



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τηλεπικοινωνιών
Μονάδα Επεξεργασίας Σήματος και Βιοϊατρικής Τεχνολογίας

Αναγνώριση Συναισθημάτων από Περιφερειακά
Φυσιολογικά Σήματα σε Δυαδικές Συζητήσεις

Διπλωματική Εργασία
του
Σωτήριου Μόσχου
ΑΕΜ: 9030

Επιθετικός: Λεόντιος Χατζηλεοντιάδης
Καθηγητής, Α.Π.Θ.
Ηλίας Κίτσας
Ε.ΔΙ.Π., Α.Π.Θ.

Θεσσαλονίκη, 17 Ιουλίου 2022



Aristotle University of Thessaloniki
Faculty of Engineering
Department of Electrical and Computer Engineering
Specialization of Telecommunications
Signal Processing and Biomedical Technology Unit

Emotion Recognition from Peripheral Physiological Signals in Dyadic Conversations

Diploma Thesis
of
Sotirios Moschos
Department ID: 9030

Supervisors: Leontios Hadjileontiads
Professor, A.U.Th.
Ilias Kitsas
L.TE.S., A.U.Th.

Thessaloniki, July 17, 2022

Περίληψη

Ο στόχος της έρευνας γύρω από την αυτόματη αναγνώριση συναισθημάτων είναι να εμπλουτίσει στις μηχανές συναισθηματική νοημοσύνη, επιτρέποντάς τους να κατανοήσουν τα ανθρώπινα συναισθήματα. Η αναγνώριση συναισθημάτων μπορεί να προσεγγιστεί χρησιμοποιώντας σήματα εξωτερικευμένων διαδικασιών, όπως η ομιλία και οι εκφράσεις του προσώπου, ή εσωτερικευμένων διαδικασιών, αναλύοντας το ηλεκτρο-εγκεφαλογράφημα. Τα περισσότερα από τα μοντέλα αυτόματης αναγνώρισης συναισθημάτων βασίζονται σε σύνολα δεδομένων από εικόνες, ήχο, βίντεο και εγκεφαλικά σήματα και συλλέγονται σε εργαστηριακά περιβάλλοντα. Σήμερα, οι τεχνολογικές εξελίξεις αναφορικά με τις φορητές συσκευές έχουν επιτρέψει τη συνεχή παρακολούθηση των φυσιολογικών σημάτων. Υπάρχει μια αναδυόμενη ανάγκη για μοντέλα αυτόματης αναγνώρισης συναισθημάτων που επικεντρώνονται στη χρήση μόνο περιφερειακών φυσιολογικών σημάτων. Στην παρούσα διπλωματική εργασία, παρουσιάζεται μια μέθοδος ταξινόμησης συναισθημάτων που βασίζεται σε νευρωνικά δίκτυα αξιοποιώντας περιφερειακά φυσιολογικά σήματα, που λαμβάνονται από φορητές συσκευές. Μια αρχιτεκτονική ταξινόμησης βασισμένη σε δίκτυο μακράς βραχυπρόθεσμης μνήμης, εμπλουτισμένο με ένα επίπεδο προσοχής, προτείνεται για την ακριβή πρόβλεψη των συναισθημάτων σε πραγματικό χρόνο σε 2 κλάσεις, που ορίζονται από τον δισδιάστατο χώρο διέγερσης-σθένους. Το σύνολο δεδομένων K-EmoCon χρησιμοποιήθηκε στην παρούσα διπλωματική εργασία και τα δεδομένα συλλέχθηκαν από 16 συνεδρίες νατουραλιστικών συζητήσεων περίπου 10 λεπτών, σε ζευγάρια μεταξύ συμμετεχόντων, για ένα κοινωνικό ζήτημα. Το σύνολο δεδομένων περιλαμβάνει επισημάνσεις συναισθημάτων και από τις τρεις διαθέσιμες οπτικές γωνίες: του συμμετέχοντα, του συνομιλητή, των εξωτερικών παρατηρητών και αναλύθηκε ο αντίκτυπός τους στην απόδοση του μοντέλου. Τα αποτελέσματα δείχνουν ότι η μέθοδος μας έχει μετρημένη ακρίβεια $\approx 91\%$ για την δυαδική ταξινόμηση, ξεπερνώντας σε επίδοση μια παραδοσιακή προσέγγιση μηχανικής μάθησης και μια απλοποιημένη αρχιτεκτονική ταξινόμησης βασισμένη σε δίκτυο μακράς βραχυπρόθεσμης μνήμης, στο ίδιο σύνολο δεδομένων.

Λέξεις Κλειδιά: Επιβλεπόμενη Μάθηση, Αναγνώριση Συναισθήματος, Φυσιολογικά Σήματα, K-EmoCon, Εξαγωγή Χαρακτηριστικών, Δίκτυο Μακράς Βραχυπρόθεσμης Μνήμης, Επίπεδο Προσοχής

Abstract

The goal of automated emotion recognition research is to endow machines with emotional intelligence, allowing them to understand human emotions. Recognizing emotions can be approached using signals of external manifestations, such as speech and facial expressions, or internal manifestations, by analyzing EEG. Most of the affective computing models are based on images, audio, videos and brain signals, collected in a laboratory environment. Nowadays, advancements in mobile computing and wearable technologies have enabled the continuous monitoring of physiological signals. There is an emerging necessity of affective computing models that focus on utilizing only peripheral physiological signals for emotion recognition. In this study, an emotion classification method based on neural networks and exploiting peripheral physiological signals, obtained by wearable devices, is presented. An Attention-based Long Short-Term Memory classification architecture is proposed to accurately predict emotions in real-time into binary levels of the 2 dimensional arousal-valence space. K-EmoCon dataset was used throughout the project and the data were gathered from 16 sessions of approximately 10-minute long paired debates between 2 participants, on a social issue. The dataset includes emotion annotations from all three available perspectives: self, debate partner, external observers and their impact on classification performance was analyzed. Results show that our method has a measured accuracy of $\approx 91\%$ for binary classification, outperforming a traditional machine learning approach and a simplified Long Short-Term Memory classification architecture on the same task.

Keywords: Supervised Learning, Emotion Recognition, Physiological Signals, K-EmoCon, Feature Extraction, Long Short-Term Memory Networks, Attention Mechanism

Ευχαριστίες

Με την ολοκλήρωση της διπλωματικής μου εργασίας, θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε όλους όσους συνέβαλλαν στην εκπόνησή της.

Ευχαριστώ θερμά τους επιβλέποντες καθηγητές μου, κύριο Λεόντιο Χατζηλεοντιάδη και Ηλία Κίτσα, για την εμπιστοσύνη που μου έδειξαν εξ' αρχής, αναθέτοντάς μου το συγκεκριμένο θέμα, την επιστημονική τους καθοδήγηση, τις υποδείξεις τους, την υποστήριξή τους και το αμείωτο ενδιαφέρον που έδειξαν από την αρχή μέχρι το τέλος της διπλωματικής μου εργασίας.

Θα ήθελα επίσης να εκφράσω την ευγνωμοσύνη μου στην οικογένειά μου για όλη τη στήριξη, τη συμπαράσταση και την κατανόησή τους, καθ' όλη τη διάρκεια των σπουδών μου.

Τέλος, θα ήθελα να ευχαριστήσω τους κοντινούς μου φίλους οι οποίοι με τη παρουσία τους έκαναν τη φοιτητική μου ζωή πιο ευχάριστη.

Αναγνώριση Συναισθημάτων από Περιφερειακά Φυσιολογικά Σήματα σε Δυαδικές Συζητήσεις

Σωτήριος Μόσχος
moschoss@ece.auth.gr

17 Ιουλίου 2022

Περιεχόμενα

Κατάλογος Σχημάτων	3
Κατάλογος Πινάκων	5
1 Εισαγωγή	6
1.1 Κίνητρο	6
1.2 Περιγραφή του Προβλήματος	7
1.3 Καινοτομία και Συνεισφορά	8
1.4 Δομή της Διπλωματικής Έργασίας	8
2 Θεωρητικό υπόβαθρο	9
2.1 Μηχανική Μάθηση	9
2.2 Ταξινόμηση	10
2.2.1 Μηχανές Διανυσμάτων Υποστήριξης	11
2.2.2 Προσεγγίσεις βασισμένες σε Δέντρα Απόφασης	12
2.2.2.1 Τυχαία Δάση	13
2.2.2.2 Η τεχνική XGBoost	13
2.2.3 Προβλήματα Αλγορίθμων Ταξινόμησης	14
2.3 Βαθιά Μάθηση	15
2.3.1 Νευρωνικά Δίκτυα Τροφοδοσίας προς τα εμπρός	15
2.3.2 Εκμάθηση με βάση το Δυναμικό	17
2.3.2.1 Βελτιστοποίηση με βάση το Δυναμικό	18
2.3.2.2 Συνάρτηση Κόστους	19
2.3.2.3 Συναρτήσεις Ενεργοποίησης	20
2.3.2.4 Αλγόριθμος οπισθοδιαδοσης	21
2.3.3 Νευρωνικά Δίκτυα Ανατροφοδότησης	22
2.3.3.1 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης	24
2.4 Μετρικές Αξιολόγησης Μοντέλων	26
2.5 Δυναμικά Χαρακτηριστικά Συναισθημάτων	28
3 Βιβλιογραφική Επισκόπηση των Νεότερων Τάσεων της Έρευνας	29
3.1 Θεωρίες συναισθήματος και επισκόπηση των νεότερων εξελίξεων στο πεδίο	29
4 Μεθοδολογία	32
4.1 Περιγραφή Συνόλου Δεδομένων	32
4.1.1 Συλογή Συνόλου Δεδομένων	33
4.1.2 Περιεχόμενα Συνόλου Δεδομένων	35
4.2 Προ-επεξεργασία Συνόλου Δεδομένων	37

4.2.1	Εξαγωγή Χαρακτηριστικών από Φυσιολογικά Σήματα Ενδιαφέροντος	38
4.2.2	Δειγματοληπτικός Συγχρονισμός Φυσιολογικών Σημάτων	41
4.3	Προτεινόμενες Αρχιτεκτονικές Νευρωνικών	
Δικτύων	43
4.3.1	Αρχιτεκτονική ενός επιπέδου μακράς βραχυπρόθεσμης μνήμης	43
4.3.2	Εμπλουτισμός αρχιτεκτονικής με ένα επίπεδο προσοχής	44
5	Θέματα πειραματικής υλοποίησης και αποτελέσματα	46
5.1	Σύγκριση Συναισθηματικών Επισημάνσεων Διαφορετικών Οπτικών Γωνιών	46
5.2	Αξιολόγηση με Διασταυρωμένη Επικύρωση	57
5.3	Παράμετροι πειραμάτων και υπερπαράμετροι μοντέλων	58
5.4	Αποτελέσματα	60
5.4.1	Μηχανές Διανυσμάτων Υποστήριξης	60
5.4.2	XGBoost	60
5.4.3	Αρχιτεκτονική ενός επιπέδου μακράς βραχυπρόθεσμης μνήμης .	61
5.4.4	Εμπλουτισμένη αρχιτεκτονική με ένα επίπεδο προσοχής	61
6	Συμπεράσματα και Μελλοντικές Προεκτάσεις	62
6.1	Σχολιασμός αποτελεσμάτων	62
6.2	Περιορισμοί	63
6.3	Μελλοντικές προεκτάσεις	64
6.3.1	Επεκτάσεις Μηχανισμού Προσοχής	64
6.3.2	Δυναμικά Χαρακηριστικά Συναισθημάτων ως επιπλέον πληροφορία	64
Α'	Διαγράμματα	65
Β'	Ακρωνύμια και συντομογραφίες	67
Γ'	Ονοματολογία	69
	Βιβλιογραφία	70

Κατάλογος Σχημάτων

2.1	Μια απλή γραμμική διανυσματική μηχανή υποστήριξης	11
2.2	Ένα παράδειγμα δομής δέντρου αποφάσεων.	12
2.3	Ένα παράδειγμα δομής τυχαίων δασών.	13
2.4	Συχνοί λόγοι εμφάνισης προβλημάτων σε αλγόριθμους ταξινόμησης. . .	14
2.5	Παράδειγμα δικτύου MLP.	16
2.6	Παράδειγμα δομής επιπέδων ενός νευρωνικού δικτύου.	17
2.7	5 βήματα από τα αριστερά στα δεξιά του αλγορίθμου κατάβασης δυναμικού. 18	18
2.8	Παραδείγμα πιο συχνά χρησιμοποιούμενων συναρτήσεων ενεργοποίησης κρυφών επιπέδων.	21
2.9	Γράφος υπολογισμού για τον υπολογισμό παραγώγων χρησιμοποιώντας τον κανόνα της αλυσίδας	22
2.10	Τεχνητό νευρωνικό δίκτυο με επαναλαμβανόμενες συνδέσεις.	22
2.11	Αναπαράσταση λειτουργιών RNN.	23
2.12	Αρχιτεκτονικές ανατροφοδοτούμενων νευρωνικών δικτύων.	24
2.13	Παραλλαγές αρχιτεκτονικών κελιών ανατροφοδοτούμενων νευρωνικών δικτύων.	26
2.14	Πίνακας Σύγχυσης δυαδικού ταξινομητή.	27
3.1	Δισδιάστο μοντέλο του Russell με την τοποθέτηση συναισθηματικών καταστάσεων κατά προσέγγιση [1]	31
3.2	Τροχός συναισθημάτων του Plutchik.	31
4.1	Η εικόνα στα αριστερά δείχνει ένα ζευγάρι συμμετεχόντων που κάθονται σε ένα τραπέζι και προετοιμάζονται για μια συζήτηση. Δύο κινητά τηλέφωνα τοποθετήμενα σε τρίποδα στη μέση του τραπεζιού κατέγραψαν τις εκφράσεις του προσώπου των συμμετεχόντων και τις κινήσεις του άνω μέρους του σώματός τους, όπως φαίνεται στο χαρακτηριστικό παράδειγμα της δεξιάς εικόνας.	32
4.2	Σπιγμιότυπο από την καταγραφή της συζήτησης 2 συμμετεχόντων ταυτόχρονα από 2 κινητά τηλέφωνα.	33
4.3	Φορητές συσκευές καταγραφής τοποθετημένες σε έναν από τους συμμετέχοντες κατά τη διάρκεια συνεδριών συλλογής δεδομένων.	34
4.4	Υποδειγματοληψία BVP	42
4.5	Αναδειγματοληψία ECG	42
4.6	Αρχιτεκτονική ενός επιπέδου μακράς βραχυπρόθεσμης μνήμης	44
4.7	Εμπλουτισμός με ένα επίπεδο προσοχής	45
5.1	Οπτική γωνία συμμετέχοντα	47
5.2	Οπτική γωνία συνομηλιτή	48

5.3 Οπτική γωνία εξωτερικών παρατηρητών	48
5.4 Κατανομές μέσης τιμής θετικής συναισθηματικής κατάστασης	49
5.5 Κατανομές μέσης τιμής αρνητικής συναισθηματικής κατάστασης	50
5.6 Κατανομές διακύμανσης θετικής συναισθηματικής κατάστασης	51
5.7 Κατανομές διακύμανσης αρνητικής συναισθηματικής κατάστασης	52
5.8 Κατανομές μέσης τετραγωνικής διαδοχικής διαφοράς θετικής συναισθηματικής κατάστασης	53
5.9 Κατανομές μέσης τετραγωνικής διαδοχικής διαφοράς αρνητικής συναισθηματικής κατάστασης	54
5.10 Κατανομές κλίσης αυτοπαλινδρόμησης θετικής συναισθηματικής κατάστασης	55
5.11 Κατανομές κλίσης αυτοπαλινδρόμησης αρνητικής συναισθηματικής κατάστασης	56
5.12 Παράδειγμα διασταυρωμένης επικύρωσης 5-Πτυχών	58
 A'.1 Διάγραμμα στάδιου προεπεξεργασίας	65
A'.2 Διάγραμμα στάδιου εκπαίδευσης	65
A'.3 Διάγραμμα στάδιου πρόβλεψης	65
A'.4 Διάγραμμα εκπαίδευσης αρχιτεκτονικών νευρωνικών με διασταυρωμένη επικύρωση Κ-Πτυχών	66

Κατάλογος Πινάκων

4.1	Φυσιολογικά σήματα που συλλέγονται με φορητές συσκευές στο K-EmoCon [2], με τους αντίστοιχους ρυθμούς δειγματοληψίας και τα εύρη των σημάτων.	34
4.2	Σύνοψη περιεχομένων συνόλου δεδομένων.	35
4.3	Σύνοψη κατηγοριών συναισθημάτων που χρησιμοποιήθηκαν για επισήμανση.	36
4.4	Ταύτοτητα συμμετέχοντος και λόγος αποκλεισμού από την συνέχεια της ανάλυσης.	38
5.1	Αποτελέσματα SVM για δυαδική ταξινόμηση διέγερσης.	60
5.2	Αποτελέσματα SVM για δυαδική ταξινόμηση σθένους.	60
5.3	Αποτελέσματα XGBoost για δυαδική ταξινόμηση διέγερσης.	60
5.4	Αποτελέσματα XGBoost για δυαδική ταξινόμηση σθένους.	60
5.5	Αποτελέσματα αρχιτεκτονικής ενός επιπέδου μακράς βραχυπρόθεσμης μνήμης για δυαδική ταξινόμηση διέγερσης.	61
5.6	Αποτελέσματα αρχιτεκτονικής ενός επιπέδου μακράς βραχυπρόθεσμης μνήμης για δυαδική ταξινόμηση σθένους.	61
5.7	Αποτελέσματα εμπλουτισμένης αρχιτεκτονικής με ένα επίπεδο προσοχής για δυαδική ταξινόμηση διέγερσης.	61
5.8	Αποτελέσματα εμπλουτισμένης αρχιτεκτονικής με ένα επίπεδο προσοχής για δυαδική ταξινόμηση σθένους.	61

Κεφάλαιο 1

Εισαγωγή

1.1 Κίνητρο

Ο στόχος της αυτόματης αναγνώρισης συναισθημάτων είναι να εμπλουτίσει τις μηχανές με συναισθηματική νοημοσύνη [3], επιτρέποντας την κατανόηση των ανθρώπινων συναισθημάτων. Αυτή η ικανότητα των μηχανών ενδέχεται να έχει πολλές χρήσεις, που κυμαίνονται από ιατρικές εφαρμογές, συμπεριλαμβανομένων της πρόγνωσης, της διάγνωσης και της θεραπείας ψυχικών ασθενειών [4], την ανάπτυξη πιο ανθρωποειδών εικονικών ευφυών βοηθών ικανών να αναγνωρίζουν, να κατανοούν, και πιθανότατα να εκφράζουν συναισθήματα [5], καθώς και την προσθήκη απλών συναισθηματικών δεξιοτήτων σε έξυπνες οικιακές συσκευές για την βελτίωση της κοινωνικής τους αλληλεπίδρασης με τους χρήστες [6]. Η ιδέα του να δώσεις σε μια μηχανή ένα ανθρώπινο χαρακτηριστικό, όπως η συναισθηματική νοημοσύνη, ήταν από καιρό ένας επιδιωκόμενος στόχος, και πρόσφατα αναπτύχθηκε σε δικό της κλάδο στην επιστήμη των υπολογιστών, από τότε που η Rosalind Picard δημοσίευσε το βίβλιο "Affective Computing" [7]. Μέχρι στιγμής, πληθώρα ερευνητών έχει συμβάλει στο συγκεκριμένο πεδίο, με στόχο την ανάπτυξη συστημάτων ικανών να αναγνωρίζουν συναισθήματα από πολλαπλές πηγές, με ορισμένες να αποτελούν τις εκφράσεις του προσώπου [8], τις χειρονομίες [9] και το κείμενο [10]. Πρόσφατα, με τη ραγδαία ανάπτυξη στον τομέα της Τεχνητής Νοημοσύνης, και συγκεκριμένα με την ταχύτητα εξελισσόμενη πρόοδο στα βαθιά νευρωνικά δίκτυα, το πεδίο ανθίζει με αλγόριθμους ικανούς για ακριβή ανίχνευση των ανθρώπινων συναισθημάτων.

1.2 Περιγραφή του Προβλήματος

Το πρόβλημα της αυτόματης αναγνώρισης συναισθημάτων απέχει πολύ από το να λυθεί, δεδομένης της εγγενούς πολυπλοκότητάς τους. Παρόλο που η ανίχνευση εξωτερικευμένων εκδηλώσεων συναισθημάτων, όπως οι εκφράσεις προσώπου και η ομιλία είναι εύκολη [11] και έχει αναπτυχθεί πληθώρα συστημάτων που βασίζονται στις παραπάνω πηγές, η χρήση τους περιορίζεται σε συγκεκριμένα σενάρια, καθώς ορισμένες διαμορφώσεις των μυών του προσώπου ή της χροιάς της ομιλίας δεν αποτυπώνουν το σύνολο της εσωτερικής κατάστασης ενός ατόμου. Σε καθημερινές συνθήκες ειδικότερα, οι εκφράσεις του προσώπου και η χροιά της ομιλίας συχνά ρυθμίζονται, με στόχο την αποτροπή της αποκαδικοποίησης της υποκείμενης συναισθηματικής κατάστασης [12]. Τα συναισθήματά είναι περισσότερα από μια σειρά εκφράσεων στα πρόσωπα, πιθανώς δεν είναι καν μετρήσιμα, ενώ ορίζονται μόνο στο πλαίσιο των κοινωνικών αλληλεπιδράσεων και σχέσεων που εμφανίστηκαν. Για το λόγο αυτό, τα συναισθήματα προτείνεται να αναγνωριστούν μέσω φυσιολογικών σημάτων, όπως του ηλεκτροκαρδιογραφήματος (ECG) και του ηλεκτροεγκεφαλογραφήματος (EEG) που δημιουργούνται ασυνείδητα από το αυτόνομο νευρικό σύστημα.

Δεδομένης της πολυπλοκότητας των συναισθημάτων, η συλλογή τους σε ελεγχόμενες εργαστηριακές συνθήκες με συγκεκριμένα ερεθίσματα που έχουν σχεδιαστεί για την πρόκληση συγκεκριμένων συναισθημάτων, είναι η πιο ευρέως χρησιμοποιούμενη προσέγγιση στο πεδίο [13] [14]. Αυτή η μέθοδος παρέχει στους ερευνητές πλήρη έλεγχο στη συλλογή δεδομένων, επιτρέποντας την αξιολόγηση συγκεκριμένων συναισθηματικών συμπεριφορών και την απόκτηση λεπτομερών δεδομένων με προηγμένες τεχνικές όπως η νευροαπεικόνιση. Ωστόσο, τα δεδομένα που παράγονται από το εργαστήριο μπορεί να γενικεύονται ελάχιστα σε ρεαλιστικά σενάρια, καθώς συχνά περιέχουν έντονες εκφράσεις πρωτότυπων συναισθημάτων που σπάνια παρατηρούνται στον πραγματικό κόσμο.

Οι εξελίξεις στις τεχνολογίες φορητών συσκευών, έχουν επιτρέψει τη συνεχή παρακολούθηση φυσιολογικών σημάτων. Οι αισθητήρες φορητών συσκευών μπορούν να παρέχουν δεδομένα φυσιολογικών σημάτων, όπως ο καρδιακός ρυθμός (HR), ο παλμός όγκου του αίματος (BVP), η ηλεκτροδερμική δραστηριότητα (EDA) καθώς και η θερμοκρασία (T), δεδομένα που μπορεί να αντανακλούν συναισθηματικές αλλαγές. Επομένως, η ανάπτυξη ενός ισχυρού συστήματος αναγνώρισης συναισθημάτων που υλοποιείται σε νατουραλιστικές συνθήκες και σενάρια και αξιοποιεί φορητούς μη επεμβατικούς αισθητήρες, καθίσταται ιδιαίτερα χρήσιμο και αποτελεί τον στόχο της παρούσας διπλωματικής εργασίας.

1.3 Καινοτομία και Συνεισφορά

Στην παρούσα διπλωματική εργασία παρουσιάζεται μια εκτενής ανάλυση για την ταξινόμηση συναισθημάτων στον δισδιάστο χώρο σθένους-διέγερσης, αξιοποιώντας περιφερειακά φυσιολογικά σήματα που λαμβάνονται με τη χρήση φορητών συσκευών κατά τη διάρκεια μιας νατουραλιστικής συζήτησης μεταξύ 2 συμμετεχόντων για ένα κοινωνικό ζήτημα. Συγκεκριμένα, η μελέτη βασίζεται στα φυσιολογικά σήματα του καρδιακού ρυθμού (HR), του παλμού όγκου του αίματος (BVP), της ηλεκτροδερμικής δραστηριότητας (EDA), καθώς και της θερμοκρασίας του σώματος (T). Η πορεία της ανάλυσης ακολούθησε τα παρακάτω βήματα.

- Ανάπτυξη μοντέλων ταξινόμησης που βασίζονται στην εξαγωγή χαρακτηριστικών από τα φυσιολογικά σήματα και χρησιμοποιούν παραδοσιακούς αλγορίθμους μηχανικής μαθήσης με σκοπό την πρόβλεψη των συναισθημάτων των συμμετεχόντων.
- Ανάπτυξη μοντέλου ταξινόμησης που βασίζεται στην αρχιτεκτονική των ανατροφοδοτούμενων νευρωνικών δικτύων και επιτρέπει τις διαδοχικές τιμές των φυσιολογικών σημάτων να μοιράζονται πληροφορίες μεταξύ τους, με σκοπό την βελτίωση των μοντέλων του πρώτου βήματος.
- Ανάπτυξη μοντέλου ταξινόμησης που βασίζεται στην αρχιτεκτονική των ανατροφοδοτούμενων νευρωνικών δικτύων και την αξιοποίηση ενός επιπέδου προσοχής, με σκόπο την παροχή προτεραιότητας στις τιμές των φυσιολογικών σημάτων που προσδίδουν περισσότερη πληροφορία για την πρόβλεψη των συναισθημάτων, για την περαιτέρω βελτίωση του μοντέλου του δεύτερου βήματος.

1.4 Δομή της Διπλωματικής Εργασίας

Η διάρθρωση της υπόλοιπης διπλωματικής εργασίας οργανώνεται στα ακόλουθα κεφάλαια :

- Στο **Κεφάλαιο 2** γίνεται η παρουσίαση των απαραίτητου θεωρητικού υποβάθρου, με σκοπό να γίνει κατανοητή η προτεινόμενη μεθοδολογία και η συλλογιστική πορεία για την ανάπτυξη των μοντέλων.
- Στο **Κεφάλαιο 3** λαμβάνει χώρα μια βιβλιογραφική επισκόπηση η οποία στοχεύει στην ανάλυση των τελευταίων τάσεων της έρευνας γύρω από το πεδίο των συναισθημάτων.
- Στο **Κεφάλαιο 4** παρουσιάζεται το σύνολο δεδομένων που χρησιμοποιήθηκε στα πλαίσια της παρούσας διπλωματικής εργασίας και η επεξεργασία στην οποία υποβλήθηκε, οι χρησιμοποιούμενες διαδικασίες για την ανάπτυξη των μοντέλων ταξινόμησης, καθώς και οι προτεινόμενες αρχιτεκτονικές.
- Στο **Κεφάλαιο 5** συζητούνται θέματα πειραματικής υλοποίησης που αφορούν τα χρησιμοποιούμενα δεδομένα, παρατίθενται τα αποτελέσματα κάθε μοντέλου ξεχωριστά και πραγματοποιείται μια σύγκριση μεταξύ τους.
- Στο **Κεφάλαιο 6** πραγματοποιείται εξαγωγή των τελικών συμπερασμάτων, επισημαίνονται οι περιορισμοί που παρουσιάστηκαν στα πλαίσια της εργασίας και αναλύεται ένα σύνολο μελλοντικών προεκτάσεων του παρόντος έργου.

