη διάρκεια της ανάπτυξης του συστήματος για την ταξινόμηση/ταυτοποίηση των χρηστών μελετήσαμε πολλές επιμέρους συνιστώσες και καταλήξαμε σε αρκετά χρήσιμα συμπεράσματα. Διαπιστώσαμε πως τα στατιστικά που περιγράφουν την μεταβολή του καρδιακού ρυθμού μαζί με την πληροφορία για τη νόρμα της γραμμικής επιτάχυνσης αποτελούν χαρακτηριστικά ικανά να διαχωρίσουν τα διάφορα άτομα μεταξύ τους. Συγκεκριμένα η δημιουργία διανυσμάτων χαρακτηριστικών που προκύπτουν από τυχαία δειγματοληψία των πεντάλεπτων βιοδεικτών των ατόμων για ένα διάστημα συνολικής διάρκειας 10 ωρών μέσα στη μέρα αποτελεί μια ισχυρή αναπαράσταση του ψηφιακού φαινότυπου των ατόμων και μπορεί να χρησιμοποιηθεί για τη διάκριση/ταξινόμηση αυτών.

Ως προς τις αρχιτεκτονικές του δικτύου παρατηρήσαμε πως η LSTM αρχιτεκτονική υπερτερεί αυτής των CNN καθώς είναι λιγότερο ευαίσθητη στη μεταβολή του μήκους των διανυσμάτων εισόδου στο σύστημα καθώς και στη χρήση μεθόδων επαύξησης των δεδομένων για την ενίσχυση της γενίκευσης και την αποφυγή υπερπροσαρμογής του συστήματος. Τέλος, η προσθήκη χρονικής κωδικοποίησης στο διάνυσμα των χαρακτηριστικών αν και επιφέρει μια μικρή μείωση στην ακρίβεια του συστήματος, φαίνεται να μπορεί να βοηθήσει στη μοντελοποίηση του κύκλου συμπεριφοράς μέσα στη μέρα. Έτσι, καταλήξαμε σε μια αρχιτεκτονική ιδιαιτέρως σταθερή ως προς την ακρίβεια ταυτοποίησης του χρήστη, με καλά αποτελέσματα, που μπορεί να αποτελέσει μια καλή βάση για την αξιοποίηση περαιτέρω σε ένα σύστημα αναγνώρισης ψυχωτικών υποτροπών.

6

Αναγνώριση Ψυχωτικών Υποτροπών

Εκατομμύρια άνθρωποι σε όλο τον κόσμο εμφανίζουν συμπτώματα ψυχωτικών διαταραχών, με πιο συχνά αυτά που σχετίζονται με τη σχιζοφρένεια και τη διπολική διαταραχή. Οι διαταραχές αυτές είναι χρόνιες και συνοδεύονται από επαναλαμβανόμενες περιόδους υποτροπής. Η έγκαιρη πρόβλεψη των ψυχωτικών υποτροπών είναι ένα σημαντικό κλινικό ζήτημα λόγω της χρόνιας και σοβαρής τους φύσης και αξίζει ιδιαίτερης προσοχής μιας και θα μπορούσε να προσφέρει στους ασθενείς καλύτερη ποιότητα ζωής. Παρόλα αυτά, οι βαθύτερες αιτίες παραμένουν άγνωστες και δεν έχουν ανακαλυφθεί ακόμη αξιόπιστοι βιοδείκτες για τη διάγνωσή τους και την πρόβλεψη της πορείας της ψυχωτικής συμπτωματολογίας στο χρόνο.

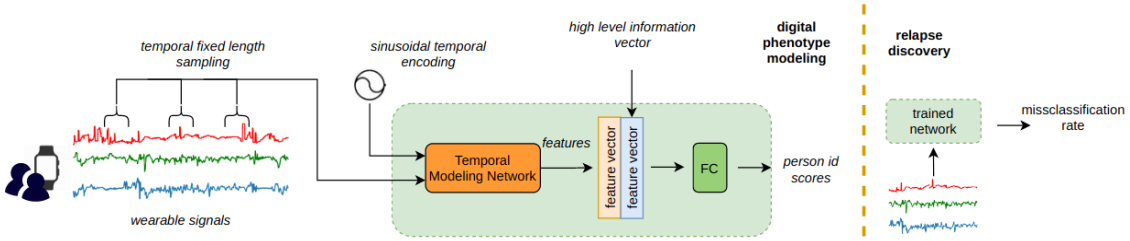
Στο Κεφάλαιο αυτό, θα εστιάσουμε στον εντοπισμό των ψυχωτικών υποτροπών με χρήση βιοσημάτων. Στο πλαίσιο αυτό θα εντάξουμε την ανάλυση που προηγήθηκε στο Κεφάλαιο 5 σχετικά με τη δημιουργία και τη μελέτη των ψηφιακών φαινοτύπων των χρηστών. **Υποστηρίζουμε ότι η ανίχνευση των μοτίβων συμπεριφοράς που προσδιορίζουν ένα άτομο μπορεί να είναι ευεργετική για την ανίχνευση ψυχωτικών υποτροπών, καθώς, κατά τη διάρκεια μιας ψυχωτικής υποτροπής, ο ασθενής τείνει να υιοθετεί διαφορετικές συμπεριφορές και άρα πιθανά αλλάζουν τα ατομικά του μοτίβα**.

Οι υποτροπές είναι συχνές στις ψυχωτικές ασθένειες και επηρεάζουν τη νόηση, τη συμπεριφορά και τα συναισθήματα με άμεσο τρόπο, ενώ έχουν σημαντικές επιπτώσεις σε ατομικό επίπεδο, συμπεριλαμβανόμενης της συμπτωματολογίας, των ανεπιθύμητων ενεργειών των φαρμάκων, της λειτουργικής ικανότητας, της επιβάρυνσης των φροντιστών και της υψηλής θνησιμότητας. Μια ψυχωτική υποτροπή είναι μια σαφής κλινική επιδείνωση του ασθενούς με την επανεμφάνιση ψυχωτικών συμπτωμάτων (παραληρητικές ιδέες, ψευδαισθήσεις) ή μανιακά, καταθλιπτικά ή μικτά επεισόδια με ψυχωτικά συμπτώματα. Ως τέλος μιας υποτροπής ορίζεται η στιγμή που τα συμπτώματα μειώνονται και ο ασθενής επιστρέφει κλινικά και λειτουργικά στην κατάσταση πριν την υποτροπή.

Έτσι η δομή του Κεφαλαίου έχει ως εξής:

- Στο πρώτο μέρος θα μελετήσουμε τον εντοπισμό των ψυχωτικών υποτροπών ως αποτέλεσμα *της μείωσης της ακρίβειας ταυτοποίησης ενός δικτύου χρονικής μοντελοποίησης που είναι εκπαιδευμένο στην ανίχνευση των προτύπων συμπεριφοράς του εκάστοτε ατόμου.*
- Στο δεύτερο μέρος αντιμετωπίζουμε το πρόβλημα *ως συνδυασμό της ανακατασκευής βιοσημάτων, υπό το πρίσμα της αναγνώρισης ανωμαλιών (anomaly detection), και της ταυτοποίησης του φαινοτύπου των ασθενών* και δείχνουμε την αυξημένη απόδοση στον εντοπισμό των υποτροπών.

6.1 Αναγνώριση Ψυχωτικών Υποτροπών μέσω Ταυτοποίησης Χρήστη



Σχήμα 6.1: Το προτεινόμενο σύστημα για τον εντοπισμό ψυχωτικών υποτροπών μέσω της ταυτοποίησης του χρήστη. Βασιζόμενοι στο σύστημα που προτάθηκε στο 5.4.2, κατά τη διάρκεια της εκπαίδευσης το μοντέλο μας μαθαίνει τα πρότυπα συμπεριφοράς διαφορετικών χρηστών. Στη συνέχεια εξετάζει το ποσοστό εσφαλμένης ταξινόμησης σε διαφορετικές περιόδους και έτσι εντοπίζονται οι περίοδοι όπου ο χρήστης βρίσκεται σε υποτροπή.

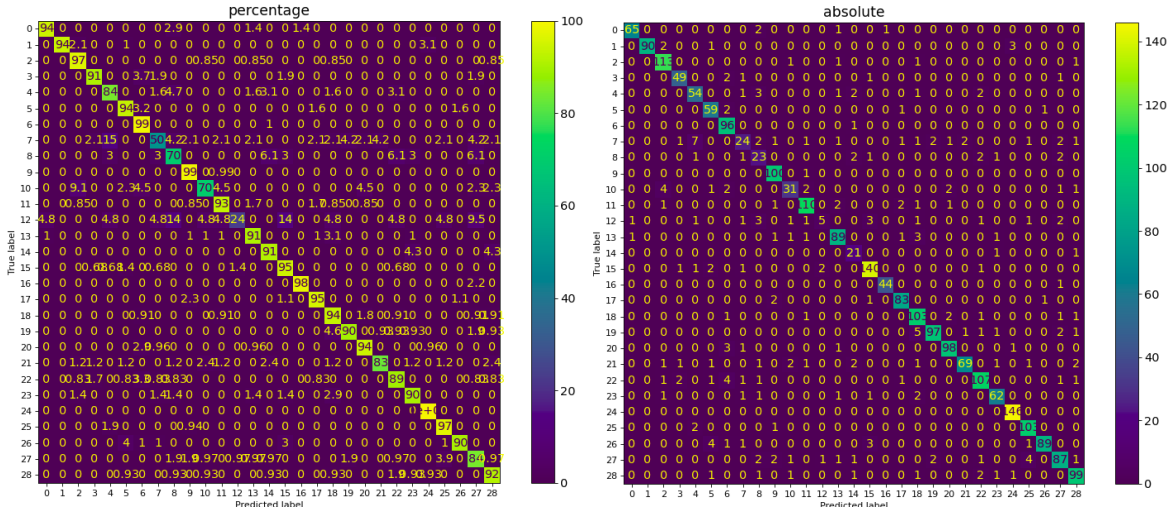
Αξιοποιώντας το σύστημα αναγνώρισης ψηφιακών φαινοτύπων που αναπτύξαμε στο Κεφάλαιο 5 προχωράμε ένα βήμα παραπέρα μελετώντας τις μεταβολές που παρουσιάζονται σε διαφορετικές περιόδους της ψυχωτικής διαταραχής. Ελέγχουμε την υπόθεσή μας ότι η ανίχνευση των μοτίβων συμπεριφοράς και δραστηριότητας που προσδιορίζουν ένα άτομο μπορεί να αποτελεί το κλειδί για την ανίχνευση ψυχωτικών υποτροπών, καθώς ο ασθενής τείνει να υιοθετεί διαφορετικά πρότυπα συμπεριφοράς κατά τις περιόδους υποτροπής. Έτσι, αναμένεται η μείωση της ακρίβειας ενός νευρωνικού δικτύου που είναι εκπαιδευμένο στην ανίχνευση των προτύπων συμπεριφοράς ενός ατόμου, όταν το άτομο αυτό βιώνει μια ψυχωτική υποτροπή.

Στην ενότητα αυτή παρουσιάζουμε μια νέα μέθοδο για την ανακάλυψη ψυχωτικών υποτροπών θέτοντας το πρόβλημα της ανίχνευσης υποτροπής ως ένα πρόβλημα λανθασμένης ταξινόμησης σε νευρωνικά δίκτυα που είναι εκπαιδευμένα στην ταυτοποίηση ατόμων. Η μεθοδολογία που προτείνουμε για την ανίχνευση περιόδων υποτροπής, μέσω της ταυτοποίησης του ατόμου, επεκτείνει το προτεινόμενο σύστημα της Ενότητας 5.4.2 και περιγράφεται στο Σχήμα 6.1. Πλέον στην αξιολόγηση του συστήματος, εξετάζουμε δεδομένα του χρήστη κατά τη διάρκεια ενός ψυχωτικού επεισοδίου αλλά και πριν από αυτό. Έτσι, διερευνούμε το πως οι αλλαγές στη συμπεριφορά του χρήστη αντανακλάται στην πιθανότητα κατανομής εξόδου και στην ακρίβεια ταξινόμησης του δικτύου.

6.1.1 Διάκριση Διαφορετικών Περιόδων Ψυχωτικών Διαταραχών

Δεδομένης της καλής απόδοσης στην αναγνώριση φαινοτύπων του συστήματος που παρουσιάσαμε στην ενότητα 5.4.2 εστιάζουμε τώρα στον τελικό στόχο: τη διάκριση των περιόδων ψυχωτικών διαταραχών. Συγκεκριμένα, θέλουμε να διερευνήσουμε εάν οι πληροφορίες που εξάγονται από το δίκτυο μπορούν να χρησιμοποιηθούν αποτελεσματικά για την ανίχνευση μιας περιόδου ψυχωτικής υποτροπής και ακόμα καλύτερα μιας περιόδου που προηγείται της υποτροπής με σκοπό την καλύτερη ενημέρωση του ιατρικού προσωπικού για την πρόοδο της διαταραχής.

Για να γίνει αυτό, εκτός από την τυπική ακρίβεια αναγνώρισης στο πλήρες σύνολο των 29 ασθενών (e-Prevention subset A) κατά τη διάρκεια κανονικών περιόδων, αναφέρουμε επίσης την απόδοση στο μειωμένο σύνολο δεδομένων έντεκα ασθενών που παρουσίασαν υποτροπές κατά τη διάρκεια της συλλογής δεδομένων. Για αυτό το υποσύνολο, χωρίζουμε τα αποτελέσματα



Σχήμα 6.2: Ενδεικτικοί πίνακες σύγχυσης για την ταυτοποίηση των φαινοτύπων των 29 χρηστών κατά την περίοδο ύφεσης. Αριστερά δίνεται η επί τοις εκατό ακρίβεια του δικτύου για τη σωστή ταξινόμηση του φαινότυπου, ενώ ο δεξιά πίνακας δίνει το πλήθος των δειγμάτων που ταξινομεί ο αλγόριθμος. Τα παραπάνω αποτελέσματα αναφέρονται στην αναγνώριση του πλήρους δικτύου (με temporal encogding), χρήση της επιπρόσθετης πληροφορίας των ωρών ύπνου και της ημέρας της εβδομάδας.

ταξινόμησης στις τρεις περιόδους μιας ψυχωτικής διαταραχής: περίοδος ύφεσης, περίοδος πριν την υποτροπή και περίοδος υποτροπής. Για ευκολία στην παρουσίαση των αποτελεσμάτων χρησιμοποιούνται οι όροι normal, pre-relapse και relapse για τις τρεις περιόδους.

Τα πειραματικά αποτελέσματα παρουσιάζονται στον Πίνακα 6.1 ανάλογα με την περίοδο της διαταραχής. Αναφέρουμε μόνο τα αποτελέσματα της εξισορροπημένης ακρίβειας καθώς είναι πιο αντιπροσωπευτικά για τη μη ισορροπημένη φύση των δεδομένων των υποτροπών.

	Metrics	**Balanced Accuracy (29 patients)**	**Balanced Accuracy (11 patients)**		
	Extra Features	**Normal**	**Normal**	**Pre-Rel.**	**Relapse**
Base	None	90.13	87.53	75.15	74.83
	Hours of sleep (HS)	**90.19**	**88.37**	77.44	72.52
	Day of Week (DW)	89.40	86.46	75.88	70.12
	HS + DW	89.76	87.65	76.88	70.35
+ Temporal encoding	None	89.04	85.93	75.02	71.53
	Hours of sleep (HS)	88.85	85.65	76.81	70.20
	Day of Week (DW)	88.07	85.16	75.20	68.84
	HS + DW	88.67	**86.00**	**74.28**	**69.86**

Πίνακας 6.1: Μελέτη επιπρόσθετων χαρακτηριστικών (πληροφορίες ύπνου και ημέρας) για την περίπτωση της αρχιτεκτονικής LSTM τόσο σε ολόκληρη τη συλλογή 29 ασθενών (e-Prevention subset A) όσο και στο υποσύνολο των 11 ασθενών που αντιμετώπισαν υποτροπές. Η εξισορροπημένη ακρίβεια (balanced accuracy) κάθε πειράματος και περιόδου αναφέρονται για τις τρεις εξεταζόμενες περιόδους: ύφεση (normal), προ-υποτροπή (pre-relapse) και υποτροπή (relapse).

Για να αποκτήσουμε μια καλύτερη εικόνα της δυνατότητας αναγνώρισης του κάθε χρήστη παρουσιάζουμε στο Σχήμα 6.2 τα confusion matrices για τους 29 ασθενείς τόσο ως προς την

επί τοις εκατό ακρίβεια αναγνώρισης (αριστερά) όσο και ως προς το πλήθος των εξεταζόμενων δειγμάτων στο σύνολο ελέγχου για ένα από τα 5-folds των πειραμάτων μας. Το μοντέλο που χρησιμοποιείται εδώ είναι το πλήρες δικτύο (με temporal encogding), χρήση της επιπρόσθετης πληροφορίας των ωρών ύπνου και της ημέρας της εβδομάδας. Στο Σχήμα 6.2 παρατηρείται πως για 25 ασθενείς το ποσοστό σωστής ταξινόμησης είναι από 84% έως και 100% για την περίοδο ύφεσης της ασθένειας. Όμως για δύο ασθενείς το ποσοστό είναι στο 70%, για έναν άλλο στο 50% και για έναν ακόμα στο 24%. Όπως μπορούμε να δούμε στον πίνακα του πλήθους των δειγμάτων, οι ασθενείς αυτοί έχουν στο σύνολο ελέγχου από 21 έως 47 δείγματα-ημέρες όταν οι περισσότεροι ασθενείς έχουν περισσότερες από 70 ημέρες στο σύνολο ελέγχου. Η παραπάνω διαφορά μας δίνει την αναλογία των δειγμάτων στο σύνολο εκπαίδευσης, πράγμα που δικαιολογεί ως ένα βαθμό την αρκετά χαμηλότερη απόδοση του δικτύου στους ασθενείς αυτούς. Στη συνέχεια θα μελετήσουμε ενδελεχώς την απόδοση του δικτύου σε κάθε ασθενή που παρουσίασε κάποια ψυχωτική υποτροπή κατά τη διάρκεια του έργου (ID ασθενών 0-10).

Όσον αφορά την περίοδο πριν από την υποτροπή, στη μελέτη μας, η περίοδος αυτή (pre-relapse) ορίζεται ως τέσσερις εβδομάδες πριν από την καταγεγραμμένη έναρξη της υποτροπής με βάση τις επισημειώσεις των κλινικών γιατρών. Στη συνέχεια θα αναφερθούμε συνοπτικά στην επιλογή αυτή.

	Metrics	**Mean Probability (11 patients)**			**Median Probability (11 patients)**		
	Extra Features	**Normal**	**Pre-Rel.**	**Relapse**	**Normal**	**Pre-Rel.**	**Relapse**
Base	None	0.8685	0.8243	0.8227	0.9708	0.9096	0.9433
	Hours of sleep (HS)	0.8794	0.8215	0.8124	0.9873	0.9109	0.9472
	Day of Week (DW)	0.8632	0.8156	0.7756	0.9859	0.9332	0.8898
	HS + DW	0.8744	0.8076	0.7902	0.9909	0.8809	0.8980
+ Temporal encoding	None	0.8564	0.8288	0.8027	0.9465	0.9120	0.8947
	Hours of sleep (HS)	0.8579	0.8076	0.7812	0.9895	0.8832	0.8724
	Day of Week (DW)	0.8486	0.8032	0.7712	0.9687	0.8506	0.8791
	HS + DW	**0.8634**	**0.7954**	**0.7593**	**0.9690**	**0.8734**	**0.8498**

Πίνακας 6.2: Μελέτη επιπρόσθετων χαρακτηριστικών (πληροφορίες ύπνου και ημέρας) για την περίπτωση της αρχιτεκτονικής LSTM τόσο σε ολόκληρη τη συλλογή 29 ασθενών (e-Prevention subset A) όσο και στο υποσύνολο των 11 ασθενών που αντιμετώπισαν υποτροπές. Η μέση και διάμεση πιθανότητα κάθε πειράματος και περιόδου αναφέρονται για τις τρεις εξεταζόμενες περιόδους: ύφεση (normal), προ-υποτροπή (pre-relapse) και υποτροπή (relapse).

Για να κατανοήσουμε καλύτερα την ικανότητα ανίχνευσης ενός τέτοιου συστήματος, αναφέρουμε στον Πίνακα 6.2 για τα ίδια πειράματα, τη μέση και διάμεση πιθανότητα (mean and median probability) ταξινόμησης των χρηστών σε κάθε περίοδο. Αυτή η μέτρηση είναι πιο ενδεικτική των αποτελεσμάτων του μοντέλου, επειδή μια μέρα θα μπορούσε να ταξινομηθεί στον σωστό χρήστη, αλλά με μικρότερη πιθανότητα ταξινόμησης για παράδειγμα, στην περίοδο πριν από την υποτροπή ή κατά την υποτροπή. Με άλλα λόγια, ακόμα κι αν το άτομο έχει αναγνωριστεί σωστά, μια σταθερή μείωση της πιθανότητας εντοπισμού του ατόμου κατά τη διάρκεια μιας περιόδου μπορεί να είναι μια πολύ χρήσιμη ένδειξη επιμέρους αλλαγής των προτύπων συμπεριφοράς.