Κεφάλαιο 2

Θεωρητικό υπόβαθρο

Στο παρόν κεφάλαιο θα τεθούν οι απαραίτητες θεωρητικές βάσεις, ώστε να μπορεί ο αναγνώστης να κατανοήσει με ευκολία την προτεινόμενη μεθοδολογία στο Κεφάλαιο 4. Σε πρώτο στάδιο, θα παρουσιαστούν οι θεωρητικές βάσεις του πεδίου της Μηχανικής Μάθησης και συγκεκριμένων μεθοδολογιών που χρησιμοποιήθηκαν στα πλαίσια της παρούσας διπλωματικής εργασίας, όπως της ταξινόμησης, ενώ θα δωθεί ιδιαίτερη προσοχή στους αλγορίθμους που επιλέχθηκαν για την δημιουργία των μοντέλων. Έπειτα, θα τεθούν οι θεωρητικές βάσεις του πεδίου της Βαθιάς Μάθησης, θα αναλυθεί το σύνολο των αρχιτεκτονικών που χρησιμοποιήθηκαν για την δημιουργία των μοντέλων και στη συνέχεια θα περιγραφούν οι μετρικές για τον υπολογισμό της απόδοσής τους. Τέλος, θα παρουσιαστεί το σύνολο των δυναμικών χαρακτηριστικών συναισθηματικών καταστάσεων που χρησιμοποιήθηκαν.

2.1 Μηχανική Μάθηση

Η μηχανική μάθηση (Machine Learning) είναι ένας κλάδος της τεχνητής νοημοσύνης (Artificial Intelligence) που σχετίζεται με την υλοποίηση υπολογιστικών αλγορίθμων που βελτιώνουν την απόδοση σε ένα σύστημα με βάση την «εμπειρία». Με άλλα λόγια, ένα σύστημα μηχανικής μάθησης μαθαίνει από τα διαθέσιμα δεδομένα που παρέχονται σε αυτό [15] [16], προσδιορίζοντας μοτίβα που βρίσκονται στα δεδομένα και προχωρώντας στη λήψη αποφάσεων με μηδαμινή ανθρώπινη παρέμβαση. Ο συγκεκριμένος κλάδος έχει αναπτυχθεί ραγδαία τα τελευταία χρόνια στο πλαίσιο της επιστήμης των δεδομένων και της ανάλυσής τους.

Είναι δυνατό να διαχωρίσουμε το σύνολο των αλγορίθμων μηχανικής μάθησης σε 4 κατηγορίες: **Επιβλεπόμενη Μάθηση** (Supervised Learning), **Μη-Επιβλεπόμενη Μάθηση** (Unsupervised Learning), **Ημι-Επιβλεπόμενη Μάθηση** (Semi-supervised Learning), **Ενισχυμένη Μάθηση** (Reinforcement Learning) [17]. Παρακατώ, θα αναλυθεί εν συντομίᾳ καθεμία από τις 4 κατηγορίες.

- **Η επιβλεπόμενη μάθηση** ορίζεται ως η διαδικασία εκμάθησης μιας συνάρτησης που αντιστοιχίζει μια είσοδο σε μια έξοδο και η οποία βασίζεται σε ζεύγη δειγμάτων δεδομένων εισόδου-εξόδου [18]. Τα δεδομένα εκπαίδευσης είναι επισημανσμένα με μια ετικέτα (label), η οποία αποτέλει την τιμή εξόδου για την αντίστοιχη τιμή εισόδου και χρησιμοποιούνται για την διαδικασία εκμάθησης της χαρακτηριστι-

κής σύναρτησης αντιστοίχησης. Οι πιο κοινές τεχνικές επιβλεπόμενης μάθησης είναι η ταξινόμηση (classification) και η παλίνδρομηση (regression).

- Η **μη-επιβλεπόμενη μάθηση** ορίζεται ως η διαδικασία ανάλυσης δεδομένων τα οποία δεν είναι επισημανμένα με μια ετικέτα, δεν υπάχει δηλαδή τιμή εξόδου για μια συγκεκριμένη είσοδο. Η συγκεκριμένη διαδικασία χρησιμοποιείται ευρέως για την εξαγωγή γενετικών χαρακτηριστικών, τον προσδιορισμό και την ομαδοποιήση σημαντικών τάσεων από ένα σύνολο δεδομένων [18].
- Η **ημι-επιβλεπόμενη μάθηση** αποτελεί μια «υθδριδική» διαδικασία εκμάθησης που βασίζεται στις μεθόδους της επιβλεπόμενης και μη-επιβλεπόμενης μάθησης καθώς λειτουργεί τόσο με επισημασμένα, όσο και με μη-επισημασμένα με ετικέτες σύνολα δεδομένων [18] [19]. Στον πραγματικό κόσμο, σύνολα δεδομένων επισημασμένα με ετικέτες είναι σπάνια έως δυσεύρετα, ενώ σύνολα δεδομένων χωρίς ετικέτες είναι πολυάριθμα, πράγμα που καθιστά την ημι-επιβλεπόμενη μάθηση χρήσιμη [17]. Ο απώτερος στόχος ενός ημι-επιβλεπόμενου μοντέλου μάθησης είναι να παρέχει ένα πιο ακριβές αποτέλεσμα για την πρόβλεψη τιμών εξόδου από το αποτελέσμα που παρέχει ένα μοντέλο που χρησιμοποιεί μόνο τα επισημασμένα με ετικέτες δεδομένα, εκμεταλλευόμενο και μη-επισημασμένα δεδομένα.
- Η **ενισχυμένη μάθηση** είναι θεμελιωδός διαφορετική από τις προηγούμενες προσεγγίσεις καθώς δεν χρειάζεται σύνολα δεδομένων για εκπαίδευση, αλλά αντ' αυτού μαθαίνει από μια διαδικασία «δοκιμής και ανατροφοδότησης» [20] [21]. Ο αλγόριθμος λαμβάνει ανατροφοδότηση από την ίδια την αναλυτική διαδικασία με τη μορφή «ανταμοιθών», όταν η δράση συμβάλλει στην επίτευξη του στόχου, ή αντίστροφα με την μορφή «κυρώσεων», όταν η ενέργεια δεν συμβάλλει στην επίτευξη του στόχου. Ο αλγόριθμος στοχεύει στη μεγιστοποίηση της «ανταμοιθής» και ελαχιστοποίηση της «κύρωσης» [22].

2.2 Ταξινόμηση

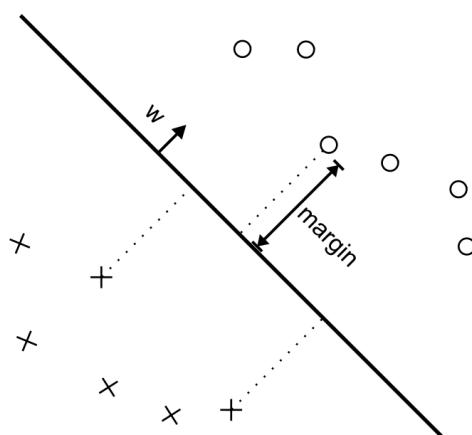
Η ταξινόμηση θεωρείται ως μια επιβλεπόμενη μέθοδος μάθησης και σχετίζεται με την μοντελοποίηση ενός προβλήματος «πρόβλεψης», όπου μια τιμή εξόδου, η οποία συχνά αναφέρεται ως ετικέτα προβλέπεται από μια τιμή εισόδου από το σύνολο δεδομένων [17]. Μαθηματικά, επικεντρώνεται στην εκμάθηση μιας συνάρτησης (f), με σκοπό την αντιστοίχηση τιμών εισόδου (x) σε τιμές εξόδου (y), ο τύπος των οποίων είναι κατηγορικός. Κατά τη διάρκεια της εκμάθησης χρησιμοποιείται ένα υποσύνολο του συνολικού συνόλου δεδομένων, το οποίο ονομάζεται σύνολο δεδομένων εκπαίδευσης, όπου οι ετικέτες εμφανίζονται στο μοντέλο, ώστε να συσχετίζει μοτίβα στις τιμές εισόδου με συγκεκριμένες κατηγορικές τιμές εξόδου. Με την ολοκλήρωση της εκμάθησης, το μοντέλο μπορεί να αξιολογηθεί σε ένα μικρό υποσύνολο του συνολικού συνόλου δεδομένων, το οποίο ονομάζεται σύνολο δεδομένων δοκιμής και η απόδοσή του καθορίζεται με χρήση κατάλληλων μετρικών. Όταν το μοντέλο προβλέπει τις τιμές των κατηγορικών τιμών εξοδού του συνόλου δεδομένων δοκιμής με υψηλή ακρίβεια, είναι δείγμα καλής γενίκευσης του μοντέλου. Παρακάτω, παρατίθενται τα πιο κοινά προβλήματα ταξινόμησης.

- Η **Δυαδική ταξινόμηση** αναφέρεται στη διαδικασία ταξινόμησης κατά την οποία οι τιμές εισόδου δύναται να αντιστοιχηθούν σε μια τιμή από ένα σύνολο κατηγοριών τιμών εξόδου, η πληθικότητα του οποίου είναι ίση με 2, όπως «αληθές και ψευδές» ή «ναι και όχι» [17].
- Η **Ταξινόμηση πολλαπλών κλάσεων** αναφέρεται στην διαδικασία ταξινόμησης κατά την οποία οι τιμές εισόδου δύναται να αντιστοιχηθούν με μια τιμή από ένα σύνολο κατηγοριών τιμών εξόδου, η πληθικότητα του οποίου είναι μεγαλύτερη του 2 [17].
- Η **Ταξινόμηση πολλαπλών ετικετών** αποτελεί μια γενικέυση της ταξινόμησης πολλαπλών κλάσεων καθώς οι τιμές εισόδου δύναται να αντιστοιχηθούν με ένα υποσύνολο του συνόλου των πολλαπλών κατηγοριών τιμών εξόδου [17].

Στις ακόλουθες υποενότητες πραγματοποιείται ανάλυση των αλγορίθμων ταξινόμησης που χρησιμοποιήθηκαν κατά την πειραματική υλοποίηση της παρούσας διπλωματικής εργασίας.

2.2.1 Μηχανές Διανυσμάτων Υποστήριξης

Οι μηχανές διανυσμάτων υποστήριξης (Support Vector Machines) [23] έχουν ισχυρά θεωρητικά θεμέλια, ενώ εφαρμόζονται άριστα σε εμπορικές εφαρμογές. Στην περίπτωση της δυαδικής ταξινόμησης μας δίνεται το σύνολο δεδομένων εκπαίδευσης, $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, τα οποία είναι διανύσματα σε κάποιο χώρο $\mathbf{X} \subseteq \mathbb{R}^d$, καθώς και οι ετικέτες τους $\{y_1, \dots, y_n\}$, όπου $y_i \in \{-1, 1\}$. Στην απλούστερη μορφή τους, οι μηχανές διανυσμάτων υποστήριξης είναι υπερεπίπεδα που χωρίζουν τα δεδομένα εκπαίδευσης με το μέγιστο περιθώριο (margin), όπως αποτυπώνεται στο Σχήμα 2.1. Όλα τα διανύσματα που βρίσκονται στη μία πλευρά του υπερεπίπεδου επισημαίνονται ως -1 και όλα τα διανύσματα που βρίσκονται από την άλλη πλευρά επισημαίνονται ως 1. Τα διανύσματα από το σύνολο των δεδομένων εκπαίδευσης που βρίσκονται πιο κοντά στο υπερεπίπεδο ονομάζονται διανύσματα υποστήριξης.



Σχήμα 2.1: Μια απλή γραμμική διανυσματική μηχανή υποστήριξης.

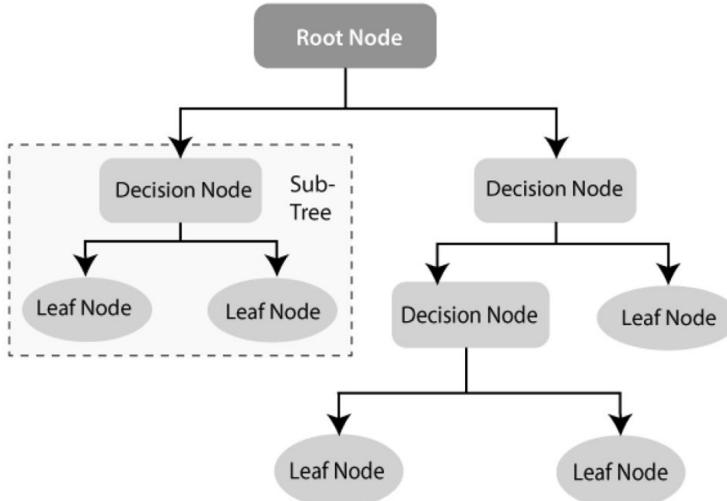
2.2.2 Προσεγγίσεις βασισμένες σε Δέντρα Απόφασης

Το δέντρο αποφάσεων (Decision Tree) [24] αποτελεί μια ευρέως διαδεδομένη μη παραμετρική επιβλεπόμενη μέθοδο μάθησης και είναι βασικό δομικό στοιχείο για πολλούς αλγόριθμους ταξινόμησης. Η ιδέα πίσω από το δέντρο αποφάσεων είναι αρκετά διαισθητική και εκπροσωπεύται καλύτερα σε οπτική μορφή, όπως απεικονίζεται στο Σχήμα 2.2. Πιο αναλυτικά, το δέντρο απόφασης αποτελείται από διαδοχικούς διακλαδιζόμενους κόμβους, όπου ο αρχικός κόμβος καλείται ρίζα του δέντρου, καθώς από αυτόν «γεννιούνται» νέοι κόμβοι, ενώ οι τερματικοί κόμβοι καλούνται φύλλα και αντιπροσωπεύουν μια κλάση. Κάθε ενδιάμεσος κόμβος χωρίζει το σύνολο των δεδομένων εισόδου σε δύο ή περισσότερα υποσύνολα, εξετάζοντας μία συγκεκριμένη συνθήκη που αφορά τις διακριτές τιμές των χαρακτηριστικών τους. Τα πιο ευρέως χρησιμοποιούμενα κριτήρια διαχωρισμού των κόμβων του δέντρου είναι ο δείκτης GINI, που απεικονίζεται στην εξίσωση 2.1 και η εντροπία που απεικονίζεται στην εξίσωση 2.2, όπου ο όρος p_j αντιπροσωπεύει την πιθανότητα ταξινόμησης στην κλάση j .

$$GINI = 1 - \sum_j p_j^2 \quad (2.1)$$

$$Entropy = - \sum_j p_j \log_2 p_j \quad (2.2)$$

Η δημιουργία υπο-κόμβων αυξάνει την ομοιογένειά τους σε σχέση με την επιθυμητή κλάση. Επομένως, μια εγγραφή από ένα σύνολο δεδομένων ταξινομείται με βάση το κριτήριο διαχωρισμού κάθε κόμβου, ξεκινώντας από την ρίζα του δέντρου απόφασης μέχρι ένα κόμβο που αποτελεί φύλλο και περιέχει μια κατηγορική τιμή εξόδου.

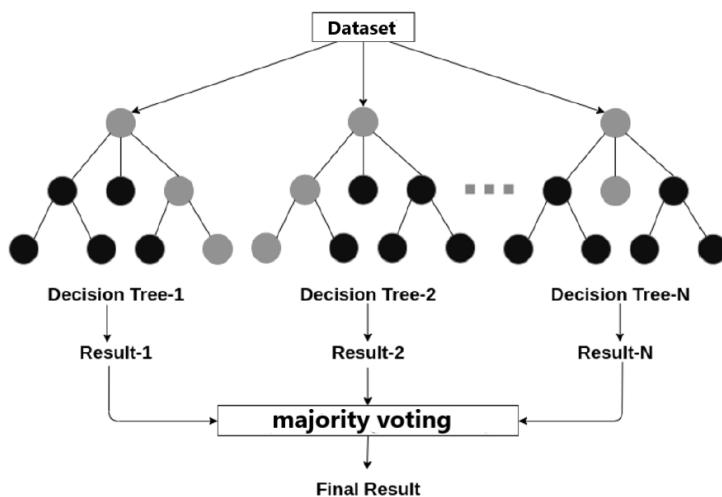


Σχήμα 2.2: Ένα παράδειγμα δομής δέντρου αποφάσεων.

Επί του παρόντος, τα δέντρα απόφασης δεν χρησιμοποιούνται ποτέ στην αρχική μορφή τους για την επίλυση προβλημάτων ταξινόμησης. Ο κύριος λόγος είναι το γεγονός ότι είναι επιρρεπή στην υπερεκπαίδευση, που αποτελεί ένα από τα κύρια προβλήματα των αλγορίθμων ταξινόμησης. Παρόλα αυτά, τα δέντρα απόφασης έγιναν το δομικό στοιχείο για δύο ευρέως χρησιμοποιούμενες τεχνικές, οι οποίες αναλύονται παρακάτω.

2.2.2.1 Τυχαία Δάση

Στην περίπτωση των τυχαίων δασών (Random Forests), κατασκευάζεται παράλληλα ένα πολύ μεγάλο πλήθος από δέντρα απόφασης, όπου το κάθε δέντρο υπερεκπαιδεύεται σε ένα διαφορετικό υποσύνολο του συνόλου δεδομένων, ξεχωριστό για κάθε δέντρο. Ωστόσο, το πρόβλημα της υπερεκπαίδευσης ελαχιστοποιείται συνδυάζοντας τα αποτελέσματα από το σύνολο των δέντρων. Επομένως, το μοντέλο που προκύπτει είναι στο σύνολο των περιπτώσεων πιο ακριβές και με μεγαλύτερη ικανότητα γενίκευσης από ένα μοντέλο που βασίζεται εξόλοκλήρου σε ένα δέντρο απόφασης [25] [26]. Στο Σχήμα 2.3 απεικονίζεται ένα χαρακτηριστικό παράδειγμα τυχαίων δασών.



Σχήμα 2.3: Ένα παράδειγμα δομής τυχαίων δασών.

2.2.2.2 Η τεχνική XGBoost

Η τεχνικής της ενίσχυσης (Boosting) στηρίζεται στην διαδοχική δημιουργία δέντρων απόφασης που έχουν χαμηλή προγνωστική ισχύ, δηλαδή τα καταφέρνουν ελαφρώς καλύτερα από την τυχαία εικασία. Σε καθέναν αυτούς τους ταξινομητές προσδίδεται ένα συγκεκριμένο σταθμισμένο υποσύνολο του αρχικού συνόλου δεδομένων, ενώ μεγαλύτερα βάρη δίνονται σε υποσύνολα που είχαν ταξινομηθεί εσφαλμένα νωρίτερα. Στη συνέχεια, οι προβλέψεις των ταξινομητών συνδυάζονται με μηχανισμούς ψηφοφορίας.

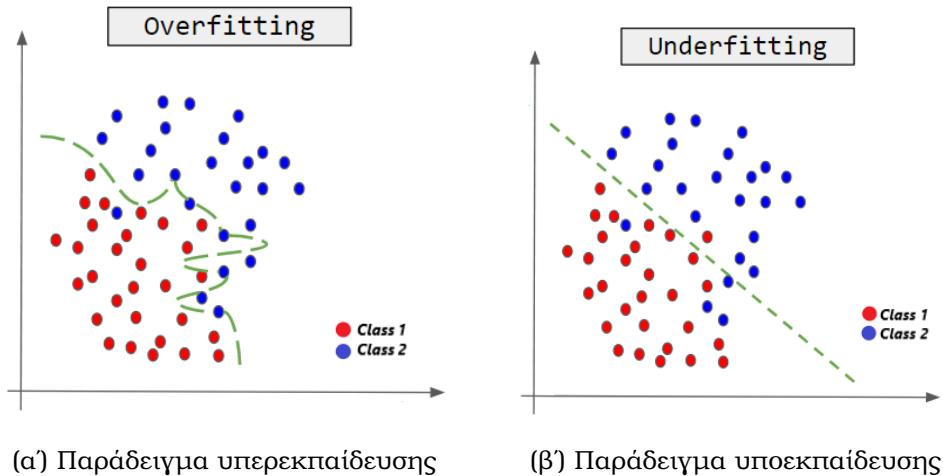
Η ενίσχυση κλίσης (Gradient Boosting) χρησιμοποιεί την κλίση της συνάρτησης απώλειών από τους επιμέρους ταξινομητές. Σε κάθε στάδιο ενίσχυσης, οι ταξινομητές χρησιμοποιούνται για την ελαχιστοποίηση της συνάρτησης απώλειας του τρέχοντος μοντέλου. Η τεχνική XGBoost λαμβάνει υπόψη την κλίση δεύτερης τάξης της συνάρτησης απώλειας και προηγμένες τεχνικές κανονικοποίησης με σκοπό την αποφυγή της υπερκπαίδευσης του μοντέλου και την επίτευξη μεγαλύτερης γενίκευσης [27].

2.2.3 Προβλήματα Αλγορίθμων Ταξινόμησης

Μια πληθώρα προβλημάτων ενδέχεται να εμφανιστεί κατά την διαδικασία δημιουργίας ενός μοντέλου ταξινόμησης. Πιο συγκεκριμένα, ένα μοντέλο ενδέχεται είτε να παρουσίασει χαμηλές τιμές μετρικών αξιολόγησης στα σύνολα δεδομένων εκπαίδευσης και ελέγχου, είτε να παρουσίασει πολύ υψηλές τιμές μετρικών αξιολόγησης στα δεδομένα εκπαίδευσης και να αδυνατεί να γενικέυσει στο σύνολο δεδομένων ελέγχου. Παρακάτω, παρατίθενται οι πιο συχνοί λόγοι εμφάνισης των παραπάνω προβλημάτων.

Υποεκπαίδευση: Το φαινόμενο της υποεκπαίδευσης εμφανίζεται όταν ένα μοντέλο ταξινόμησης αδυνατεί να συλλάβει επαρκώς την υποκείμενη δομή του συνόλου δεδομένων εκπαίδευσης με αποτέλεσμα να μην μπορεί να παράξει αξιόπιστες προβλέψεις τόσο σε αυτό, όσο και στο σύνολο δεδομένων ελέγχου. Σε αυτή την περίπτωση, το μοντέλο χαρακτηρίζεται από υψηλή διακύμανση (Variance) και χαμηλή πόλωση (Bias).

Υπερεκπαίδευση: Το φαινόμενο της υπερεκπαίδευσης εμφανίζεται όταν ένα μοντέλο ταξινόμησης προσδιόριζει το σύνολο των μοτίβων που βρίσκονται στο σύνολο δεδομένων εκπαίδευσης σε σημείο που μοντελοποιεί και το θόρυβο που εμπεριέχεται σε αυτό. Ως εκ τούτου, το μοντέλο αδυνατεί να γενικέυσει στο σύνολο δεδομένων ελέγχου με αποτέλεσμα να μην μπορεί να παράξει αξιόπιστες προβλέψεις. Σε αυτή την περίπτωση, το μοντέλο χαρακτηρίζεται από υψηλή πόλωση (Bias) και χαμηλή διακύμανση (Variance).



Σχήμα 2.4: Συχνοί λόγοι εμφάνισης προβλημάτων σε αλγόριθμους ταξινόμησης.

Ανισορροπία Κλάσεων: Το φαινόμενο της ανισορροπίας κλάσεων εμφανίζεται όταν ο αριθμός των δειγμάτων στο σύνολο δεδομένων εκπαίδευσης για κάθε ετικέτα δεν είναι ισορροπημένος, η κατανομή δηλαδή των δειγμάτων στις κλάσεις έχει μεγάλη απόκλιση και οδηγεί στο πρόβλημα της μη ισορροπημένης ταξινόμησης. Για την αντιμετώπισή του, χρησιμοποιούνται τεχνικές όπως η υπερδειγματοληψία, η υποδειγματοληψία και η ανάθεση βαρών στα δείγματα.

2.3 Βαθιά Μάθηση

Η βαθιά μάθηση (Deep Learning), η οποία με τη σειρά της αποτελεί υποσύνολο της γενικότερης έννοιας της μηχανική μάθησης, βασίζεται κυρίως σε νευρωνικά δίκτυα. Το κύριο χαρακτηριστικό των νευρωνικών δικτύων είναι η ικανότητα εξαγωγής πληροφοριών από ένα σύνολο δεδομένων χωρίς προηγούμενη γνώση και η ικανότητα σχηματισμού πολύπλοκων δομών και σχέσεων από το σύνολο δεδομένων.

Τα νευρωνικά δίκτυα τροφοδοσίας προς τα εμπρός (Feedforward Neural Networks), τα οποία είναι γνωστά στην βιβλιογραφία ως MLP, αποτελούν την ιδρυτική και πιο βασική κατηγορία μοντέλων βαθιάς μάθησης. Τα νευρωνικά δίκτυα χρησιμοποιούνται ως προσεγγιστές συναρτήσεων, όπως για παράδειγμα στην περίπτωση χρησιμοποίησης ενός νευρωνικού δικτύου για την εκμάθηση μιας διαδικασίας ταξινόμησης, $y = f^*(\mathbf{x})$, που αντιστοιχίζει ένα διάνυσμα εισόδου \mathbf{x} σε μια ετικέτα y , ένα MLP θα προσεγγίσει μια συνάρτηση που ορίζει την αντιστοίχηση $y = f^*(\mathbf{x}; \theta)$ μεταξύ εισόδου και εξόδου, μαθαίνοντας το σύνολο των παραμέτρων θ , που μας δίνουν την καλύτερη προσέγγιση. Τα νευρωνικά δίκτυα θεωρητικά μπορούν να αντιπροσωπεύσουν οποιαδήποτε συνάρτηση και για το λόγο αυτό χαρακτηρίζονται ως καθολικοί προσεγγιστές συναρτήσεων [28]. Ο λόγος που τα προαναφερόμενα μοντέλα ονομάζονται νευρωνικά δίκτυα τροφοδοσίας προς τα εμπρός είναι επειδή οι πληροφορίες «ρέουν» προς τα εμπρός μέσω του δικτύου, από το διάνυσμα εισόδου \mathbf{x} , μέσω των ενδιάμεσων υπολογισμών που χρησιμοποιούνται για τον καθορισμό της αντιστοίχισης $y = f^*(\mathbf{x}; \theta)$, μέχρι την ετικέτα y .

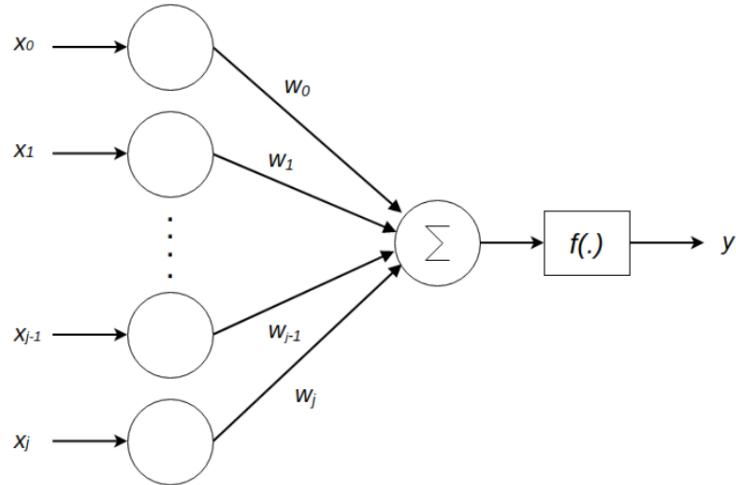
Υπάρχουν πολλοί διαφορετικοί τύποι αρχιτεκτονικών τεχνητών νευρωνικών δικτύων, όπως τα CNN, που είναι μια κατηγορία νευρωνικών δικτύων βασισμένα στην πράξη της συνέλιξης ή τα RNN, που είναι ειδικά σχεδιασμένα για σύνολα δεδομένων που αποτελούν αποτέλεσμα μιας διαδοχικής διεργασίας, λόγω των επαναλαμβανόμενων συνδέσεων που έχουν. Παρακάτω παρατίθενται πληροφορίες για το δομικό στοιχείο ενός νευρωνικού δικτύου, περισσότερες λεπτομέριες σχετικά με την διαδικασία της μάθησης καθώς και η κύρια αρχιτεκτονική που χρησιμοποιήθηκε στα πλαίσια της διπλωματικής εργασίας.

2.3.1 Νευρωνικά Δίκτυα Τροφοδοσίας προς τα εμπρός

Το βασικό δομικό στοιχείο όλων των νευρωνικών δικτύων είναι ο τεχνητός νευρώνας, γνωστός και ως perceptron. Ένας τεχνητός νευρώνας αποτελεί ουσιαστικά μια μαθηματική συνάρτηση, όπως απεικονίζεται στην εξίσωση 2.3.

$$y = f(\sum_{j=0}^n w_j x_j) \quad (2.3)$$

Το συγκεκριμένο δομικό στοιχείο λαμβάνει μία ή περισσότερες τιμές εισόδου x_i και υπολογίζει μια τιμή εξόδου y , λαμβάνοντας υπόψη ορισμένα βάρη w_j . Εστω ότι έχουμε ένα διάνυσμα τιμών εισόδου $\mathbf{x} = \{x_0, x_1, \dots, x_n\}$ και το αντίστοιχο διάνυσμα βαρών $\mathbf{w} = \{w_0, w_1, \dots, w_n\}$. Η έξοδος δίνεται εφαρμόζοντας τα παραπάνω διανύσματα στην εξίσωση 2.3. Ο όρος που προκύπτει για $j = 0$ ονομάζεται όρος πόλωσης (bias term), όπου $x_0 = 1$ και $w_0 = \text{bias}$.

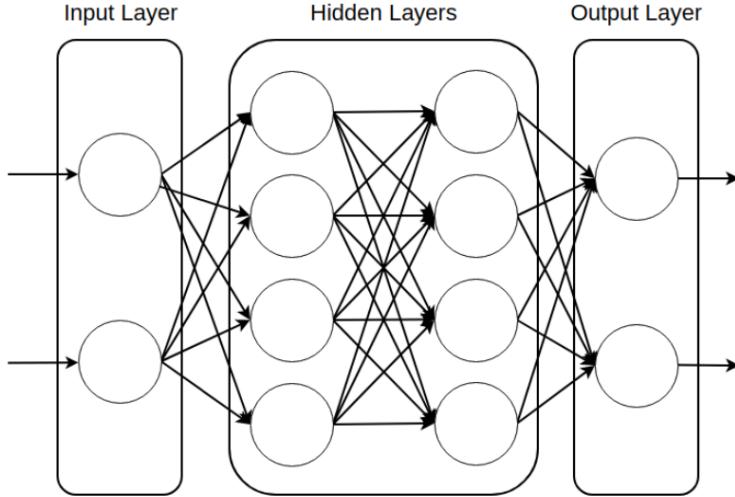


Σχήμα 2.5: Παράδειγμα δίκτυου MLP.

Ένα νευρωνικό δίκτυο τροφοδοσίας προς τα εμπρός αποτελείται από ένα σύνολο συνδεδεμένων τεχνητών νευρώνων, σχηματίζοντας έτσι ένα δίκτυο οργανωμένο σε επίπεδα. Για παράδειγμα, εάν ένα δίκτυο έχει τρία επίπεδα, όπως απεικονίζεται στην εξίσωση 2.4, το πρώτο επίπεδο είναι γνωστό ως επίπεδο εισόδου και αντιπροσωπεύεται από την $f^{(1)}$, το δεύτερο επίπεδο είναι γνωστό ως κρυφό επίπεδο και αντιπροσωπεύεται από την $f^{(2)}$ και το τρίτο επίπεδο είναι το επίπεδο εξόδου και αντιπροσωπεύεται από την $f^{(3)}$. Αυτή η αλυσιδωτή δομή είναι συχνά χρησιμοποιούμενη στα τεχνητά νευρωνικά δίκτυα και μας δίνει το βάθος του μοντέλου.

$$f(x) = f^{(3)}(f^{(2)}((f^{(1)}))) \quad (2.4)$$

Ο στόχος ενός νευρωνικού δίκτυου είναι να βρει την ακριβέστερη συνάρτηση $f(\mathbf{x})$ που προσεγγίζει καλύτερα την αληθή συνάρτηση αντιστοίχησης $f^*(\mathbf{x})$. Για να επιτευχθεί αυτό, το δίκτυο εκπαιδεύεται με προσεγγίσεις της $f^*(x)$, για διαφορετικές τιμές εισόδου x , όπου η κάθε τιμή συνοδεύται από την αντίστοιχη ετικέτα $y \approx f^*(x)$. Η διαδικασία αυτή ονομάζεται επιβλεπόμενη μάθηση, όπως αναλύθηκε παραπάνω. Το δίκτυο θα αλλάξει τις παραμέτρους των ενδιάμεσων επιπέδων για να βρει την ακριβέστερη αντιστοίχιση, ενώ η δομή των επιπέδων ενός νευρωνικού δίκτυου απεικονίζεται στο Σχήμα 2.6.



Σχήμα 2.6: Παράδειγμα δομής επιπέδων ενός νευρωνικού δικτύου.

Η συνάρτηση f που εισάγεται στην εξίσωση 2.3, χρησιμοποιείται για την εισαγωγή μιας μη-γραμμικότητας στους υπολογισμούς, καθώς χωρίς αυτή θα ήταν αδύνατο να αναπαραστήσουμε τις μη γραμμικές αλληλεπιδράσεις μεταξύ της εισόδου και της εξόδου και το δίκτυο θα περιοριζόταν στην αντιστοίχηση γραμμικών σχέσεων. Η στρατηγική συνοψίζεται στην εφαρμογή της γραμμικής αντιστοίχισης σε κάποια συνάρτηση $\Phi(\mathbf{x})$, αντί για το \mathbf{x} , όπου το Φ είναι ένας μη γραμμικός μετασχηματισμός. Το ερώτημα εδώ είναι ποια συνάρτηση μη γραμμικού μετασχηματισμού θα πρέπει να εφαρμοστεί, όπου το σύνολο των πιο δημοφιλών μη γραμμικών μετασχηματισμών αναλύεται στην υποενότητα 2.4.3. Εφαρμόζοντας τον μετασχηματισμό που παρουσιάζεται στην εξίσωση 2.5, οι παράμετροι θ θα χρησιμοποιηθούν για την εκμάθηση της ακριβέστερης συνάρτησης Φ και των παραμέτρων \mathbf{W} , με σκοπό την αντιστοίχιση της εισόδου \mathbf{x} , από την $\Phi(\mathbf{x})$, στην επιθυμητή έξοδο y .

$$y = f(\mathbf{x}; \theta) = \Phi(\mathbf{x}; \theta)^T \mathbf{W} \quad (2.5)$$

2.3.2 Εκμάθηση με βάση το Δυναμικό

Ο σχεδιασμός και η εκπαίδευση ενός νευρωνικού δικτύου χρησιμοποιεί ως αλγόριθμο βελτιστοποίησης την μείωση κλίσης (Gradient Descent). Ένα από τα μεγάλα μειονεκτήματα των τεχνητών νευρωνικών δικτύων είναι ότι η μη γραμμικότητα που επιβάλλεται από το σύνολο των συναρτήσεων ενεργοποίησης προκαλεί τη συνάρτηση κόστους να γίνει μη κυρτή. Αυτό σημαίνει ότι τα νευρωνικά δίκτυα συνήθως εκπαιδεύονται χρησιμοποιώντας έναν επαναληπτικό βελτιστοποιητή που βασίζεται στο δυναμικό, γεγονός που οδηγεί τη συνάρτηση κόστους του δικτύου σε χαμηλή τιμή, ενώ η συνολική σύγκλιση δεν είναι εγγυημένη. Οι αλγόριθμοι κατάβασης δυναμικού που εφαρμόζονται σε μη κυρτές συναρτήσεις κόστους δεν έχουν καμία εγγύηση σύγκλισης και επηρεάζονται από τις τιμές των αρχικών παραμέτρων. Στα νευρωνικά δίκτυα, οι παράμετροι που πρόκειται να βελτιστοποιηθούν είναι τα βάρη και οι όροι πόλωσης των διαφορετικών επιπέδων.

2.3.2.1 Βελτιστοποίηση με βάση το Δυναμικό

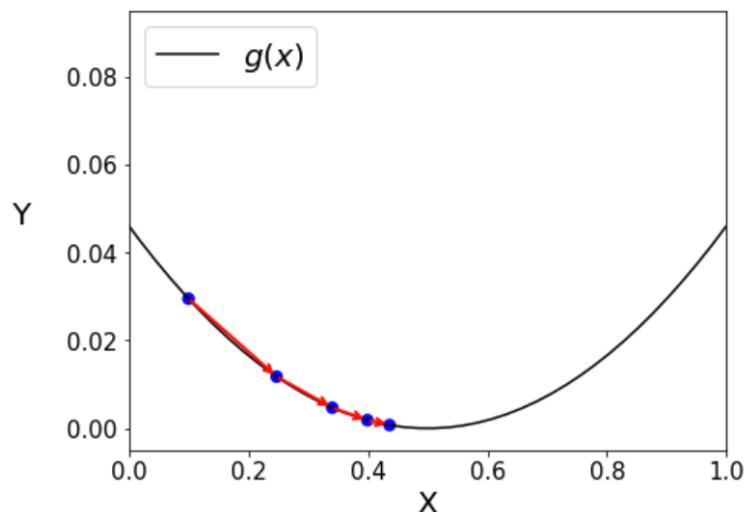
Η εργασία βελτιστοποίησης περιλαμβάνει είτε την ελαχιστοποίηση, είτε τη μεγιστοποίηση μιας συνάρτησης $g(\mathbf{x})$, με τροποποίηση των παραμέτρων \mathbf{x} . Η συνάρτηση που πρόκειται να βελτιστοποιηθεί ονομάζεται αντικειμενική συνάρτηση, συνάρτηση κόστους, συνάρτηση απώλειας ή συνάρτηση σφάλματος. Το πρόβλημα βελτιστοποίησης απεικονίζεται στην εξίσωση 2.6, όπου μας ενδιαφέρει η τιμή του \mathbf{x} που ελαχιστοποιεί τη συνάρτηση g .

$$\mathbf{x}^* = \arg_{\mathbf{x}} \min(g(\mathbf{x})) \quad (2.6)$$

Το πρόβλημα βελτιστοποίησης επιλύεται με την εφαρμογή του αλγόριθμου κατάβασης δυναμικού, ο οποίος αποτελεί έναν πρώτης τάξης επαναληπτικό αλγόριθμο βελτιστοποίησης, για την εύρεση μιας ελάχιστης (ή μέγιστης) τιμής μιας συνάρτησης. Ο αλγόριθμος εκτελείται βήμα-βήμα ανάλογα με την αρνητική τιμή του δυναμικού της συνάρτησης, με σκοπό την ενημέρωση των παραμέτρων. Για παράδειγμα, σε μια συνάρτηση $y = g(x)$, όπου το x και το y είναι πραγματικοί αριθμοί, το δυναμικό της $g(x)$ δίνει την κλίση της συνάρτησης g στο σημείο x . Πραγματοποιώντας πολλά βήματα στην αρνητική κατεύθυνση της κλίσης, θα οδηγήσει την συνάρτηση προς ένα ελάχιστο, όπως αποτυπώνεται στην εξίσωση 2.7, όπου το a συμβολίζεται ως ο ρυθμός εκμάθησης, ελέγχει το μέγεθος κάθε βήματος προς την αρνητική κατεύθυνση και έχει μεγάλη επιρροή κατά την εκπαίδευση νευρωνικών δικτύων, με ένα παράδειγμα να απεικονίζεται στο Σχήμα 2.7.

$$g_{n+1}(x) \approx g(x) - ag_n(x) \quad (2.7)$$

Κάποια προβλήματα που εμφανίζονται όταν χρησιμοποιείται η κατάβαση δυναμικού ως αλγόριθμος βελτιστοποίησης σχετίζονται με τη μη κυρτή φύση των συναρτήσεων που παράγονται από τα νευρωνικά δίκτυα, γεγονός που μπορεί να οδηγήσει σε πολύ αργό ρυθμός σύγκλισης προς το ελάχιστο ή και γενικότερη αδυναμία σύγκλισης. Τέτοιου είδους προβλήματα μπορούν να ελαχιστοποιηθούν με τη χρήση τεχνικών που έχουν εισαχθεί σε παραλλαγές του αλγορίθμου κατάβασης δυναμικού όπως οι παραλλαγές **Adam** [29], **Adadelta** [30], **Adagrad** [31] μεταξύ άλλων. Μια εκτενής σύγκριση των παραπάνω τεχνικών παρουσιάζεται στο [32].



Σχήμα 2.7: 5 βήματα από τα αριστερά στα δεξιά του αλγορίθμου κατάβασης δυναμικού.

2.3.2.2 Συνάρτηση Κόστους

Οι συναρτήσεις κόστους που χρησιμοποιούνται στα νευρωνικά δίκτυα είναι παρόμοιες με τις συναρτήσεις κόστους για άλλα παραμετρικά μοντέλα, όπου στις περισσότερες περιπτώσεις ορίζεται μια κατανομή $p(y|x; \theta)$ και χρησιμοποιείται η αρχή της μέγιστης πιθανότητας. Το παραπάνω αποτελεί την ίδια διαδικασία με την χρήση ως συνάρτησης κόστους της διασταυρούμενης εντροπίας μεταξύ των ετικετών εκπαίδευσης και των προβλέψεων του μοντέλου. Η Διασταυρούμενη εντροπία (Cross-entropy) είναι μια μετρική που μας δείχνει την ανομοιότητα μεταξύ της πραγματικής πιθανότητας p_i και της πιθανότητας εξόδου του μοντέλου q_i , όπως απεικονίζεται στην εξίσωση 2.8.

$$H(p, q) = - \sum_i p_i \log(q_i) \quad (2.8)$$

Πολλές συναρτήσεις κόστους μπορούν να χρησιμοποιηθούν όταν εκπαιδεύεται ένα νευρωνικό δίκτυο και η επιλογή σχετίζεται με την φύση τους προβλήματος. Τα περισσότερα νευρωνικά δίκτυα εκπαιδεύονται χρησιμοποιώντας την αρχή της μέγιστης πιθανότητας, δηλαδή η συναρτήση κόστους είναι απλώς η απώλεια διασταυρούμενης εντροπίας, όπως περιγράφεται στην εξίσωση 2.9, όπου το \mathbf{E} αντιπροσωπεύει την αναμενόμενη τιμή.

$$J(\theta) = -\mathbf{E}_{\mathbf{x}, \mathbf{y} \sim \tilde{p}_{data}} \log(p_{model}(\mathbf{y}|\mathbf{x})) \quad (2.9)$$

Ένα παράδειγμα διαφορετικής συνάρτησης κόστους είναι το μέσο τετραγωνικό σφάλμα (Mean Squared Error), που εισάγεται στην εξίσωση 2.10, το οποίο οδηγεί σε λύση του προβλήματος βελτιστοποίησης που απεικονίζεται στην εξίσωση 2.11.

$$J(\theta) = \mathbf{E}_{\mathbf{x}, \mathbf{y} \sim p_{data}} \|\mathbf{y} - f(\mathbf{x})\|^2 \quad (2.10)$$

$$f^*(\mathbf{x}) = \arg_{\mathbf{x}} \min \mathbf{E}_{\mathbf{x}, \mathbf{y} \sim p_{data}} \|\mathbf{y} - f(\mathbf{x})\|^2 \quad (2.11)$$

Ένα ακόμη παρεμφερές παράδειγμα συνάρτησης κόστους είναι το μέσο απόλυτο σφάλμα (Mean Absolute Error), που εισάγεται στην εξίσωση 2.12, το οποίο οδηγεί σε λύση του προβλήματος βελτιστοποίησης που απεικονίζεται στην εξίσωση 2.13.

$$J(\theta) = \mathbf{E}_{\mathbf{x}, \mathbf{y} \sim p_{data}} \|\mathbf{y} - f(\mathbf{x})\| \quad (2.12)$$

$$f^*(\mathbf{x}) = \arg_{\mathbf{x}} \min \mathbf{E}_{\mathbf{x}, \mathbf{y} \sim p_{data}} \|\mathbf{y} - f(\mathbf{x})\| \quad (2.13)$$

Ένα σύνηθες ζήτημα που προκύπτει κατά τη σχεδίαση των νευρωνικών δικτύων είναι το γεγονός ότι πρέπει η κλίση της συνάρτησης κόστους να έχει την κατάλληλη τιμή ώστε να χρησιμεύει ως καλός οδηγός για τον αλγόριθμο εκμάθησης. Το συγκεκριμένο πρόβλημα σχετίζεται με την επιλογή των συναρτήσεων ενεργοποίησης, που χρησιμοποιούνται για μας δώσουν την έξοδο από τα κρυφά επίπεδα και το επίπεδο εξόδου και ενδέχεται να μας οδηγήσουν σε πολύ μικρή ή πολύ μεγάλη τιμή της κλίσης της συνάρτησης κόστους (Vanishing/Exploding Gradients). Επομένως, οι συναρτήσεις ενεργοποίησης πρέπει να επιλέγονται προσεκτικά.

2.3.2.3 Συναρτήσεις Ενεργοποίησης

Επίπεδο Εξόδου: Ο ρόλος του επιπέδου εξόδου είναι να παρέχει ένα πρόσθετο μετασχηματισμό στα χαρακτηριστικά που δημιουργούνται από το νευρωνικό δίκτυο, $\mathbf{h} = f(\mathbf{x}; \theta)$. Υπάρχει ένα ευρύ φάσμα από συναρτήσεις που δύναται να χρησιμοποιηθούν ως συναρτήσεις μετασχηματισμού στο επίπεδο εξόδου, με τις πιο ευρέως χρησιμοποιούμενες να αναλύονται παρακάτω:

- Γραμμική μονάδα (Linear Unit) που απεικονίζεται στην εξίσωση 2.14. Αυτή η μονάδα χρησιμοποιείται συνήθως για την παραγωγή του μέσου όρου μιας Gaussian υπό συνθήκη κατανομής, $p(\mathbf{y}|\mathbf{x}) = N(\mathbf{y}; \hat{\mathbf{y}}, \mathbf{I})$, όπου η μεγιστοποίηση της λογαριθμικής πιθανότητας ισοδυναμεί με την ελαχιστοποίηση του μέσου τετραγώνου σφάλματος.

$$f(x_i) = x_i \quad (2.14)$$

- Σιγμοειδής μονάδα (Sigmoid Unit) που απεικονίζεται στην εξίσωση 2.15 και είναι κατάλληλη για προβλήματα δυαδικής ταξινόμησης όπου το μοντέλο αντιπροσωπεύεται από μια κατανομή Bernoulli.

$$f(x_i) = \sigma(x_i) = \frac{1}{1 + e^{-x_i}} \quad (2.15)$$

- Softmax μονάδα (Softmax Unit) που απεικονίζεται στην εξίσωση 2.16 και χρησιμοποιείται για την αναπαράσταση μια κατανομής πιθανότητας σε μια διακριτή μεταβλητή με n πιθανές τιμές.

$$f(x_i) = softmax(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \quad (2.16)$$

Κρυφά Επίπεδα: Αναφορικά με την επιλογή μιας συνάρτησης ενεργοποίησης για τα κρυφά επίπεδα είναι πολύ δύσκολο να προβλεφθεί εκ των προτέρων ποια θα λειτουργήσει καλύτερα. Σχετικά με την επιλογή, εκτελείται μια συνεχής διαδικασία «δοκιμής και σφάλματος», όπου επιλέγεται μια συνάρτηση ενεργοποίησης, η οποία διαισθητικά ενδέχεται να λειτουργεί καλά, εκπαιδεύεται ένα δίκτυο με τη συγκεκριμένη συνάρτηση και αξιολογείται η απόδοσή της σε ένα σύνολο επικύρωσης (Validation set), το οποίο είναι υποσύνολο του συνολικού συνόλου δεδομένων. Παρόμοια με τις μονάδες εξόδου υπάρχει ένα ευρύ φάσμα συναρτήσεων που μπορούν να επιλεγούν. Οι πιο συχνά χρησιμοποιούμενες αναφέρονται παρακάτω και απεικονίζονται στο Σχήμα 2.8:

- Γραμμική μονάδα (Linear Unit)

$$f(x_i) = x_i \quad (2.17)$$

- Διορθωμένη Γραμμική μονάδα (Rectified Linear Unit)

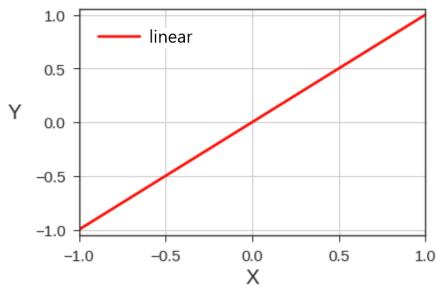
$$f(x_i) = max\{0, x_i\} \quad (2.18)$$

- Λογιστική Σιγμοειδής μονάδα (Logistic sigmoid)

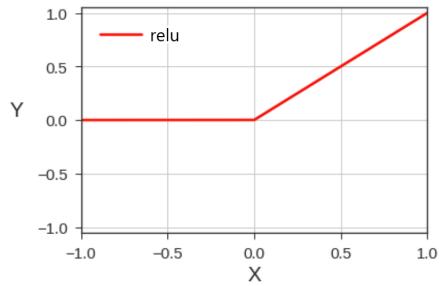
$$f(x_i) = \sigma(x_i) \quad (2.19)$$

- Μονάδα Υπερβολική Εφαπτομένης (Hyperbolic tangent)

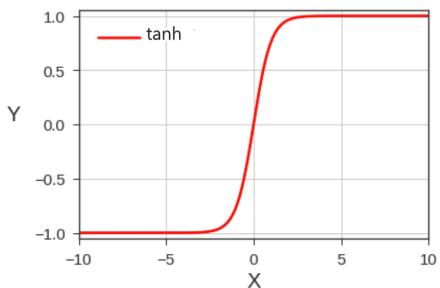
$$f(x_i) = tanh(x_i) \quad (2.20)$$



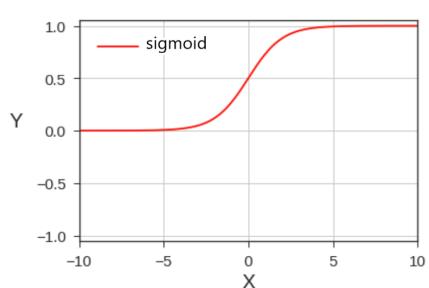
(α) Linear function



(β) ReLu function



(γ) Hyperbolic tangent function



(δ) Sigmoid function

Σχήμα 2.8: Παραδείγμα πιο συχνά χρησιμοποιούμενων συναρτήσεων ενεργοποίησης κρυφών επιπέδων.

2.3.2.4 Αλγόριθμος οπισθοδίαδοσης

Κατά την εκπαίδευση ενός νευρωνικού δικτύου υπάρχουν δύο βήματα: το βήμα προς τα εμπρός και το βήμα προς τα πίσω. Κατά το πρώτο βήμα, οι πληροφορίες «ρέουν» από το επίπεδο εισόδου έως το επίπεδο εξόδου, όπου και υπολογίζεται η τιμή της συνάρτησης κόστους $J(\theta)$. Κατά το δεύτερο βήμα, οι πληροφορίες από αυτήν την τιμή θα «ρέουν» προς τα πίσω μέσω του δικτύου, προκειμένου να υπολογιστεί το δυναμικό. Ο υπολογισμός του δυναμικού πραγματοποιείται με την εφαρμογή του αλγορίθμου οπισθοδίαδοσης (backpropagation), ο οποίος υπολογίζει τις κλίσεις που χρησιμοποιούνται στο βήμα εκμάθησης, ενώ για την ενημέρωση των παραμέτρων του δικτύου χρησιμοποιείται ο αλγόριθμος κατάβασης δυναμικού.

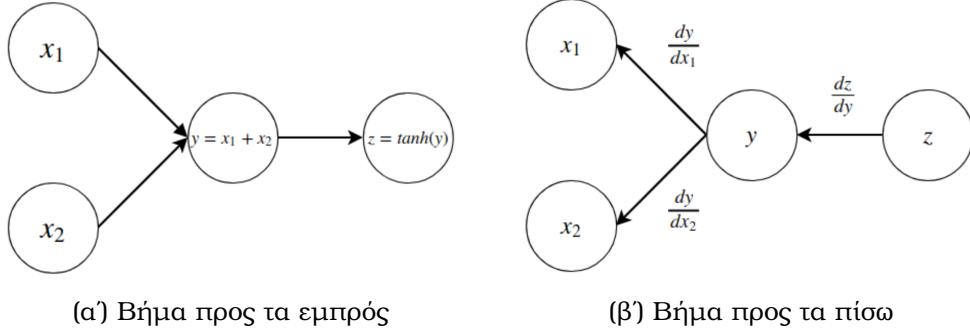
Ο αλγόριθμος οπισθοδίαδοσης βασίζεται στη θεωρία των γραφημάτων υπολογισμού και στον κανόνα της αλυσίδας. Κάθε κόμβος ενός υπολογιστικού γραφήματος αντιπροσωπεύει μια μεταβλητή, η οποία μπορεί να είναι οποιουδήποτε τύπου. Οι πράξεις στο υπολογιστικό γράφημα είναι συναρτήσεις ανάμεσα σε μία ή περισσότερες μεταβλητές, όπου έχουμε ένα σύνολο επιτρεπόμενων πράξεων. Για παράδειγμα, αν η μεταβλητή y υπολογίζεται εφαρμόζοντας μια πράξη σε μια μεταβλητή x , τότε σχεδιάζουμε μια κατευθυνόμενη ακμή από το x στο y . Η δεύτερη έννοια είναι ο κανόνας της αλυσίδας που μας δίνει τη δυνατότητα να υπολογίσουμε τιμές παραγώγων μέσω του γραφήματος, όπως στην εξίσωση 2.21, όπου $y = g(x)$ και $z = f(g(x)) = f(y)$.

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx} \quad (2.21)$$

Η παραπάνω εξίσωση μπορεί να γενικευτεί και πέρα από την περίπτωση βαθμωτών μεταβλητών. Έστω ότι $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$ είναι διανύσματα, η συνάρτηση g αντιστοιχίζει

τιμές από το \mathbb{R}^m στο \mathbb{R}^n και η συνάρτηση f αντιστοιχίζει τιμές από το \mathbb{R}^n στο \mathbb{R} και ο κανόνας της αλυσίδας απεικονίζεται στην εξίσωση 2.22, όπου $\mathbf{y} = g(\mathbf{x})$ και $z = f(\mathbf{y})$.