Οι Πίνακες 6.1 και 6.2, εκτός από την καταγραφή της απόδοσης του δικτύου σχετικά με το πρόβλημα ανίχνευσης της υποτροπής, παρέχουν επίσης πληροφορίες για τη μελέτη της επίδρασης ενός συνόλου χαρακτηριστικών υψηλού επιπέδου, δηλαδή για το ποσοστό ύπνου και για την ημέρα της εβδομάδας, όπως αναφέρεται στην Ενότητα 5.2.1. Θεωρούμε ότι η ποσότητα του ύπνου είναι ένα κρίσιμο χαρακτηριστικό, επειδή συνήθως κυμαίνεται κατά τη διάρκεια των

υποτροπών, ενώ οι πληροφορίες της ημέρας μπορεί να βοηθήσει το δίκτυο να διακρίνει τα μοτίβα της ίδιας ημέρας σε διαφορετικές εβδομάδες.

Συνολικά, εξάγουμε τα ακόλουθα συμπεράσματα από τα παραπάνω αποτελέσματα:

1. Όσον αφορά την ακρίβεια, η απλούστερη έκδοση του δικτύου (Base) με επιπλέον χαρακτηριστικό αυτό των ωρών ύπνου επιτυγχάνει τα υψηλότερα αποτελέσματα. Ωστόσο, όταν εξετάζουμε το σύνολο των 11 ασθενών, ενδιαφερόμαστε κυρίως για τον εντοπισμό των αλλαγών στην ακρίβεια κατά τη σύγκριση των περιόδων πριν της υποτροπής, της υποτροπής και των φυσιολογικών περιόδων. Παρατηρούμε ότι γενικά, η προσθήκη χαρακτηριστικών υψηλού επιπέδου διευρύνει το χάσμα μεταξύ της φάσης ύφεσης της ασθένειας και της φάσης της υποτροπής όσον αφορά την ακρίβεια.

2. Η μέση και η διάμεση πιθανότητα ταξινόμησης των χρηστών σε κάθε περίοδο είναι πράγματι πιο ενδεικτική των αποτελεσμάτων του μοντέλου.

3. Συμπεραίνουμε ότι όλα τα πιθανά συμπληρωματικά χαρακτηριστικά (χρονική κωδικοποίηση, ύπνος και πληροφορίες ημέρας της εβδομάδας) συμβάλλουν στην ανίχνευση ψυχωτικών υποτροπών επειδή αποτυπώνουν διαισθητικά ένα πλήρες προφίλ των καθημερινών δραστηριοτήτων ενός ατόμου.

Χρονική διάρκεια περιόδου πριν την υποτροπή

Όπως αναφέραμε, στην εργασία μας έχουμε θεωρήσει ως διάρκεια της pre-relapse περιόδου τις τέσσερις εβδομάδες (28 ημέρες) πριν την ημερομηνία έναρξης της ψυχωτικής υποτροπής. Στη βιβλιογραφία δεν υπάρχει κάποιος ορισμός για το χρονικό διάστημα πριν την υποτροπή κατά τον οποίο θα μπορούσαν να παρουσιαστούν ενδείξεις ότι ο ασθενής πλησιάζει σε μια ψυχωτική υποτροπή. Παρ' όλα αυτά ο ορισμός μιας τέτοιας περιόδου θα μπορούσε να βοηθήσει στην πρόβλεψη της υποτροπής, την πιο άμεση αντιμετώπιση από τους θεράποντες ιατρούς και την ταχύτερη επαναφορά του ασθενούς στη φάση της ύφεσης. Στην παρούσα διατριβή φυσικά δεν είναι στόχος μας ο ορισμός της περιόδου αυτής, καθώς αποτελεί ιατρικό πρόβλημα, αλλά η μελέτη της ύπαρξης ενδείξεων στον ψηφιακό φαινότυπο των ασθενών κάποιο χρονικό διάστημα πριν την υποτροπή.

Η επιλογή των τεσσάρων εβδομάδων στη δική μας προσέγγιση έγινε ύστερα από μια προκαταρκτική μελέτη. Στον Πίνακα 6.3 παρουσιάζονται τα συγκριτικά αποτελέσματα για διαφορετικές διάρκειες της pre-relapse περιόδου χρησιμοποιώντας την CNN εκδοχή του προτεινόμενου συστήματος με χρήση temporal encoding και την επαύξηση των δεδομένων με προσθήκη θορύβου και εφαρμογή τυχαίας μάσκας. Παρατηρούμε πως δεν παρουσιάζονται ιδιαίτερες μεταβολές στην ακρίβεια ταυτοποίησης του συνολικού συστήματος στο σύνολο των 29 ασθενών. Ως προς τη σύγκριση στα αποτελέσματα των 11 ασθενών που παρουσίασαν τουλάχιστον μια υποτροπή, μεταξύ των 28 και των 42 ημερών θα μπορούσαμε να πούμε πως είναι αποδεκτές και οι δύο επιλογές. Παρ' όλα αυτά, επιλέξαμε τη διάρκεια των 28 ημερών καθώς το σύστημα παρουσιάζει μεγαλύτερη ακρίβεια αναγνώρισης στην περίοδο της υποτροπής. Ως προς τις 21 ημέρες, διάρκεια περιόδου που συναντάται στη βιβλιογραφία π.χ. [Torous et al.; 2016], δεν επιλέχθηκε καθώς κρίναμε πως ο όγκος των δεδομένων στη βάση μας για την pre-relapse περίοδο δεν ήταν αρκετός οπότε τα συμπεράσματα για τη συγκεκριμένη περίοδο θα ήταν πιο επισφαλή. Τέλος, όπως φαίνεται και από τον πίνακα, στην επιλογή των 56 ημερών το ποσοστό ταυτοποίησης στην pre-relapse περίοδο έχει ανέβει σημαντικά πλησιάζοντας το ποσοστό των κανονικών περιόδων, πράγμα που θα μπορούσαμε να αποδώσουμε στο ότι μια pre-relapse περίοδος 56 ημερών περιλαμβάνει αρκετά δεδομένα «κανονικότητας» της δραστηριότητας του ασθενούς.

Pre-relapse Duration (days)	Balanced Accuracy (%)			
	29 patients	11 patients		
	Normal	Normal	Pre-Rel.	Relapse
21	85.03	81.22	66.56	62.35
28	85.25	82.38	69.03	65.01
42	84.06	79.44	69.88	56.62
56	86.07	80.67	75.57	65.30

Πίνακας 6.3: Μελέτη διάρκειας της pre-relapse περιόδου για την περίπτωση της αρχιτεκτονικής CNN τόσο σε ολόκληρη τη συλλογή 29 ασθενών (e-Prevention subset A) όσο και στο υποσύνολο των 11 ασθενών που αντιμετώπισαν υποτροπές. Παρουσιάζεται η εξισορροπημένη ακρίβεια (balanced accuracy) κάθε πειράματος και περιόδου.

6.1.2 Ταυτοποίηση Χρήστη ανά Άτομο και ανά Περίοδο

Για να κατανοήσουμε περαιτέρω την ικανότητα ανίχνευσης υποτροπής, διεξάγουμε τώρα μια μελέτη ανά άτομο για κάθε περίοδο (ύφεση, προ-υποτροπή και υποτροπή), όπως παρουσιάζεται στον Πίνακα 6.4. Αφού είδαμε πως η πιθανότητα ταυτοποίησης του ατόμου είναι μια αποδοτική μετρική για το πρόβλημά μας, στον πίνακα χρησιμοποιούμε τη μέση τιμή της πιθανότητας αναγνώρισης ανά άτομο για κάθε περίοδο. Τα αναφερόμενα αποτελέσματα αντιστοιχούν στο μοντέλο LSTM με όλες τις πρόσθετες πληροφορίες (χρονική κωδικοποίηση, πληροφορίες ύπνου, ημέρα της εβδομάδας) που πέτυχαν σημαντική διάκριση μεταξύ διαφορετικών περιόδων στον Πίνακα 6.2.

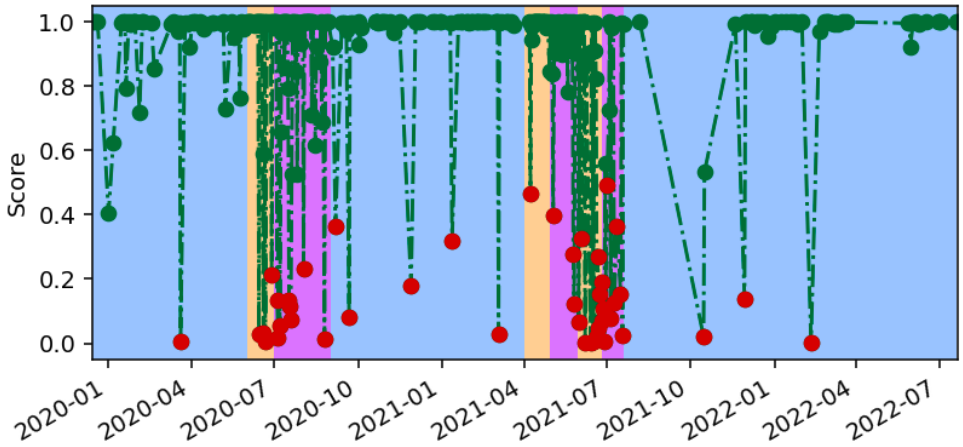
Στον Πίνακα 6.4, παρέχουμε επίσης την απόλυτη πτώση της μέσης πιθανότητας αναγνώρισης για κάθε χρήστη, για ευκολία. Από τα στοιχεία του πίνακα προκύπτει ότι έχουμε σημαντική πτώση στη μέση πιθανότητα αναγνώρισης του δικτύου, τόσο κατά τις περιόδους πριν την υποτροπή όσο και κατά τις περιόδους υποτροπής. Αυτό δείχνει ότι το δίκτυο δεν μπορεί να διακρίνει καλά τον χρήστη από τον ψηφιακό φαινότυπο όταν μπήκε σε υποτροπή και ακριβώς πριν από αυτήν (προ-υποτροπή). Αυτό το μοτίβο απόλυτης μείωσης της πιθανότητας μπορεί να φανεί στους περισσότερους χρήστες, με ορισμένες εξαιρέσεις, όπως ο χρήστης #6, που είχε υψηλές πιθανότητες αναγνώρισης σε όλες τις περιόδους του. Σημειώστε ότι η διαθεσιμότητα μεγαλύτερου όγκου δεδομένων και κυρίως μεγαλύτερου συνόλου ασθενών θα μπορούσε να βοηθήσει περαιτέρω στη διάκριση μεταξύ τους και κατά συνέπεια στον εντοπισμό διαφορετικών ψυχωτικών περιόδων. Στην επόμενη υποενότητα συμπεριλαμβάνουμε επίσης μια στατιστική ανάλυση των σκορ ανά ασθενή.

Η προτεινόμενη λογική του εντοπισμού των υποτροπών μέσω της αναγνώρισης φαινοτύπου μπορεί επίσης να απεικονιστεί στο Σχήμα 6.3, όπου φαίνονται τα αποτελέσματα αναγνώρισης για τον χρήστη #1 σε διάστημα σχεδόν δυόμισι ετών. Παρατηρούμε ότι, για τον συγκεκριμένο ασθενή, η πλειονότητα των ημερών σε περιόδους ύφεσης οδηγεί σε σωστές ταυτοποιήσεις, ενώ κατά τις υποτροπές παρουσιάζεται υψηλό ποσοστό λανθασμένων ταξινομήσεων του χρήστη, υποδηλώνοντας κάποια σημαντική αλλαγή-μη ομαλότητας του διανύσματος χαρακτηριστικών που προκύπτει από τα βιοσήματά του.

Στο Σχήμα 6.4 παρουσιάζονται τα αποτελέσματα αναγνώρισης για δύο ακόμα ασθενείς. Στην περίπτωση του ασθενή #2, εύκολα παρατηρείται η συμφωνία του διαγράμματος με τις τιμές του Πίνακα 6.4 όπου είδαμε να υπάρχει σημαντικά στατιστική διαφορά μόνο μεταξύ pre-relapse και normal περιόδων. Πράγματι στο διάγραμμα βλέπουμε την «επιτυχή» εύρεση της pre-relapse περιόδου ενώ για τις λανθασμένες ταξινομήσεις της normal περιόδου, προφανώς αντιμετωπίζονται ως outliers μιας και το πλήθος τους είναι συγκριτικά πολύ μικρότερο του

	Probability			Absolute Change		
ID	normal	pre-rel	relapse	normal → pre-rel	normal → relapse	pre-rel → relapse
0	0.971	0.944	0.826	**−0.027**	**−0.145**	−0.118
1	0.952	0.928	0.872	**−0.024**	**−0.080**	**−0.056**
2	0.963	0.913	0.997	**−0.051**	0.034	0.085
3	0.853	0.784	0.726	**−0.070**	**−0.127**	**−0.058**
4	0.905	0.718	0.765	**−0.187**	**−0.140**	0.047
5	0.905	1.000	0.794	0.095	**−0.112**	−0.206
6	0.966	0.983	0.974	0.017	0.008	−0.009
7	0.659	0.672	0.496	0.014	**−0.163**	**−0.176**
8	0.693	0.407	0.541	**−0.286**	**−0.152**	0.134
9	0.961	0.986	1.000	0.025	0.039	0.014
10	0.703	0.406	0.358	**−0.297**	−0.346	−0.049
all	0.866	0.795	0.759	−0.0723	−0.107	−0.035

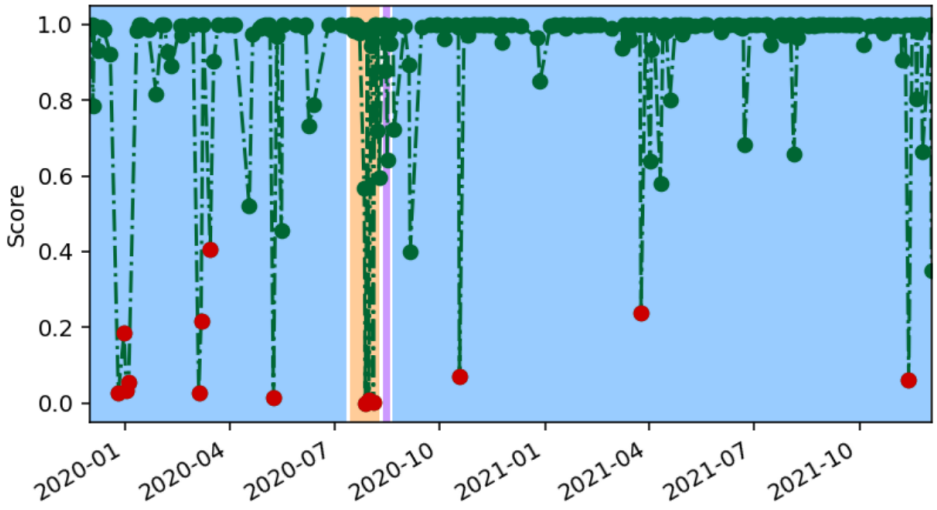
Πίνακας 6.4: Ανά χρήστη και ανά περίοδο (ύφεση, προ-υποτροπή και υποτροπή) η μέση πιθανότητα αναγνώρισης. Δείχνουμε για ευκολία την απόλυτη αλλαγή στη μέση πιθανότητα μεταξύ όλων των συνδυασμών μεταβάσεων ανάμεσα στις περιόδους. Οι τιμές της απόλυτης αλλαγής ($x \rightarrow y$) με έντονους χαρακτήρες υποδηλώνουν ότι οι τιμές κατά τη φάση x ήταν στατιστικά σημαντικές από τις βαθμολογίες κατά τη φάση y του ασθενούς.



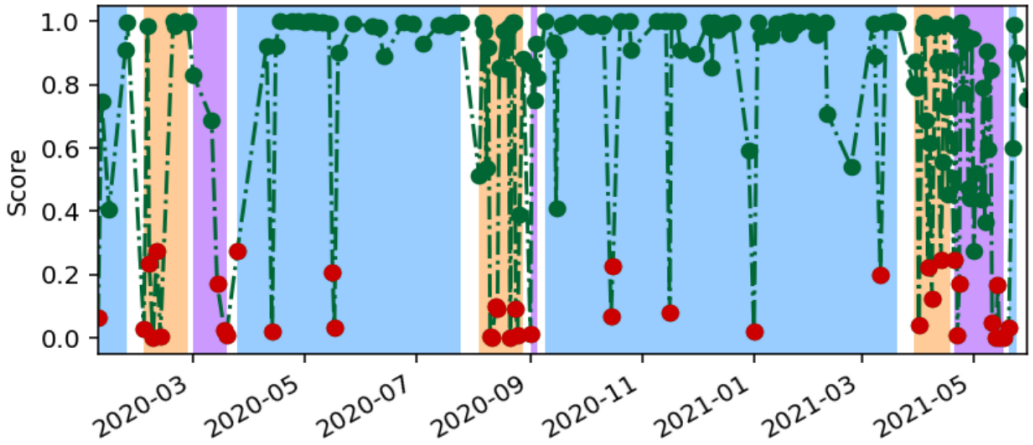
Σχήμα 6.3: Οπτικοποίηση προβλέψεων αναγνώρισης του ασθενούς #1 : οι σωστές ταυτοποιήσεις του χρήστη απεικονίζονται με πράσινο χρώμα, ενώ οι λανθασμένες σημειώνονται με κόκκινο. Οι τρεις περίοδοι ύφεσης σκιάζονται με γαλάζιο χρώμα, οι τρεις περίοδοι υποτροπής με μωβ και οι τρεις περίοδοι πριν από την υποτροπή με πορτοκαλί.

πλήθους των normal ημερών (ταξινομήσεων-score). Αντίστοιχα, στο Σχήμα 6.4 (β) παρουσιάζονται τα αποτελέσματα για τον ασθενή #3 που σύμφωνα και με τον Πίνακα 6.4 παρατηρούνται στατιστικά σημαντικές μεταβολές μεταξύ όλων των συνδυασμών των 3 περιόδων.

Ως τώρα η προσέγγισή μας καταλήγει σε μια μοναδική βαθμολογία ταξινόμησης- ταυτοποίησης για κάθε άτομο ανά ημέρα. Για να κάνουμε ένα πρώτο αδρό φιλτράρισμα μεμονωμένων ακραίων τιμών και να βγάλουμε μια εκτίμηση για την τάση τριών ημερών, χρησιμοποιούμε στατιστικά στοιχεία για κινούμενα παράθυρα τριών ημερών. Έτσι μπορούμε να καταλήξουμε σε μια πιο ομαλή γραμμή για την προσωπική βαθμολογία και να εντοπίσουμε περιόδους με χαμηλότε-



(α) Οπτικοποίηση προβλέψεων αναγνώρισης για τον ασθενή #2



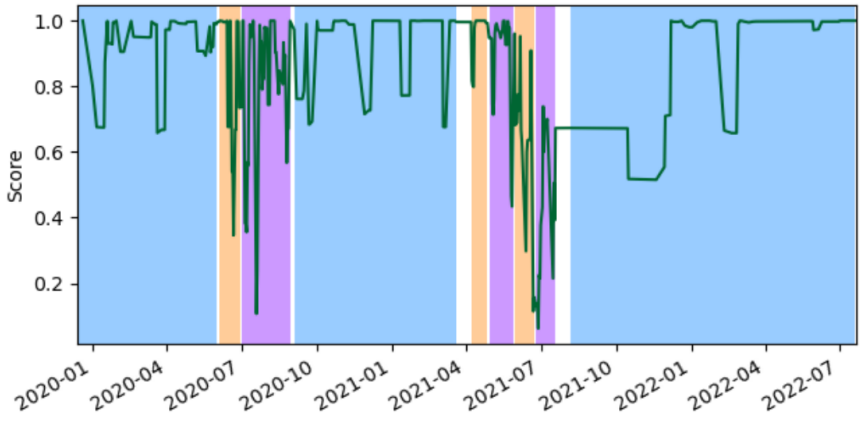
(β) Οπτικοποίηση προβλέψεων αναγνώρισης για τον ασθενή #3

Σχήμα 6.4: Οπτικοποίηση προβλέψεων αναγνώρισης για δύο ασθενείς: οι σωστές ταυτοποιήσεις του χρήστη απεικονίζονται με πράσινο χρώμα, ενώ οι λανθασμένες σημειώνονται με κόκκινο. Οι τρεις περίοδοι ύφεσης σκιάζονται με γαλάζιο χρώμα, οι τρεις περίοδοι υποτροπής με μωβ και οι τρεις περίοδοι πριν από την υποτροπή με πορτοκαλί.