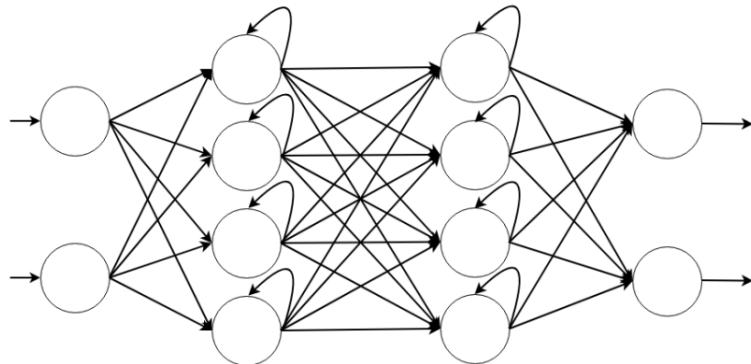
$$\frac{\partial z}{\partial x_i} = \sum_j \frac{\partial z}{\partial y_j} \frac{\partial y_j}{\partial x_i} \quad (2.22)$$



Σχήμα 2.9: Γράφος υπολογισμού για τον υπολογισμό παραγώγων χρησιμοποιώντας τον κανόνα αλυσίδας

2.3.3 Νευρωνικά Δίκτυα Ανατροφοδότησης

Τα νευρωνικά δίκτυα ανατροφοδότησης (RNN) είναι μια οικογένεια τεχνητών νευρωνικών δικτύων που έχουν επαναλαμβανόμενες συνδέσεις, όπως στο Σχήμα 2.10, και ως εκ τούτου είναι κατάλληλα για την επεξεργασία διαδοχικών δεδομένων.



Σχήμα 2.10: Τεχνητό νευρωνικό δίκτυο με επαναλαμβανόμενες συνδέσεις.

Ένα βασικό χαρακτηριστικό των ανατροφοδοτούμενων νευρωνικών δικτύων είναι η χρήση κοινών παραμέτρων σε κάθε βήμα της ακολουθίας των δεδομένων, επιτρέποντας στο μοντέλο να γενικεύει ικανοποιητικά σε ακολουθίες δεδομένων που δεν ανήκουν στα δεδομένα εκπαίδευσης. Η κοινή χρήση παραμέτρων είναι επίσης σημαντική όταν μια συγκεκριμένη πληροφορία μπορεί να εμφανιστεί σε πολλά διαφορετικά βήματα εντός της ακολουθίας και βασίζεται στην υπόθεση ότι το ίδιο σύνολο παραμέτρων δύναται να χρησιμοποιηθεί για διαφορετικά χρονικά βήματα, δηλαδή η κατανομή της υπό συνθήκη πιθανότητας των μεταβλητών τη χρονική στιγμή $t + 1$ δεδομένων των μεταβλητών τη χρονική στιγμή t είναι στάσιμη, που σημαίνει ότι η σχέση μεταξύ του προηγούμενου

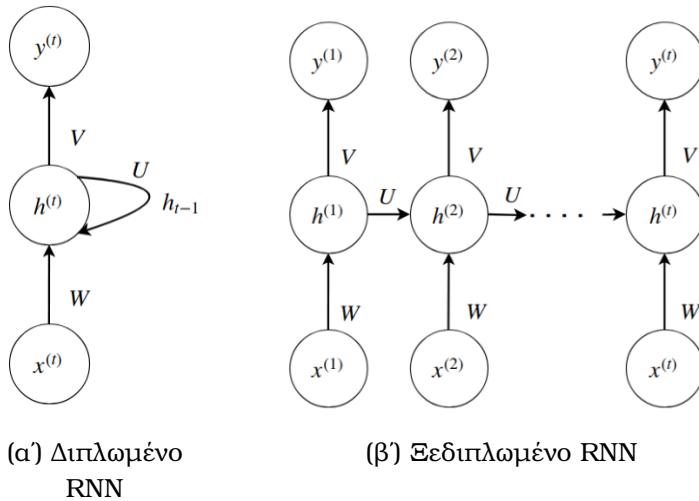
χρονικού βήματος και του επόμενου χρονικού βήματος δεν αλλάζει με το χρόνο.

Ένας καλός τρόπος εισαγωγής των RNN είναι η χρήση γραφημάτων υπολογισμού. Ένα υπολογιστικό γράφημα είναι ένας τρόπος για να επισημοποιήσει τη δομή ενός συνόλου υπολογισμών, όπως αυτοί που εμπλέκονται σε ένα RNN, με τη χρήση ενός γραφήματος που έχει μια επαναλαμβανόμενη δομή. Η κλασική μορφή ενός δυναμικού συστήματος αναπαρίσταται από την εξίσωση 2.23, όπου το $s^{(t)}$ αντιπροσωπεύει το σύστημα την χρονική στιγμή t , και αποτελεί μια επαναλαμβανόμενη εξίσωση αφού η κατάσταση την χρονική στιγμή t εξαρτάται από το προηγούμενο χρονικό βήμα $t - 1$ και δύναται εφικτό να ξεδιπλωθεί σε ένα κατευθυνόμενο άκυκλο υπολογιστικό γράφημα με αναδρομική εφαρμογή της εξίσωσης 2.23.

$$s^{(t)} = f(s^{(t-1)}; \theta) \quad (2.23)$$

Για ένα RNN η αρχή είναι ακριβώς η ίδια με την διαφορά ότι σε αντίθεση με την κατάσταση $s^{(t)}$, τα RNN έχουν μια κρύφη κατάσταση $h^{(t)}$ και συνήθως ένα σήμα $x^{(t)}$ που αντιπροσωπεύει την είσοδο, όπως φαίνεται στην εξίσωση 2.24. Τα RNN έχουν συνήθως πρόσθετα επίπεδα στην είσοδο και την έξοδο με σκοπό την προσθήκη πρόσθετων μετασχηματισμών στα δεδομένα. Ένα παράδειγμα ενός ξεδιπλωμένου RNN απεικονίζεται στο Σχήμα 2.11.

$$h^{(t)} = f(h^{(t-1)}, x^{(t)}; \theta) \quad (2.24)$$



Σχήμα 2.11: Αναπαράσταση λειτουργιών RNN.

Όταν το ανατροφοδοτούμενο νευρωνικό δίκτυο εκπαιδεύται για την πρόβλεψη της τιμής του επόμενου βήματος βάσει τιμών προηγούμενων βημάτων μαθαίνει να χρησιμοποιεί την κρυφή κατάσταση $h^{(t)}$ ως μια σύνοψη των χακτηριστικών που προκύπτουν από προηγούμενες ακολουθίες, έως την τρέχουσα ακολουθία t , διαδικασία ώστοσο που περιλαμβάνει απώλειες καθώς αντιστοιχίζει ένα μεταβλητό μήκος ακολουθίας σε ένα μεμονωμένο διάνυσμα εξόδου. Η προαναφερόμενη διαδικασία έχει δύο πλεονεκτήματα, με το πρώτο να είναι ότι ανεξάρτητα από το μήκος της ακολουθίας, η κρυφή κατάσταση θα έχει πάντα το ίδιο μέγεθος επειδή εκφράζεται ως η μετάβαση από μια κατάσταση στην επόμενη και το δεύτερο να αποτελεί ότι είναι δυνατό να χρησιμοποιηθεί η ίδια συνάρτηση f για όλες τις μεταβάσεις. Αυτοί οι δύο παράγοντες καθιστούν δυνατή την εκμάθηση ενός μεμονωμένου μοντέλου που λειτουργεί σε όλα τα χρονικά βήματα και

μήκη ακολουθίας εκμεταλλεύομενο τις κοινές παραμέτρους για όλες τις ακολουθίες. Σε ένα βήμα προς τα εμπρός ενός ανατροφοδοτούμενου νευρωνικού δικτύου, για όλες τις χρονικές στιγμές μιας ακολουθίας, εφαρμόζονται οι εξισώσεις 2.25-2.28, όπου ως συνάρτηση ενεργοποίησης χρησιμοποιείται η υπερθολική εφαπτομένη, το g αποτελεί την μονάδα εξόδου και οι παράμετροι είναι οι όροι πόλωσης \mathbf{b} και \mathbf{c} , οι πίνακες $\mathbf{U}, \mathbf{V}, \mathbf{W}$ αντίστοιχα.

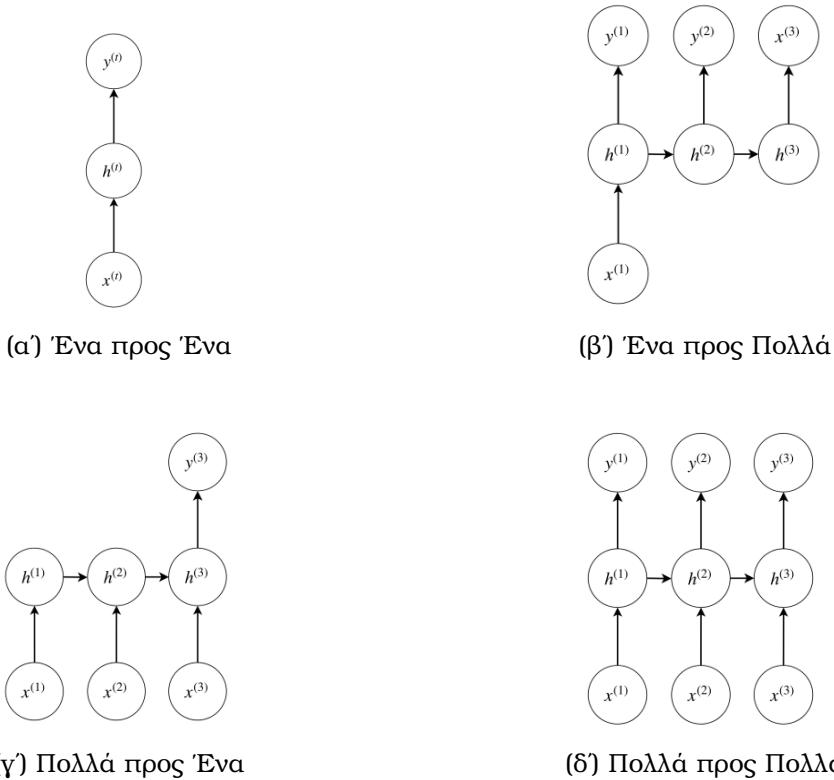
$$\alpha(t) = \mathbf{b} + \mathbf{Wh}^{(t-1)} + \mathbf{Ux}^{(t)} \quad (2.25)$$

$$\mathbf{h}^{(t)} = \tanh(\alpha^{(t)}) \quad (2.26)$$

$$\mathbf{o}^{(t)} = \mathbf{c} + \mathbf{Vh}^{(t)} \quad (2.27)$$

$$y^{(t)} = g(\mathbf{o}^{(t)}) \quad (2.28)$$

Υπάρχει μια μεγάλη ποικιλία σχέσεων ανατροφοδοτούμενων νευρωνικών δικτύων που μπορούν να δημιουργηθούν και το σύνολο αυτών των σχέσεων απεικονίζεται στο Σχήμα 2.12.



Σχήμα 2.12: Αρχιτεκτονικές ανατροφοδοτούμενων νευρωνικών δικτύων.

2.3.3.1 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης

Το δίκτυο LSTM παρουσιάστηκε για πρώτη φόρα το 1997 [33] και σχεδιάστηκε ειδικά για την αντιμετώπιση μακροπρόθεσμων εξαρτήσεων σε διαδοχικά δεδομένα. Η βασική ιδέα πίσω από το LSTM είναι η εισαγωγή της κατάστασης κυψέλης (cell state) C_t , όπου το LSTM έχει τη δυνατότητα να προσθέτει ή να αφαιρεί πληροφορίες σε αυτή την κατάσταση, η οποία λειτουργεί ως μνήμη. Η διαδικασία πρόσθεσης ή αφαίρεσης πληροφοριών ελέγχεται από διαδικασίες που πραγματοποιούνται στις 3 πύλες που αναλύονται παρακάτω.

Forget Gate: Αυτή η πύλη έχει το καθήκον να αποφασίσει ποιες πληροφορίες πρέπει να αφαιρεθούν από την κατάσταση κυψέλης, η οποία αντιπροσωπεύει τη «γνώση» που έχει αποκτηθεί από όλα τα προηγούμενα χρονικά βήματα. Ανάλογα με την διεργασία, όταν επεξεργαζόμαστε νέα δεδομένα δυνάται να θέλουμε να διατηρήσουμε ή να αφαιρέσουμε πληροφορίες από προηγούμενα χρονικά βήματα.

$$\mathbf{f}^{(t)} = \sigma(\mathbf{W}_f \times [\mathbf{h}^{(t-1)}, \mathbf{x}^{(t)}] + \mathbf{b}_f) \quad (2.29)$$

Input Gate: Αυτή η πύλη έχει ως στόχο να επιλέξει ποιες πληροφορίες θα αποθηκευτούν στην κατάσταση κυψέλης και διαχωρίζεται σε 2 μέρη. Το πρώτο ονομάζεται επιπέδο εισόδου την πύλης και είναι ένα σιγμοειδές επίπεδο που αποφασίζει ποιες τιμές θα χρησιμοποιηθούν για την ανανέωση της κατάστασης κυψέλης και το δεύτερο είναι ένα επίπεδο υπερβολικής εφαπτομένης που δημιουργεί ένα διάνυσμα με νέες πληροφορίες, υποψήφιες για την τροφοδότηση της κατάστασης κυψέλης. Αυτή η πύλη εφαρμόζει έναν μετασχηματισμό σε μια είσοδο $x^{(t)}$ και στην κρυφή κατάσταση της προηγούμενης χρονικής στιγμής $\mathbf{h}^{(t-1)}$. Ο βασικός στόχος είναι να δοθεί στο δίκτυο η δυνατότητα επιλογής των σχετικών πληροφοριών που θα αποθηκευτούν στην κατάσταση της κυψέλης.

$$\mathbf{i}^{(t)} = \sigma(\mathbf{W}_i \times [\mathbf{h}^{(t-1)}, \mathbf{x}^{(t)}] + \mathbf{b}_i) \quad (2.30)$$

$$\tilde{\mathbf{C}}^{(t)} = \tanh(\mathbf{W}_c \times [\mathbf{h}^{(t-1)}, \mathbf{x}^{(t)}] + \mathbf{b}_c) \quad (2.31)$$

Updated Cell State: Η κατάσταση της κυψέλης θα ενημερωθεί σύμφωνα με τα προαναφερόμενα βήματα. Πιο αναλυτικά, θα πολλαπλασιαστεί η κατάσταση της κυψέλης με την forget gate και στη συνέχεια θα προστεθούν οι σχετικές πληροφορίες για την ενημέρωσή της.

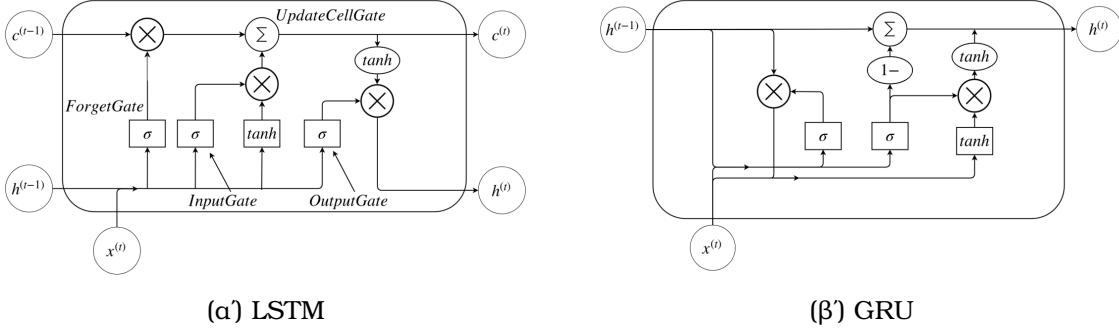
$$\mathbf{C}^{(t)} = \mathbf{f}^{(t)} \otimes \mathbf{C}^{(t-1)} + \mathbf{i}^{(t)} \otimes \tilde{\mathbf{C}}^{(t)} \quad (2.32)$$

Output Gate: Η πύλη εξόδου έχει ως στόχο τον έλεγχο των πληροφοριών που θα περάσουν στο επόμενο χρονικό βήμα μέσω της κρυφής κατάστασης $\mathbf{h}^{(t)}$ και εξαρτάται από μια φιλτραρισμένη έκδοση της τρέχουσας κατάστασης του κελιού. Η φιλτραρισμένη αυτή έκδοση υπολογίζεται από ένα σιγμοειδές επίπεδο που αποφασίζει ποιες πληροφορίες από την κατάσταση της κυψέλης πρόκειται να περάσουν στην έξοδο και από ένα επίπεδο υπερβολικής εφαπτομένης που θα περιορίσει τις τιμές της κατάστασης της κυψέλης μεταξύ 1 και -1.

$$\mathbf{o}^{(t)} = \sigma(\mathbf{W}_o \times [\mathbf{h}^{(t-1)}, \mathbf{x}^{(t)}] + \mathbf{b}_o) \quad (2.33)$$

$$\mathbf{h}^{(t)} = \mathbf{o}^{(t)} \otimes \tanh(\mathbf{C}^{(t)}) \quad (2.34)$$

Παραλλαγές: Από το LSTM αναπτύχθηκαν παρόμοιες ιδέες, με την πιο γνωστή να αποτελεί το Gated Recurrent Unit. Το GRU συνδυάζει τις πύλες Forget, Input σε μια ενιαία πύλη ενώ η κρυφή κατάσταση και η κατάσταση κυψέλης ομαδοποιούνται. Το GRU αποτελεί ένα πιο απλό μοντέλο από το LSTM, επιτυχγάνει ωστόσο παρόμοια αποτελέσματα. Το Σχήμα 2.13 απεικονίζει την αρχιτεκτονική ενός LSTM κελιού και ενός GRU κελιού. Κελιά όπως το Lattice Recurrent Unit [34] ή το Statistical Recurrent Unit [35] έχουν προταθεί επίσης.



Σχήμα 2.13: Παραλλαγές αρχιτεκτονικών κελιών ανατροφοδοτούμενων νευρωνικών δικτύων.

2.4 Μετρικές Αξιολόγησης Μοντέλων

Με την ολοκλήρωση της εκμάθησης, ένα μοντέλο ταξινόμησης δύναται να αξιολογηθεί σε ένα μικρό υποσύνολο του συνολικού συνόλου δεδομένων, το οποίο ονομάζεται σύνολο δεδομένων δοκιμής, όπου και προκύπτουν οι τελικές προβλέψεις για τις ετικέτες των δειγμάτων του συνόλου. Με σκοπό την αξιολόγηση ενός ταξινομητή, έχει αναπτυχθεί ένα σύνολο μετρικών, καθεμιά από τις οποίες μας παρέχει και μια ξεχωριστή πληροφορία αναφορικά με την επίδοσή του. Το σύνολο των μετρικών που χρησιμοποιήθηκαν στα πλαίσια της παρούσας εργασίας, καθώς και ο πίνακας από τον οποίο «αντλούν» αριθμητικούς όρους, αναλύονται παρακάτω.

Πίνακας σύγχυσης: Ο Πίνακας Σύγχυσης είναι ένας τετραγωνικός πίνακας, όπου οι διαστάσεις σου είναι ίσες με την πληθικότητα των ετικετών ενός συνόλου δεδομένων και αποτελεί μια γραφική απεικόνιση της απόδοσης ενός μοντέλου ταξινόμησης πάνω σε ένα σύνολο δεδομένων δοκιμής. Στο Σχήμα 2.14 απεικονίζεται η μορφή του πίνακα σύγχυσης στο πρόβλημα της δυαδικής ταξινόμησης. Ορίζοντας την μια κλάση ως θετική και την άλλη ως αρνητική, προκύπτουν από τον πίνακα οι ακόλουθοι αριθμητικοί όροι.

- **TP** (True Positive), το σύνολο των σωστών προβλέψεων για δείγματα που ανήκουν στην θετική κλάση.
- **TN** (True Negative), το σύνολο των σωστών προβλέψεων για δείγματα που ανήκουν στην αρνητική κλάση.
- **FP** (False Positive), το σύνολο των λανθασμένων προβλέψεων για δείγματα που ανήκουν στην αρνητική κλάση, αλλά προβλέφθηκαν στη θετική.
- **FN** (False Negative), το σύνολο των λανθασμένων προβλέψεων για δείγματα που ανήκουν στη θετική κλάση, αλλά προβλήφθηκαν στην αρνητική.

		Predicted class	
		(+)	(-)
Actual class	(+)	<i>TP</i>	<i>FN</i>
	(-)	<i>FP</i>	<i>TN</i>

Σχήμα 2.14: Πίνακας Σύγχυσης δυαδικού ταξινομητή.

Μετρικές Αξιολόγησης: Με βάση τον παραπάνω πίνακα δύναται να οριστούν οι παρακάτω μετρικές αξιολόγησης.

- **Accuracy** που ορίζεται ως ο λόγος του συνόλου των σωστών προβλέψεων προς τον συνολικό αριθμό δειγμάτων.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.35)$$

- **Precision** που ορίζεται ως ο λόγος του συνόλου των σωστών θετικών προβλέψεων προς τον συνολικό αριθμό των θετικών προβλέψεων.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.36)$$

- **Recall** που ορίζεται ως ο λόγος του συνόλου των σωστών θετικών προβλέψεων προς τον συνολικό αριθμό των θετικών δειγμάτων.

$$\text{Precision} = \frac{TP}{TP + FN} \quad (2.37)$$

- **F1-score** που ορίζεται ως ο αρμονικός μέσος όρος των μετρικών Precision και Recall.

$$F1 - score = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (2.38)$$

Ενώ το F1-score αποτελεί μια κατάλληλη μετρική για προβλήματα δυαδικής ταξινόμησης, δίνει προτεραιότητα στη θετική κλάση. Επομένως χρησιμοποιούμε και την μετρική **AUROC**, που ορίζεται ως το εμβαδόν κάτω από την καμπύλη χαρακτηριστικής λειτουργίας του δέκτη (ROC) [36]. Η καμπύλη χαρακτηριστικής λειτουργίας του δέκτη ορίζεται από τα ποσοστά των FP και TP και απεικονίζει τις σχετικές αντισταθμίσεις μεταξύ αυτών.

2.5 Δυναμικά Χαρακτηριστικά Συναισθημάτων

Το πρόσφατο αυξανόμενο ενδιαφέρον για την ανάλυση της δυναμικής των συναισθημάτων προκάλεσε την δημιουργία χαρακτηριστικών, καθένα από τα οποία ισχυρίζεται ότι αξιολογεί μια μοναδική δυναμική πτυχή της συναισθηματικής μας ζωής [37]. Παρακάτω, παρατίθενται τα χαρακτηριστικά που χρησιμοποιήθηκαν για την ανάλυση θετικών (PA) και αρνητικών (NA) συναισθηματικών καταστάσεων.

Μέση Τιμή PA & NA: Καταγράφει το μέσο επίπεδο των θετικών ή αρνητικών συναισθηματικών καταστάσεων και απεικονίζεται στην εξίσωση 2.39, όπου T είναι ο συνολικός αριθμός συναισθηματικών επισημάνσεων, t είναι μια συγκεκριμένη επισήμανση και j ένα συγκεκριμένο άτομο.

$$M_{PA_j} = \frac{1}{T_j} \sum_{t=1}^{T_j} PA_{t_j} \quad (2.39)$$

Διακύμανση PA & NA: Καταγράφει την συναισθηματική απόκλιση από τα μέσα επίπεδα θετικών ή αρνητικών συναισθηματικών καταστάσεων, προσδιορίζει την συναισθηματική μεταβλητότητα και απεικονίζεται στην εξίσωση 2.40.

$$s.d_{PA_j} = \sqrt{\frac{\sum_{t=1}^{T_j} (PA_{t_j} - M_{PA_j})^2}{T_j - 1}} \quad (2.40)$$

Μέση τετραγωνική διαδοχική διαφορά PA & NA: Καταγράφει τη μέση αλλαγή στην εντάση των συναισθηματικών καταστάσεων για 2 διαδοχικές επισημάνσεις, προσδιορίζει την συναισθηματική αστάθεια και απεικονίζεται στην εξίσωση 2.41.

$$MSSD_{PA_j} = \sqrt{\frac{\sum_{t=2}^{T_j} (PA_{t_j} - PA_{t_j-1})^2}{T_j - 1}} \quad (2.41)$$

Κλίση αυτοπαλινδρόμησης PA & NA: Αποτελεί την αυτοπαλινδρομική κλίση σε ένα AR μοντέλο πρώτης τάξης στο οποίο η εντάση των συναισθηματικών καταστάσεων τη στιγμή $t-1$ προβλέπει την αντίστοιχη ένταση τη στιγμή t . Καταγράφει τον βαθμό στον οποίο μια συναισθηματική κατάσταση μεταφέρεται από μια στιγμή στην επόμενη, προσδιορίζει την συναισθηματική αδράνεια και απεικονίζεται στην εξίσωση 2.43, όπου b ένας σταθερός όρος, g μια σταθερή επίδραση, r μια τυχαία επίδραση και ϵ το σφάλμα.

$$PA_{t_j} = (b^g + b_j^r) + (AR_{PA}^g + AR_{pa_j}^r)PA_{t_j-1} + \epsilon_{t_j} \quad (2.42)$$

$$AR_{PA_j} = AR_{PA}^g + AR_{PA_j}^r \quad (2.43)$$

Κεφάλαιο 3

Βιβλιογραφική Επισκόπηση των Νεότερων Τάσεων της Έρευνας

3.1 Θεωρίες συναισθήματος και επισκόπηση των νεότερων εξελίξεων στο πεδίο

Παρά τη συχνή εμφάνιση των συναισθημάτων στην καθημερινή ζωή ενός ανθρώπου, δεν υπάρχει ομοφωνία στην επιστημονική κοινότητα σχετικά με την επιστημονικό ορισμό των συναισθημάτων. Ένας λόγος σχετικά με την αδυναμία σχηματισμού επιστημονικού ορισμού των συναισθημάτων είναι ότι συχνά συγχέονται με τη διάθεση, την ιδιοσυγκρασία, την προσωπικότητα και άλλα ψυχολογικά χαρακτηριστικά ενός ανθρώπου. Η πλειονότητα των ερευνητών υποστηρίζει την διάκριση μεταξύ των συναισθημάτων και των ψυχολογικών χαρακτηριστικών καθώς τα συναισθήματα είναι σκόπιμα, δηλαδή επηρεάζονται δυναμικά από μια συνθήκη, ενώ οι διαθέσεις που αποτελούν ένα ψυχολογικό χαρακτηριστικό δεν είναι εκ προθέσεως, με πιο χαρακτηριστικό παράδειγμα την κατάθλιψη [38] [39]. Οι διαθέσεις επίσης τείνουν επίσης να διαρκούν μεγαλύτερη διάρκεια και να επηρεάζουν τις γνωστικές μας στρατηγικές, τις γνωστικές μας δεξιότητες και την γνωστική μας ανάπτυξη σε αντίθεση με τα συναισθήματα, τα οποία προκατέχουν τις πράξεις μας [40]. Η αίσθηση ενός αντικειμένου ή μιας κατάστασης είναι επίσης ιδιότητες αυτού και όχι καταστάσεις ενός ατόμου.

Η βασική Θεωρία Συναισθημάτων (Basic Theory of Emotions) του Ekman [41] υποστηρίζει την ύπαρξη 6 βασικών και καθολικών συναισθημάτων: **ευτυχία, έκπληξη, φόβος, λύπη, θυμός, και αηδία** [42]. Ο όρος «βασική» υποδηλώνει ότι τα προαναφερθέντα έξι συναισθήματα είναι ξεχωριστά και διακριτά όσον αφορά τη φυσιολογία, ενώ παράλληλα χαρακτηρίζονται και ως «καθολικά», καθώς υπάρχουν μεταξύ των διαφορετικών πολιτισμών και της κουλτούρας που διακατέχει τον καθένα ξεχωριστά.