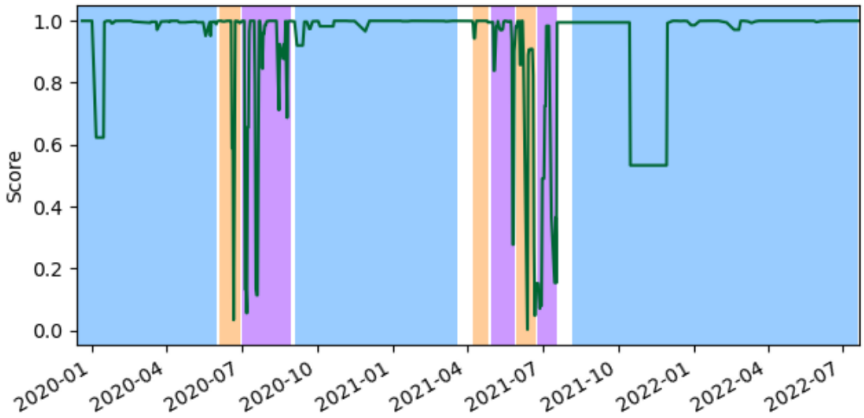
ρες βαθμολογίες. Στο Σχήμα 6.5 παρουσιάζουμε φιλτράρισμα με χρήση (α) της μέσης τιμής πρόβλεψης τριών ημερών, (β) τη διάμεσο της πρόβλεψης για τον ασθενή #1. Παρατηρούμε η χρήση τέτοιων φίλτρων σε περιόδους με λίγες προβλέψεις, π.χ. η περίοδος της δεύτερης υποτροπής, η περίοδος πριν από αυτή καθώς και η περίοδος μετά την τελευταία υποτροπή, εξαλείφουν και συγχέουν την όποια διαθέσιμη πληροφορία.

6.1.3 Στατιστική Ανάλυση των Πιθανοτήτων Ταυτοποίησης

Πραγματοποιήσαμε επίσης περαιτέρω στατιστική ανάλυση στα αποτελέσματα του εκπαιδευμένου δικτύου. Πρώτον, συλλέγουμε για κάθε ασθενή όλες τις τιμές-πιθανότητες αναγνώρισης για κάθε διαφορετική περίοδο και πραγματοποιούμε κατά ζεύγη single tail Mann-Whitney



(α) Μέση τιμή πρόβλεψης τριών ημερών



(β) Διάμεση τιμή πρόβλεψης τριών ημερών

Σχήμα 6.5: Οπτικοποίηση της μέσης τιμής (α) και της διάμεσης τιμής (β) για την πρόβλεψη τριών ημερών για τον ασθενή #1.

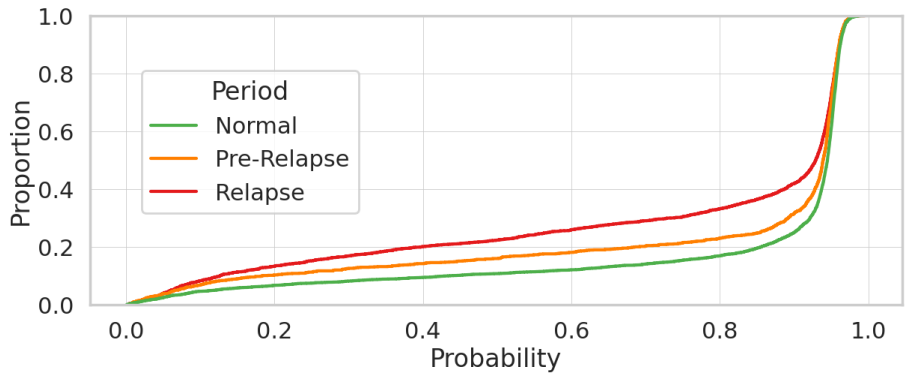
U-test με διόρθωση Bonferroni [1] [Mann and Whitney; 1947]. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 6.4. Κάθε τιμή απόλυτης μεταβολής που είναι γραμμένη με έντονους χαρακτήρες υποδηλώνει ότι οι πιθανότητες κατά την περίοδο στα αριστερά του βέλους (στο όνομα της στήλης) ήταν στατιστικά σημαντικές σχετικά με τις τιμές κατά τη διάρκεια της περιόδου που αναγράφεται στα δεξιά του βέλους. Μολονότι παρατηρούμε πως υπάρχουν ορισμένες περιπτώσεις (ασθενείς #6 και #9) όπου οι μεταβολές που εντοπίσαμε δεν είναι στατιστικά σημαντικές, στη μεγάλη πλειονότητα των ασθενών οι αλλαγές που ανιχνεύονται μεταξύ των περιόδων υποτροπής και των περιόδων ύφεσης, καθώς και ανάμεσα στις περιόδους πριν τις υποτροπές και τις περιόδους ύφεσης είναι στατιστικά σημαντικές και στην περίπτωση τριών ασθενών, οι τιμές κατά την περίοδο πριν από την υποτροπή είναι σημαντικά μεγαλύτερες από τις βαθμολογίες κατά την περίοδο της υποτροπής.

Επιπλέον, συλλέγουμε όλες τις τιμές-πιθανότητες αναγνώρισης σε όλους τους ασθενείς και τις περιόδους. Το πλήθος των τιμών που προκύπτει είναι $N = 7942$, με 4012 τιμές για τις περιόδους ύφεσης, 1360 για τις περιόδους πριν τις υποτροπές και 2570 για τις περιόδους υποτροπών.

[1] Η χρήση ενός single-tailed U-test Mann-Whitney ελέγχου με διόρθωση Bonferroni μπορεί να βοηθήσει μιας και θέλουμε να συγκρίνετε ανά δύο όλες τις ανεξάρτητες ομάδες των βαθμολογιών και να προσδιορίσουμε εάν η μία ομάδα έχει σταθερά σημαντικά υψηλότερες τιμές από την άλλη.

Αρχικά, δείχνουμε στο Σχήμα 6.6 για κάθε περίοδο την εμπειρική αθροιστική κατανομή των πιθανοτήτων (empirical Cumulative Distribution Function - eCDF). Αυτή η καμπύλη δείχνει για κάθε τιμή p στον άξονα x, το ποσοστό όλων των βαθμολογιών που είναι χαμηλότερες από p. Εύκολα διακρίνεται πως οι πιθανότητες αναγνώρισης κατά τις περιόδους υποτροπής των χρηστών τείνουν να να είναι χαμηλότερες από τις τιμές των πιθανοτήτων κατά τις περιόδους πριν από την υποτροπή και τις περιόδους ύφεσης. Σε όλα τα σημεία p της καμπύλης βλέπουμε πως το ποσοστό των τιμών για την καμπύλη υποτροπής που είναι χαμηλότερο από p, είναι μεγαλύτερο σε σύγκριση με το ποσοστό πριν την υποτροπή και αυτό των περιόδων ύφεσης. Το ίδιο ισχύει όταν συγκρίνουμε τις βαθμολογίες των περιόδων πριν από την υποτροπή με τις βαθμολογίες κατά τις κανονικές περιόδους. Τα συμπεράσματα αυτά συνάδουν με την ανάλυση της προηγούμενης υποενότητας.

Επίσης διεξήγαμε δοκιμές single tail U-test Mann-Whitney με διόρθωση Bonferroni για το σύνολο των πιθανοτήτων ανά περίοδο, οι οποίες έδειξαν ότι α) οι πιθανότητες αναγνώρισης κατά την κανονική περίοδο ήταν μεγαλύτερες από αυτές κατά την περίοδο υποτροπής ($U = 4102847.5, p < 0.001$), β) οι πιθανότητες κατά τη διάρκεια της κανονικής περιόδου ήταν μεγαλύτερες από αυτές κατά την περίοδο πριν από την υποτροπή ($U = 2413996.5, p < 0.001$), γ) οι πιθανότητες πριν την υποτροπή ήταν μεγαλύτερες από αυτές κατά την περίοδο υποτροπής ($U = 1578681, p < 0.001$).

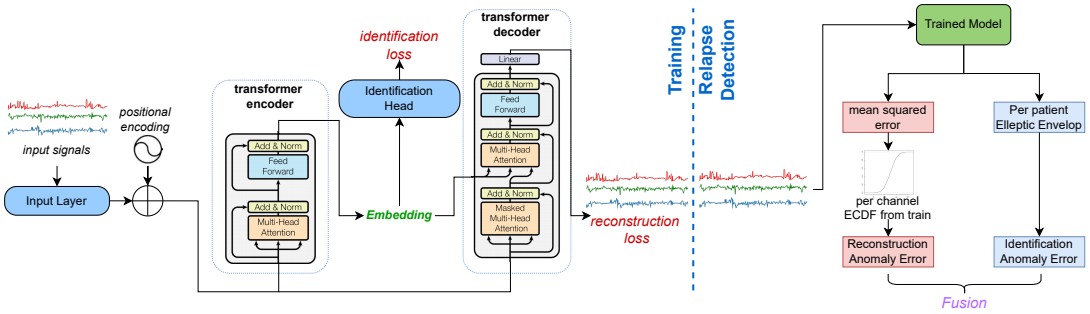


Σχήμα 6.6: Εμπειρική αθροιστική κατανομή πιθανοτήτων (eCDF) των τιμών ταυτοποίησης κατά τη διάρκεια των περιόδων ύφεσης, των περιόδων πριν την υποτροπή και των περιόδων υποτροπής. Όπως φαίνεται, οι βαθμολογίες κατά τις περιόδους υποτροπής και πριν της υποτροπής λαμβάνουν χαμηλότερες τιμές πιο συχνά, σε σύγκριση με τις περιόδους ύφεσης.

Η παραπάνω μελέτη φαίνεται πως ανοίγει το δρόμο προς τη συσχέτιση των ψυχωτικών διαταραχών με την αναγνώριση φαινοτύπου, βασιζόμενη στην αντίληψη ότι οι ψυχωτικές υποτροπές αλλάζουν το καθημερινό προφίλ συμπεριφοράς του χρήστη και έτσι δημιουργούν σύγχυση σε ένα σύστημα ταυτοποίησης. Η εκτεταμένη πειραματική μας ανάλυση επαλήθευσε αυτή την έννοια, καθώς ανακαλύψαμε σημαντικές αλλαγές στην κατανομή των τιμών εξόδου των δικτύων κατά την περίοδο υποτροπής και προ-υποτροπής, καθώς και πτώση στο ποσοστό επιτυχούς ταξινόμησης του δικτύου.

6.2 Αναγνώριση Ψυχωτικών Υποτροπών με Συνδυασμό Ανίχνευσης Ανωμαλιών και Ταυτοποίησης Χρήστη

Σε αυτή την ενότητα παρουσιάζουμε ένα νέο σύστημα που ενισχύει τις παραδοσιακές αρχιτεκτονικές των αυτόματων κωδικοποιητών (autoencoders) που χρησιμοποιούνται για την ανί-



Σχήμα 6.7: Επισκόπηση προτεινόμενου συστήματος. Ένας transformer autoencoder εκπαιδεύεται αρχικά με χρήση συνάρτησης σφάλματος για την ανακατασκευή και αναγνώρισης (identification and reconstruction loss). Στο επίπεδο της αξιολόγησης (relapse detection), το τελικό σκορ ανωμαλίας υπολογίζεται χρησιμοποιώντας το σφάλμα ανακατασκευής eCDF και το σκορ αναγνώρισης όπως προκύπτει από την ελλειπτική περιβάλλουσα κάθε χρήστη (elliptic envelop).

χνευση ανωμαλιών. Πιο συγκεκριμένα, προτείνουμε την ενσωμάτωση του εντοπισμό υποτροπών που βασίζεται σε αυτόματο κωδικοποιητή με μεθόδους ταυτοποίησης χρήστη [Efthymiou et al.; 2023, Retsinas et al.; 2020]. Συνοπτικά, το σύστημα παρουσιάζεται στο Σχήμα 6.7. Στις επόμενες ενότητες:

1. Επεκτείνουμε τις κλασικές αρχιτεκτονικές των αυτόματων κωδικοποιητών για ανίχνευση ανωμαλιών (anomaly detection) αξιοποιώντας επιπρόσθετα στοιχεία αναγνώρισης ασθενών υλοποιώντας μια καθολική αρχιτεκτονική για όλους τους ασθενείς. Εκπαιδεύουμε το σύστημα έχοντας ως ενιαίο στόχο την εκμάθηση της ανακατασκευής σημάτων και της ταυτοποίησης ασθενών. Αυτή η επιλογή έχει ως αποτέλεσμα την ουσιαστική βελτίωση στην ανίχνευση υποτροπών, ακόμη και όταν δεν λαμβάνονται υπόψη, σε επίπεδο αξιολόγησης, οι προβλέψεις του δικτύου για την ταυτοποίηση του χρήστη.

2. Αξιοποιούμε τις αναπαραστάσεις των χαρακτηριστικών που ήδη έχει μάθει το δίκτυο και εκπαιδεύουμε ειδικά-ατομικά μοντέλα για να κατασκευάσουμε ένα *σφάλμα ανωμαλίας ταυτοποίησης (identification anomaly error)* που βελτιώνει περαιτέρω την ανίχνευση των υποτροπών σε σύγκριση με το κλασικό *σφάλμα ανωμαλίας ανακατασκευής (reconstruction anomaly error).*

3. Τέλος, εξερευνούμε τη σύμμειξη (fusion) των δύο σφαλμάτων *(identification anomaly error, recontruction anomaly error)* κατασκευάζοντας ένα μοναδικό σφάλμα *από κοινού σφάλμα (joint anomaly error)* που επιτυγχάνει σημαντικά υψηλότερη απόδοση εντοπισμού των υποτροπών σε σύγκριση με τη χρήση κάθε σφάλματος ξεχωριστά.

6.2.1 Σύνολο δεδομένων

Στην παρούσα ενότητα χρησιμοποιούμε για την εκπαίδευση και την αξιολόγηση του συστήματός μας το σύνολο δεδομένων του Track 2 του e-Prevention Grand Challenge I [SPGC e-Prevention I; 2023] με σκοπό την άμεση σύγκριση με άλλες εργασίες που δημοσιεύτηκαν πρόσφατα. Το σύνολο αυτό περιλαμβάνει βιοσήματα έξυπνων ρολογιών που φορούσαν 10 ασθενείς με ψυχωτικές διαταραχές, περίπου για διάρκεια 6 μηνών, και όλοι παρουσίασαν τουλάχιστον μια ψυχωτική υποτροπή. Τα εγγεγραμμένα σήματα από το ρολόι του κάθε χρήστη χωρίζονται

Patients	Train Set	Validation Set		Test Set	
		Non-Relapse	Relapse	Non-Relapse	Relapse
C0	248	31	9	31	10
C1	179	22	57	23	57
C2	204	25	13	26	13
C3	168	21	17	21	17
C4	176	22	3	23	4
C5	217	27	22	28	22
C6	210	26	4	27	5
C7	230	29	93	29	94
C8	105	13	3	14	4
C9	169	21	73	22	74
Total	1906	237	294	244	300

Πίνακας 6.5: Στατιστικά στοιχεία των δεδομένων του συνόλου (Track 2, e-Prevention Grand Challenge I) που χρησιμοποιείται στην παρούσα ενότητα, ως προς τις ημέρες των συνόλων εκπαίδευσης, επικύρωσης και ελέγχου ανά ασθενή.

σε ξεχωριστές ημέρες και τα δεδομένα για κάθε ημέρα περιλαμβάνουν συνεχή σήματα γραμμική επιτάχυνση (από το επιταχυνσιόμετρο), γωνιακή ταχύτητα (από το γυροσκόπιο), καρδιακός ρυθμός και RR-interval (από φωτοπληθυσμογραφία - PPG). Οι τιμές των σημάτων δίνονται ως η μέση τιμή των καταγραφών διαστημάτων των 5 δευτερολέπτων, με σκοπό να μετριαστεί η επίδραση του θορύβου κάθε αισθητήρα στις εργασίες ταξινόμησης, όπως επισημαίνεται στο [Retsinas et al.; 2020]. Το τελικό σύνολο εκπαίδευσης του συνόλου δεδομένων περιλαμβάνει δεδομένα που αποκτήθηκαν μόνο όταν οι ασθενείς ήταν σε φάση ύφεσης, ενώ το σύνολο επικύρωσης και το σύνολο ελέγχου περιλαμβάνουν μέρες ύφεσης αλλά και υποτροπής της ασθένειας, περισσότερα στοιχεία για τα δεδομένα του συνόλου παρουσιάζονται στον Πίνακα 6.5.

6.2.2 Αρχιτεκτονική Συστήματος

Ορμώμενοι από δύο διαφορετικές αλλά αποτελεσματικές, προσεγγίσεις, δηλαδή την ανίχνευση υποτροπής ως πρόβλεψη ανωμαλίας [Calcagno et al.; 2023, Panagiotou et al.; 2022] και την ανίχνευση υποτροπής ως αστοχία ταυτοποίησης του ασθενή [Efthymiou et al.; 2023], προτείνουμε έναν απρόσκοπτο τρόπο συνδυασμού τους σε ένα ενιαίο αποτελεσματικό πλαίσιο. Αυτός ο συνδυασμός είναι πολύπλευρος: συμπεριλαμβάνουμε και τα δύο σχήματα στη φάση της εκπαίδευσης, με δύο ξεχωριστές συναρτήσεις κόστους που σχηματίζουν μια απώλεια πολλαπλών εργασιών, καθώς και κατά την αξιολόγηση, με μια διερευνημένη σειρά πιθανών προσεγγίσεων σύντηξης. Η αρχιτεκτονική του συστήματός μας εμπνέεται από το [Calcagno et al.; 2023], το οποίο πέτυχε τα κορυφαία αποτελέσματα κατά τη διάρκεια του διαγωνισμού ICASSP2023 e-Prevention, και το [Panagiotou et al.; 2022] που αξιοποίησε πρώτο τους transformer autoencoders για την αναγνώριση των υποτροπών. Στις εργασίες αυτές, η ανίχνευση υποτροπής χαρακτηρίζεται ως ανίχνευση ανωμαλίας. Πιο συγκεκριμένα, οι Calcagno et al. [Calcagno et al.; 2023], υλοποιούν μια αρχιτεκτονική που βασίζεται σε αυτόματο κωδικοποιητή και εκπαιδεύεται σε δεδομένα κατά τη διάρκεια περιόδων ύφεσης. Μετά την εκπαίδευση, υπολογίζεται μια εμπειρική συνάρτηση κατανομής (eCDF) για το σφάλμα ανακατασκευής στα δεδομένα εκπαίδευσης. Τέλος, κατά το χρόνο συμπερασμάτων, η τιμή της ECDF για το σφάλμα ανακατασκευής χρησιμοποιείται ως σκορ ανωμαλίας.

Αναλυτικότερα, ως προς την αρχιτεκτονική του συστήματος (6.7), το μοντέλο μας είναι ένας

transformer autoencoder με τα ακόλουθα χαρακτηριστικά:

- Οι είσοδοί του προβάλλονται σε ένα embedding χώρο μέσω ενός γραμμικού επιπέδου.
- Αξιοποιείται κωδικοποίηση θέσης, δηλαδή temporal encoding όπως είδαμε και στο 5.4.2, για να αποτυπωθεί η χρονική συσχέτιση των δειγμάτων.
- Και οι δύο υπομονάδες κωδικοποίησης (encoder) και αποκωδικοποίησης (decoder), δηλαδή τα βασικά στοιχεία αυτού του συστήματος αυτόματου κωδικοποιητή, είναι μετασχηματιστές πολλαπλών επιπέδων και πολλαπλών κεφαλών (multi-layer, multi-headed). Συγκεκριμένα, ο decoder λαμβάνει ως είσοδο τόσο την αρχική είσοδο του κωδικοποιητή, όσο και τις ενδιάμεσες αναπαραστάσεις (embeddings) που προκύπτουν από τον κωδικοποιητή. Στην ενότητα 6.2.5 παρουσιάζεται ο πειραματισμός που έγινε σχετικά με τις επιλογές των υπερπαραμέτρων του μετασχηματιστή.
- Στο τέλος, εφαρμόζεται γραμμικός μετασχηματισμός της εξόδου του αποκωδικοποιητή που οδηγεί σε μια ακολουθία χαρακτηριστικών του ίδιου μεγέθους με την είσοδο.

Στις προηγούμενες εργασίες μας [Efthymiou et al.; 2023, Panagiotou et al.; 2022] καθώς και στην [Calcagno et al.; 2023], εξετάστηκαν πολλές οικογένειες νευρωνικών δικτύων. Στην τελευταία επιλέχθηκε το καλύτερο μοντέλο ανά ασθενή μέσω της απόδοσής του στο σύνολο επικύρωσης, με την πλειοψηφία των μοντέλων να βασίζονται σε transformer αρχιτεκτονική. Σε αντίθεση με τα παραπάνω, σε αυτή την εργασία, εστιάζουμε μόνο στην αρχιτεκτονική με transformer autoencoders, με στόχο να αναδείξουμε την εξέχουσα θέση του, αποφεύγοντας παράλληλα την πιο δαπανηρή λύση πολλαπλών εξατομικευμένων δικτύων.

Προτείνουμε ένα «καθολικό» μοντέλο ικανό να προβλέπει ανωμαλίες σε διαφορετικά άτομα. Το κίνητρο μιας τέτοιας επιλογής είναι διπλό:

- η δημιουργία μοτίβων διαφορετικών ασθενών για την ανίχνευση ανωμαλιών. Ενδέχεται να μπορούμε να εξάγουμε συμπεράσματα για πιθανή υποτροπή από τα δεδομένα άλλων ασθενών, και
- η αποτελεσματικότητα αποθήκευσης και χρήσης. Έχουμε μόνο ένα μόνο μοντέλο προς εκπαίδευση και αποθήκευση για όλους τους ασθενείς.

6.2.3 Ενσωμάτωση της Ταυτοποίησης Ατόμων

Έχοντας περιγράψει το σύστημα ανίχνευσης ανωμαλιών, που λειτουργεί ως η ραχοκοκαλιά της προτεινόμενης εργασίας, προχωράμε ενσωματώνοντας την έννοια της ταξινόμησης ατόμου/φαινοτύπου στο υπάρχον πλαίσιο, ακολουθώντας την επιτυχία των προηγούμενων εργασιών μας [Retsinas et al.; 2020, Efthymiou et al.; 2023]. Αρχικά, πρέπει να εκπαιδεύσουμε το υπάρχον δίκτυο ώστε να είναι σε θέση να αναγνωρίζει τους διαφορετικούς ασθενείς. Ο απλούστερος τρόπος για να γίνει αυτό είναι συσχετίζοντας την ταυτότητα του ατόμου με τις ενδιάμεσες αναπαραστάσεις του encoder, ο οποίος κωδικοποιεί όλες τις πιθανές πληροφορίες πριν από την ανακατασκευή του αρχικού σήματος. Για το σκοπό αυτό, επαυξάνουμε την αρχική αρχιτεκτονική του autoencoder με ένα επιπλέον identity head που προβλέπει για κάθε δείγμα την πιθανότητα να ανήκει σε έναν συγκεκριμένο ασθενή του συνόλου δεδομένων. Αυτή η επιπλέον εργασία ταξινόμησης μπορεί να πραγματοποιηθεί με ένα απλό cross-entropy loss, που λειτουργεί πρακτικά ως βοηθητικός όρος στη συνολική συνάρτηση κόστους. Στη συνέχεια, μια επιπλέον απώλεια cross entropy υπολογίζεται από την πιθανότητα και την πραγματική ετικέτα του ατόμου. Συνολικά, ο autoencoder εκπαιδεύεται πλέον με το ακόλουθο κριτήριο:

$$\mathcal{L}(\mathbf{x}, y) = \mathcal{L}_{MSE}(d(e(\mathbf{x})), \mathbf{x}) + \lambda \mathcal{L}_{CE}(c(e(\mathbf{x})), y) \tag{6.1}$$

, όπου **x** είναι το σήμα εισόδου, y η ταυτότητα του ατόμου, ενώ οι e, d και c είναι συναρτήσεις, που υλοποιούνται ως νευρωνικά δίκτυα, και αντιπροσωπεύουν τον κωδικοποιητή, τον αποκωδικοποιητή και την κεφαλή ταξινόμησης, αντίστοιχα. Επιπλέον, $\mathcal{L}_{MSE}$ είναι το μέσο τετραγωνικό σφάλμα του κόστους (mean squared error loss), $\mathcal{L}_{CE}$ το cross-entropy loss και λ μια υπερπαράμετρος που ελέγχει τη συμβολή του επιπλέον σφάλματος ταξινόμησης και επομένως ρυθμίζει την ισορροπία μεταξύ των δύο συναρτήσεων κόστους, δηλαδή τη συμβολή των δύο στρατηγικών για τον εντοπισμό της υποτροπής.

Εκτός από την προφανή πτυχή του τελικού στόχου, η κοινή εκπαίδευση της ανακατασκευής σήματος και της ταυτοποίησης ατόμων θα μπορούσε να βελτιώσει και τις δύο επιμέρους εργασίες, τονίζοντας τη συσχέτισή τους και προωθώντας περαιτέρω τη δημιουργία πολύ αποτελεσματικών ενδιάμεσων αναπαραστάσεων από τον κωδικοποιητή.

6.2.4 Κριτήρια Αξιολόγησης

Η ανάπτυξη ενός συστήματος δύο ταυτόχρονων εργασιών οδηγεί στις ακόλουθες ερωτήσεις: «ποιο κριτήριο να χρησιμοποιήσουμε για την αξιολόγηση;» και «πώς μπορούμε να συνδυάσουμε και τις δύο μετρήσεις που σχετίζονται με την εργασία;».

Αρχικά, δεδομένου ότι η ανίχνευση υποτροπής γίνεται ανά ημέρα, για να δημιουργήσουμε τις αντίστοιχες βαθμολογίες ανωμαλιών, πρώτα εξάγουμε ακολουθίες 10 χαρακτηριστικών από διαστήματα δεδομένων των πέντε λεπτών και τις συνδυάζουμε σε παράθυρα τεσσάρων ωρών. Χρησιμοποιούμε όλα τα διαθέσιμα παράθυρα διάρκειας τεσσάρων ωρών σε επικαλυπτόμενα διαστήματα τριών ωρών για να λάβουμε μια ισχυρή αναπαράσταση.

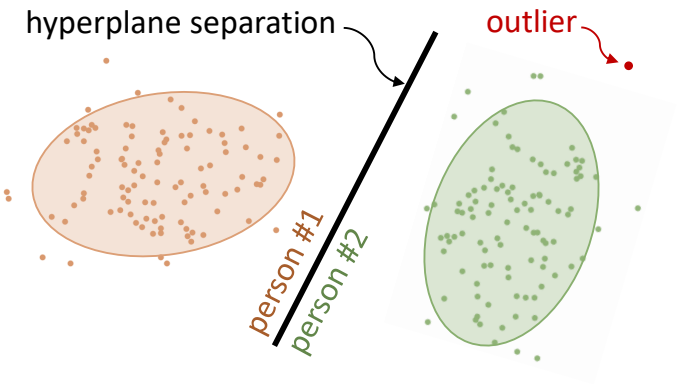
Ακολουθώντας το σκεπτικό του εντοπισμού ανωμαλιών, ως σκορ ανωμαλίας χρησιμοποιείται η τιμή που προκύπτει από την κανονικοποίηση του σφάλματος ανακατασκευής κάθε καναλιού πληροφορίας μέσω κατανομών ECDF που προκύπτουν από όλα τα δεδομένα εκπαίδευσης. Από την άλλη πλευρά, ακολουθώντας το σκεπτικό της ταξινόμησης φαινοτύπων, ένας απλός τρόπος για να οριστεί ένα σκορ ανωμαλίας είναι να χρησιμοποιηθεί απευθείας η πρόβλεψη της πιθανότητας ταξινόμησης p του αντίστοιχου χρήστη για κάθε δείγμα (στην πραγματικότητα, το $1 - p$ της πιθανότητας αυτής). Πιο απλά, η μείωση της πιθανότητας να συσχετίζεται με ισχυρή ένδειξη για ύπαρξη πιθανής ανωμαλίας. Ωστόσο, ο προκαταρκτικός πειραματισμός μας έδειξε ότι αυτή η μέτρηση οδήγησε σε χαμηλές επιδόσεις και δεν ήταν καλός δείκτης για τις μεταβολές στη συμπεριφορά των χρηστών.

Αυτή η μείωση της απόδοσης μπορεί να εξηγηθεί αν σκεφτούμε τον χώρο ταξινόμησης, όπου οι τάξεις διαχωρίζονται με υπερεπίπεδα, όπως απεικονίζεται στο Σχήμα 6.8. Σύμφωνα με αυτήν την υπόθεση, η εισαγωγή μιας ακραίας τιμής, με την έννοια ότι η νέα πρόβλεψη δεν είναι κοντά σε μια υπάρχουσα, μπορεί να ταξινομηθεί πιστά σε μια συγκεκριμένη κατηγορία με υψηλή πιθανότητα, ακολουθώντας αυτή τη λογική διαχωρισμού υπέρ-επίπεδου. Από την άλλη πλευρά, εάν μοντελοποιήσουμε την κατανομή κάθε κλάσης, μέσω της εκπαίδευσης του προτεινόμενου συστήματος, μπορούμε ενδεχομένως να ανιχνεύσουμε ανωμαλίες που αντιστοιχούν σε αλλαγές συμπεριφοράς. Για το σκοπό αυτό προτείναμε την ακόλουθη διαδικασία:

1. εκπαιδεύουμε επιπρόσθετα για κάθε χρήστη μια ελλειπτική περιβάλλουσα (elliptic envelop) σε χαρακτηριστικά που εξάγονται από τον κωδικοποιητή για κάθε δείγμα.

2. Στη συνέχεια, λαμβάνουμε τη βαθμολογία που προκύπτει από το elliptic envelop για κάθε δείγμα ως *identification anomaly error*.

Αυτή η παραλλαγή θα μπορούσαμε να πούμε πως είναι μια μετα-επεξεργασία που προσαρμόζει ελαφρά το καθολικό μοντέλο στα ατομικά δεδομένα των ασθενών. Στη συνέχεια θα δούμε πως αυτή η μέθοδος ενίσχυσε τα αποτελέσματα ανίχνευσης υποτροπής.



Σχήμα 6.8: Παράδειγμα της αδυναμίας του ταξινομητή να αντιληφθεί ένα ακραίο στοιχείο (outlier) ως ανωμαλία. Εδώ, εικονίζεται το υπερεπίπεδο που χωρίζει τις δύο κλάσεις - ασθενείς. Όταν εισάγεται ένα ακραίο στοιχείο, το οποίο μπορεί να υποδεικνύει μια ξαφνική αλλαγή στη συμπεριφορά, ταξινομείται με μεγάλη πιθανότητα να είναι το άτομο #1, όπως και συμβαίνει στην πραγματικότητα. Ωστόσο, εάν δεν ληφθεί υπόψη ένα μοντέλο της κατανομής των χαρακτηριστικών του χρήστη, αυτή η λογική δεν μπορεί να βοηθήσει το μοντέλο να αντιληφθεί πιθανές κρίσιμες ακραίες τιμές.

Όπως είδαμε παραπάνω, το προτεινόμενο πλαίσιο καταλήγει σε δύο διακριτές βαθμολογίες ανωμαλιών. Για να προωθήσουμε περαιτέρω την αποτελεσματικότητα της μεθόδου μας, εξετάζουμε διαφορετικές τεχνικές για τη συγχώνευση των σφαλμάτων ανωμαλίας ανακατασκευής και αναγνώρισης (reconstruction anomaly error and the identification anomaly error). Έτσι εξετάσαμε τη σύμμειξη των βαθμολογιών μέσω:

1. ενός απλού μη γραμμικού συνδυασμού τους (γινόμενο),
2. προσαρμογή ενός μη γραμμικού μοντέλου στις δύο βαθμολογίες,
3. προσαρμογή ενός μη γραμμικού μοντέλου στη βαθμολογία ανωμαλίας ανακατασκευής ανά κανάλι πληροφορίας και τη βαθμολογία ταυτοποίησης.

6.2.5 Πειραματική Ανάλυση

Πειραματική διάταξη (setup)

Τα μοντέλα μας εκπαιδεύονται και αξιολογούνται χρησιμοποιώντας ως διανύσματα, ακολουθίες εισόδου 10 χαρακτηριστικών που εξάγονται από πεντάλεπτα διαστήματα δεδομένων. Αυτά τα χαρακτηριστικά είναι η μέση τιμή του μέτρου της γραμμικής και γωνιακής επιτάχυνσης, η μέση τιμή των RR-intervals και του καρδιακού ρυθμού (HRV), ο κύριος άξονας της έλλειψης Poincare, οι κανονικοποιημένες χαμηλές και υψηλές δυνάμεις του περιοδογραφήματος Lomb-Scargle, η χρονική κωδικοποίηση του χρόνου εγγραφής και το ποσοστό των έγκυρων δειγμάτων στα δεδομένα των 5 λεπτών. Κατά τη διάρκεια της αξιολόγησης, επειδή κάθε μέρα έχει διαφορετικό αριθμό παραθύρων, χρησιμοποιούμε τη μέση βαθμολογία ανωμαλίας για την τελική βαθμολογία ανωμαλίας ανά ημέρα. Για κάθε πείραμά μας, εκπαιδεύουμε την εκάστοτε εκδοχή του συστήματος τρεις φορές για 80 εποχές στο σύνολο δεδομένων εκπαίδευσης των ημερών χωρίς υποτροπή, και χρησιμοποιούμε το σύνολο επικύρωσης για να επιλέξουμε το καλύτερο μοντέλο. Ως μετρικές της ανίχνευσης της υποτροπής χρησιμοποιούμε το εμβαδόν κάτω από την καμπύλη λειτουργικού χαρακτηριστικού δέκτη (Receiver Operating Characteristic -

Μοντέλο	Size	Heads	Layers	AUROC	AUPRC	Mean
Βασικό Μοντέλο	32	8	2	**0.6254**	0.6343	**0.6299**
	32	8	4	0.6206	0.6285	0.6246
	32	16	4	0.6212	**0.6368**	0.6290
	64	8	2	0.6224	0.6319	0.6271
	128	8	2	0.6221	0.6296	0.6259
Επαυξημένο Μοντέλο με Συνάρτηση Κόστους Ταυτοποίησης	32	8	2	**0.6469**	**0.6565**	**0.6517**
	32	8	4	0.6223	0.6312	0.6267
	32	16	4	0.6205	0.6300	0.6252
	64	8	2	0.6290	0.6456	0.6373
	128	8	2	0.6217	0.6305	0.6261

Πίνακας 6.6: Πειραματική μελέτη για διάφορες αρχιτεκτονικές transformers. Η αξιολόγηση που παρουσιάζεται γίνεται στο validation set με χρήση του μέσου τετραγωνικού σφάλματος ανακατασκευής (Mean Square Reconstruction Error).

AUROC), το εμβαδόν κάτω από την καμπύλη ακριβείας-ανάκλησης (Area Under the Precision-Recall Curve - AUPRC) και τον αρμονικό μέσο όρο τους (Mean). Στους παρακάτω πίνακες φαίνεται ο μέσος όρος των μετρικών αυτών για τις τρεις επαναλήψεις εκπαίδευσης και αξιολόγησης.

Μελέτη Αρχιτεκτονικής

Αρχικά, εξετάσαμε την επίδραση ενός ενιαίου μοντέλου για όλους τους χρήστες με τη χρήση της προτεινόμενης συνάρτησης κόστους ταυτοποίησης (identification loss) και χωρίς αυτή, κάτω από διαφορετικές αρχιτεκτονικές επιλογές (hiden size of the model - encoder/decoder, #heads, #layers). Η αξιολόγηση πραγματοποιείται χρησιμοποιώντας μόνο το σφάλμα ανακατασκευής και τα αποτελέσματα παρουσιάζονται στον Πίνακα 6.6. Ως βασικό μοντέλο αναφέρεται πρακτικά η εκδοχή του συστήματος εντοπισμού ανωμαλιών (όπως παρουσιάστηκε στο [Calcagno et al.; 2023] αλλά εκπαιδεύοντας μόνο ένα μοντέλο για όλους τους χρήστες) με χρήση αποκλειστικά του $\mathcal{L}_{MSE}$ ως συνάρτηση κόστους. Το μοντέλο που αναφέρεται ως επαυξημένο μοντέλο αναφέρεται αυτό που για την εκπαίδευσή του συνυπολογίζει τα δύο losses (ανακατασκευής και ταυτοποίησης) σύμφωνα με την εξίσωση 6.1. Παρατηρούμε ότι η εισαγωγή της συνάρτησης κόστους για την ταυτοποίση του χρήστη, παρέχει μια σταθερή αύξηση της πρόβλεψης της υποτροπής. Το καλύτερο μοντέλο (32 hidden size / 8 #heads / 2 # layers) ξεπερνά τα υπόλοιπα βασικά μοντέλα και επιτυγχάνει αποτελέσματα συγκρίσιμα με τη νικητήρια μέθοδο του e-Prevention Challenge [Calcagno et al.; 2023].

Στρατηγικές αξιολόγησης & επιλογές μετρικών

Στη συνέχεια διερευνούμε την απόδοση του σκορ που προκύπτει από το identification error στην ανίχνευση υποτροπής, καθώς και τη διαμόρφωση μιας βαθμολογίας που προκύπτει και από τα δύο σκορ ανωμαλίας. Ο Πίνακας 6.7 παρουσιάζει αποτελέσματα ανίχνευσης υποτροπής με βάση τρεις διαφορετικές βαθμολογίες ανωμαλίας που προκύπτουν από:

1. το *σφάλμα ανακατασκευής (reconstruction anomaly error)*,
2. το *σφάλμα αναγνώρισης (identification anomaly error)*
3. το *γινόμενο* τους (product - fusion).

		Reconstruction Error			Identification Error			Product Error		
Set	Best Model	AUROC	AUPRC	Mean	AUROC	AUPRC	Mean	AUROC	AUPRC	Mean
Valid.	Reconstruction	0.6129	0.6121	0.6125	0.5688	0.5772	0.5730	0.6061	0.6205	0.6133
	Identification	0.5494	0.5632	0.5563	0.5922	0.6039	0.5981	0.5920	0.6111	0.6016
	Product	0.5808	0.5874	0.5841	0.6007	0.6074	0.6040	**0.6229**	**0.6396**	**0.6312**
Test	Reconstruction	0.6449	0.6504	0.6477	0.6195	0.6184	0.6190	0.6570	0.6516	0.6543
	Identification	0.5816	0.5981	0.5899	0.6537	0.6549	0.6543	0.6520	0.6581	0.6550
	Product	0.5976	0.6082	0.6029	0.6545	0.6587	0.6566	**0.6637**	**0.6692**	**0.6649**

Πίνακας 6.7: Μετρικές για την ανίχνευση υποτροπών (γραμμές) όταν επιλέγουμε ως τελική βαθμολογία ανωμαλίας: 1) το σφάλμα ανακατασκευής, 2) το σφάλμα αναγνώρισης, 3) το γινόμενό τους. Κάθε υπερστήλη δείχνει σύμφωνα με ποιο κριτήριο επιλέχθηκε το τελικό μοντέλο, σύμφωνα με το validation set.

Είναι σημαντικό ότι κατά τη φάση εκπαίδευσης, η επιλογή του τελικού μοντέλου μας επηρεάζεται από την απόδοσή του στην ανίχνευση της υποτροπής στο σύνολο επικύρωσης. Έτσι έχουμε τρεις διαφορετικούς τρόπους να υπολογίσουμε τη μετρική για την ανίχνευση των υποτροπών. Επομένως, δεν είναι δίκαιο να επιλέγουμε το καλύτερο μοντέλο επικύρωσης σύμφωνα με μια συγκεκριμένη βαθμολογία και στη συνέχεια να αξιολογούμε το σύστημά μας χρησιμοποιώντας μια άλλη επιλογή βαθμολόγησης. Για το σκοπό αυτό, ανάλογα με το κριτήριο που χρησιμοποιείται, επιλέχθηκαν τρία διαφορετικά μοντέλα από το σύνολο επικύρωσης και αξιολογήθηκαν με βάση και τις τρεις βαθμολογίες. Όπως παρατηρούμε, η χρήση του προτεινόμενου σφάλματος ταυτοποίησης έχει ως αποτέλεσμα υψηλότερη απόδοση σε σύγκριση με τη χρήση του σφάλματος ανακατασκευής. Επιπλέον, τόσο για το σύνολο επικύρωσης όσο και για το σύνολο δοκιμής, η χρήση του πολλαπλασιαστικού κριτηρίου βελτιώνει περαιτέρω το αποτέλεσμα σχεδόν σε κάθε περίπτωση. Συνεπώς αντιλαμβανόμαστε πως η χρήση και των δύο losses κατά τη διάρκεια της εκπαίδευσης του συστήματος αλλά και κάποιου συνδυασμού των μετρικών του σφάλματος ανακατασκευής και του σφάλματος ταυτοποίησης επιφέρουν βελτίωση των αποτελεσμάτων του συστήματος.