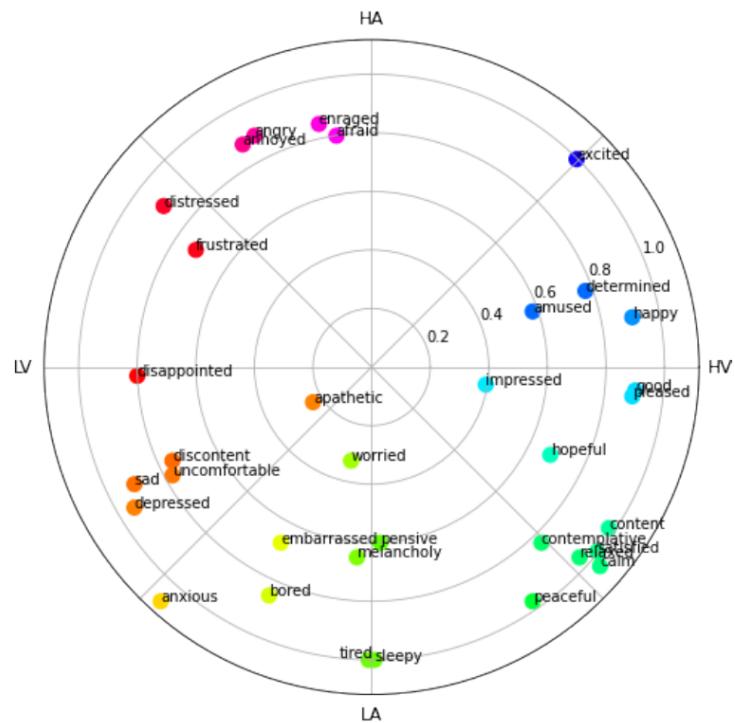
Άλλες θεωρίες μοντελοποιούν τα συναισθήματα χρησιμοποιώντας έναν συγκεκριμένο αριθμό διαστάσεων, όπως το μοντέλο του Russell (Russell's Circumplex Model of Affect) [43], που χρησιμοποιεί ως διαστάσεις τις συναισθηματικές καταστάσεις την διέγερσης (arousal) και τους οθένους (valence). Παρόμοιες προσεγγίσεις με την παραπάνω ακολουθούν ο δισδιάστατος τροχός συναισθημάτων του Plutchik (Plutchik's Wheel of Emotions) [44] και το τετραδιάστατο μοντέλο των συναισθημάτων της κλεψύδρας (H-

ourglass Model of Emotions) [45] που χρησιμοποιεί ως διαστάσεις τις ακόλουθες: **ευ-χαρίστηση, ευαισθησία, έφεση και προσοχή**. Το πιο ευρέως χρησιμοποιούμενο μεταξύ των μοντέλων από αυτά που χρησιμοποιούν διαστάσεις για την κατηγοριοποίηση των συναισθημάτων είναι το μοντέλο του Russel, λόγω της δεδομένης λακωνικότητας του δισδιάστατου χώρου και της ευκολίας χρήσης για απλές έρευνες. Απεικονίζεται στο Σχήμα 3.1.

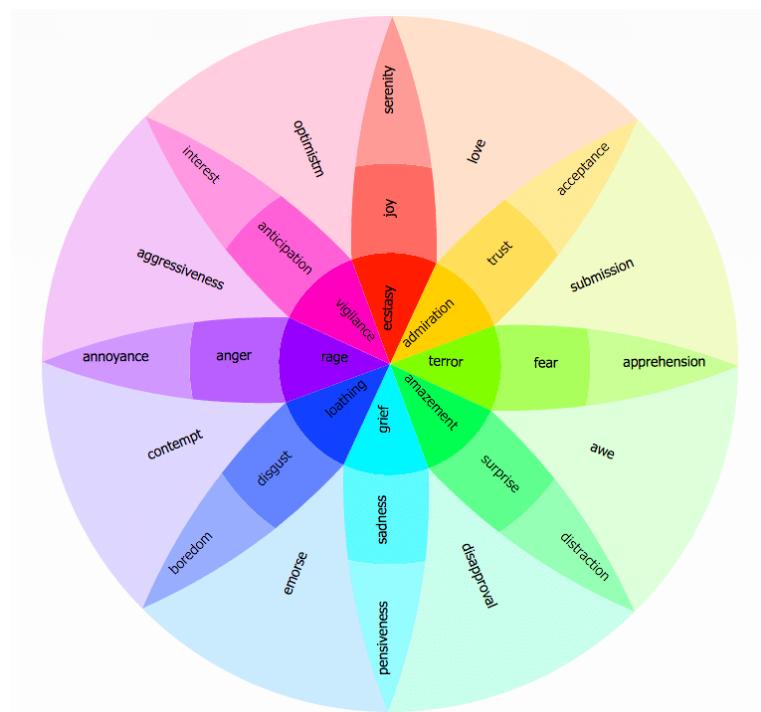
Ωστόσο, τα τελευταία ευρήματα στην έρευνα για τα συναισθήματα υποδηλώνουν ότι τα συναισθήματα δεν περιορίζονται σε έξι διακριτές κατηγορίες, ούτε πολλές μετρήσιμες διαστάσεις. Μια πρόσφατη έρευνα που σχετίζεται με τις εκφράσεις του προσώπου μας δείχνει τα συναισθήματα είναι σύνθετα [46], όπως για παράδειγμα το γεγονός πως κάποιος μπορεί να νιώθει ταυτόχρονα λυπημένος και χαρούμενος σε ορισμένες περιστάσεις. Μια άλλη έρευνα προτείνει ότι τα συναισθήματα είναι τακτικά παρά απόλυτα, δηλαδή υπερτονίζει το γεγόνος ότι υπάρχει μια σύνδεση μεταξύ τους [47], παρά το γεγονός ότι στην βιβλιογραφία οι επισημάνσεις των συναισθημάτων γίνονται με απόλυτο τρόπο.

Ακόμα πιο πρόσφατες έρευνες υποδηλώνουν την ύπαρξη περισσότερων από 20 κατηγοριών συναισθημάτων πέρα τα βασικά έξι [48] [49] που αναφέρονται στην έρευνα του Ekman, συμπεριλαμβανομένων των βασικών έξι και πιο αφηρημένων συναισθημάτων όπως η **υπερηφάνεια**, το **δέος** και η **αγάπη**, ενώ προτείνεται ο σχηματισμός μιας πολυδιάστατης ταξινόμησης συναισθηματικών εμπειριών και εκφράσεων [50] [51], χρησιμοποιώντας μεθόδους που βασίζονται σε δεδομένα (data-driven) για να συλλάβουν μυριάδες συναισθηματικές συμπεριφορές, εμπειρίες, και εκφράσεις [52]. Ωστόσο, πρέπει να σημειωθεί ότι αυτές οι εξελίξεις δεν απορρίπτουν προηγούμενες θεωρίες, καθώς προτείνουν ένα συμπληρωματικό γενικότερο πλαίσιο κατηγοριοποίησης των συναισθημάτων [53], υποστηρίζοντας ότι τα βασικά συναισθήματα δεν είναι μεμονωμένες συναισθηματικές καταστάσεις αλλά μια αποτελούν μια οικογένεια σχετικών συναισθηματικών καταστάσεων.

Τέλος, η συζήτηση παραμένει σε εξέλιξη, με την εμφάνιση της έννοιας της κοινωνικής λειτουργικότητας [54] [55], η οποία προοδιορίζει τα συναισθήματα ως μέσα για την εξυπηρέτηση διαφόρων σκοπών όπως η επιβίωση και η αναπαραγωγή και ως μέσα μετάδοσης πληροφοριών, όπως η επιβράβευση συμπεριφορών και η πρόκληση συγκεκριμένων συμπεριφορών κατά τις διαπροσωπικές αλληλεπιδράσεις.



Σχήμα 3.1: Δισδιάστο μοντέλο του Russell με την τοποθέτηση συναισθηματικών καταστάσεων κατά προσέγγιση [1]



Σχήμα 3.2: Τροχός συναισθημάτων του Plutchik.

Κεφάλαιο 4

Μεθοδολογία

4.1 Περιγραφή Συνόλου Δεδομένων

Το σύνολο δεδομένων που χρησιμοποιήθηκε στα πλαίσια της παρούσας διπλωματικής εργασίας είναι το K-EmoCon [2]. Το παραπάνω πολυτροπικό σύνολο δεδομένων αποκτήθηκε από 32 άτομα που συμμετείχαν σε 16 ζεύγη νατουραλιστικών συζητήσεων, οι οποίες σχετίζονταν με ένα κοινωνικό ζήτημα. Αποτελείται από δεδομένα προερχόμενα από φυσιολογικά σήματα, τα οποία συλλέχθηκαν από 3 φορητές συσκευές καταγραφής, οπτικοακουστικά πλάνα των συμμετεχόντων κατά την διάρκεια της συζήτησης και συνέχεις επισημάνσεις συναισθημάτων. Το συγκεκριμένο σύνολο δεδομένων συμβάλλει σημαντικά στην τρέχουσα βιβλιογραφία που απευθύνεται στην αναγνώριση συναισθημάτων, καθώς αποτελεί το πρώτο σύνολο δεδομένων με επισημάνσεις συναισθημάτων στο πλαίσιο μιας νατουραλιστικής συζήτησης από όλες τις πιθανές οπτικές πλευρές: του συμμετέχοντα, του συνομιλητή, καθώς και των εξωτερικών παρατηρητών που δεν συμμετέχουν στη συζήτηση, παραμόνο την παρακολούθιούν. Στη συνέχεια, θα αναλύσουμε τη διαδικασία συλλογής των δεδομένων καθώς και τα περιεχόμενά τους.



Σχήμα 4.1: Η εικόνα στα αριστερά δείχνει ένα ζευγάρι συμμετεχόντων που κάθονται σε ένα τραπέζι και προετοιμάζονται για μια συζήτηση. Δύο κινητά τηλέφωνα τοποθετήμενα σε τρίποδα στη μέση του τραπεζιού κατέγραψαν τις εκφράσεις του προσώπου των συμμετεχόντων και τις κινήσεις του άνω μέρους του σώματός τους, όπως φαίνεται στο χαρακτηριστικό παράδειγμα της δεξιάς εικόνας.



(α') Συμμετέχων 7



(β') Συμμετέχων 8

Σχήμα 4.2: Στιγμιότυπο από την καταγραφή της συζήτησης 2 συμμετεχόντων ταυτόχρονα από 2 κινητά τηλέφωνα.

4.1.1 Συλογή Συνόλου Δεδομένων

Το K-EmoCon [2] είναι ένα δημοσίως διαθέσιμο πολυτροπικό σύνολο δεδομένων που αποκτήθηκε από 32 συμμετέχοντες που συμμετέχουν σε ζεύγη συζητήσεων διάρκειας 10 λεπτών αναφορικά με ένα κοινωνικό ζήτημα. Επιτρέπει τη μελέτη των συναισθημάτων, τη διαδοχή των συναισθηματικών καταστάσεων στο πλαίσιο μιας νατουραλιστικής συζήτησης η οποία λαμβάνει χώρα σε ένα κοινωνικό περιβάλλον. Για τη συλλογή των δεδομένων, οι συμμετέχοντες χωρίστηκαν τυχαία σε ζευγάρια και συμμετείχαν σε μια νατουραλιστική συζήτηση πρόσωπο με πρόσωπο για την προσφυγική κρίση στην Jeju Yemeni για χρονικό διάστημα περίπου 10 λεπτών. Όλες οι συζήτησείς για το σύνολο των ζευγαριών πραγματοποιήθηκαν σε αίθουσα με ελεγχόμενη θερμοκρασία και φωτισμό για την ελαχιστοποίηση των περιβαλλοντικών διακυμάνσεων στις συζητήσεις, με σκοπό οι συναισθηματικές εναλλαγές να σχετιζόνται όσο το δυνατόν λιγότερο με περιβαλλοντικές αλλαγές στο περιβάλλον της αίθουσας. Όπως απεικονίζεται στα σχήματα 4.1 και 4.2 δύο συμμετέχοντες κάθισαν σε ένα τραπέζι ο ένας απέναντι στον άλλο, με κάμερες στη μέση του τραπεζιού, οι οποίες κατέγραφαν τις εκφράσεις του προσώπου τους, καθώς και τις κινήσεις του άνω μέρους του σώματός τους, μαζί με τις ομιλίες κατά τη διάρκεια της συζήτησης.

Οι συμμετέχοντες φορούσαν ένα σύνολο φορητών συσκευών καταγραφής φυσιολογικών βιοσημάτων κατά τη διάρκεια των συζητήσεων. Η επιλογή των φορητών συσκευών καθιστά το σύνολο δεδομένων κατάλληλο για την μελέτη της διαδοχής των συναισθηματικών καταστάσεων κατά τη διάρκεια νατουραλιστικών συζητήσεων, σε αντίθεση με την πλειονότητα των συνόλων δεδομένων με επισημάνσεις συναισθημάτων που συλλέγονται σε ελεγχόμενο εργαστηριακό περιβάλλον με συσκευές υψηλής ιατρικής ακρίβειας.

Το σχήμα 4.3 αποτυπώνει λεπτομερείς πληροφορίες αναφορικά με τον τρόπο χρήσης του συνόλου των φορητών συσκευών που χρησιμοποιήθηκαν για την κατασκευή του K-EmoCon [2], ενώ ο πίνακας 4.1 αναπαριστά το σύνολο των συλλεγόμενων δεδομένων, τον ρυθμό δειγματοληψίας τους, καθώς και το εύρος τιμών τους.

Συσκευή	Δεδομένα	Δειγματοληψία	Εύρος
Empatica E4	3-axis acceleration	32Hz	[-2g, 2g]
Empatica E4	BVP (PPG)	64Hz	n/a
Empatica E4	EDA	4Hz	[0.01µS, 100µS]
Empatica E4	Heart rate (from BVP)	1Hz	n/a
Empatica E4	IBI (from BVP)	n/a	n/a
Empatica E4	Body temperature	4Hz	[-40 °C, 115 °C]
NeuroSky MindWave	Brainwave (EEG Fp1)	125Hz	n/a
NeuroSky MindWave	Attention & Meditation	1Hz	[0, 100]
Polar H7	HR (ECG)	1Hz	n/a

Πίνακας 4.1: Φυσιολογικά σήματα που συλλέγονται με φορητές συσκευές στο K-EmoCon [2], με τους αντίστοιχους ρυθμούς δειγματοληψίας και τα εύρη των σημάτων.



Σχήμα 4.3: Φορητές συσκευές καταγραφής τοποθετημένες σε έναν από τους συμμετέχοντες κατά τη διάρκεια συνεδριών συλλογής δεδομένων.

Το σύνολο των συζητήσεων παρατηρήθηκε και ρυθμίστηκε από έναν συντονιστή, ενώ κάθε συζήτηση είχε διαρκεία κατά προσέγγιση 10 λεπτά. Κάθε συμμετέχων μιλούσε εναλλάξ για δύο συνεχόμενα λεπτά και ο συντονιστής φροντίζε να τηρείται τόσο αυτό το χρονικό διάστημα, όσο και το συνολικό χρονικό διάστημα των 10 λεπτών με κάποια ευελιξία. Όταν η διαδικασία της συζήτησης ολοκληρωνόταν, οι συμμετέχοντες προχωρούσαν σε επισήμανση τόσο των δικών τους συναισθημάτων παρακολουθώντας τα πλάνα που απεικόνιζαν τον εαυτό τους, όσο και των συναισθημάτων του συνομιλητή τους παρακολουθώντας τα πλάνα που απεικόνιζαν τον συνομιλητή, ανά 5 δευτερόλεπτα από την αρχή μέχρι το τέλος της συζήτησης. Επιπροσθέτως, πέντε εξωτερικοί παρατηρητές συμμετείχαν στην διαδικασία της επισήμανσης των συναισθημάτων των συμμετεχόντων κατά τη διάρκεια των συζητήσεων.

4.1.2 Περιεχόμενα Συνόλου Δεδομένων

Το σύνολο δεδομένων που προκύπτει περιλαμβάνει δεδομένα από 16 ζεύγη συζητήσεων, η συνολική διάρκεια των οποίων αθροίζεται σε 172 λεπτά και 92 δευτερόλεπτα. Τα δεδομένα περιλαμβάνουν φυσιολογικά σήματα, οπτικοακουστικές εγγραφές συζητήσεων και συνεχείς επισημάνσεις συναισθημάτων ανά 5 δευτερόλεπτα από τρεις διαφορετικές οπτικές γωνίες. Ο πίνακας 4.2 συνοψίζει τα περιεχόμενα του συνόλου δεδομένων, ενώ ο πίνακας 4.3 αποτυπώνει αναλυτικά τις διαφορετικές κατηγορίες συναισθημάτων που χρησιμοποιήθηκαν για επισήμανση.

Σύνοψη συνόλου δεδομένων	
Αριθμός συμμετεχόντων	32 (20 άνδρες & 12 γυναίκες)
Ηλικία συμμετεχόντων	19 έως 36, (mean = 23.8 years, stdev. = 3.3 years)
Συνολική διάρκεια	172.92 min, (mean = 10.8 min, stdev. = 1.04 min)
Βιοσήματα	3-axis Acc. (32Hz), BVP (64Hz), EDA (4Hz), heart rate (1Hz), IBI (n/a), body temperature (4Hz), EEG (8 band, 32Hz), ECG (1Hz)
Ακουστικές ροές	172.92 min (from 16 debate sessions)
Οπτικές ροές	223.35 min (from 21 participants)
Επισημάνσεις συναισθημάτων	Self: 4.159 Partner: 4.159 5 external observers: 20.803

Πίνακας 4.2: Σύνοψη περιεχομένων συνόλου δεδομένων.

Κατηγορίες συναισθημάτων	Περιγραφή κατηγορίας	Κλίμακα
Arousal/Valence	Two affective dimensions from Russell's circumplex model of affect [56]	1: very low, 2: low, 3: neutral, 4: high, 5: very high
Cheerful/Happy Angry/Nervous/ Sad	Emotions states describing a subjective stress state [57]	1: very low, 2: low, 3: high, 4: very high
Boredom/Confusion/ Engaged concentration/ Frustation/Surprise/ Delight/None	Commonly used Baker Rodrigo Ocumpaugh Monitoring Protocol educationally relevant affective categories [58]	Choose one
Confrustion/Contempt/ Dejection/Disgust/ Pride/Sorrow/ Eureka/None	Less Commonly used Baker Rodrigo Ocumpaugh Monitoring Protocol educationally relevant affective categories [58]	Choose one

Πίνακας 4.3: Σύνοψη κατηγοριών συναισθημάτων που χρησιμοποιήθηκαν για επισήμανση.

Από το σύνολο των συλλεγόμενων δεδομένων, εστιάζουμε σε φυσιολογικά σήματα για την αναγνώριση των συναισθημάτων. Συγκεκριμένα, χρησιμοποιούμε δεδομένα που συλλέγονται με τη συσκευή Empatica E4 (BVP, EDA, HST) και τη συσκευή Polar H7 (ECG) για την πρόβλεψη της έντασης των συναισθημάτων σθένους και διέγερσης (Arousal & Valence), όπως διακρίνεται στον πίνακα 4.3. Η συγκεκριμένη επιλογή εξυπηρετεί τον στόχο μας για την ανάπτυξη ενός μοντέλου αναγνώρισης συναισθημάτων σε εφαρμογές παρακολούθησης της καθημερινής ζωής και εφαρμογές HMI, καθώς τα περιφερειακά σήματα που παρέχονται από μη επεμβατικές φορητές συσκευές είναι περισσότερα κατάλληλα για χρήση.

4.2 Προ-επεξεργασία Συνόλου Δεδομένων

Τα δεδομένα του K-EmoCon [2] έχουν συγχρονιστεί χρονικά μεταξύ τους καθώς έχουν μετατραπεί όλες οι χρονικές σημάνσεις από τη ζώνη UTC +9 στη ζώνη UTC +0, ενώ τα ακατέργαστα δεδομένα έχουν αποκοπεί έτσι ώστε να περιλαμβάνουν μόνο τα τμήματα των δεδομένων που αντιστοιχούν στις συζητήσεις και στις βασικές μετρήσεις. Σχετικά με το οπτικοακουστικό υλικό, έγινε επιπλέον περικοπή για την εξαγωγή του υλικού που αντιστοιχεί σε συζητήσεις σε αντίθεση με τα φυσιολογικά σήματα που αποκόπηκαν από την αρχή της συνεδρίας συλλογής δεδομένων έως το τέλος της συζήτησης, καθώς τα αρχικά 2 λεπτά αμέσως μετά την έναρξη μιας συνεδρίας αντιστοιχούν σε μια βασική μέτρηση για την εξαγωγή φυσιολογικών δεδομένων σε μια ουδέτερη συναισθηματική κατάσταση.

Η προεπεξεργασία των ακατέργαστων φυσιολογικών δεδομένων και η εξαγωγή χαρακτηριστικών από αυτά αποτελούν τον πυρήνα της ανάπτυξης ενός μοντέλου αναγνώρισης συναισθημάτων. Σε ένα πρώτο πλαίσιο, υλοποιήθηκε διαχωρισμός των φυσιολογικών σημάτων σε 2 χρονικές περιόδους, αυτή των βασικών μετρήσεων και αυτή των συζητήσεων. Στην περίπτωση της υπάρξης μεγαλύτερου αριθμού συναισθηματικών επισημάνσεων από τον συμμετέχοντα σε σύγκριση με τις επισημάνσεις από τον συνομιλητή ή/και τους εξωτερικούς παρατηρητές για ένα σύνολο συμμετεχόντων, υλοποιήθηκε περικοπή των μεγαλύτερων αρχείων επισημάνσεων από την αρχή του αρχείου, καθώς οι επιπλέον επισημάνσεις στην αρχή είναι πιθανώς από λανθασμένες συναισθηματικές επισημάνσεις κατά τις βασικές μετρήσεις.

Σε ένα δεύτερο πλαίσιο, πραγματοποιήθηκε χρονικός συγχρονισμός μεταξύ των φυσιολογικών σημάτων ενδιαφέροντος μεταξύ τους. Τα συγχρονισμένα πλέον φυσιολογικά σήματα συγχρονίστηκαν και με το σύνολο των συναισθηματικών επισημάνσεων από όλες τις οπτικές γωνίες.

Τέλος, παρόλο που τα φυσιολογικά σήματα στο σύνολο δεδομένων είναι ως επί το πλείστον χωρίς σφάλματα, με τα περισσότερα από τα αρχεία δεδομένων να είναι συμπληρωμένα πάνω από το 95% από το σύνολο των πιναθών καταγραφών, εντοπίζονται περιπτώσεις συμμετεχόντων όπου ένα τμήμα δεδομένων λείπει λόγω προβλημάτων που σχετίζονται με σφάλμα της συσκευής ή ανθρώπινο. Στην περίπτωση που τα προαναφερθέντα σφάλματα αφορούν συσκευές καταγραφής των φυσιολογικών σημάτων ενδιαφέροντος, οι συμμετέχοντες αποκλείστηκαν από τη συνέχεια της ανάλυσης με αποτέλεσμα από το σύνολο των 32 συμμετεχόντων, ο αριθμός αυτός να μειωθεί στους 21, όπως αποτυπώνεται αναλυτικά και στον πίνακα 4.4.

Συμμετεχών	Λόγος αποκλεισμού
P2, P3, P6, P7	Empatica E4 malfunction
P17, P20	Empatica E4 poor contact between the device and the participant's skin, resulting to missing EDA values
P12, P18, P21, P29, and P30	Polar HR malfunction

Πίνακας 4.4: Ταύτοτητα συμμετέχοντος και λόγος αποκλεισμού από την συνέχεια της ανάλυσης.

Στις επόμενες υπο-ενότητες, συζητάμε τις διαδικασίες προ-επεξεργασίας των φυσιολογικών σημάτων και συναισθημάτων επισημάνσεων για την εκπαίδευση των μοντέλων αναγνώρισης συναισθημάτων που αναπτύχθηκαν.

4.2.1 Εξαγωγή Χαρακτηριστικών από Φυσιολογικά Σήματα Ενδιαφέροντος

Για την εξαγωγή χαρακτηριστικών από τα φυσιολογικά σήματα, χρησιμοποιήσαμε την εργαλειοθήκη PyTEAP [59], η οποία αποτελεί μια παραλλαγή της εργαλειοθήκης TEAP [60], γραμμένη στη γλώσσα προγραμματισμού Python. Η συγκεκριμένη εργαλειοθήκη χρησιμοποιείται ευρέως για τη μελέτη συναισθηματικών καταστάσεων μέσω κατάλληλων χαρακτηριστικών που εξάγονται από φυσιολογικά σήματα. Ενώ η αρχική έκδοση του TEAP είναι γραμμένη σε MATLAB, χρησιμοποιήσαμε την παραλλαγή της σε Python, καθώς η συγκεκριμένη γλώσσα προγραμματισμού υποστηρίζει εργαλειοθήκες κατάλληλες για την ανάπτυξη μοντέλων μηχανικής και βαθιάς μάθησης, όπως το PyTorch [61].

Τα φυσιολογικά σήματα που υποστηρίζει η εργαλειοθήκη PyTEAP συμπεριλαμβάνουν το ηλεκτροκαρδιογράφημα (ECG), τον παλμό όγκου του αίματος (BVP), την ηλεκτροδρεμική δραστηριότητα (EDA), την θερμοκρασία του ανθρώπινου δέρματος (T), το ηλεκτρομυογράφημα (EMG), την αναπνοή (RES) και το ηλεκτροεγκεφαλογράφημα (EEG).

Παρακάτω αναλύονται τα χρησιμοποιούμενα φυσιολογικά σήματα, καθώς και τα χαρακτηριστικά που εξήχθηκαν μέσω του PyTEAP. Το ελάχιστο μήκος ενός σήματος που υποστηρίζει το PyTEAP για την εξαγωγή χαρακτηριστικών είναι 25 δευτερόλεπτα. Η εργαλειοθήκη εφαρμόζει επίσης αυτόματα φίλτρα χαμηλής διέλευσης για την αφαίρεση των διαταραχών υψηλής συχνότητας στα σήματα που συλλέγονται σε υψηλές συχνότητες δειγματοληψίας. Για το K-EmoCon, ωστόσο, το φιλτράρισμα εφαρμόζεται μόνο για το BVP, που έχει ρυθμό δειγματοληψίας 64 Hz, στο 1/8 του αρχικού ρυθμού. Για τα υπόλοιπα σήματα οι συχνότητες δειγματοληψίας είναι ήδη πολύ χαμηλές για οποιοδήποτε φιλτράρισμα.

Ηλεκτροδερμική Δραστηριότητα (EDA): Η ηλεκτροδερμική δραστηριότητα αποτελεί ένα μέτρο της ηλεκτρικής αντίστασης στην επιφάνεια του δέρματος [62]. Αυτή η τιμή μειώνεται σε καταστάσεις χαλάρωσης, αυξάνεται σε καταστάσεις προσπάθειας, ενώ αλλάζει αισθητά με την ποσότητα της έκκρισης ιδρώτα, που είναι υπό τον έλεγχο του συμπαθητικού νευρικού συστήματος. Το συγκεκριμένο τμήμα του νευρικού συστήματος μεσολαβεί στην απόκριση «μάχης ή φυγής» που ενεργοποιείται στην αίσθηση απειλής και σχετίζεται με τις συναισθηματικές καταστάσεις του φόβου και του άγχους. Επομένως, με την μέτρηση του σήματος της ηλεκτροδερμικής δραστηριότητας και κατόπιν την εξαγωγή και ανάλυση χαρακτηριστικών, δύναται να εντοπίσουμε τη χρονική στιγμή εκδήλωσης των παραπάνω συναισθηματικών καταστάσεων.

Η ηλεκτροδερμική δραστηριότητα χαρακτηρίζεται από 2 τύπους αλλαγών, με την πρώτη να σχετίζεται με αργά μεταβαλλόμενα επίπεδα αγωγιμότητας, χωρίς ιδιαίτερη παρέμβαση από εξωτερικά ερεθίσματα. Η δεύτερη είναι η φασική απόκριση αγωγιμότητας του δέρματος και χαρακτηρίζεται από γρήγορες διακυμάνσεις στην αγωγιμότητα υπό την επίδραση κάποιου εξωτερικού παράγοντα. Οι αλλαγές αυτές αναφέρονται ως αποκρίσεις αγωγιμότητας του δέρματος. Παρακάτω, παρατίθενται τα εξαγόμενα χαρακτηριστικά.