Εποπτευόμενη Σύμμειξη

Ως τώρα επιλέξαμε να ακολουθήσουμε τεχνικές μη επιβλεπόμενης μάθησης ώστε να μπορέσουμε να συγκριθούμε με τα αποτελέσματα του e-Prevention challenge. Για την προσέγγιση όμως του προβλήματος του εντοπισμού των ψυχωτικών υποτροπών πηγαίνουμε ένα βήμα πιο πέρα και διερευνούμε μια εποπτευόμενη σύμμειξη των βαθμολογιών ανωμαλίας, εκτελώντας παλινδρόμηση κορυφογραμμής (ridge regression) με μη αρνητικούς συντελεστές στις βαθμολογίες ανωμαλιών του συνόλου επικύρωσης και των αντίστοιχων label, και στη συνέχεια πραγματοποιούμε τη δοκιμή στο σύνολο ελέγχου. Η επιλογή των μη αρνητικών συντελεστών δικαιολογείται καθώς θέλουμε απλώς να υπολογίσουμε κάποια βάρη μεταξύ των μετρικών ώστε να επιτύχουμε έναν καλύτερο συνδυασμό μεταξύ τους και να μπορούμε να δούμε τη συνεισφορά των εκάστοτε μετρικών. Πιο συγκεκριμένα, εξετάζουμε δυο διαφορετικές εκδοχές:

1. ένα μη γραμμικό μοντέλο για τις δύο βαθμολογίες ανακατασκευής και ταυτοποίησης (score regression),

2. ένα μη γραμμικό μοντέλο για τις βαθμολογίες ανακατασκευής του κάθε καναλιού πληροφορίας/χαρακτηριστικού και της βαθμολογίας ταυτοποίησης (score regression per channel).

Και οι δύο αυτές εκδοχές αξιοποιούν τα δεδομένο του συνόλου επικύρωσης στο σύνολό τους χωρίς να εισερχόμαστε σε διαχωρισμό των δεδομένων από ασθενή σε ασθενή.

Models	AUROC	AUPRC	Mean
Score regression	0.6094	0.7698	0.6896
Score regression per channel	0.6650	0.7873	0.7262

Πίνακας 6.8: Σύγκριση αποτελεσμάτων στο σύνολο ελέγχου για τις περιπτώσεις επιβλεπόμενης σύμμειξης των βαθμολογιών ανωμαλίας με χρήση ridge regression.

Τα αποτελέσματά μας, που εμφανίζονται στον Πίνακα 6.8 δείχνουν ότι η χρήση παλινδρόμησης για την εύρεση ενός συνδυασμού των βαθμολογιών ανωμαλίας, από αυτή του απλού γινομένου, δίνει καλύτερα αποτελέσματα. Ακόμα περισσότερη βελτίωση παρατηρούμε όταν μέσω της παλινδρόμησης κορυφογραμμής συσχετίζουμε τις αναπαραστάσεις που προκύπτουν από κάθε κανάλι πληροφορίας, δηλαδή κάθε χαρακτηριστικού ξεχωριστά.

Ακόμα, στον Πίνακα 6.9 παρουσιάζεται η σύγκριση με τις άλλες μεθόδους SoTA στο σύνολο δεδομένων e-Prevention Challenge I. Όπως παρατηρούμε η ταυτοποίηση του φαινοτύπου του χρήστη βοηθά το αρχικό μοντέλο στην καλύτερη αναγνώριση της κατάστασης του χρήση (ύφεση ή υποτροπή). Ακόμα περισσότερο, η αξιοποίηση των label του συνόλου επικύρωσης και η χρήση παλινδρόμησης κορυφογραμμής συνεισφέρει σημαντικά στη βελτίωση των μοντέλων διαμορφώνοντας μια σχέση μεταξύ των καναλιών πληροφορίας και της κατάστασης του ασθενούς.

Methods	AUROC	AUPRC	Mean
Avramidis et al. [Avramidis et al.; 2023]	0.5839	0.6263	0.6051
Hamieh et al. [Hamieh et al.; 2023]	0.6072	0.6347	0.6209
Calcagno et al. [Calcagno et al.; 2023]	0.6469	0.6509	0.6489
Product (Ours)	**0.6637**	**0.6692**	**0.6649**
Score regression per channel (Ours)	**0.6650**	**0.7873**	**0.7262**

Πίνακας 6.9: Τελικά συγκριτικά αποτελέσματα στο σύνολο e-Prevention Challenge I.

Τέλος, για καλύτερη κατανόηση στη συνεισφορά του συγκεκριμένου μοντέλου έναντι του προηγούμενου καλύτερου, παρουσιάζουμε στον Πίνακα 6.10 τις τιμές των μετρικών για κάθε ασθενή ξεχωριστά, για την εργασία μας και την εργασία των Calcagno et al. Σύμφωνα με τα αποτελέσματα στην περίπτωση των ασθενών C2 και C4, το ατομικό μοντέλο των [Calcagno et al.; 2023] αδυνατεί να αναγνωρίσει την κατάσταση του ασθενούς σε αντίθεση με το μοντέλο μας που καταφέρνει να αξιοποιήσει την πληροφορία που προκύπτει από τα δεδομένα και των υπολοίπων ασθενών και να αποδώσει ως ένα βαθμό. Αντίθετα στην περίπτωση των C5, C6 όπου η άλλη ομάδα επέλεξε μια εξατομικευμένη αρχιτεκτονική, διαφορετική από αυτή που αξιοποιήσαμε στο καθολικό μας μοντέλο, παρατηρούμε πως αποδίδει η επιλογή αυτή πετυχαίνοντας καλά αποτελέσματα. Τέλος, στην περίπτωση του ασθενή C8 βλέπουμε πως και οι δύο μέθοδοι αποτυγχάνουν στην αναγνώριση της κατάστασης του ασθενή και η περίπτωση αυτή χρήζει περαιτέρω διερεύνησης. Παρ' ότι δεν ήταν στόχος του διαγωνισμού η εκτίμηση της κατάστασης σε επίπεδο ασθενή, αλλά στο σύνολό τους στο σετ ελέγχου, ο Πίνακας αυτός μας δίνει μια καλή διαίσθηση για τη διαφορετικότητα των ατόμων ως προς την ψηφιακή τους ταυτότητα και πως αυτή μπορεί να διακριθεί από κάποιο καθολικό ή εξατομικευμένο μοντέλο.

Συμπερασματικά, στην ενότητα αυτή παρουσιάσαμε ένα νέο πλαίσιο για την ανίχνευση υποτροπής που συνδυάζει τις παραδοσιακές αρχιτεκτονικές αυτόματου κωδικοποιητή για την ανακατασκευή σημάτων και τη λογική της ταυτοποίησης των ατόμων μέσω των φαινοτύπων τους, που αναπτύξαμε προηγουμένως. Τα πειράματά μας έδειξαν ότι η διπλή προσέγγισή μας

	[Calcagno et al.; 2023]			[Efthymiou et al.;]		
Patients	AUROC	AUPRC	Mean	AUROC	AUPRC	Mean
C0	0.7516	0.5939	0.6728	0.7581	0.7027	**0.7303**
C1	0.7338	0.8408	**0.7873**	0.6396	0.8655	0.7525
C2	0.3758	0.2745	0.3251	0.5385	0.5473	**0.5429**
C3	0.8011	0.8114	**0.8063**	0.7568	0.7999	0.7778
C4	0.5761	0.1638	0.3699	0.6268	0.5335	**0.5801**
C5	0.6802	0.6608	**0.6705**	0.487	0.5859	0.5364
C6	0.6741	0.5170	**0.5956**	0.5888	0.4242	0.5010
C7	0.8107	0.9365	**0.8736**	0.7257	0.9217	0.8237
C8	0.5536	0.2269	0.3902	0.4048	0.3972	**0.4010**
C9	0.4945	0.7940	0.6442	0.6173	0.8958	**0.7565**

Πίνακας 6.10: Συγκριτικά αποτελέσματα για την προτεινόμενη μέθοδο και τη νικητήρια μέθοδο του e-Prevention Challenge I, για κάθε ασθενή.

αποδίδει καλύτερα από κάθε μέθοδο ξεχωριστά και ξεπερνά τις προηγούμενες μεθόδους. Σημαντικό είναι ακόμα να σημειώσουμε πως η προσέγγισή μας αφορά τη χρήση ενός συνολικού μοντέλου για όλους τους χρήστες και όχι εξατομικευμένα μοντέλα κάτι το οποίο είναι αρκετά σημαντικό καθώς βοηθά στην ύπαρξη ενός βασικού μοντέλου για την εκτίμηση των υποτροπών ακόμα και για ασθενείς με λίγα δεδομένα.

6.3 Συμπεράσματα Κεφαλαίου

Στόχος αυτού του Κεφαλαίου ήταν ο εντοπισμός ψυχωτικών υποτροπών με χρήση βιοσημάτων από ασθενείς με ψυχωτικές διαταραχές μέσα από τον εντοπισμό των αλλαγών στα μοτίβα των βιοσημάτων τους. Έτσι αναπτύξαμε και παρουσιάσαμε δύο διαφορετικά συστήματα αναγνώρισης των υποτροπών.

Αρχικά, μελετήσαμε τον εντοπισμό των ψυχωτικών υποτροπών ως αποτέλεσμα της μείωσης της ακρίβειας ταυτοποίησης ενός δικτύου χρονικής μοντελοποίησης που είναι εκπαιδευμένο στην ανίχνευση των προτύπων συμπεριφοράς του εκάστοτε ατόμου. Επεκτείνοντας το σύστημα που προτάθηκε στο προηγούμενο Κεφάλαιο, αντιμετωπίζουμε το πρόβλημα ως ένα πρόβλημα λανθασμένης ταξινόμησης των βιοσημάτων των ασθενών κατά τη διάρκεια της υποτροπής. Για τη μελέτη μας διακρίναμε τρεις περιόδους, αυτές της ύφεσης της διαταραχής, της υποτροπής της καθώς και μια περιόδου πριν από την υποτροπή με σκοπό να ερευνήσουμε για τυχόν ενδείξεις πριν τη διάγνωση των ιατρών για την υφιστάμενη υποτροπή. Έτσι πειραματιστήκαμε εκτενώς ως προς προσθήκη υψηλού επίπεδου χαρακτηριστικών ώστε να ενισχύσουμε την πληροφορία που δέχεται το σύστημα.

Αξιολογήσαμε το σύστημά μας τόσο μέσω της απλής όσο και της εξισορροπημένης ακρίβειας αλλά και ως προς τη μέση και τη διάμεση πιθανότητα ταξινόμησης των ασθενών ανά περίοδο. Ύστερα από εκτεταμένο πειραματισμό διαπιστώσαμε πως όλα τα πιθανά συμπληρωματικά χαρακτηριστικά (χρονική κωδικοποίηση, ύπνος και πληροφορίες ημέρας της εβδομάδας) συμβάλλουν στην ανίχνευση των υποτροπών ενώ βοήθησαν και στη διεύρυνση του χάσματος της ακρίβειας μεταξύ της φάσης ύφεσης και της υποτροπής. Στη συνέχεια μελετήσαμε εκτενώς την απόδοση του δικτύου σε επίπεδο χρήστη και σε επίπεδο περιόδων και διαπιστώσαμε πως έχουμε στατιστικά σημαντική πτώση της πιθανότητας από τις περιόδους ύφεσης σε αυτές της προ-υποτροπής και της υποτροπής για του περισσότερους ασθενείς ενώ γενικά δεν μπορούμε να διακρίνουμε σημαντική διαφορά ανάμεσα στις περιόδους πριν την υποτροπή και

κατά τη διάρκειά της. Τέλος, πραγματοποιήσαμε στατιστική ανάλυση των δεδομένων όλων των ασθενών για τις τρεις περιόδους και είδαμε ότι οι πιθανότητες αναγνώρισης κατά την κανονική περίοδο ήταν μεγαλύτερες από αυτές της περιόδου προ-υποτροπής και αυτές από της περιόδου υποτροπής.

Στο δεύτερο μέρος, αναπτύξαμε ένα σύστημα που βασίζεται σε transformer autoencoders και αντιμετωπίζουμε το πρόβλημα ως συνδυασμό της ανακατασκευής βιοσημάτων, υπό το πρίσμα της αναγνώρισης ανωμαλιών, και της ταυτοποίησης του φαινοτύπου των ασθενών. Σ' αυτή τη λογική, δημιουργήσαμε δύο σφάλματα εκτίμησης του δικτύου ένα για την ανακατασκευή των σημάτων και ένα για την ταυτοποίηση των χρηστών μέσω από την εκπαίδευση μιας ελλειπτικής περιβάλλουσας για κάθε ασθενή. Στη συνέχεια μελετήσαμε διαφορετικές τεχνικές σύμμειξης των επιμέρους σφαλμάτων καταλήγοντας ως βέλτιστου στην προσαρμογή ενός μη γραμμικού μοντέλου στη βαθμολογία ανωμαλίας ανακατασκευής ανά κανάλι πληροφορίας και τη βαθμολογία ταυτοποίησης με χρήση παλινδρόμηση κορυφογραμμής. Αξιολογήσαμε το σύστημά μας μέσω εκτενούς πειραματικής ανάλυσης και πετύχαμε καλύτερα αποτελέσματα από αυτά των state-of-the-art μεθόδων.

Ως προς τη σύγκριση των δύο συστημάτων θα μπορούσαμε να πούμε πως το πρώτο έχει μελετηθεί και εκπαιδευτεί σε ένα μεγάλο μέρος της βάσης e-Prevention (δεδομένα 29 ασθενών, ατομικής διάρκειας από 6 μήνες έως 2.5 χρόνια) λαμβάνοντας υπ' όψιν τρεις διαφορετικές περιόδους της ψυχωτικής υποτροπής και με χρήση τεχνικών επαύξησης δεδομένων για την αποφυγή της υπερπροσαρμογής του μοντέλου. Το δεύτερο σύστημά μας εκπαιδεύτηκε στα δεδομένα του e-Prevention Challenge I (δεδομένα 10 ασθενών, ατομικής διάρκειας περίπου 6 μηνών) με στόχο να αξιοποιήσει τη λογική της ταυτοποίησης των ατόμων, που όπως δείξαμε στο πρώτο σύστημα φαίνεται να μπορεί να αναδείξει τις μεταβολές στις διαφορετικές περιόδους και να μπορεί να συγκριθεί με τις υπόλοιπες μεθόδους της βιβλιογραφίας. Θεωρούμε πως οι παραπάνω μελέτες μας φέρνουν ένα βήμα πιο κοντά στην ανίχνευση των ψυχωτικών υποτροπών και ανοίγουν το δρόμο για περαιτέρω έρευνα.

7

Συνεισφορές και Επεκτάσεις

Στην παρούσα διατριβή μελετήσαμε εκτεταμένα διαφορετικές εκφάνσεις των ανθρώπινων δράσεων και δραστηριοτήτων μέσα από την επεξεργασία πολύμορφων σημάτων καταγραφής. Με την εξέλιξη της μηχανικής μάθησης αλλά και της τεχνολογίας των αισθητήρων επεκτείνεται διαρκώς ο τρόπος με τον οποίο μπορούμε να καταγράφουμε και να αντιλαμβανόμαστε την ανθρώπινη δραστηριότητα μηχανικά, δίνοντάς μας νέες δυνατότητες και προοπτικές. Έτσι στο πρώτο μέρος της διατριβής μελετήσαμε την αναγνώριση της δραστηριότητας μέσω οπτικών αισθητήρων επικεντρώνοντας στην αναγνώριση παιδικών κινήσεων, ενώ στο δεύτερο αξιοποιήσαμε την ψηφιακή αποτύπωση της καθημερινής δραστηριότητας του ατόμου, μέσα από πολλαπλούς μη οπτικούς αισθητήρες, για να εντοπίσουμε σημαντικές μεταβολές στην καθημερινότητα ατόμων με ψυχωτικές διαταραχές.

7.1 Συνεισφορές

Στο Μέρος I αναλύσαμε διεξοδικά την ανάπτυξη πολυτροπικών συστημάτων αντίληψης σε εφαρμογές αλληλεπιδράσεων παιδιών και ρομπότ. Πιο συγκεκριμένα:

- Μελετήσαμε κλασικές και πιο σύγχρονες μεθόδους όρασης υπολογιστών για την ανάπτυξη συστημάτων αναγνώρισης κινήσεων και χειρονομιών παιδιών από μεμονωμένες αλλά και από πολλαπλές κάμερες. Προτείναμε μια προσέγγιση πολλαπλών όψεων που βελτίωσε την απόδοση των μεθόδων μονής όψης ενώ διαπιστώσαμε τη σημαντικότητα της εκπαίδευσης των συστημάτων σε δεδομένα παιδιών όταν προορίζονται για αναγνώριση παιδικών δράσεων.

- Αναπτύξαμε το ChildBot, ένα ολοκληρωμένο, αυτόνομο σύστημα αλληλεπίδρασης παιδιών και ρομπότ ενσωματώνοντας πολλαπλούς αισθητήρες, πολυάριθμες μονάδες αντίληψης και διαφορετικά ρομπότ. Το ChildBot διαθέτει καινοτόμες μονάδες αντίληψης για πολυτροπική κατανόηση σκηνής που έχουν αναπτυχθεί ύστερα από εκτενείς μελέτες και προσαρμοστεί σε πραγματικές συνθήκες αλληλεπίδρασης παιδιών και ρομπότ. Ο οπτικοακουστικός εντοπισμός των ομιλητών, ο εντοπισμός και η ιχνηλασία αντικειμένων, η οπτική αναγνώριση δράσεων και η αναγνώριση απομακρυσμένης ομιλίας είναι απαραίτητα για την ανάλυση και την παρακολούθηση της ανθρώπινης συμπεριφοράς κατά την εξέλιξη μιας αλληλεπίδρασης.

- Ορίσαμε και υλοποιήσαμε ενδεικτικά σενάρια χρήσης προκειμένου να αναδείξουμε τη μεγάλη γκάμα εφαρμογών στις οποίες μπορεί να χρησιμοποιηθεί το ChildBot. Συλλέξαμε και δημιουργήσαμε τη βάση BabyRobot με πολυάριθμες αλληλεπιδράσεις μεταξύ τριών

ρομπότ και άνω των 50 παιδιών. Η βάση αυτή επέτρεψε μια εκτενή αντικειμενική αξιολόγηση των δυνατοτήτων των συστημάτων που αναπτύξαμε στο Μέρος I κατά τη διάρκεια πραγματικών σεναρίων αλληλεπίδρασης.

- Στο TeachBot, ένα ευφυές σύστημα αλληλεπίδρασης παιδιού-ρομπότ, αναπτύξαμε και ενσωματώσαμε ισχυρές αρχιτεκτονικές που βασίζονται σε βαθιά νευρωνικά δίκτυα για την αντίληψη των ενεργειών των παιδιών και την αποκωδικοποίηση της συναισθηματικής τους κατάστασης μέσω οπτικών πληροφοριών, οι οποίες αξιολογήθηκαν σε δύο βάσεις δεδομένων παιδιών, και αποδείχθηκε ότι ξεπέρασαν τις state-of-the-art μεθόδους με χαμηλό υπολογιστικό κόστος. Ακόμα, ενσωματώσαμε υπάρχουσες τεχνολογίες αναγνώρισης και σύνθεσης ομιλίας για την περαιτέρω διευκόλυνση της ευχάριστης και φυσικής αλληλεπίδρασης μεταξύ παιδιών και ρομπότ και μέσω της ομιλίας.

- Καινοτομήσαμε υιοθετώντας μεθόδους επαυξητικής μάθησης στο σύστημα αναγνώρισης δράσεων που αναπτύχθηκε στο TeachBot. Επεκτείναμε memory-replay μεθόδους με χρήση χρονικής δειγματοληψίας των βίντεο και διαπιστώσαμε πως το προτεινόμενο δίκτυό πέτυχε το χαμηλότερο ποσοστό λήθης σε όλες τις περιπτώσεις που μελετήσαμε και την υψηλότερη ακρίβεια αναγνώρισης στις περισσότερες από αυτές. Επιπλέον, πραγματοποιήσαμε μελέτες σχετικά με το μέγεθος της μνήμης και τον αντίκτυπό της στην ακρίβεια, τη λήθη, τον χρόνο εκπαίδευσης, αποδεικνύοντας την αποτελεσματικότητα του προτεινόμενου συστήματος.

Η έρευνα που παρουσιάστηκε στο Μέρος II φέρνει πιο κοντά τη συσχέτιση των ψυχωτικών υποτροπών με την αποτύπωσή τους στον ψηφιακό φαινότυπο ασθενών με ψυχωτικές διαταραχές. Πιο συγκεκριμένα οι συνεισφορές της διατριβής συνοψίζονται ως εξής:

- Μελετήσαμε εκτενώς τη δημιουργία ψηφιακών φαινοτύπων μέσω ισχυρών αναπαραστάσεων των σημάτων που συλλέγονται από έξυπνα ρολόγια. Πραγματοποιήσαμε στατιστική ανάλυση των εξαγόμενων χαρακτηριστικών για τον εντοπισμό σημαντικών διαφορών μεταξύ μιας ομάδας ελέγχου και μιας ομάδας ασθενών με ψυχωτικές διαταραχές. Τα ευρήματά μας έδειξαν ότι οι ασθενείς τείνουν να συμπεριφέρονται με μεγαλύτερη μεταβλητότητα και να παρουσιάζουν μεγάλες ακραίες τιμές στην κινητική τους συμπεριφορά μέσα στη μέρα ενώ κατά τη διάρκεια του ύπνου καταγράφεται η αντίθετη συμπεριφορά.

- Σχεδιάσαμε και αναπτύξαμε ένα σύστημα ταυτοποίησης ατόμων βασισμένο στην αναγνώριση των ψηφιακών τους φαινοτύπων. Διεξήγαμε εκτενείς μελέτες ως προς την αρχιτεκτονική του δικτύου, τα χαρακτηριστικά του φαινοτύπου, τις στρατηγικές επαύξησης των δεδομένων με σκοπό να ενισχυθεί η ικανότητα γενίκευσης του μοντέλου. Αξιολογήσαμε τα προτεινόμενα μοντέλα και επαληθεύσαμε την αποτελεσματικότητά τους μέσω των υψηλών ποσοστών ταξινόμησης.

- Αντιμετωπίσαμε το πρόβλημα της ανίχνευσης ψυχωτικών υποτροπών θέτοντας το ως ένα πρόβλημα λανθασμένης ταξινόμησης σε νευρωνικά δίκτυα που είναι εκπαιδευμένα για αναγνώριση ατόμων. Διαπιστώσαμε χαμηλότερη πιθανότητα ταξινόμησης-ταυτοποίησης των ατόμων σε περιόδους υποτροπής της νόσου και λίγο πριν από αυτήν. Επαληθεύσαμε τις αλλαγές στην κατανομή των βαθμολογιών-πιθανοτήτων που προκύπτουν για κάθε άτομο στις διάφορες περιόδους της ψυχωτικής διαταραχής με ενδελεχή στατιστική ανάλυση.

- Παρουσιάσαμε ένα νέο πλαίσιο που ενισχύει τις παραδοσιακές αρχιτεκτονικές αυτόματων κωδικοποιητών που χρησιμοποιούνται για την ανίχνευση ανωμαλιών ενσωματώνοντάς επιπλέον στοιχεία ταυτοποίησης των ασθενών. Ορίζοντας στο προτεινόμενο σύ-

στημα ως κοινό στόχο την ανακατασκευή των σημάτων και την ταυτοποίηση των ασθενών παρατηρήθηκε βελτίωση στην ανίχνευση των υποτροπών μέσω ενός καθολικού μοντέλου. Εφαρμόζοντας περαιτέρω παλινδρόμηση κορυφογραμμής στα δεδομένα του συνόλου επικύρωσης καταφέρνουμε να λάβουμε βελτιωμένη απόδοση στην εκτίμηση της φάσης της διαταραχής των ασθενών ακόμα και σε κάποιους ασθενείς που τα εξατομικευμένα μοντέλα της βιβλιογραφίας δεν μπορούσαν να το επιτύχουν.

- Εκπαιδεύσαμε τα μοντέλα μας και επικυρώσαμε τις υποθέσεις μας ανιχνεύοντας ψυχωτικές υποτροπές σε ένα από τα μεγαλύτερα σύνολα δεδομένων βιομετρικών φορητών δεδομένων που έχουν συλλεχθεί ποτέ. Το e-Prevention dataset περιλαμβάνει δεδομένα από 60 άτομα (38 με ψυχωτικές διαταραχές) και πάνω από 20.000 ημέρες καταγραφής δεδομένων, με έως και 2,5 χρόνια συνεχής παρακολούθησης των ασθενών.

7.2 Επεκτάσεις

Περνώντας σε μια νέα εποχή όπου τα ρομπότ πλέον μαθαίνουν και προσαρμόζονται στο περιβάλλον τους είναι σημαντική η δημιουργία αντιληπτικών συστημάτων που θα είναι ευθυγραμμισμένα σ' αυτό τον σκοπό και θα καθιστούν τα ρομπότ πολύ πιο ευέλικτα και ικανά να χειρίζονται πολύπλοκα σενάρια του πραγματικού κόσμου. Η κατεύθυνση που δόθηκε στο Μέρος I, με την αξιοποίηση επαυξητικών μεθόδων φαίνεται πως εντάσσεται σ' αυτή τη λογική. Έτσι θα μπορούσαμε να προτείνουμε:

- την επέκταση της μονάδας αναγνώρισης δράσεων σε αλληλεπιδράσεις ανθρώπων και ρομπότ με τη συμμετοχή ενηλίκων και τη μελέτη του συστήματος σε ένα περιβάλλον πιο απαιτητικό με προσθήκη ολοένα και μεγαλύτερων συνόλων κλάσεων προς ταξινόμηση.

- την οπτική αναγνώριση υψηλότερου επιπέδου - πιο σύνθετων δράσεων που σχετίζονται με ανθρώπινες λειτουργίες όπως η στοχοπροσήλωση κατά τη διάρκεια μιας αλληλεπίδρασης και η συμμετοχή σε αυτό, σε κατευθύνσεις παρόμοιες με την εργασία μας [Anagnostopoulou et al.; 2022].

- την προσαρμογή των αλληλεπιδράσεων σε συγκεκριμένες ομάδες, όπως παιδιά που παρουσιάζουν Διαταραχή Ελλειμματικής Προσοχής και Υπερκινητικότητας ή βρίσκονται στο φάσμα του αυτισμού, και την αναγνώριση δράσεων για την αξιοποίηση ρομποτικών συστημάτων από τους θεραπευτές τους (π.χ. [Anagnostopoulou et al.; 2021]).

Για το κομμάτι της αναγνώρισης της καθημερινής δραστηριότητας των ανθρώπων και την αξιοποίησή τους σε εφαρμογές ηλεκτρονικής υγείας θα επικεντρωθούμε σε κατευθύνσεις που μπορούν να υλοποιηθούν άμεσα ως συνέχεια της παρούσας έρευνας στα δεδομένα του e-Prevention, όπως:

- Η έρευνα των ψηφιακών φαινοτύπων σε επίπεδο συνεχόμενων ή απλά ακολουθιακών δεδομένων με σκοπό την καλύτερη μοντελοποίηση των μεταβολών που παρουσιάζονται κατά τη διάρκεια μερικών ημερών.

- Η μελέτη για την αξιοποίηση των δεδομένων ύπνου πιο στοχευμένα καθώς ο ύπνος αποτελεί μια συνιστώσα που μεταβάλλεται αρκετά όταν οι ασθενείς βρίσκονται σε ένα ψυχωτικό επεισόδιο [Avramidis et al.; 2023].

- Η ανάπτυξη ενός συστήματος για τον εντοπισμό των μη ψυχωτικών υποτροπών που είναι ήδη καταγεγραμμένες στη βάση.

- Η δημιουργία ενός πολυμεσικού συστήματος αναγνώρισης υποτροπών (ψυχωτικών ή μη) μέσω της παράλληλης αξιοποίησης οπτικών δεδομένων, από τα tablet που χρησιμοποιούσαν οι ασθενείς στις συνεντεύξεις με το ιατρικό προσωπικό, και των δεδομένων των ρολογιών.
- Η παράλληλη αξιοποίηση και εισαγωγή των ιατρικών επισημειώσεων για την πορεία των ασθενών στο input των συστημάτων αναγνώρισης υποτροπών.

Κατάλογος σχημάτων

Κατάλογος πινάκων

Ακρωνύμια

BN	Batch Normalization	Κανονικοποίηση σε Τμήματα
BoVW	Bag-of-Visual-Words	Μέθοδος Συνόλου Οπτικών Λέξεων
CNN	Convolutional Neural Network	Συνελικτικό Νευρωνικό Δίκτυο
CRI	Child-Robot Interaction	Αλληλεπίδραση Παιδιού-Ρομπότ
DNN	Deep Neural Network	Βαθύ Νευρωνικό Δίκτυο
DT	Dense Trajectories	Πυκνές Τροχιές
ECDF	Empirical Cumulative Distribution Function	Εμπειρική Αθροιστική Συνάρτηση Κατανομής
FC	Fully-Connected	Πλήρως συνδεδεμένο
HOF	Histogram of Optical Flow	Ιστόγραμμα Οπτικής Ροής
HOG	Histogram of Oriented Gradients	Ιστόγραμμα Προσανατολισμένων Κλίσεων
HRI	Human-Robot Interaction	Αλληλεπίδραση Ανθρώπου-Ρομπότ
IL	Incremental Learning	Επαυξητική Μάθηση
LSTM	Long-Short Term Memory	Μακρά βραχύχρονη μνήμη
MBH	Motion Boundary Histogram	Ιστόγραμμα Ορίου Κίνησης
MOS	Mean Opinion Score	Μέση Βαθμολογία Γνώμης
MSE	Mean-Squared Error	Μέσο Τετραγωνικό Σφάλμα
RMSSD	Root Mean Squared of Successive RR interval Differences	Τετραγωνική Ρίζα της μέσης τιμής των τετράγωνων των διαφορών των διαδοχικών παλμών
RNN	Recurrent Neural Network	Επαναληπτικό Νευρωνικό Δίκτυο
RR	RR-intervals	Χρονικό διάστημα μεταξύ Διαδοχικών παλμών
SDRR	Standard Deviation of RR intervals	Τυπική απόκλιση των RR-intervals
SGD	Stochastic Gradient Descent	Στοχαστική Μέθοδος Καθόδου Κλίσης
SVM	Support Vector Machine	Μηχανή Υποστήριξης Διανυσμάτων
TSN	Temporal Segment Network	Δίκτυο Χρονικής Δειγματοληψίας
TTS	Text-To-Speech	Σύνθεση Φωνής από Κείμενο
VLAD	Vector of Locally Aggregated Descriptors	Διάνυσμα των Τοπικά Συγκεντρωμένων Περιγραφητών

Παράρτημα Α΄

Λίστα Δημοσιεύσεων

Δημοσιεύσεις σε Διεθνή Περιοδικά

[A1] A. Zlatintsi, P. P. Filntisis, N. Efthymiou, C. Garoufis, G. Retsinas, T. Sounapoglou, I. Maglogiannis, P. Tsanakas, N. Smyrnis, and P. Maragos, "Person identification and relapse detection from continuous recordings of biosignals challenge: Overview and results," *IEEE Open Journal of Signal Processing, to be published.*

[A2] G. Retsinas, N. Efthymiou, D. Anagnostopoulou, and P. Maragos, "Mushroom detection and 3D pose estimation from multi-view point clouds," *Sensors*, vol. 23, 2023.

[A3] E. Kalisperakis, T. Karantinos, M. Lazaridi, V. Garyfalli, P. P. Filntisis, A. Zlatintsi, N. Efthymiou, A. Mantas, L. Mantonakis, T. Mougiakos, I. Maglogiannis, P. Tsanakas, P. Maragos, and N. Smyrnis, "Smartwatch digital phenotypes predict positive and negative symptom variation in a longitudinal monitoring study of patients with psychotic disorders," *Frontiers in Psychiatry*, vol. 14, 2023.

[A4] A. Zlatintsi, P. P. Filntisis, C. Garoufis, N. Efthymiou, P. Maragos, A. Menychtas, I. Maglogiannis, P. Tsanakas, T. Sounapoglou, E. Kalisperakis, T. Karantinos, M. Lazaridi, V. Garyfalli, A. Mantas, L. Mantonakis, and N. Smyrnis, "E-Prevention: Advanced support system for monitoring and relapse prevention in patients with psychotic disorders analyzing long-term multimodal data from wearables and video captures," *Sensors*, vol. 22, p. 7544, 2022.

[A5] N. Efthymiou, P. P. Filntisis, P. Koutras, A. Tsiami, J. Hadfield, G. Potamianos, and P. Maragos, "ChildBot: Multi-robot perception and interaction with children," *Robotics and Autonomous Systems*, vol. 150, p. 103975, 2022.

[A6] N. Efthymiou, P. P. Filntisis, G. Potamianos, and P. Maragos, "Visual robotic perception system with incremental learning for child-robot interaction scenarios," *Technologies*, vol. 9, p. 86, 2021.

[A7] P. P. Filntisis, N. Efthymiou, P. Koutras, G. Potamianos, and P. Maragos, "Fusing body posture with facial expressions for joint recognition of affect in child-robot interaction," *IEEE Robotics and automation letters*, vol. 4, pp. 4011--4018, 2019.

[A8] A. Zlatintsi, P. Koutras, G. Evangelopoulos, N. Malandrakis, N. Efthymiou, K. Pastra, A. Potamianos, and P. Maragos, "Cognimuse: A multimodal video database annotated with saliency, events, semantics and emotion with application to summarization," *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, pp. 1--24, 2017.

Δημοσιεύσεις σε Διεθνή Συνέδρια

[B1] N. Efthymiou, G. Retsinas, P. Filntisis, and P. Maragos, "Augmenting transformer autoencoders with phenotype classification for robust detection of psychotic relapses," in *Proc. ICASSP, under review.*

[B2] G. Retsinas, N. Efthymiou, and P. Maragos, "Mushroom segmentation and 3D pose estimation from point clouds using fully convolutional geometric features and implicit pose encoding," in *Proc. CVPRW - Agriculture-Vision Workshop*, 2023.

[B3] D. Anangostopoulou, G. Retsinas, N. Efthymiou, P. Filntisis, and P. Maragos, "A realistic synthetic mushroom scenes dataset," in *Proc. CVPRW - Agriculture-Vision Workshop*, 2023.

[B4] N. Efthymiou, G. Retsinas, P. Filntisis, C. Garoufis, A. Zlatintsi, E. Kalisperakis, V. Garyfalli, T. Karantinos, M. Lazaridi, N. Smyrnis, and P. Maragos, "From digital phenotype identification to detection of psychotic relapses," in *Proc. ICHI*, 2023.

[B5] E. Fekas, A. Zlatintsi, P. Filntisis, C. Garoufis, N. Efthymiou, and P. Maragos, "Relapse prediction from long-term wearable data using self-supervised learning and survival analysis," in *Proc. ICASSP*, 2023.

[B6] D. Anagnostopoulou, N. Efthymiou, C. Papailiou, and P. Maragos, "Child engagement estimation in heterogeneous child-robot interactions using spatiotemporal visual cues," in *Proc. IROS*, 2022.

[B7] C. Garoufis, A. Zlatintsi, P. Filntisis, N. Efthymiou, E. Kalisperakis, T. Karantinos, V. Garyfalli, M. Lazaridi, N. Smyrnis, and P. Maragos, "Towards unsupervised subject-independent speech-based relapse detection in patients with psychosis using variational autoencoders," in *Proc. EUSIPCO*, 2022.

[B8] M. Panagiotou, A. Zlatintsi, P. Filntisis, A. Roumeliotis, N. Efthymiou, and P. Maragos, "A comparative study of autoencoder architectures for mental health analysis using wearable sensors data," in *Proc. EUSIPCO*, 2022.

[B9] N. Efthymiou, P. Filntisis, G. Potamianos, and P. Maragos, "A robotic edutainment framework for designing child-robot interaction scenarios," in *Proc. PETRA*, 2021.

[B10] P. P. Filntisis, N. Efthymiou, G. Potamianos, and P. Maragos, "An audiovisual child emotion recognition system for child-robot interaction applications," in *Proc. EUSIPCO*, 2021.

[B11] D. Anagnostopoulou, N. Efthymiou, C. Papailiou, and P. Maragos, "Engagement estimation during child robot interaction using deep convolutional networks focusing on asd children," in *Proc. ICRA*, 2021.

[B12] C. Garoufis, A. Zlatintsi, P. P. Filntisis, N. Efthymiou, E. Kalisperakis, V. Garyfalli, T. Karantinos, L. Mantonakis, N. Smyrnis, and P. Maragos, "An unsupervised learning approach for detecting relapses from spontaneous speech in patients with psychosis," in *Proc. BHI*, 2021.

[B13] P. P. Filntisis, N. Efthymiou, G. Potamianos, and P. Maragos, "Emotion understanding in videos through body, context, and visual-semantic embedding loss," in *Proc. ECCVW*, 2020.

[B14] G. Retsinas, P. P. Filntisis, N. Efthymiou, E. Theodosis, A. Zlatintsi, and P. Maragos, "Person identification using deep convolutional neural networks on short-term signals from wearable sensors," in *Proc. ICASSP*, 2020.

[B15] I. Maglogiannis, A. Zlatintsi, A. Menychtas, D. Papadimatos, P. P. Filntisis, N. Efthymiou, G. Retsinas, P. Tsanakas, and P. Maragos, "An intelligent cloud-based platform for effective monitoring of patients with psychotic disorders," in *Proc. AIAI*, 2020.

[B16] N. Efthymiou, P. Koutras, P. P. Filntisis, G. Potamianos, and P. Maragos, "Multi-view fusion for action recognition in child-robot interaction," in *Proc. ICIP*, 2018.

[B17] J. Hadfield, P. Koutras, N. Efthymiou, G. Potamianos, C. S. Tzafestas, and P. Maragos, "Object assembly guidance in child-robot interaction using rgb-d based 3d tracking," in *Proc. IROS*, 2018.

[B18] A. Tsiami, P. P. Filntisis, N. Efthymiou, P. Koutras, G. Potamianos, and P. Maragos, "Far-field audio-visual scene perception of multi-party human-robot interaction for children and adults," in *Proc. ICASSP*, 2018.

[B19] A. Tsiami, P. Koutras, N. Efthymiou, P. P. Filntisis, G. Potamianos, and P. Maragos, "Multi3: Multi-sensory perception system for multi-modal child interaction with multiple robots," in *Proc. ICRA*, 2018.

[B20] A. Zlatintsi, P. Koutras, N. Efthymiou, P. Maragos, A. Potamianos, and K. Pastra, "Quality evaluation of computational models for movie summarization," in *Proc. QoMEX*, 2015.

Βιβλιογραφία

[Abbaspur-Behbahani et al.; 2022] Abbaspur-Behbahani, S., Monaghesh, E., Hajizadeh, A., and Fehresti, S. (2022). Application of mobile health to support the elderly during the COVID-19 outbreak: A systematic review. *Health policy and technology*, 11(1):100595.

[Adler et al.; 2020] Adler, D. A., Ben-Zeev, D., Tseng, V. W., Kane, J. M., Brian, R., et al. (2020). Predicting early warning signs of psychotic relapse from passive sensing data: an approach using encoder-decoder neural networks. *JMIR mHealth and uHealth*, 8:e19962.

[Al-Nuaimi et al.; 2017] Al-Nuaimi, A. H., Jammeh, E., Sun, L., and Ifeachor, E. (2017). Higuchi fractal dimension of the electroencephalogram as a biomarker for early detection of alzheimer's disease. In *Proc. EMBC*. IEEE.

[Aljundi et al.; 2018] Aljundi, R., Babiloni, F., Elhoseiny, M., Rohrbach, M., and Tuytelaars, T. (2018). Memory aware synapses: Learning what (not) to forget. In *Proc. ECCV*.

[Anagnostopoulou et al.; 2021] Anagnostopoulou, D., Efthymiou, N., Papailiou, C., and Maragos, P. (2021). Engagement estimation during child robot interaction using deep convolutional networks focusing on asd children. In *Proc. ICRA*.

[Anagnostopoulou et al.; 2022] Anagnostopoulou, D., Efthymiou, N., Papailiou, C., and Maragos, P. (2022). Child engagement estimation in heterogeneous child-robot interactions using spatiotemporal visual cues. In *Proc. IROS*.

[Arandjelovic and Zisserman; 2013] Arandjelovic, R. and Zisserman, A. (2013). All about VLAD. In *Proc. CVPR*.

[Asadzadeh and Kalankesh; 2021] Asadzadeh, A. and Kalankesh, L. R. (2021). A scope of mobile health solutions in COVID-19 pandemics. *Informatics in medicine unlocked*, 23:100558.

[Aung et al.; 2017] Aung, M. H., Matthews, M., and Choudhury, T. (2017). Sensing behavioral symptoms of mental health and delivering personalized interventions using mobile technologies. *Depression and Anxiety*, 34:603--609.

[Avramidis et al.; 2023] Avramidis, K., Adsul, K., Bose, D., and Narayanan, S. (2023). Signal processing grand challenge 2023 - e-Prevention: Sleep behavior as an indicator of relapses in psychotic patients. In *Proc. ICASSP*.

[Baldi; 2012] Baldi, P. (2 July 2012). Autoencoders, unsupervised learning, and deep architectures. In *Proc. ICML Workshop on Unsupervised and Transfer Learning*.