- **Κορυφές ανά δευτερόλεπτο**, ο μέσος αριθμός κορυφών άνω των 100 ohms ανά δευτερόλεπτο.
- **Μέσο πλάτος**, το μέσο πλάτος κορυφών από το σαγματικό σημείο που προηγείται της κορυφής.
- **Μέσος χρόνος ανόδου**, ο μέσος χρόνος για το σήμα να φτάσει τις κορυφές από τα σαγματικά σημεία σε δευτερόλεπτα.
- **Στατιστικές ροπές**, ο μέσος όρος και η τυπική απόκλιση του σήματος.

Θερμοκρασία Δέρματος (T): Η θερμοκρασία του δέρματος υπολογίζεται σε βαθμούς κελσίου και μπορεί να χρησιμοποιηθεί για να αναγνωρίσει εάν ένα άτομο είναι σε μια χαλαρή κατάσταση ή όχι. Τα διεσταλμένα αγγεία θα προσδώσουν μια πιο ζεστή θερμοκρασία στο σημείο μέτρησης κατά τη χαλάρωση ή πιο κρύα κατά τη διάρκεια του στρες ή του άγχους, όπου τα αγγεία συστέλλονται. Γενικά, αποτελεί ένα σχετικά αργό δείκτη αλλαγών συναισθηματικών καταστάσεων [63]. Παρακάτω, παρατίθενται τα εξαγόμενα χαρακτηριστικά.

- **Φασματική πυκνότητα ισχύος**, που υπολογίζεται για $f \in \{[0, 0.1], [0.1, 0.2]\} \text{ Hz}$ και δίνεται από τον ακόλουθο τύπο.

$$PSD = \log(P_x(f)) \quad (4.1)$$

- **Στατιστικές ροπές**, ο μέσος όρος, η τυπική απόκλιση, η κύρτωση και η λοξότητα του σήματος.

Παλμός όγκου αίματος (BVP): Ο παλμός όγκου του αίματος αποτελεί ένα μέτρο για τον προσδιορισμό της ποσότητας του αίματος που ρέει στα αγγεία και για την μέτρησή του χρησιμοποιείται ένα φωτοπληθυσμογράφημα (PPG). Σχετίζεται στενά με το IBI, το οποίο μετρά τις χρονικές αποστάσεις μεταξύ δύο διαδοχικών καρδιακών παλμών. Ο παλμός όγκου του αίματος χρησιμοποιείται συχνά για την αναγνώριση συναισθημάτων, καθώς αυξάνεται κατά τη διάρκεια των συναισθηματικών καταστάσεων του θυμού ή του στρες και μειώνεται κατά τη διάρκεια της θλίψης και της χαλάρωσης. Παρακάτω, παρατίθενται τα εξαγόμενα χαρακτηριστικά.

- **Διάστημα μεταξύ παλμών (IBI)**, όπου υπολογίζεται η μέση τιμή του διαστήματος μεταξύ 2 παλμών, καθώς και η τυπική απόκλιση του συγκεκριμένου διαστήματος, η οποία συχνά αναφέρεται και ως μεταβλητότητα καρδιακού ρυθμού (HRV).
- **Εντροπία πολλαπλής κλίμακας (MSE)**, όπου εξετάζεται η χρονοσειρά σε 5 διαφορετικά επίπεδα χρονικών αναλύσεων.
- **Ισχύς ταχογράμματος**, που υπολογίζεται σε 3 επίπεδα συχνοτήτων, ενώ υπολογίζεται και ο λόγος μεταξύ της ισχύος του μεσαίου επιπέδου ως προς το άθροισμα της ισχύος του υψηλού και του χαμηλού επιπέδου, όπως απεικονίζεται παρακάτω.

$$\text{Tachogram Power} = \log(P_x^{LF}(f)), \quad f \in [0.01, 0.08] \text{ Hz} \quad (4.2)$$

$$\text{Tachogram Power} = \log(P_x^{MF}(f)), \quad f \in [0.08, 0.15] \text{ Hz} \quad (4.3)$$

$$\text{Tachogram Power} = \log(P_x^{HF}(f)), \quad f \in [0.15, 0.4] \text{ Hz} \quad (4.4)$$

$$Ratio = \frac{\log(P_x^{MF}(f))}{\log(P_x^{LF}(f)) + \log(P_x^{HF}(f))} \quad (4.5)$$

- **Φασματική πυκνότητα ισχύος**, που υπολογίζεται για 4 επίπεδα συχνοτήτων, $f \in \{[0, 0.1], [0.1, 0.2], [0.2, 0.3], [0.3, 0.4]\} \text{ Hz}$, καθώς ο λόγος μεταξύ της ισχύος των χαμηλών συχνοτήτων, $f \in \{[0.08, 0.15]\} \text{ Hz}$, προς την ισχύ των υψηλών συχνοτήτων, $f \in \{[0.15, 0.4]\} \text{ Hz}$.

$$PSD = \log(P_x(f)) \quad (4.6)$$

$$Ratio = \frac{\log(P_x^{LF}(f))}{\log(P_x^{HF}(f))} \quad (4.7)$$

- **Στατιστικές ροπές**, όπου υπολογίζεται ο μέσος όρος του σήματος.

Ηλεκτροκαρδιογράφημα (ECG): Στο K-EmoCon, το ηλεκτροκαρδιογραφήμα ισοδυναμεί με τους παλμούς ανά λεπτό (BPM) και υπολογίζονται ο μέσος όρος και η τυπική απόκλιση του σήματος.

Τα παραπάνω χαρακτηριστικά δύναται να εξαχθούν για σήματα ελάχιστης διάρκειας 25 δευτερολέπτων και είναι στο σύνολό τους 30. Στη συνέχεια, πραγματοποιήθηκε κανονικοποίηση των χαρακτηριστικών ώστε να έχουν μηδενικό μέσο όρο και μοναδιαία διακύμανση (z -standardization), όπως απεικονίζεται στην εξίσωση 4.8, όπου μ είναι η μέση τιμή και σ η διακύμανση ενός χαρακτηριστικού.

Η κανονικοποίηση των χαρακτηριστικών είναι μια κοινή απαίτηση ενός μεγάλου αριθμού ταξινομητών. Για παράδειγμα, πολλά στοιχεία που χρησιμοποιούνται στην αντικειμενική συνάρτηση ενός αλγορίθμου εκμάθησης, όπως ο πυρήνας RBF των Μηχανών Διανυσματικής Υποστήριξης, υποθέτουν ότι όλα τα χαρακτηριστικά είναι κεντραρισμένα γύρω από το 0 και έχουν διακύμανση ίδιας τάξης μεγέθους. Εάν ένα χαρακτηριστικό έχει διακύμανση που είναι τάξεις μεγέθους μεγαλύτερη από άλλα, μπορεί να κυριαρχεί στην αντικειμενική συνάρτηση, με αποτέλεσμα την αδυναμία εκμάθησης των υπολοίπων.

$$z = \frac{x - \mu}{\sigma} \quad (4.8)$$

4.2.2 Δειγματοληπτικός Συγχρονισμός Φυσιολογικών Σημάτων

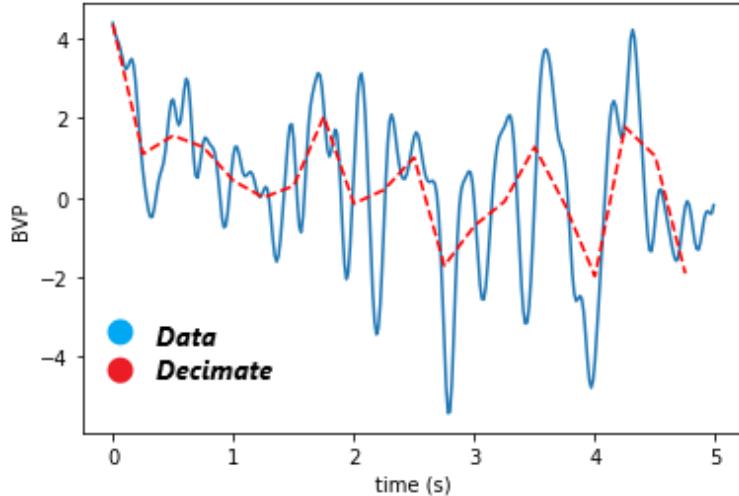
Μια σημαντική διαφορά και πλεονέκτημα των νευρωνικών δικτύων σε σύγκριση με τα παραδοσιακά μοντέλα μηχανικής μάθησης είναι ότι επιτρέπουν την παράκαμψη των βημάτων εξαγωγής χαρακτηριστικών, καθώς υλοποιούν την συγκεκριμένη διαδικασία αυτομάτα. Με σκοπό την εισαγωγή των ακατέργαστων δεδομένων του K-EmoCon στις αρχιτεκτονικές νευρωνικών δικτύων που αναλύονται παρακάτω, εφαρμόστηκε ελάχιστη προ-επεξεργασία στα φυσιολογικά σήματα ενδιαφέροντος. Συγκεκριμένα, εφαρμόστηκε δειγματοληπτικός συγχρονισμός μεταξύ τους, καθώς και κανονικοποίηση του κάθε σήματος ξεχωριστά με βάση τη μέση τιμή μ και την διακύμανση σ .

Πιο αναλυτικά, οι ρυθμοί δειγματοληψίας διαφέρουν μεταξύ των φυσιολογικών σημάτων. Για το λόγο αυτό, καθίσταται αναγκαίο να πραγματοποιηθεί υποδειγματοληψία για το σήμα του παλμού όγκου αίματος (BVP) από την αρχική τιμή των 64 Hz στα 4 Hz, καθώς και μετατροπή του ρυθμού δειγματοληψίας από την αρχική τιμή του 1 Hz στα 4Hz για το σήμα του ηλεκτροκαρδιογραφήματος (ECG). Η παραπάνω διαδικασία κρίνεται απαραίτητη για την αντιστοίχηση των ρυθμών δειγματοληψίας στα 4 Hz, που είναι ο ρυθμός δειγματοληψίας για τα σήματα της ηλεκτροδερμικής δραστηριότητας (EDA) και της θερμοκρασίας του δέρματος (T). Τα φυσιολογικά σήματα είναι αναγκαίο να έχουν το ίδιο μήκος από τη στιγμή που πρόκειται να εισαχθούν σε ένα βαθύ νευρωνικό δίκτυο.

Μετά το τέλος της επαναδειγματοληψίας, τα 4 μονοδιάστατα φυσιολογικά σήματα στοιβάζονται για να σχηματιστούν έναν πίνακα διάστασης $4 \times$ (ρυθμός δειγματοληψίας \times διάρκεια χρονικού παραθύρου). Για παράδειγμα, για ένα χρονικό παράθυρο 5 δευτερολέπτων σχηματίζεται ένας πίνακας διάστασης 4×20 . Τέλος, υλοποιείται κανονικοποίηση του κάθε σήματος ξεχωριστά με βάση τη μέση τιμή μ και την διακύμανση σ του σήματος, χρησιμοποιώντας την εξίσωση 4.8.

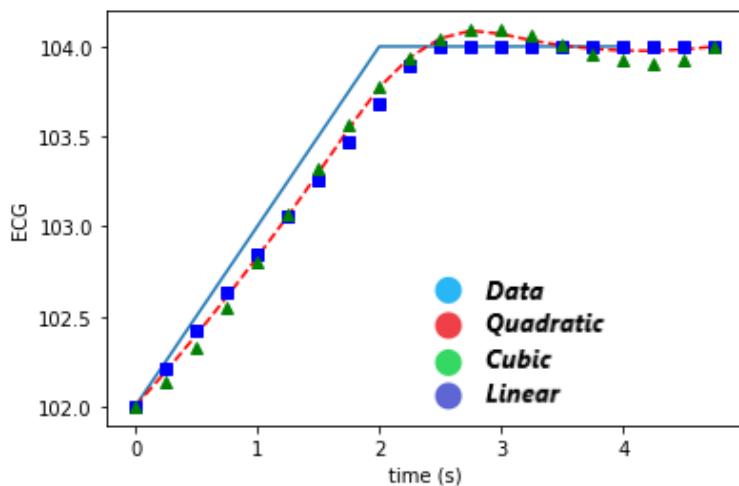
Παρακάτω, αναλύονται οι τεχνικές που χρησιμοποιήθηκαν για την υποδειγματοληψία του σήματος του παλμού όγκου αίματος (BVP) και την αναδειγματοληψία του σήματος του ηλεκτροκαρδιογραφήματος (ECG).

Υποδειγματοληψία: Για την επίτευξη της υποδειγματοληψίας, χρησιμοποιήθηκε η εργαλειοθήκη SciPy [64] και συγκεκριμένα η μέθοδος decimate. Η μέθοδος αυτή επιτελεί την υποδειγματοληψία του σήματος μετά την εφαρμογή ενός φίλτρου κατά της παραμόρφωσης. Στην περίπτωσή μας χρησιμοποιήθηκε ένα φίλτρο Chebyshev 8ής τάξης, το οποίο αποτελεί και την προεπιλογή της μεθόδου, ενώ μια γραφική απεικόνιση της υποδειγματοληψίας αναπαρίσταται στο Σχήμα 4.4.



Σχήμα 4.4: Υποδειγματοληψία BVP

Αναδειγματοληψία: Για την μετατροπή του ρυθμού δειγματοληψίας από την αρχική τιμή του 1 Hz στα 4 Hz, χρησιμοποιήθηκε η εργαλειοθήκη SciPy και συγκεκριμένα η μέθοδος interp1d. Η συγκεκριμένη μέθοδος παρέχει πληθώρα διαφορετικών ειδών παρεμβολής που δύναται να χρησιμοποιηθούν. Στο Σχήμα 4.5, απεικονίζεται ένα υποσύνολο από τα διαφορετικά είδη παρεμβολής. Το είδος που επιλέχθηκε είναι το quadratic, όπου η επιλογή έιναι ευρετικά μετά από οπτική επιθεώρηση μεγάλου αριθμού σχημάτων.



Σχήμα 4.5: Αναδειγματοληψία ECG

4.3 Προτεινόμενες Αρχιτεκτονικές Νευρωνικών Δικτύων

Τα ανατροφοδοτούμενα νευρωνικά δίκτυα σχεδιάστηκαν ειδικά για την αντιμετώπιση προβλημάτων όπου τα δεδομένα είναι διαδοχικά και είναι ικανά να κωδικοποιούν το διαδοχικό τους περιθάλλον στην κρυφή κατάσταση $\mathbf{h}(t)$, η οποία θεωρητικά μπορεί να αποθηκεύσει απείρως μακροπρόθεσμες εξαρτήσεις.

Στην πράξη ωστόσο, τα RNN έχουν παρουσιάσει προβλήματα για την καταγραφή μακροπρόθεσμων εξαρτήσεων, κυρίως λόγω του προβλήματος της πολύ μικρής τιμής που λαμβάνει η κλίση της συνάρτησης κόστους για αυτές τις εξαρτήσεις (Vanishing Gradients). Με σκοπό την αντιμετώπιση αυτού του προβλήματος, το LSTM και άλλες παραλλαγές, όπως το GRU, έχουν προταθεί, αποκτώντας μεγάλη δημοτικότητα.

Στις επομένες 2 υποενότητες, παρουσιάζονται οι αρχιτεκτονικές ανατροφοδοτούμενων νευρωνικών δικτύων που αναπτύχθηκαν και αξιολογήθηκαν στα πλαίσια της διπλωματικής εργασίας. Και οι 2 αρχιτεκτονικές βασίζονται σε ένα επίπεδο μακράς βραχυπρόθεσμης μνήμης, με σκοπό την μοντελοποίηση της χρονικής εξάρτησης των συναισθηματικών καταστάσεων, σε αντίθεση με παραδοσιακές προσεγγίσεις μηχανικής μάθησης που βασίζονται στην εξαγωγή χρονικά ασυσχέτιστων χαρακτηριστικών από την χρονοσειρά.

4.3.1 Αρχιτεκτονική ενός επιπέδου μακράς βραχυπρόθεσμης μνήμης

Στόχος της παρούσας αρχιτεκτονικής είναι να προβλέψει την συναισθηματική κατάσταση ενός συμμετέχοντα για ένα χρονικό παράθυρο διάρκειας ελάχιστης διάρκειας 25 δευτερολέπτων, εκμεταλλευόμενοι την ακολουθία των επαναδειγματολειπτημένων φυσιολογικών σημάτων στο επιλεγόμενο χρόνικο παράθυρο. Απότερος σκοπός αποτελεί η μοντελοποίηση της εξάρτησης της συναισθηματικής κατάστασης από το σύνολο των τιμών της ακολουθίας μέσω της κρυφής κατάστασης $\mathbf{h}(t)$ του τελευταίου χρονικού βήματος του παραθύρου. Πιο αναλύτικα, η συγκεκριμένη αρχιτεκτονική απεικονίζεται στο Σχήμα 4.6 και σχηματίζεται σε 3 επίπεδα.

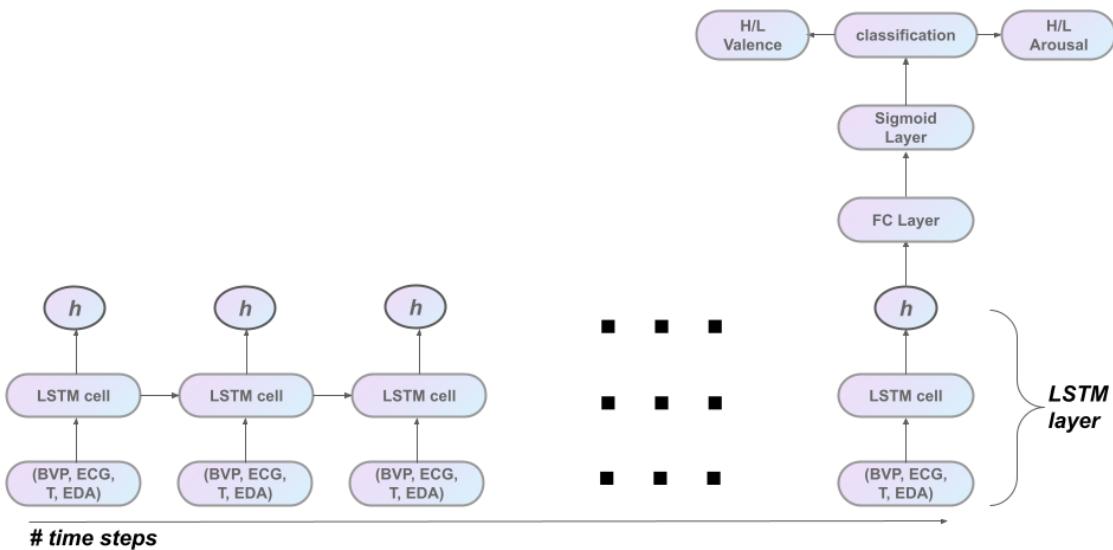
Επίπεδο Μακράς Βραχυπρόθεσμης μνήμης: Η διάσταση εισόδου του συγκεκριμένου επιπέδου ισούται με το σύνολο των τιμών από τα φυσιολογικά σήματα ενδιαφέροντος για ένα χρονικό βήμα, με αποτέλεσμα να είναι ίση με 4. Το σύνολο των χρονικών βημάτων ισούται με το γίνομενο του ρυθμού δειγματοληψίας και της διάρκειας του χρονικού παραθύρου που επιθυμούμε να γίνει η πρόβλεψη, όπως απεικονίζεται στην εξίσωση 4.9.

$$\text{time steps} = 4 \text{ (Hz)} \times \text{time window length} \quad (4.9)$$

Από τη σύνολο των κρυφών κατάστασεων, επίλεγουμε αυτή του τελευταίου χρονικού βήματος καθώς περιέχει πληροφορίες για όλα τα προηγούμενα χρονικά βήματα, μοντελοποιώντας με αυτό τον τρόπο την εξάρτηση της συναισθηματικής κατάστασης από όλο το σύνολο των βημάτων.

Πλήρως συνδεδεμένο επίπεδο: Η διάσταση εισόδου του συγκεκριμένου επιπέδου ισούται με την διάσταση του διανύσματος της κρυφής κατάστασης του προηγούμενου επιπέδου, ενώ ως έξοδος παρέχεται μια τιμή, η οποία αποτελεί την ακατέργαστη τιμή πρόβλεψης (logit).

Σιγμοειδές επίπεδο: Στόχος του συγκεκριμένου επιπέδου είναι να μετατρέψει την μη-κανονικοποιήμενη τιμή εξόδου του πλήρως συνδεδεμένου επιπέδου σε μια τιμή στο διάστημα $[0, 1]$. Η συγκεκριμένη τιμή μας δίνει την πιθανότητα ταξινόμησης στην κλάση που έχουμε ορίσει ως θετική (1) και για το λόγο αυτό η πρόβλεψη πραγματοποιείται με τη σύγκριση με ένα κατώφλι ίσο με 0.5 και κατόπιν ταξινόμηση είτε στην θετική, είτε στην αρνητική κλάση.



Σχήμα 4.6: Αρχιτεκτονική ενός επιπέδου μακράς βραχυπρόθεσμης μνήμης

4.3.2 Εμπλουτισμός αρχιτεκτονικής με ένα επίπεδο προσοχής

Στόχος της παρούσας αρχιτεκτονικής είναι να προβλέψει την συναισθηματική κατάσταση ενός συμμετέχοντα για ένα χρονικό παράθυρο ελάχιστης διάρκειας 25 δευτερολέπτων, εκμεταλλευόμενοι την ακολουθία των επαναδειγματολειπτημένων φυσιολογικών σημάτων στο επιλεγόμενο χρόνικο παράθυρο. Απότερο σκοπό αποτελεί η μοντελοποίηση της εξάρτησης της συναισθηματικής κατάστασης από το σύνολο των τιμών της ακολουθίας και συγκεκριμένα η εστίαση στις τιμές με το μεγαλύτερο πληροφοριακό περιεχόμενο, εμπλουτίζουμε την αρχιτεκτονική της υποενότητας 4.3.1 με ένα επιπλέον επίπεδο προσοχής.

Τα νευρωνικά δίκτυα που συμπλεριλαμβάνουν τον μηχανισμό προσοχής στην αρχιτεκτονική τους έχουν επιδείξει επιτυχία σε ένα ευρύ φάσμα επιστημονικών πεδίων [65] [66]. Στη συγκεκριμένη διπλωματική εργασία προτείνεται η εισαγωγή του μηχανισμού σε αρχιτεκτονικές νευρωνικών δικτύων που έχουν ως στόχο την αναγνώριση συναισθήματων από φυσιολογικά σήματα προερχόμενα από φορητές συσκευές καταγραφής.

Παρακάτω, πραγματοποιείται μια ανάλυση του επιπέδου προσοχής, ενώ η συνολική αρχιτεκτονική απεικονίζεται στο Σχήμα 4.7 και σχηματίζεται σε 4 επίπεδα.¹

Επίπεδο Προσοχής: Έστω ότι ο πίνακας \mathbf{H} αποτελείται από το σύνολο των διανυσμάτων των κρυφών καταστάσεων $[\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_T]$ που παράγονται από το επίπεδο μακράς βραχυπρόθεσμης μνήμης, όπου T είναι το μήκος των χρονικών βημάτων. Ο συγκεκριμένος πίνακας αποτελεί την είσοδο στο επίπεδο πρόσοχης, ενώ το διάνυσμα εξόδου του επιπέδου σχηματίζεται εφαρμόζοντας διαδοχικά τις εξισώσεις 4.10-4.13 και αποτελεί την είσοδο στο πλήρως συνδεδεμένο επίπεδο.

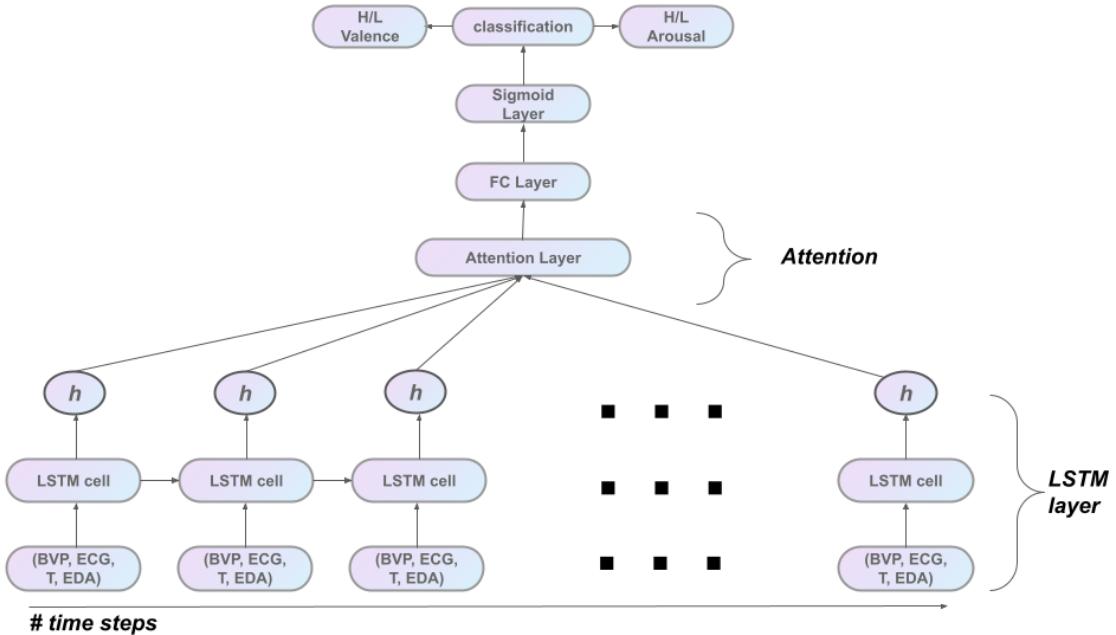
Ο Πίνακας $\mathbf{H} \in \mathbb{R}^{d^w \times T}$, όπου το d^w αποτελεί την διάσταση των διανυσμάτων των κρυφών καταστάσεων, ενώ το διάνυσμα \mathbf{w} αποτελείται από παραμέτρους που πρόκειται να εκπαιδευτούν. Οι διαστάσεις των \mathbf{w} , α , \mathbf{r} , \mathbf{h}^* είναι d^w , T , d^w και d^w , αντίστοιχα.

$$\mathbf{M} = \tanh(\mathbf{H}) \quad (4.10)$$

$$\alpha = \text{softmax}(\mathbf{w}^T \mathbf{M}) \quad (4.11)$$

$$\mathbf{r} = \mathbf{H} \alpha^T \quad (4.12)$$

$$\mathbf{h}^* = \tanh(\mathbf{r}) \quad (4.13)$$



Σχήμα 4.7: Εμπλουτισμός με ένα επίπεδο προσοχής

Κεφάλαιο 5

Θέματα πειραματικής υλοποίησης και αποτελέσματα

5.1 Σύγκριση Συναισθηματικών Επισημάνσεων Διαφορετικών Οπτικών Γωνιών

Τα υπάρχοντα σύνολα δεδομένων που αφορούν το πεδίο της συναισθηματικής ανάλυσης περιέχουν επισημάνσεις συναισθημάτων από τον συμμετέχοντα στο πείραμα και εξωτερικούς παρατηρητές που το παρακολουθούν. Παράλληλα, ακόμα και σε περιπτώσεις όπου στο πείραμα συμμετέχουν και συμπληρωματικά ατόμα (συνομιλητές ή συνεργατές), παραλείπεται η συμμετοχή τους στην διαδικασία επισήμανσης.