[Barnett et al.; 2018] Barnett, I., Torous, J., Staples, P., Sandoval, L., Keshavan, M., et al. (2018). Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology*, 43:1660--1666.

[Bauer et al.; 2008] Bauer, M., Wilson, T., Neuhaus, K., Sasse, J., Pfennig, A., et al. (2008). Self-reporting software for bipolar disorder: validation of chronorecord by patients with mania. *Psychiatry research*, 159:359--366.

[Belouadah and Popescu; 2019] Belouadah, E. and Popescu, A. (2019). Il2m: Class incremental learning with dual memory. In *Proc. ICCV*.

[Belpaeme et al.; 2012] Belpaeme, T., Baxter, P., Read, R., Wood, R., Cuayáhuitl, H., et al. (2012). Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, 1:33--53.

[Belpaeme et al.; 2013] Belpaeme, T., Baxter, P., De Greeff, J., Kennedy, J., Read, R., Looije, R., et al. (2013). Child-robot interaction: Perspectives and challenges. In *Proc. ICSR*.

[Belpaeme et al.; 2015] Belpaeme, T., Kennedy, J., Baxter, P., Vogt, P., Krahmer, E. E., et al. (2015). L2TOR-second language tutoring using social robots. In *Proc. ICSR*.

[Belpaeme et al.; 2018] Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: A review. *Science robotics*, 3:eaat5954.

[Belpaeme; 2020] Belpaeme, T. (2020). The wizard is dead, long live data: towards autonomous social behaviour using data-driven methods. In *Companion Publication of ICMI*.

[Ben-Zeev et al.; 2017] Ben-Zeev, D., Brian, R., Wang, M., et al. (2017). CrossCheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. *Psychiatric Rehabilitation J.*, 40:266.

[Beniczky et al.; 2021] Beniczky, S., Wiebe, S., Jeppesen, J., Tatum, W. O., Brazdil, M., et al. (2021). Automated seizure detection using wearable devices: A clinical practice guideline of the international league against epilepsy and the international federation of clinical neurophysiology. *Clinical Neurophysiology*, 132:1173--1184.

[Benjamini and Hochberg; 1995] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: Series B (Methodological)*, 57:289--300.

[Berry and Springer; 1993] Berry, D. S. and Springer, K. (1993). Structure, motion, and preschoolers' perceptions of social causality. *Ecological Psychology*, 5:273--283.

[Bertelsen et al.; 2008] Bertelsen, M., Jeppesen, P., Petersen, L., Thorup, A., Øhlenschlæger, J., et al. (2008). Five-year follow-up of a randomized multicenter trial of intensive early intervention vs standard treatment for patients with a first episode of psychotic illness: the opus trial. *Archives of General Psychiatry*, 65:762--771.

[Bickel et al.; 2020] Bickel, V. T., Conway, S. J., Tesson, P., Manconi, A., Loew, S., et al. (2020). Deep learning-driven detection and mapping of rockfalls on mars. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:2831--2841.

[Blum and Magill; 2008] Blum, J. and Magill, E. (2008). M-psychiatry: Sensor networks for psychiatric health monitoring. In *Proc. Symposium The Convergence of Telecommunications, Networking and Broadcasting*.

[Boccanfuso et al.; 2016] Boccanfuso, L., Barney, E., Foster, C., Ahn, Y. A., Chawarska, K., et al. (2016). Emotional robot to examine different play patterns and affective responses of children with and without ASD. In *Proc. HRI*.

[Böhm et al.; 2019] Böhm, B., Karwiese, S. D., Böhm, H., Oberhoffer, R., et al. (2019). Effects of mobile health including wearable activity trackers to increase physical activity outcomes among healthy children and adolescents: systematic review. *JMIR mHealth and uHealth*, 7:e8298.

[Boletsis et al.; 2015] Boletsis, C., McCallum, S., and Landmark, B. F. (2015). The use of smartwatches for health monitoring in home-based dementia care. In *Proc. Int'l Conf. Human Aspects IT Aged Population*.

[Brennan et al.; 2001] Brennan, M., Palaniswami, M., and Kamen, P. (2001). Do existing measures of poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE Trans. on Biomedical Engineering*, 48:1342--1347.

[Brown et al.; 2020] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877--1901.

[Cai et al.; 2018] Cai, G., Lin, Z., Dai, H., Xia, X., Xiong, Y., et al. (2018). Quantitative assessment of parkinsonian tremor based on a linear acceleration extraction algorithm. *Biomedical Signal Processing and Control*, 42:53--62.

[Calcagno et al.; 2023] Calcagno, S., Mineo, R., Giordano, D., and Spampinato, C. (2023). Ensemble and personalized transformer models for subject identification and relapse detection in e-Prevention challenge. In *Proc. ICASSP*.

[Canzian and Musolesi; 2015] Canzian, L. and Musolesi, M. (2015). Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proc. UbiComp*.

[Cao et al.; 2017] Cao, Z., Simon, T., Wei, S., and Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proc. CVPR*.

[Carreira and Zisserman; 2017] Carreira, J. and Zisserman, A. (2017). Quo vadis, action recognition? a new model and the Kinetics dataset. In *Proc. CVPR*.

[Castro et al.; 2018] Castro, F. M., Marín-Jiménez, M. J., Guil, N., Schmid, C., and Alahari, K. (2018). End-to-end incremental learning. In *Proc. ECCV*.

[Cella et al.; 2018] Cella, M., Okruszek, L., Lawrence, M., Zarlenga, V., He, Z., et al. (2018). Using wearable technology to detect the autonomic signature of illness severity in schizophrenia. *Schizophrenia Research*, 195:537--542.

[Chalvatzaki et al.; 2020] Chalvatzaki, G., Koutras, P., Tsiami, A., Tzafestas, C. S., and Maragos, P. (2020). i-Walk intelligent assessment system: Activity, mobility, intention, communication. In *Proc. ECCVW ACVR*.

[Chandra et al.; 2019] Chandra, S., Dillenbourg, P., and Paiva, A. (2019). Children teach handwriting to a social robot with different learning competencies. *Journal of Social Robotics*, 12:721--748.

[Chang and Lin; 2011] Chang, C. and Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27.

[Chapman et al.; 2017] Chapman, J. J., Roberts, J. A., Nguyen, V. T., and Breakspear, M. (2017). Quantification of free-living activity patterns using accelerometry in adults with mental illness. *Scientific reports*, 7:43174.

[Chaspari; 2022] Chaspari, T. (2022). Sensor integration for behavior monitoring. *Elsevier*.

[Chen; 2020] Chen, Y. (2020). Crowd behaviour recognition using enhanced butterfly optimization algorithm based recurrent neural network. *Multimedia Research*, 3:20.

[Chiang et al.; 2018] Chiang, M. L., Feng, J., Zeng, W. L., Fang, C. Y., and Chen, S. W. (2018). A vision-based human action recognition system for companion robots and human interaction. In *Proc. ICCC*.

[Churamani et al.; 2020] Churamani, N., Kalkan, S., and Gunes, H. (2020). Continual learning for affective robotics: Why, what and how? In *Proc. RO-MAN*.

[Dalal and Triggs; 2005] Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proc. CVPR*.

[Davison et al.; 2020] Davison, D. P., Wijnen, F. M., Charisi, V., van der Meij, J., Evers, V., et al. (2020). Working with a social robot in school: a long-term real-world unsupervised deployment. In *Proc. HRI*.

[De Choudhury et al.; 2013] De Choudhury, M., Gamon, M., Counts, S., and Horvitz, E. (2013). Predicting depression via social media. In *Proc. ICWSM*.

[Dehghan et al.; 2019] Dehghan, M., Zhang, Z., Siam, M., Jin, J., Petrich, L., et al. (2019). Online object and task learning via human robot interaction. In *Proc. ICRA*.

[Deng et al.; 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., et al. (2009). ImagNet: A large-scale hierarchical image database. In *Proc. CVPR*.

[Devi et al.; 2023] Devi, D. H., Duraisamy, K., Armghan, A., Alsharari, M., Aliqab, K., et al. (2023). 5G technology in healthcare and wearable devices: A review. *Sensors*, 23(5):2519.

[Di Dio et al.; 2020] Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., et al. (2020). Shall i trust you? from child-robot interaction to trusting relationships. *Frontiers in Psychology*, 11:469.

[Domingo-Lopez et al.; 2022] Domingo-Lopez, D. A., Lattanzi, G., Schreiber, L. H., Wallace, E. J., Wylie, R., et al. (2022). Medical devices, smart drug delivery, wearables and technology for the treatment of diabetes mellitus. *Advanced Drug Delivery Reviews*, page 114280.

[Efthymiou et al.;] Efthymiou, N., Retsinas, G., Filntisis, P., and Maragos, P. Augmenting transformer autoencoders with phenotype classification for robust detection of psychotic relapses. In *Proc. ICASSP, under review*.

[Efthymiou et al.; 2018] Efthymiou, N., Koutras, P., Filntisis, P. P., Potamianos, G., and Maragos, P. (2018). Multi- view fusion for action recognition in child-robot interaction. In *Proc. ICIP*.

[Efthymiou et al.; 2021a] Efthymiou, N., Filntisis, P., Potamianos, G., and Maragos, P. (2021a). A robotic edutainment framework for designing child-robot interaction scenarios. In *Proc. PETRA*.

[Efthymiou et al.; 2021b] Efthymiou, N., Filntisis, P. P., Potamianos, G., and Maragos, P. (2021b). Visual robotic perception system with incremental learning for child-robot interaction scenarios. *Technologies*, 9:86.

[Efthymiou et al.; 2022] Efthymiou, N., Filntisis, P. P., Koutras, P., Tsiami, A., Hadfield, J., Potamianos, G., and Maragos, P. (2022). ChildBot: Multi-robot perception and interaction with children. *Robotics and Autonomous Systems*, 150:103975.

[Efthymiou et al.; 2023] Efthymiou, N., Retsinas, G., Filntisis, P., Garoufis, C., Zlatintsi, A., Kalisperakis, E., Garyfalli, V., Karantinos, T., Lazaridi, M., Smyrnis, N., and Maragos, P. (2023). From digital phenotype identification to detection of psychotic relapses. In *Proc. ICHI*.

[Eldele et al.; 2021] Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C., et al. (2021). Time-series representation learning via temporal and contextual contrasting. In *Proc. IJCAI*.

[Engels; 1960] Engels, F. (1960). *Dialectics of nature*. Wellred Books.

[Esteban et al.; 2017] Esteban, P. G., Baxter, P., Belpaeme, T., Billing, E., Cai, H., et al. (2017). How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder. *Paladyn, Journal of Behavioral Robotics*, 8:18--38.

[Falconer; 2004] Falconer, K. (2004). *Fractal geometry: Mathematical foundations and applications*. John Wiley & Sons.

[Feichtenhofer et al.; 2017] Feichtenhofer, C., Pinz, A., and Wildes, R. P. (2017). Spatiotemporal multiplier networks for video action recognition. In *Proc. CVPR*.

[Fekas et al.; 2023] Fekas, E., Zlatintsi, A., Filntisis, P., Garoufis, C., Efthymiou, N., and Maragos, P. (2023). Relapse prediction from long-term wearable data using self-supervised learning and survival analysis. In *Proc. ICASSP*.

[Filippini et al.; 2020] Filippini, C., Spadolini, E., Cardone, D., Bianchi, D., Preziuso, M., et al. (2020). Facilitating the child-robot interaction by endowing the robot with the capability of understanding the child engagement: the case of mio amico robot. *International Journal of Social Robotics*, 13:677--689.

[Filntisis et al.; 2020a] Filntisis, P. P., Efthymiou, N., Potamianos, G., and Maragos, P. (2020a). Emotion understanding in videos through body, context, and visual-semantic embedding loss. In *Proc. ECCVW*.

[Filntisis et al.; 2020b] Filntisis, P. P., Zlatintsi, A., Efthymiou, N., Kalisperakis, E., Karantinos, T., et al. (2020b). Identifying differences in physical activity and autonomic function patterns between psychotic patients and controls over a long period of continuous monitoring using wearable sensors. *arXiv preprint arXiv:2011.02285*.

[Fotso; 2018] Fotso, S. (2018). Deep neural networks for survival analysis based on a multi-task framework. *arXiv e-prints*, pages arXiv--1801.

[Furhat Robotics;] Furhat Robotics. http://furhatrobotics.com.

[Gaebel et al.; 1993] Gaebel, W., Frick, U., Köpcke, W., Linden, M., Müller, P., et al. (1993). Early neuroleptic intervention in schizophrenia: are prodromal symptoms valid predictors of relapse? *The British Journal of Psychiatry*, 163:8--12.

[Garoufis et al.; 2022] Garoufis, C., Zlatintsi, A., Filntisis, P., Efthymiou, N., Kalisperakis, E., Karantinos, T., Garyfalli, V., Lazaridi, M., Smyrnis, N., and Maragos, P. (2022). Towards unsupervised subject-independent speech-based relapse detection in patients with psychosis using variational autoencoders. In *Proc. EUSIPCO*.

[Gat et al.; 1998] Gat, E., Bonnasso, R. P., Murphy, R., et al. (1998). On three-layer architectures. *Artificial Intelligence and Mobile Robots*, 195:210.

[Geurts et al.; 2006] Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63.

[Goodfellow et al.; 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

[Google Speech-to-Text; 2021] Google Speech-to-Text (2021). https://cloud.google.com/speech-to-text.

[Google Text-to-Speech; 2021] Google Text-to-Speech (2021). https://cloud.google.com/text-to-speech.

[Gordon et al.; 2015] Gordon, G., Breazeal, C., and Engel, S. (2015). Can children catch curiosity from a social robot? In *Proc. HRI*.

[Gravenhorst et al.; 2015] Gravenhorst, F., Muaremi, A., Bardram, J., Grünerbl, A., Mayora, O., et al. (2015). Mobile phones as medical devices in mental disorder treatment: an overview. *Personal and Ubiquitous Computing*, 19:335--353.

[Gunn III and Lester; 2013] Gunn III, J. F. and Lester, D. (2013). Using google searches on the internet to monitor suicidal behavior. *Journal of affective disorders*, 148:411--412.

[Hadfield et al.; 2018] Hadfield, J., Koutras, P., Efthymiou, N., Potamianos, G., Tzafestas, C. S., et al. (2018). Object assembly guidance in child-robot interaction using RGB-D based 3D tracking. In *Proc. IROS*.

[Hall et al.; 2016] Hall, L., Hume, C., and Tazzyman, S. (2016). Five degrees of happiness: Effective smiley face Likert scales for evaluating with children. In *Proc. IDC*.

[Hamieh et al.; 2023] Hamieh, S., Heiries, V., Al Osman, H., and Godin, C. (2023). Relapse detection in patients with psychotic disorders using unsupervised learning on smartwatch signals. In *Proc. ICASSP*.

[Harel; 1987] Harel, D. (1987). Statecharts: A visual formalism for complex systems. *Science of computer programming*, 8:231--274.

[Hegelstad et al.; 2012] Hegelstad, W. t. V., Larsen, T. K., Auestad, B., Evensen, et al. (2012). Long-term follow-up of the TIPS early detection in psychosis study: effects on 10-year outcome. *American Journal of Psychiatry*, 169:374--380.

[Heilbron et al.; 2015] Heilbron, F. C., Escorcia, V., Ghanem, B., and Niebles, J. C. (2015). ActivityNet: A large-scale video benchmark for human activity understanding. In *Proc. CVPR*.

[Henry et al.; 2010] Henry, B. L., Minassian, A., Paulus, M. P., Geyer, M. A., and Perry, W. (2010). Heart rate variability in bipolar mania and schizophrenia. *J. psychiatric research*, 44:168--176.

[Henson et al.; 2020] Henson, P., Barnett, I., Keshavan, M., and Torous, J. (2020). Towards clinically actionable digital phenotyping targets in schizophrenia. *NPJ Schizophrenia*, 6:1--7.

[Higuchi; 1988] Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D*, 31(2):277--283.

[Insel; 2007] Insel, T. R. (2007). The arrival of preemptive psychiatry. *Early Intervention in Psychiatry*, 1:5--6.

[Jain et al.; 2015] Jain, S. H., Powers, B. W., Hawkins, J. B., and Brownstein, J. S. (2015). The digital phenotype. *Nature biotechnology*, 33:462--463.

[Jégou et al.; 2010] Jégou, H., Douze, M., Schmid, C., and Pérez, P. (2010). Aggregating local descriptors into a compact image representation. In *Proc. CVPR*.

[Kalisperakis et al.; 2023] Kalisperakis, E., Karantinos, T., Lazaridi, M., Garyfalli, V., Filntisis, P. P., Zlatintsi, A., Efthymiou, N., Mantas, A., Mantonakis, L., Mougiakos, T., Maglogiannis, I., Tsanakas, P., Maragos, P., and Smyrnis, N. (2023). Smartwatch digital phenotypes predict positive and negative symptom variation in a longitudinal monitoring study of patients with psychotic disorders. *Frontiers in Psychiatry*, 14.

[Kanda et al.; 2007] Kanda, T., Sato, R., Saiwaki, N., and Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human-robot interaction. *IEEE Transaction on robotics*, 23:962--971.

[Kanda et al.; 2012] Kanda, T., Shimada, M., and Koizumi, S. (2012). Children learning with a social robot. In *Proc. HRI*.

[Karpathy et al.; 2014] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., et al. (2014). Large-scale video classification with convolutional neural networks. In *Proc. CVPR*.

[Katzman et al.; 2018] Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., et al. (2018). DeepSurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Medical Research Methodology*, 18.

[Kazemi et al.; 2019] Kazemi, S. M., Goel, R., Eghbali, S., Ramanan, J., Sahota, J., Thakur, S., Wu, S., Smyth, C., Poupart, P., and Brubaker, M. (2019). Time2Vec: Learning a vector representation of time. *arXiv preprint arXiv:1907.05321*.

[Kennedy et al.; 2015] Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2015). Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions. In *Proc. ICSR*.

[Kennedy et al.; 2016] Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2016). Social robot tutoring for child second language learning. In *Proc. HRI*.

[Kennedy et al.; 2017] Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., et al. (2017). Child speech recognition in human-robot interaction: evaluations and recommendations. In *Proc. HRI*.

[Kesić and Spasić; 2016] Kesić, S. and Spasić, S. Z. (2016). Application of higuchi's fractal dimension from basic to clinical neurophysiology: a review. *Computer methods and programs in biomedicine*, 133:55--70.

[Khoa et al.; 2012] Khoa, T. Q. D., Ha, V. Q., and Toi, V. V. (2012). Higuchi fractal properties of onset epilepsy electroencephalogram. *Computational and mathematical methods in medicine*, 2012.

[Kirkpatrick et al.; 2017] Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., et al. (2017). Overcoming catastrophic forgetting in neural networks. *Proc. of the national academy of sciences*, 114.

[Komatsubara et al.; 2019] Komatsubara, T., Shiomi, M., Kaczmarek, T., Kanda, T., and Ishiguro, H. (2019). Estimating children's social status through their interaction activities in classrooms with a social robot. *Intl. Journal of Social Robotics*, 11:35--48.

[Koutsouleris et al.; 2011] Koutsouleris, N., Davatzikos, C., Bottlender, R., Patschurek-Kliche, K., Scheuerecker, J., et al. (2011). Early recognition and disease prediction in the at-risk mental states for psychosis using neurocognitive pattern classification. *Schizophrenia Bulletin*, 38:1200--1215.

[Kuehne et al.; 2011] Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., and Serre, T. (2011). HMDB: a large video database for human motion recognition. In *Proc. ICCV*.

[Laptev et al.; 2008] Laptev, I., Marszalek, M., Schmid, C., and Rozenfeld, B. (2008). Learning realistic human actions from movies. In *Proc. CVPR*.

[Lesort et al.; 2020] Lesort, T., Lomonaco, V.and Stoian, A., Maltoni, D., Filliat, D., and Díaz-Rodríguez, N. (2020). Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges. *Information fusion*, 58:52--68.

[Li and Hoiem; 2017] Li, Z. and Hoiem, D. (2017). Learning without forgetting. *IEEE Trans. on pattern analysis and machine intelligence*, 40:2935--2947.

[Lin et al.; 2019] Lin, J., Gan, C., and Han, S. (2019). TSM: Temporal shift module for efficient video understanding. In *Proc. CVPR*.

[Liu et al.; 2021a] Liu, F., Wu, X., Ge, S., Fan, W., and Zou, Y. (2021a). Exploring and distilling posterior and prior knowledge for radiology report generation. In *Proc. CVPR*.

[Liu et al.; 2008] Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation forest. In *Proc. ICDM*.

[Liu et al.; 2022] Liu, H., Zhou, Y., Liu, B., Zhao, J., Yao, R., et al. (2022). Incremental learning with neural networks for computer vision: a survey. *Artificial Intelligence Review*.