Το K-EmoCon [2] αποτελεί το πρώτο σύνολο δεδομένων που εμπεριέχει επισημάνσεις συναισθημάτων από όλες τις διαθέσιμες οπτικές γωνίες: του συμμετέχοντα, του συνονομιλητή και των εξωτερικών παρατηρητών, γεγονός που μας οδηγεί να ερευνήσουμε πως αντιλαμβάνεται τον μηχανισμό των συναισθημάτων κάθε μια από τις οπτικές γωνίες, μέσω των δυναμικών χαρακτηριστικών που παρουσιάστηκαν στην υποενότητα 2.5.

Πιο συγκεκριμένα, αυτό που παρουσιάζει ενδιαφέρον είναι ο τρόπος με τον οποίο επάγεται ο μηχανισμός των συναισθημάτων στον εξωτερικό παρατηρητή ο οποίος δεν συμμετέχει στη διαδικασία αλλά απλά τη βλέπει, σε σχέση με τους συμμετέχοντες που διαμορφώνουν την αλληλουχία των συναισθημάτων και κατά πόσο αυτή η συναισθηματική αδράνεια επηρεάζει την αντίληψη των νέων συναισθημάτων.

Για το λόγο αυτό, καθίσταται αναγκαίο να υπάρχουν επισημάνσεις συναισθημάτων σε ένα σύνολο που δύναται να διαχωριστεί στα υποσύνολα των θετικών (PA) και αρνητικών (NA) συναισθηματικών καταστάσεων.

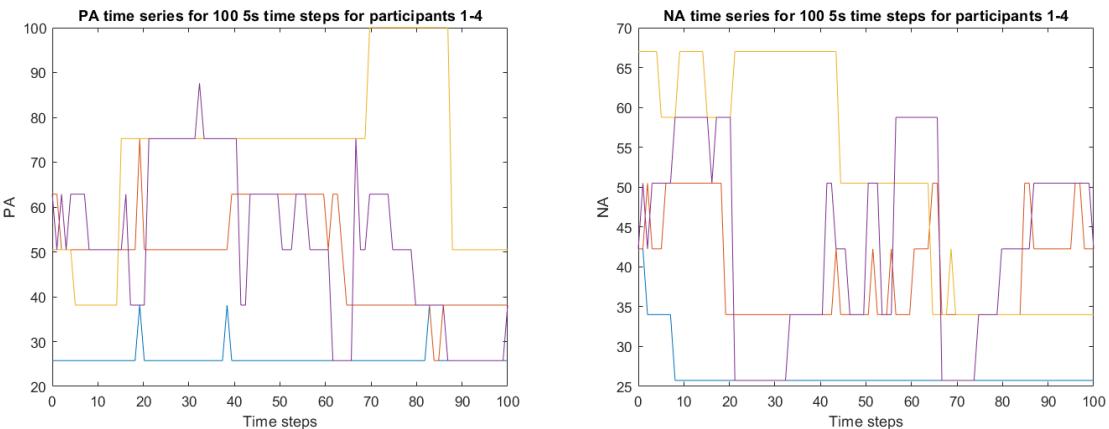
Επομένως, επιλέχθηκε το σύνολο των συναισθηματικών καταστάσεων που περιγράφουν μια υποκειμενική κατάσταση άγχους [57], τα οποία είναι μετρημένα σε μια κλίμακα από το 1 έως το 4 και διαχωρίζονται σε επιμέρους υποσύνολα ως εξής:

$$PA = \begin{cases} Cheerful \\ Happy \end{cases}$$

$$NA = \begin{cases} Angry \\ Nervous \\ Sad \end{cases}$$

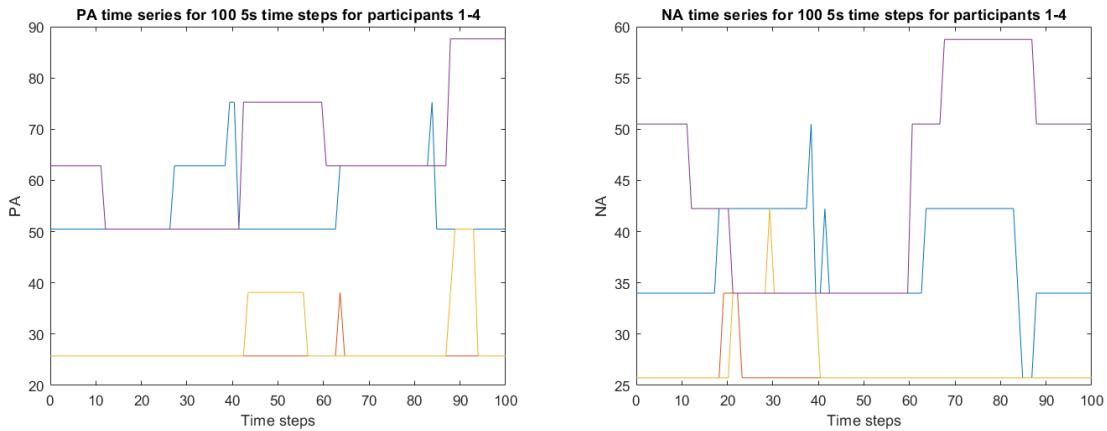
Η μοναδική τιμή της θετικής συναισθηματικής κατάστασης (PA) και της αρνητικής συναισθηματικής κατάστασης (NA) για κάθε βήμα επισήμανσης προκύπτει ως ο μέσος όρος των επιμέρους συναισθηματικών καταστάσεων που ανήκουν σε κάθε υποσύνολο για το βήμα αυτό. Στη συνέχεια, η συγκεκριμένη τιμή μετασχηματίζεται από το παρόν εύρος ([1, 4]) σε ένα μεγαλύτερο ([1, 100]). Το αποτέλεσμα είναι η δημιουργία μιας χρονοσειράς για κάθε συναισθηματική κατάσταση, συμμετέχοντα στο πείραμα και οπτική γωνία επισήμανσης.

Παρακάτω, παρατίθεται ένα παράδειγμα απεικόνισης των χρονοσειρών 4 συμμετεχόντων και 100 βημάτων επισήμανσης για κάθε συναισθηματική κατάσταση και οπτική γωνία.



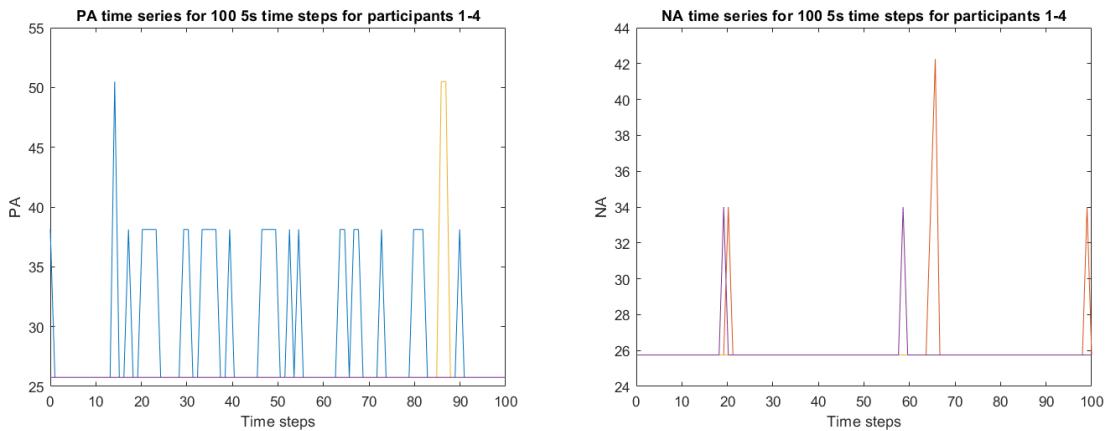
(α') χρονοσειρά θετικής συναισθηματικής κα- (β') χρονοσειρά αρνητικής συναισθηματικής κατάστασης

Σχήμα 5.1: Οπτική γωνία συμμετέχοντα



(α') χρονοσειρά θετικής συναισθηματικής κα- (β') χρονοσειρά αρνητικής συναισθηματικής κατάστασης

Σχήμα 5.2: Οπτική γωνία συνομηλιτή



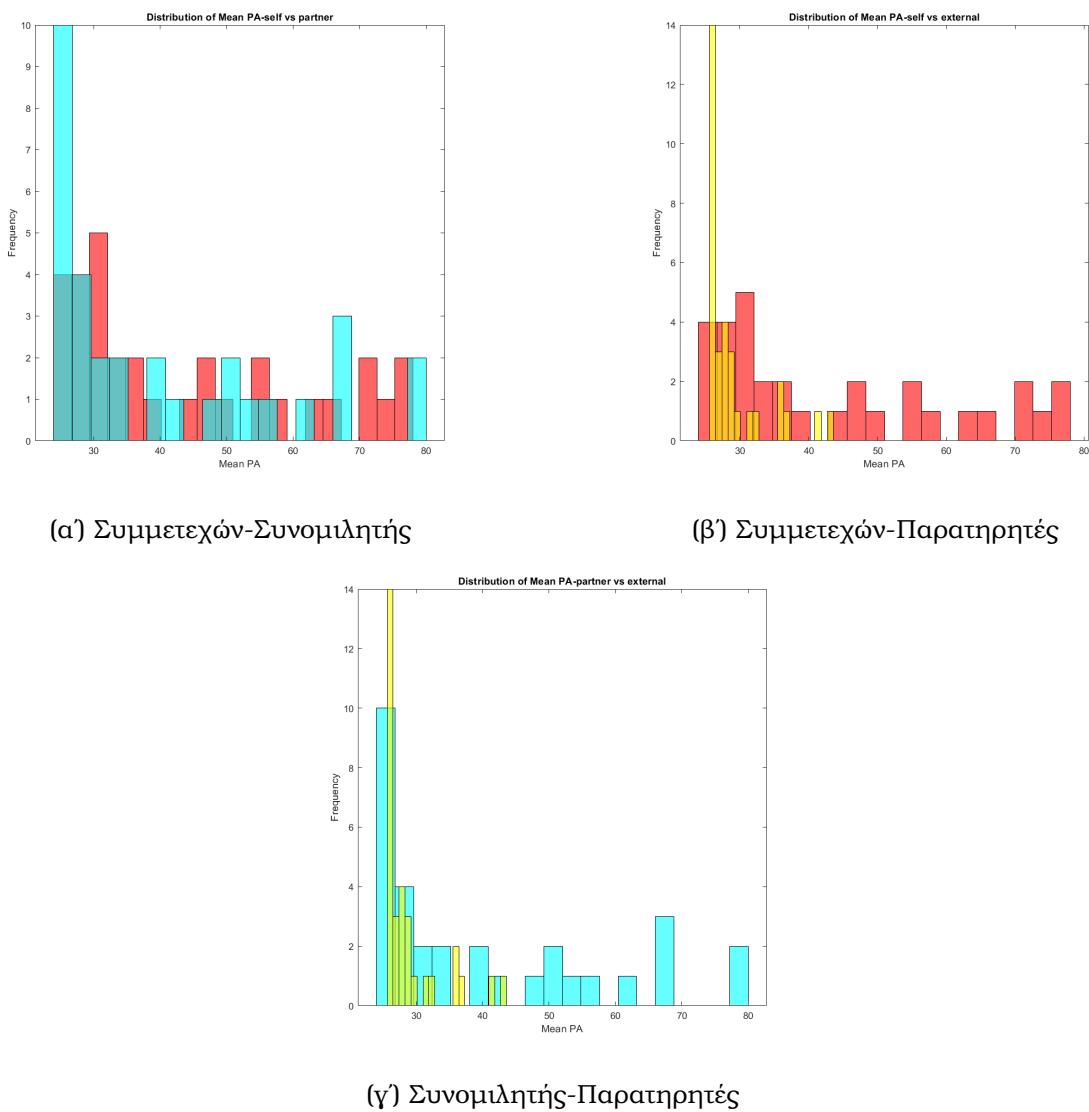
(α') χρονοσειρά θετικής συναισθηματικής κα- (β') χρονοσειρά αρνητικής συναισθηματικής κατάστασης

Σχήμα 5.3: Οπτική γωνία εξωτερικών παρατηρητών

Παρακάτω, παρατίθενται διαγράμματα που απεικονίζουν την σύγκριση μεταξύ των κατανομών κάθε δυναμικού χαρακτηριστικού μεταξύ των 32 συμμετεχόντων του συνόλου δεδομένων KEmoCon για κάθε οπτική γωνία επισήμανσης.

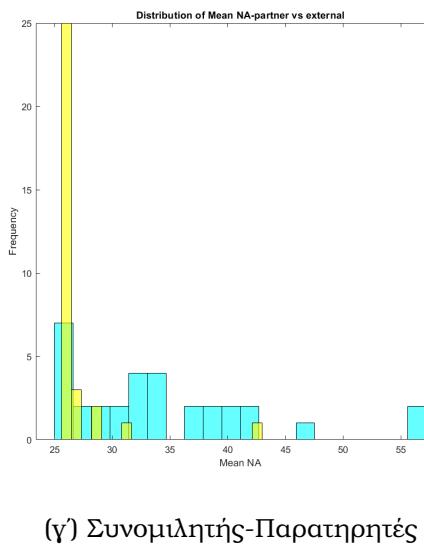
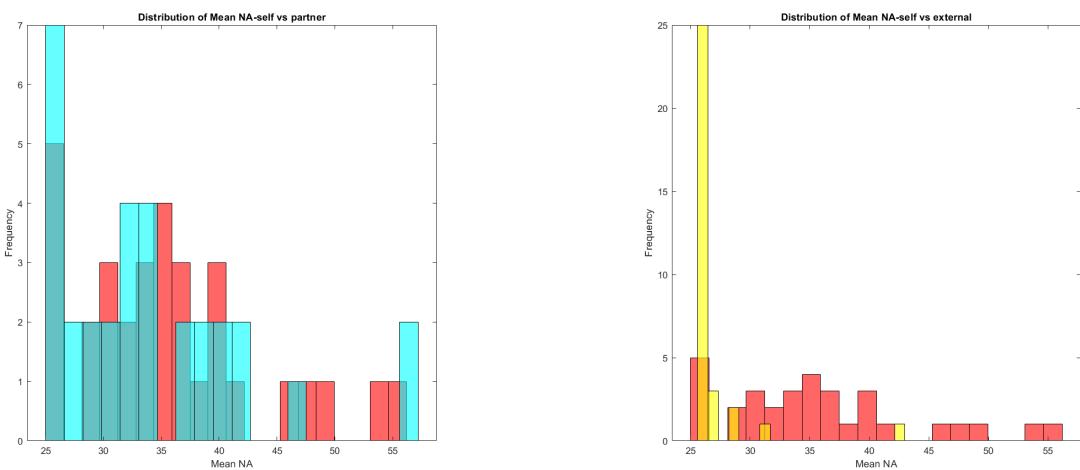
Παρατηρώντας τα παρακάτω διαγράμματα καταλήγουμε στο συμπέρασμα ότι το σύνολο των εξωτερικών παρατηρητών αντιλαμβάνεται το μέσο επίπεδο των θετικών και αρνητικών συναισθηματικών καταστάσεων διαφορετικά από τους συμμετέχοντες στη συζήτηση. Παράλληλα, παρατηρείται έντονη διαφορά στις κατανομές των συναισθηματικών καταστάσων για το δυναμικό χαρακτηριστικό της κλίσης αυτοπαλινδρόμησης μεταξύ των εξωτερικών παρατηρητών και των συμμετεχόντων. Επομένως, είτε οι εξωτερικοί παρατηρητές αδυνατούν να συλλάβουν πλήρως την αλληλουχία των συναισθημάτων καθώς δεν συμμετέχουν στη συζήτηση, είτε ένα υποσύνολο των παρατηρητών είναι πολωμένοι προς την ένταση ορισμένων επιμέρους συναισθηματικών καταστάσεων. Για το λόγο αυτό, στα πειράματα που ακολουθούν χρησιμοποιούμε τις επισημάνσεις των συμμετεχόντων στη συζήτηση.

Μέση Τιμή



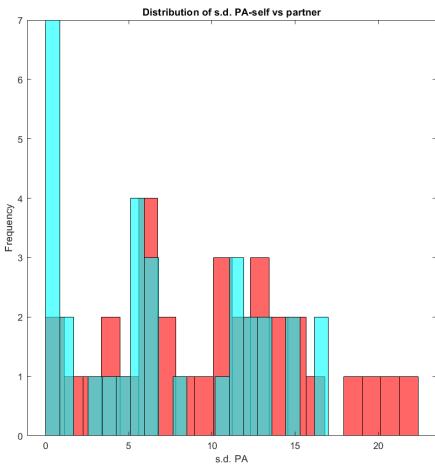
Σχήμα 5.4: Κατανομές μέσης τιμής θετικής συναισθηματικής κατάστασης

Μέση Τιμή

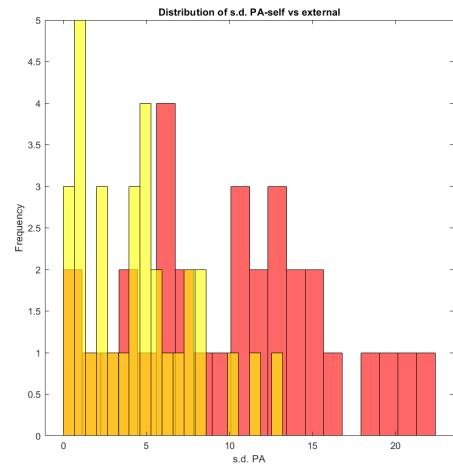


Σχήμα 5.5: Κατανομές μέσης τιμής αρνητικής συναισθηματικής κατάστασης

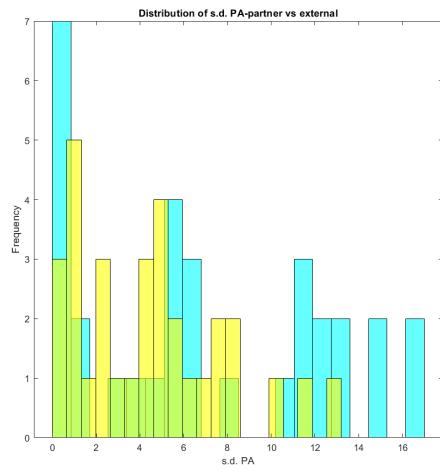
Διακύμανση



(α) Συμμετεχών-Συνομιλητής



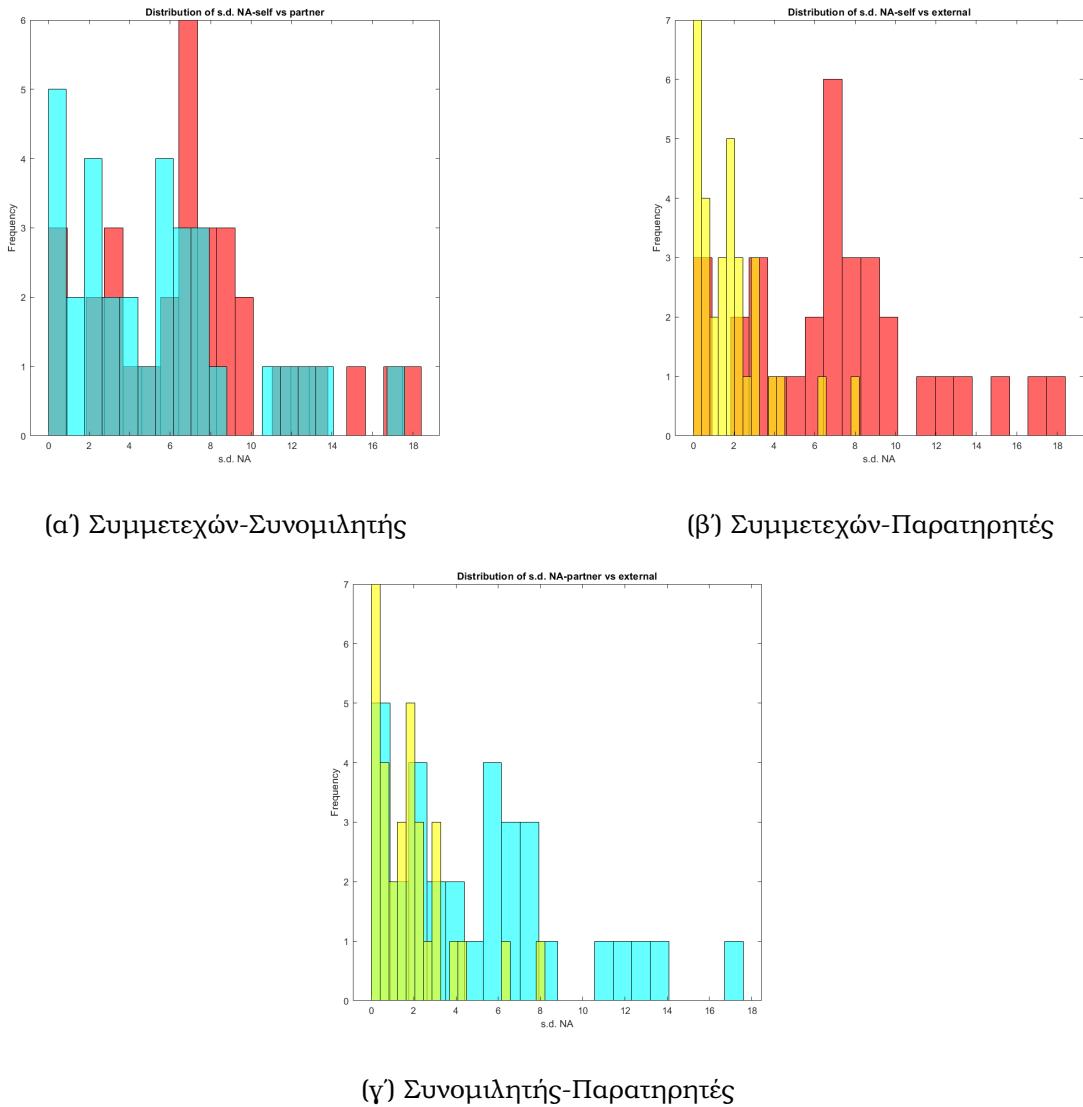
(β') Συμμετεχών-Παρατηρητές



(γ') Συνομιλητής-Παρατηρητές

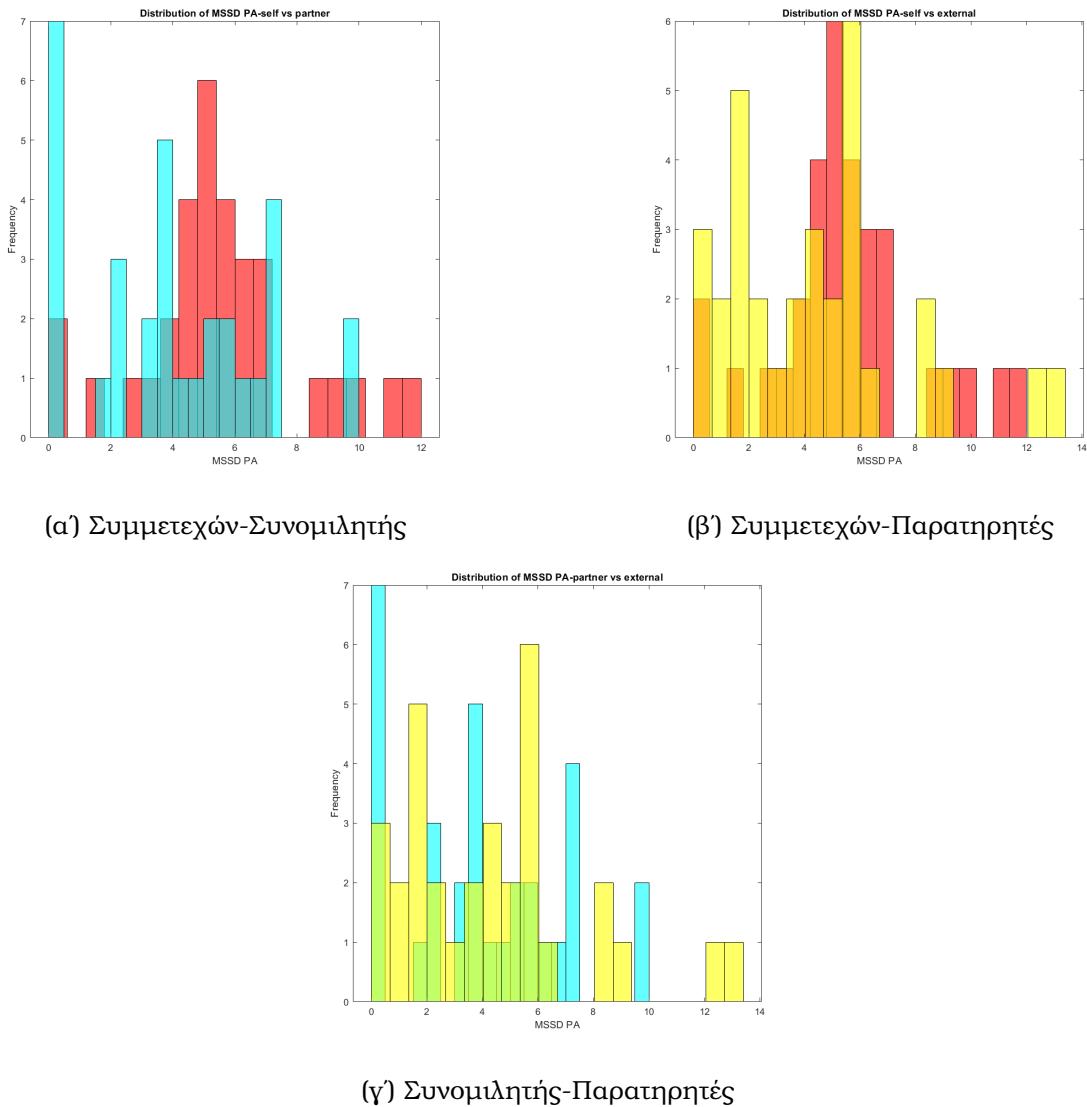
Σχήμα 5.6: Κατανομές διακύμανσης θετικής συναισθηματικής κατάστασης

Διακύμανση



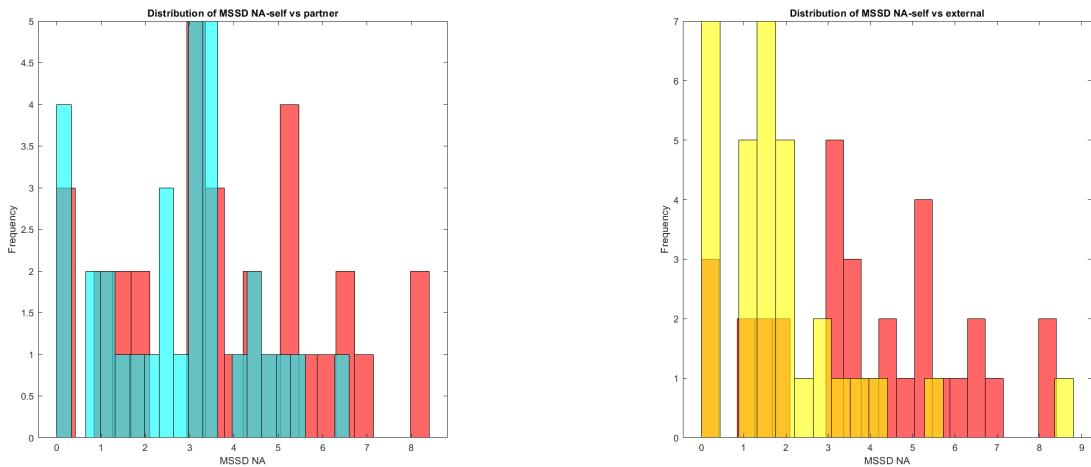
Σχήμα 5.7: Κατανομές διακύμανσης αρνητικής συναισθηματικής κατάστασης