[Liu et al.; 2019a] Liu, L., Jiang, H.and He, P., Chen, W., Liu, X., Gao, J., et al. (2019a). On the variance of the adaptive learning rate and beyond. *arXiv preprint arXiv:1908.03265*.

[Liu et al.; 2019b] Liu, W., Liao, S., Ren, W., Hu, W., and Yu, Y. (2019b). High-level semantic feature detection: A new perspective for pedestrian detection. In *CVPR*.

[Liu et al.; 2021b] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., et al. (2021b). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proc. CVPR*.

[Lv et al.; 2019] Lv, M., Xu, W., and Chen, T. (2019). A hybrid deep convolutional and recurrent neural network for complex activity recognition using multimodal sensors. *Neurocomputing*, 362:33--40.

[Maglogiannis et al.; 2020] Maglogiannis, I., Zlatintsi, A., Menychtas, A., Papadimatos, D., Filntisis, P. P., et al. (2020). An intelligent cloud-based platform for effective monitoring of patients with psychotic disorders. In *Proc. AIAI*.

[Maiorana et al.; 2022] Maiorana, E., Romano, C., Schena, E., and Massaroni, C. (2022). BIOWISH: Biometric recognition using wearable inertial sensors detecting heart activity. *arXiv preprint arXiv:2210.09843*.

[Mann and Whitney; 1947] Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of Mathematical Statistics*, pages 50--60.

[Maracani et al.; 2021] Maracani, A., Michieli, U., Toldo, M., and Zanuttigh, P. (2021). RECALL: Replay-based continual learning in semantic segmentation. In *Proc. ICCV*.

[Maragos; 1994] Maragos, P. (1994). Fractal signal analysis using mathematical morphology. In *Advances in electronics and electron physics*, volume 88, pages 199--246.

[Marinoiu et al.; 2018] Marinoiu, E., Zanfir, M., Olaru, V., and Sminchisescu, C. (2018). 3D human sensing, action and emotion recognition in robot assisted therapy of children with autism. In *Proc. CVPR*.

[Masana et al.; 2020] Masana, M., Liu, X., Twardowski, B., Menta, M., Bagdanov, A. D., et al. (2020). Class-incremental learning: survey and performance evaluation. *arXiv preprint arXiv:2010.15277*.

[Maxhuni et al.; 2016] Maxhuni, A., Muñoz-Meléndez, A., Osmani, V., Perez, H., Mayora, O., et al. (2016). Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients. *Pervasive and Mobile Computing*, 31:50--66.

[McCandless-Glimcher et al.; 1986] McCandless-Glimcher, L., McKnight, S., Hamera, E., Smith, B. L., Peterson, K. A., et al. (1986). Use of symptoms by schizophrenics to monitor and regulate their illness. *Psychiatric Services*, 37:929--933.

[McGorry et al.; 2014] McGorry, P., Keshavan, M., Goldstone, S., Amminger, P., Allott, K., et al. (2014). Biomarkers and clinical staging in psychiatry. *World Psychiatry*, 13:211--223.

[Melo et al.; 2019] Melo, F. S., Sardinha, A., Belo, D., Couto, M., Faria, M., et al. (2019). Project INSIDE: towards autonomous semi-unstructured human–robot social interaction in autism therapy. *Artificial Intelligence in Medicine*, 96:198--216.

[Mohapatra et al.; 2023] Mohapatra, P., Pandey, A., Keten, S., Chen, W., and Zhu, Q. (2023). Person identification with wearable sensing using missing feature encoding and multi-stage modality fusion. In *Proc. ICASSP*.

[Mukhopadhyay; 2014] Mukhopadhyay, S. C. (2014). Wearable sensors for human activity monitoring: A review. *IEEE Sensors Journal*, 15:1321--1330.

[NAO;] NAO. https://www.softbankrobotics.com/.

[Nasralla et al.; 2023] Nasralla, M. M., Khattak, S. B. A., Ur Rehman, I., and Iqbal, M. (2023). Exploring the role of 6g technology in enhancing quality of experience for m-health multimedia applications: A comprehensive survey. *Sensors*, 23(13):5882.

[Ni et al.; 2023] Ni, J., Young, T., Pandelea, V., Xue, F., and Cambria, E. (2023). Recent advances in deep learning based dialogue systems: A systematic survey. *Artificial intelligence review*, 56:3055–3155.

[Norman and Malla; 1995] Norman, R. M. G. and Malla, A. K. (1995). Prodromal symptoms of relapse in schizophrenia: a review. *Schizophrenia Bulletin*, 21:527--539.

[Os and Kapur; 2009] Os, J. V. and Kapur, S. (2009). Schizophrenia. *The Lancet*, 374:635--645.

[Panagiotou et al.; 2022] Panagiotou, M., Zlatintsi, A., Filntisis, P., Roumeliotis, A., Efthymiou, N., and Maragos, P. (2022). A comparative study of autoencoder architectures for mental health analysis using wearable sensors data. In *Proc. EUSIPCO*.

[Papathanasiou et al.; 2009] Papathanasiou, G., Georgoudis, G., Papandreou, M., Spyropoulos, P., Georgakopoulos, D., et al. (2009). Reliability measures of the short international physical activity questionnaire (ipaq) in greek young adults. *Hellenic J Cardiol*, 50:283--294.

[Parisi et al.; 2019] Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., and Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54--71.

[Park and Kim; 2020] Park, J. Y. and Kim, J. H. (2020). Online incremental classification resonance network and its application to human–robot interaction. *IEEE Transactions on Neural Networks and Learning Systems*, 31:1426--1436.

[Patel et al.; 2012] Patel, S., Park, H., Bonato, P., Chan, L., and Rodgers, M. (2012). A review of wearable sensors and systems with application in rehabilitation. *Journal of Neuroengineering and Rehabilitation*, 9:21.

[Peng and Schmid; 2015] Peng, X. and Schmid, C. (2015). Encoding feature maps of CNNs for action recognition. In *Proc. CVPRW, THUMOS*.

[Peng et al.; 2016] Peng, X., Wang, L., Wang, X., and Qiao, Y. (2016). Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice. *Computer Vision and Image Understanding*, 150:109--125.

[Perronnin et al.; 2010] Perronnin, F., Sánchez, J., and Mensink, T. (2010). Improving the Fisher kernel for large-scale image classification. In *Proc. ECCV*.

[Pulido et al.; 2017] Pulido, J. C., González, J. C., Suárez-Mejías, C., Bandera, A., Bustos, P., et al. (2017). Evaluating the child-robot interaction of the NAOTherapist platform in pediatric rehabilitation. *International Journal of Social Robotics*, 9:343--358.

[Rao and Georgeff; 1995] Rao, A. S. and Georgeff, M. P. (1995). BDI agents: from theory to practice. In *Proc. ICMAS*.

[Rebuffi et al.; 2017] Rebuffi, S., Kolesnikov, A., Sperl, G., and Lampert, C. H. (2017). iCaRL: Incremental classifier and representation learning. In *Proc. CVPR*.

[Retsinas et al.; 2020] Retsinas, G., Filntisis, P. P., Efthymiou, N., Theodosis, E., Zlatintsi, A., and Maragos, P. (2020). Person identification using deep convolutional neural networks on short-term signals from wearable sensors. In *Proc. ICASSP*.

[Reyes-Ortiz et al.; 2014] Reyes-Ortiz, J. L., Oneto, L., Ghio, A., Samá, A., Anguita, D., et al. (2014). Human activity recognition on smartphones with awareness of basic activities and postural transitions. In *Proc. ICANN*.

[Richman and Moorman; 2000] Richman, J. S. and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American J. Physiology-Heart and Circulatory Physiology*, 278:2039--2049.

[Robinson et al.; 2020] Robinson, N. L., Connolly, J., Hides, L., and Kavanagh, D. J. (2020). A social robot to deliver an 8-week intervention for diabetes management: Initial test of feasibility in a hospital clinic. In *Proc. ICSR*.

[Robokind. Advanced Social Robots.;] Robokind. Advanced Social Robots. http://robokind.com/.

[Robotham et al.; 2016] Robotham, D., Satkunanathan, S., Doughty, L., and Wykes, T. (2016). Do we still have a digital divide in mental health? A five-year survey follow-up. *J. Medical Internet Research*, 18:e309.

[Rusu et al.; 2016] Rusu, A. A., Rabinowitz, N. C., Desjardins, G., Soyer, H., Kirkpatrick, J., et al. (2016). Progressive neural networks. *arXiv preprint arXiv:1606.04671*.

[Saeb et al.; 2015] Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., et al. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of Medical Internet Research*, 17:e175.

[Saerbeck et al.; 2010] Saerbeck, M., Schut, T., Bartneck, C., and Janse, M. (2010). Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In *Proc. CHI*.

[Scargle; 1982] Scargle, J. D. (1982). Studies in astronomical time series analysis. II-statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835--853.

[Schizophrenia;] Schizophrenia. Schizophrenia.com. http://schizophrenia.com/.

[Senft et al.; 2018] Senft, E., Lemaignan, S., Bartlett, M., Baxter, P., Belpaeme, T., et al. (2018). Robots in the classroom: Learning to be a good tutor. In *Proc. HRI-W, R4L*.

[Shaffer and Ginsberg; 2017] Shaffer, F. and Ginsberg, J. (2017). An overview of heart rate variability metrics and norms. *Frontiers Public Health*, 5.

[Sharkey; 2016] Sharkey, A. (2016). Should we welcome robot teachers? *Ethics and Information Technology*, 18:283--297.

[Shin et al.; 2017] Shin, H., Lee, J., Kim, J., and Kim, J. (2017). Continual learning with deep generative replay. In *Proc. NIPS*.

[Shiomi et al.; 2015] Shiomi, M., Kanda, T., Howley, I., Hayashi, K., and Hagita, N. (2015). Can a social robot stimulate science curiosity in classrooms? *Intl. J. Social Robotics*, 7:641--652.

[Simonyan and Zisserman; 2014] Simonyan, K. and Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. In *Proc. NIPS*.

[Skantze and Al Moubayed; 2012] Skantze, G. and Al Moubayed, S. (2012). IrisTK: a statechart-based toolkit for multi-party face-to-face interaction. In *Proc. ICMI*.

[Smith; 2009] Smith, P. K. (2009). *Children and play: Understanding children's worlds*. John Wiley & Sons.

[SoftGrip;] SoftGrip. Soft Robotic Greaper. https://www.softgrip-project.eu/.

[Soomro et al.; 2012] Soomro, K., Zamir, A. R., and Shah, M. (2012). UCF101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*.

[SPGC e-Prevention I; 2023] SPGC e-Prevention I (2023). Signal Processing Grand Challenge e-Prevention I, ICASSP. https://robotics.ntua.gr/eprevention-sp-challenge/.

[Szegedy et al.; 2016] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proc. CVPR*.

[Tapus et al.; 2019] Tapus, A., Bandera, A., Vazquez-Martin, R., and Calderita, L. V. (2019). Perceiving the person and their interactions with the others for social robotics-a review. *Pattern Recognition Letters*, 118:3--13.

[Torous et al.; 2016] Torous, J., Kiang, M. V., Lorme, J., and Onnela, J.-P. (2016). New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Mental Health*, 3:e16.

[Tran et al.; 2015] Tran, D., Bourdev, L., Fergus, R., Torresani, L., and Paluri, M. (2015). Learning spatiotemporal features with 3d convolutional networks. In *Proc. ICCV*.

[Tran et al.; 2017] Tran, D., Ray, J., Shou, Z., Chang, S. F., and Paluri, M. (2017). ConvNet architecture search for spatiotemporal feature learning. *arXiv preprint arXiv:1708.05038*.

[Tsiami et al.; 2018a] Tsiami, A., Filntisis, P. P., Efthymiou, N., Koutras, P., Potamianos, G., et al. (2018a). Far-field audio-visual scene perception of multi-party human-robot interaction for children and adults. In *Proc. ICASSP*.

[Tsiami et al.; 2018b] Tsiami, A., Koutras, P., Efthymiou, N., Filntisis, P. P., Potamianos, G., et al. (2018b). Multi3: Multi-sensory perception system for multi-modal child interaction with multiple robots. In *Proc. ICRA*.

[Turkle et al.; 2006] Turkle, S., Breazeal, C., Dasté, O., and Scassellati, B. (2006). Encounters with kismet and cog: Children respond to relational artifacts. *Digital media: Transformations in human communication*, 120.

[Valenza et al.; 2014] Valenza, G., Nardelli, M., Lanat`a, A., Gentili, C., Bertschy, G., et al. (2014). Wearable monitoring for mood recognition in bipolar disorder based on history-dependent long-term heart rate variability analysis. *IEEE Jour. Of Biomedical and Health Informatics*, 18:1625--1635.

[Valipour et al.; 2017] Valipour, S., Perez, C., and Jagersand, M. (2017). Incremental learning for robot perception through HRI. In *Proc. IROS*.

[Van de Ven et al.; 2020] Van de Ven, G. M., Siegelmann, H. T., and Tolias, A. S. (2020). Brain-inspired replay for continual learning with artificial neural networks. *Nature communications*, 11:4069.

[Vantuch; 2018] Vantuch, T. (2018). Analysis of time series data.

[Vaswani et al.; 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., et al. (2017). Attention is all you need. In *Proc. NIPS*.

[Viet Tuyen et al.; 2018] Viet Tuyen, N. T., Jeong, S., and Chong, N. Y. (2018). Emotional bodily expressions for culturally competent robots through long term human-robot interaction. In *Proc. IROS*.

[Villalba-Diez et al.; 2019] Villalba-Diez, J., Schmidt, D., Gevers, R., Ordieres-Meré, J., Buchwitz, M., et al. (2019). Deep learning for industrial computer vision quality control in the printing industry 4.0. *Sensors*, 19:3987.

[Voss et al.; 2006] Voss, A., Baier, V., Schulz, S., and Bar, K. (2006). Linear and nonlinear methods for analyses of cardiovascular variability in bipolar disorders. *Bipolar disorders*, 8:441--452.

[Vouloutsi et al.; 2016] Vouloutsi, V., Blancas, M., Zucca, R., Omedas, P., Reidsma, D., et al. (2016). Towards a synthetic tutor assistant: the EASEL project and its architecture. In *Proc. Biomimetic and Biohybrid Systems*.

[Wallkötter et al.; 2020] Wallkötter, S., Stower, R., Kappas, A., and Castellano, G. (2020). A robot by any other frame: framing and behaviour influence mind perception in virtual but not real-world environments. In *Proc. HRI*.

[Wang et al.; 2009] Wang, H., Ullah, M., Klaser, A., Laptev, I., and Schmid, C. (2009). Evaluation of local spatio-temporal features for action recognition. In *Proc. BMVC*.

[Wang et al.; 2011] Wang, H., Klaser, A., Schmid, C., and Liu, C. (2011). Action recognition by dense trajectories. In *Proc. CVPR*.

[Wang et al.; 2016a] Wang, L., Xiong, Y., Wang, Z., Qiao, Y., Lin, D., et al. (2016a). Temporal segment networks: Towards good practices for deep action recognition. In *Proc. ECCV*.

[Wang et al.; 2016b] Wang, R., Aung, M. S., Abdullah, S., Brian, R., Campbell, A. T., et al. (12--16 September 2016b). CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In *Proc. UbiComp*.

[Wang et al.; 2022] Wang, W., Sun, L., Liu, T., and Lai, T. (2022). The use of e-health during the COVID-19 pandemic: a case study in china's hubei province. *Health Sociology Review*, 31(3):215--231.

[Wang et al.; 2018] Wang, X., Girshick, R., Gupta, A., and He, K. (2018). Non-local neural networks. In *Proc. CVPR*.

[Wang et al.; 2019] Wang, Y., Cang, S., and Yu, H. (2019). A survey on wearable sensor modality centred human activity recognition in health care. *Expert Systems with Applications*, 137:167--190.

[Welling; 2009] Welling, M. (2009). Herding dynamical weights to learn. In *Proc. ICML*.

[Wen et al.; 2018] Wen, D., Wei, Z., Zhou, Y., Li, G., Zhang, X., et al. (2018). Deep learning methods to process fmri data and their application in the diagnosis of cognitive impairment: a brief overview and our opinion. *Frontiers in neuroinformatics*, 12:23.

[WHO;] WHO. World Health Organization. https://www.who.int/.

[Wickstrøm et al.; 2022] Wickstrøm, K.and Kampffmeyer, M., Mikalsen, K. Ø., and Jenssen, R. (2022). Mixing up contrastive learning: Self-supervised representation learning for time series. *Pattern Recognition Letters*, 155:54--61.

[Wiersma et al.; 1995] Wiersma, D., Nienhuis, F. J., Slooff, C. J., and Giel, R. (1995). Prodromes and precursors: Epidemiologic data for primary prevention of disorders with slow onset. *The American J. psychiatry*, 152:967.

[Wiersma et al.; 1998] Wiersma, D., Nienhuis, F. J., Slooff, C. J., and Giel, R. (1998). Natural course of schizophrenic disorders: a 15-year follow up of a dutch incidence cohort. *Schizophrenia bulletin*, 24:75--85.

[Wolfe et al.; 2018] Wolfe, E., Weinberg, J., and Hupp, S. (2018). Deploying a social robot to co-teach social emotional learning in the early childhood classroom. In *Proc. HRI*.

[Wood et al.; 2017] Wood, L., Dautenhahn, K., Robins, B., and Zaraki, A. (2017). Developing child-robot interaction scenarios with a humanoid robot to assist children with autism in developing visual perspective taking skills. In *Proc. RO-MAN*.

[Wright et al.; 2017] Wright, M. N., Dankowski, T., and Ziegler, A. (2017). Unbiased split variable selection for random survival forests using maximally selected rank statistics. *Statistics in medicine*, 36:1272--1284.

[Wu and Tu; 2023] Wu, J. and Tu, M. (2023). A person identification system for the ICASSP 2023 e-prevention Challenge. In *Proc. ICASSP*.

[Wu et al.; 2019] Wu, Q., Wang, S., Cao, J., He, B., Yu, C., et al. (2019). Object recognition-based second language learning educational robot system for chinese preschool children. *IEEE Access*, 7:7301--7312.

[Xu et al.; 2015] Xu, Z., Yang, Y., and Hauptmann, A. G. (2015). A discriminative CNN video representation for event detection. In *Proc. CVPR*.

[Yan et al.; 2014] Yan, H., Ang, M. H., and Poo, A. N. (2014). A survey on perception methods for human-robot interaction in social robots. *International Journal of Social Robotics*, 6:85--119.

[Yeung and Alwan; 2018] Yeung, G. and Alwan, A. (2018). On the difficulties of automatic speech recognition for kindergarten-aged children. *Proc. INTERSPEECH*.

[Yin et al.; 2022] Yin, W., Li, L., and Wu, F. (2022). Deep learning for brain disorder diagnosis based on fmri images. *Neurocomputing*, 469:332--345.

[Zaraki et al.; 2017] Zaraki, A., Pieroni, M., De Rossi, D., Mazzei, D., Garofalo, R., et al. (2017). Design and evaluation of a unique social perception system for human-robot interaction. *IEEE Trans. on Cognitive and Developmental Systems*, 9:341--355.

[Zhang et al.;] Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*.

[Zhang et al.; 2016] Zhang, H., Wu, P., Beck, A., Zhang, Z., and Gao, X. (2016). Adaptive incremental learning of image semantics with application to social robot. *Neurocomputing*, 173:93--101.

[Zhang et al.; 2020] Zhang, J., Wu, F., Wei, B., Zhang, Q., Huang, H., et al. (2020). Data augmentation and dense-lstm for human activity recognition using wifi signal. *IEEE Internet of Things Journal*, 8:4628--4641.

[Zhang et al.; 2022] Zhang, X., Zhao, Z., Tsiligkaridis, T., and Zitnik, M. (2022). Self-supervised contrastive pre-training for time series via time-frequency consistency. *Proc. NeurIPS*.

[Zhang et al.; 2021] Zhang, Y., Tian, Y., Wu, P., and Chen, D. (2021). Application of skeleton data and long short-term memory in action recognition of children with autism spectrum disorder. *Sensors*, 21:411.

[Zlatintsi et al.; 2022] Zlatintsi, A., Filntisis, P. P., Garoufis, C., Efthymiou, N., Maragos, P., Menychtas, A., Maglogiannis, I., Tsanakas, P., Sounapoglou, T., Kalisperakis, E., Karantinos, T., Lazaridi, M., Garyfalli, V., Mantas, A., Mantonakis, L., and Smyrnis, N. (2022). E-Prevention: Advanced support system for monitoring and relapse prevention in patients with psychotic disorders analyzing long-term multimodal data from wearables and video captures. *Sensors*, 22:7544.

[Zou et al.; 2019] Zou, H., Yang, J., Prasanna Das, H., Liu, H., Zhou, Y., et al. (2019). WiFi and vision multimodal learning for accurate and robust device-free human activity recognition. In *Proc. CVPR*.