Μέση τετραγωνική διαδοχική διαφορά



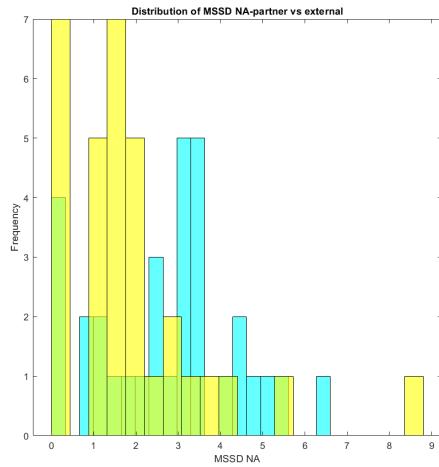
Σχήμα 5.8: Κατανομές μέσης τετραγωνικής διαδοχικής διαφοράς θετικής συναισθηματικής κατάστασης

Μέση τετραγωνική διαδοχική διαφορά



(α) Συμμετεχών-Συνομιλητής

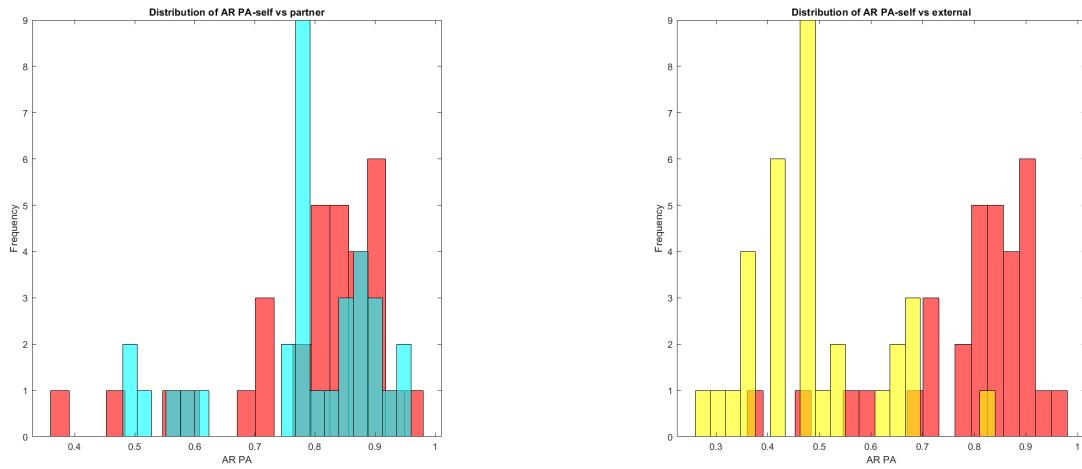
(β') Συμμετεχών-Παρατηρητές



(γ') Συνομιλητής-Παρατηρητές

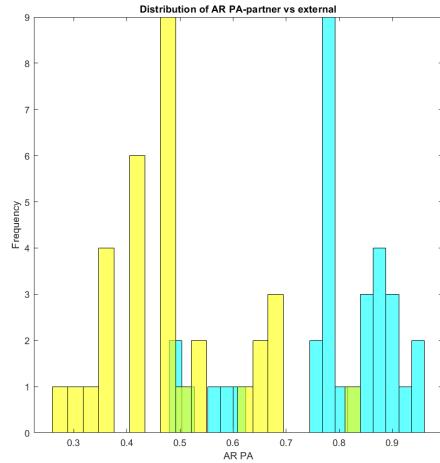
Σχήμα 5.9: Κατανομές μέσης τετραγωνικής διαδοχικής διαφοράς αρνητικής συναισθηματικής κατάστασης

Κλίση αυτοπαλινδρόμησης



(α) Συμμετεχών-Συνομιλητής

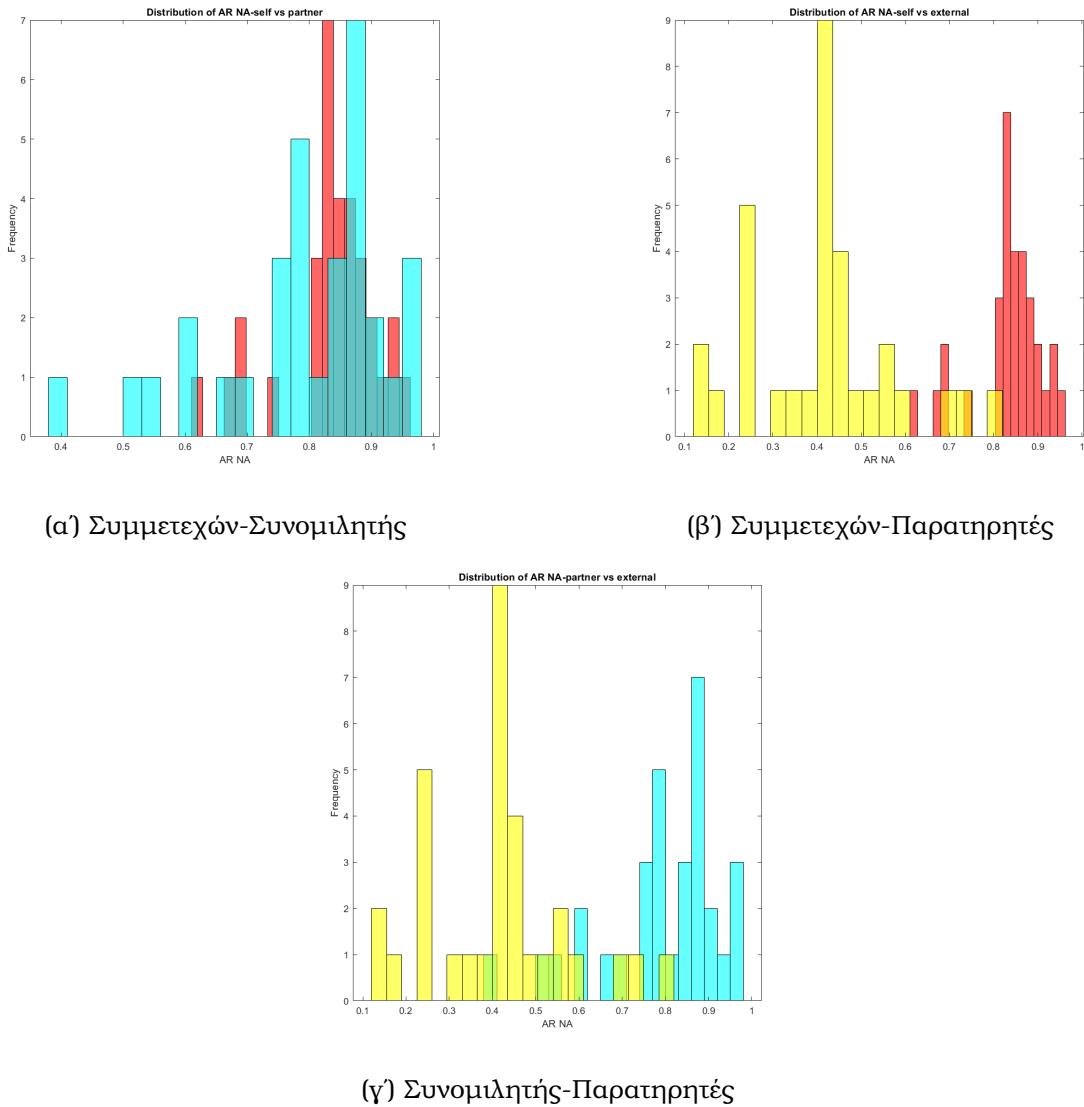
(β') Συμμετεχών-Παρατηρητές



(γ) Συνομιλητής-Παρατηρητές

Σχήμα 5.10: Κατανομές κλίσης αυτοπαλινδρόμησης θετικής συναισθηματικής κατάστασης

Κλίση αυτοπαλινδρόμησης



Σχήμα 5.11: Κατανομές κλίσης αυτοπαλινδρόμησης αρνητικής συναισθηματικής κατάστασης

5.2 Αξιολόγηση με Διασταυρωμένη Επικύρωση

Η εκπαίδευση και η αξιολόγηση των μοντέλων μηχανικής μάθησης και των αρχιτεκτονικών νευρωνικών δικτύων και συγκεκριμένα ο διαχωρισμός του συνόλου δεδομένων σε υπόσυνολα εκπαίδευσης και ελέγχου, δύναται να πραγματοποιηθεί με ένα σύνολο διαφορετικών μεθόδων. Οι πιο συχνά χρησιμοποιούμενες τεχνικές είναι η Holdout, η LOSO διασταυρωμένη επικύρωση και η διασταυρωμένη επικύρωση K-Πτυχών.

Κατά την τεχνική Holdout, το σύνολο δεδομένων χωρίζεται σε δύο αμοιβαίως αποκλειόμενα υπόσυνολα, στο σύνολο δεδομένων εκπαίδευσης και στο σύνολο δεδομένων ελέγχου. Το μοντέλο επιβλεπόμενης μάθησης εκπαιδεύεται πάνω στα δεδομένα εκπαίδευσης και στη συνέχεια αξιολογείται η απόδοση του με βάση τις προβλέψεις που πραγματοποιήθηκαν για τα δεδομένα του συνόλου ελέγχου.

Κατά την τεχνική LOSO πραγματοποιείται μια επαναληπτική διαδικασία εκπαίδευσης και ελέγχου. Πιο συγκεκριμένα, ο αριθμός των επαναλήψεων είναι ίσος με τον αριθμό των δειγμάτων στο σύνολο δεδομένων, όπου στην περίπτωσή μας θα ισούται με τον αριθμό των συμμετεχόντων. Σε κάθε επανάληψη το σύνολο ελέγχου αποτελείται από τα δεδομένα ενός ξεχωριστού συμμετέχοντος, ενώ το σύνολο των υπολοίπων απαρτίζουν τα δεδομένα εκπαίδευσης. Στο τέλος της διαδικασίας, υπολογίζονται οι μέσοι όροι των μετρικών που εξήχθησαν σε κάθε επανάληψη.

Κατά την διασταυρωμένη επικύρωση K-Πτυχών, πραγματοποιείται μια επαναληπτική διαδικασία εκπαίδευσης και ελέγχου για ένα σύνολο K επαναλήψεων. Πιο αναλυτικά, το σύνολο δεδομένων χωρίζεται σε K υποσύνολα. Σε κάθε επανάληψη τα K-1 υποσύνολα χρησιμοποιούνται ως δεδομένα εκπαίδευσης και το μοντέλο εκπαιδεύεται σε αυτά, ενώ το υποσύνολο που απομένει περιλαμβάνει τα δεδομένα ελέγχου και με βάση αυτό εξάγονται οι επιθυμητές μετρικές αξιολόγησης του μοντέλου. Στο τέλος της διαδικασίας, υπολογίζονται οι μέσοι όροι των μετρικών που εξήχθησαν σε κάθε επανάληψη και οι προκύπτοντες αριθμοί αποτελούν τον δείκτη απόδοσης του εκάστοτε μοντέλου ταξινόμησης.

Η διασταυρωμένη επικύρωση K-Πτυχών αντιμετωπίζει ικανοποιητικά το ζήτημα της αντιστάθμισης μεροληψίας-διακύμανσης, που εμφανίζεται συχνά σε μοντέλα ταξινόμησης. Πιο αναλυτικά, ένα μεγάλο ποσοστό του συνόλου δεδομένων χρησιμοποιείται για την εκπαίδευση του αλγορίθμου με αποτέλεσμα την ελαχιστοποίηση της μεροληψίας (bias) στην διαδικασία μάθησης. Παράλληλα, μειώνεται η διακύμανση (variance) καθώς όλα τα δεδομένα περνούν από την διαδικασία ελέγχου.

Στην παρούσα διπλωματική εργασία επιλέξαμε την διασταυρωμένη επικύρωση K-Πτυχών για την εκπαίδευση των μοντέλων επιβλεπόμενης μάθησης και την αξιολόγησή τους. Πιο συγκεκριμένα, για όλα τα πειράματα χρησιμοποιήσαμε διασταυρωμένη στρωματοποιημένη επικύρωση 5 Πτυχών, όπως απεικονίζεται στο Σχήμα 5.12. Ο όρος "στρωματοποιημένη" υποδηλώνει μια εκδοχή της τεχνικής κατά την οποία σε κάθε διαφορετικό σύνολο ελέγχου περιέχεται περίπου το ίδιο ποσοστό δειγμάτων ανά κλάση.

Test	Training	Training	Training	Training
Training	Test	Training	Training	Training
Training	Training	Test	Training	Training
Training	Training	Training	Test	Training
Training	Training	Training	Training	Test

Σχήμα 5.12: Παράδειγμα διασταυρωμένης επικύρωσης 5-Πτυχών

5.3 Παράμετροι πειραμάτων και υπερπαράμετροι μοντέλων

Κατηγοριοποίηση συναισθημάτων: Τα συναισθήματα των συμμετεχόντων επισημαίνονται κατά την διάρκεια της συζήτησης κάθε 5 δευτερόλεπτα από 3 διαφορετικές οπτικές γωνίες. Στην παρούσα διπλωματική εργασία, χρησιμοποιήθηκαν οι οπτικές γωνίες του συμμετέχοντα στη συζήτηση και του συνομιλητή του.

Μια κατηγορία συναισθηματικών καταστάσεων αποτελεί το δισδιάστατο συναισθηματικό μοντέλο διέγερσης-σθένους του Russell [43] και με βάση αυτό τα συναισθήματα επισημαίνονται σε μια κλίμακα Likert από το 1 έως το 5. Ανάλογα, ταξινομούνται τα συναισθήματα από το μοντέλο με βάση το επίπεδο διέγερσης και σθένους σε υψηλό (H), το οποίο θεωρούμε ως θετική κλάση (1) και χαμηλό (L), το οποίο θεωρούμε ως αρνητική κλάση (0). Πιο συγκεκριμένα, οι επισημάνσεις μετατρέπονται σε H και L σύμφωνα με τη μέση τιμή του διαστήματος της κλίμακας (2,5).

$$\text{Arousal} = \begin{cases} \text{Low (0)} & \text{if } \text{Arousal} \in [1, 2] \\ \text{High (1)} & \text{if } \text{Arousal} \in [3, 5] \end{cases}$$

$$\text{Valence} = \begin{cases} \text{Low (0)} & \text{if } \text{Valence} \in [1, 2] \\ \text{High (1)} & \text{if } \text{Valence} \in [3, 5] \end{cases}$$

Επιλογή Χρονικού Παραθύρου: Το ελάχιστο μήκος ενός σήματος που υποστηρίζει το PyTEAP για την εξαγωγή χαρακτηριστικών είναι 25 δευτερόλεπτα, ενώ τα συναισθήματα των συμμετεχόντων επισημαίνονται κατά την διάρκεια της συζήτησης κάθε 5 δευτερόλεπτα.

Στην παρούσα διπλωματική εργασία χρησιμοποιήθηκε ένα χρονικό παράθυρο διάρκειας 30 δευτερολέπτων για την μελέτη των συναισθηματικών καταστάσεων. Η συγκεκριμένη επιλογή έγινε ευρετικά, καθώς για αυτή την χρονική διαρκεία, οι μετρικές αξιολόγησης των μοντέλων ταξινόμησης παρουσίαζαν καλύτερα αποτελέσματα από ένα χρονικό παράθυρο διάρκειας 25 δευτερολέπτων. Για την σύγκριση των 2 παραθύρων, πραγματοποιήθηκε κατάλληλη επιλογή υπερπαραμέτρων για όλο το σύνολο των μοντέλων.

Παράλληλα, στην περίπτωση χρονικών παραθύρων μεγαλύτερης διάρκειας παρουσιάστηκε η ανάγκη να αυξηθεί η διάσταση της κρυφής κατάστασης του επιπέδου μακράς βραχυπρόθεσμης μνήμης, καθώς εμφανιζόταν το φαινόμενο της υποεκπαίδευσης. Η επιλογή της κατάλληλης διάστασης απαιτούσε μια μακρά διαδικασία «δοκιμής και ελέγχου» ξεχωριστά για κάθε παράθυρο, ενώ παράλληλα αυξανόταν σημαντικά ο αριθμός των παραμέτρων των μοντέλων βαθιάς μάθησης.

Παρακάτω, παρουσιάζονται οι σημαντικότεροι βέλτιστοι υπερπαράμετροι των μοντέλων για ένα χρονικό παράθυρο διάρκειας 30 δευτερολέπτων.

Υπερπαράμετροι SVM:

- C = 1, όπου αποτελεί μια παράμετρο κανονικοποίησης.
- kernel = 'rbf', όπου καθορίζει τον τύπο πυρήνα που θα χρησιμοποιηθεί στον αλγόριθμο.
- gamma = 0.001, όπου αποτελεί τον συντελεστή του πυρήνα.

Υπερπαράμετροι XGBoost:

- learning rate = 0.5, όπου καθορίζει το μέγεθος του βήματος σε κάθε επανάληψη ενίσχυσης.
- max depth = 8, όπου καθορίζει το μέγιστο βάθος κάθε δέντρου.
- num rounds = 100, όπου καθορίζει τον αριθμό των επαναλήψεων ενίσχυσης.
- objective = 'binary:logitraw', όπου καθορίζει την έξοδο του μοντέλου.

Υπερπαράμετροι αρχιτεκτονικών νευρωνικών δικτύων:

- hidden size = 100, όπου καθορίζει την διάσταση του διανύσματος στην κρυφή κατάσταση του επιπέδου μακράς βραχυπρόθεσμης μνήμης.
- max epochs = 500, όπου καθορίζει το αριθμό των μέγιστων εποχών εκπαίδευσης.
- evaluation metric = 'BCEWithLogitsLoss', όπου αποτελεί την μετρική αξιολόγησης σε κάθε εποχή.
- learning rate, όπου για τον καθορισμό της χρησιμοποιήθηκε η μέθοδος tune της εργαλειοθήκης PyTorch, η οποία βρίσκει τον βέλτιστο ρυθμό εκπαίδευσης μέσω ενός συγκεκριμένου αλγορίθμου [67].

5.4 Αποτελέσματα

Τα αποτελέσματα της εφαρμογής των ανωτέρω αρχιτεκτονικών στα δεδομένα παρουσιάζονται στη συνέχεια με μορφή πινάκων.

5.4.1 Μηχανές Διανυσμάτων Υποστήριξης

Arousal			
Perspective	Accuracy	F1-Score	AUROC
Self	0.679	0.6059	0.6257
Partner	0.6953	0.619	0.65

Πίνακας 5.1: Αποτελέσματα SVM για δυαδική ταξινόμηση διέγερσης

Valence			
Perspective	Accuracy	F1-Score	AUROC
Self	0.7565	0.70999	0.675
Partner	0.76	0.7105	0.683

Πίνακας 5.2: Αποτελέσματα SVM για δυαδική ταξινόμηση σθένους

5.4.2 XGBoost

Arousal			
Perspective	Accuracy	F1-Score	AUROC
Self	0.786	0.772	0.8672
Partner	0.812	0.7984	0.879

Πίνακας 5.3: Αποτελέσματα XGBoost για δυαδική ταξινόμηση διέγερσης

Valence			
Perspective	Accuracy	F1-Score	AUROC
Self	0.81	0.814	0.8055
Partner	0.805	0.8103	0.8164

Πίνακας 5.4: Αποτελέσματα XGBoost για δυαδική ταξινόμηση σθένους

5.4.3 Αρχιτεκτονική ενός επιπέδου μακράς βραχυπρόθεσμης μνήμης

Arousal			
Perspective	Accuracy	F1-Score	AUROC
Self	0.822	0.8235	0.882
Partner	0.8402	0.834	0.8973

Πίνακας 5.5: Αποτελέσματα αρχιτεκτονικής ενός επιπέδου μακράς βραχυπρόθεσμης μνήμης για δυαδική ταξινόμηση διέγερσης

Valence			
Perspective	Accuracy	F1-Score	AUROC
Self	0.8351	0.8202	0.8152
Partner	0.8302	0.8258	0.8243

Πίνακας 5.6: Αποτελέσματα αρχιτεκτονικής ενός επιπέδου μακράς βραχυπρόθεσμης μνήμης για δυαδική ταξινόμηση σθένους

5.4.4 Εμπλουτισμένη αρχιτεκτονική με ένα επίπεδο προσοχής

Arousal			
Perspective	Accuracy	F1-Score	AUROC
Self	0.8777	0.915	0.905
Partner	0.903	0.9323	0.9241

Πίνακας 5.7: Αποτελέσματα εμπλουτισμένης αρχιτεκτονικής με ένα επίπεδο προσοχής για δυαδική ταξινόμηση διέγερσης

Valence			
Perspective	Accuracy	F1-Score	AUROC
Self	0.9185	0.9165	0.9423
Partner	0.9154	0.9133	0.9391

Πίνακας 5.8: Αποτελέσματα εμπλουτισμένης αρχιτεκτονικής με ένα επίπεδο προσοχής για δυαδική ταξινόμηση σθένους

Κεφάλαιο 6

Συμπεράσματα και Μελλοντικές Προεκτάσεις

6.1 Σχολιασμός αποτελεσμάτων

Σε ένα πρώτο πλαίσιο, μελετήθηκε ο τρόπος με τον οποίο επάγεται ο μηχανισμός των συναισθημάτων στο σύνολο των εξωτερικών παρατηρητών, οι οποίοι δεν συμμετέχουν στη διαδικασία της συζήτησης, σε σχέση με τους συμμετέχοντες που διαμορφώνουν την αλληλουχία των συναισθημάτων, με απότερο σκόπο την επιλογή των κατάλληλων οπτικών γωνιών επισήμανσης για την εκπαίδευση και τον έλεγχο των μοντέλων ταξινόμησης. Τα αποτελέσματα υποδεικνύουν ότι το σύνολο των εξωτερικών παρατηρητών αντιλαμβάνεται ορισμένα δυναμικά χαρακτηριστικά των συναισθηματικών καταστάσεων διαφορετικά και για το λόγο αυτό στα πειράματα χρησιμοποιήθηκαν οι επισημάνσεις των συμμετεχόντων στη συζήτηση.

Σε ένα δεύτερο πλαίσιο, μελετήθηκε ένα σύνολο αλγορίθμων ταξινόμησης με σκοπό την δημιουργία ενός μοντέλου αναγνώρισης συναισθημάτων που υλοποιείται σε νατουραλιστικές συνθήκες και σενάρια και αξιοποιεί φυσιολογικά σήματα καταγεγραμμένα από φορητές μη επεμβατικές συσκευές.

Η πρώτη προσέγγιση περιλάμβανε την εξάγωγη κατάλληλων χαρακτηριστηκών από τα φυσιολογικά σήματα ενδιαφέροντος και την εκμετάλλευση παραδοσιακών αλγορίθμων μηχανικής μάθησης, με σκοπό την ταξινόμηση συναισθημάτων σε ένα χρονικό παράθυρο διάρκειας 30 δευτερολέπτων. Η αξιολόγηση των μοντέλων μηχανικής μάθησης που προέκυψαν μας οδηγεί στο συμπέρασμα ότι τα χρησιμοποιούμενα φυσιολογικά σήματα αποτελούν δείκτες για την αναγνώριση της συναισθηματικής κατάστασης ενός ατόμου.

Η δεύτερη προσέγγιση επιχειρεί να μοντελοποιήσει την εξάρτηση της συναισθηματικής κατάστασης από το σύνολο των τιμών της ακολουθίας των φυσιολογικών σημάτων ενδιαφέροντος εντός του χρονικού παραθύρου. Η υπόθεση στην οποία βασίστηκε η παραπάνω προσέγγιση είναι ότι η συναισθηματική κατάσταση ενός ατόμου μια συγκεκριμένη χρονική στιγμή σχετίζεται και επηρεάζεται από τις αντίστοιχες καταστάσεις προηγούμενων χρονικών βημάτων εντός του παραθύρου. Για το λόγο αυτό, χρησιμοποιήθηκε μια αρχιτεκτονική ανατροφοδοτούμενου νευρωνικού δικτύου που βασίζεται

σε ένα επίπεδο μακράς βραχυπρόθεσμης μνήμης. Η αξιολόγηση του μοντέλου βαθιάς μάθησης που προέκυψε παρουσιάζει καλύτερα αποτελέσματα από την πρώτη προσέγγιση, επιβεβαιώνοντας την προαναφερόμενη υπόθεση.

Η τρίτη προσέγγιση επιχειρεί να εστιάσει στο σύνολο των τιμών της ακολουθίας των φυσιολογικών σημάτων ενδιαφέροντος εντός του χρονικού παραθύρου με το μεγαλύτερο πληροφοριακό περιεχόμενο για την ταξινόμηση της συναισθηματικής κατάστασης, να υπερτονίσει δηλαδή το γεγονός ότι οι τιμές των φυσιολογικών σημάτων από το σύνολο των χρονικών βημάτων δεν παρέχουν εξίσου σημαντική πληροφορία για την ταξινόμηση της συναισθηματικής κατάστασης. Η υπόθεση στην οποία βασίστηκε η παραπάνω προσέγγιση είναι ότι η συναισθηματική κατάσταση ενός ατόμου μια συγκεκριμένη χρονική στιγμή επηρεάζεται περισσότερο από συγκεκριμένες καταστάσεις προηγούμενων χρονικών βημάτων εντός του παραθύρου σε σχέση με άλλες. Για το λόγο αυτό, η αρχιτεκτονική του ανατροφοδοτούμενου νευρωνικού δικτύου ενός επιπέδου μακράς βραχυπρόθεσμης μνήμης εμπλουτίστηκε με ένα επίπεδο προσοχής. Η αξιολόγηση του μοντέλου βαθιάς μάθησης που προέκυψε παρουσιάζει καλύτερα αποτελέσματα από τις προηγούμενες προσεγγίσεις, επιβεβαιώνοντας την υπόθεση που έγινε.

6.2 Περιορισμοί

Διαδικασία συλλογής δεδομένων: Η διαδικασία της εναλασσόμενης συζήτησης ενδέχεται να έχει προκαλέσει τους συμμετέχοντες να ρυθμίσουν ή ακόμη και να καταστέλλουν τις συναισθηματικές τους εκφράσεις, καθώς η ασυγκράτητη επίδειξη συναισθημάτων θεωρείται συχνά ανεπιθύμητη κατά τη διάρκεια μιας δυαδικής συζήτησης στο πλαίσιο που πραγματοποιήθηκε στο K-EmoCon [2]. Το γεγονός αυτό συνέβαλε σε μια ασυμφωνία μεταξύ των συναισθηματικών επισημάνσεων των συμμετεχόντων και των εξωτερικών παρατηρητών, η οποία μπορεί απαλειφθεί για πιο φυσικές δυαδικές αλληλεπιδράσεις, προσδίδοντας ακόμα μεγαλύτερη ακρίβεια στις επισημάνσεις.

Δημογραφικά στοιχεία: Τα δημογραφικά στοιχεία των συμμετεχόντων πιθανότατα να έχουν εισαγάγει μεροληψία (bias) στα δεδομένα. Όλοι οι συμμετέχοντες και οι εξωτερικοί παρατηρητές είναι νέοι (η ηλικία τους κυμαίνεται μεταξύ 19 και 36 ετών) και υψηλού μορφωτικού επιπέδου. Παράλληλα, η πλειονότητά τους είναι άτομα ασιατικής εθνότητας. Ως εκ τούτου, τα δεδομένα μας μπορεί να μην γενικεύονται καλά σε άτομα διαφορετικών εθνοτήτων ή διαφορετικών ηλικιακών ομάδων.

Πιθανή παραμόρφωση σημάτων: Ενώ τα σήματα του παλμού όγκου του αίματος (BVP) και του ηλεκτροκαρδιογραφήματος (ECG) υπέστησαν επαναδειγματοληψία στα 4 Hz από 64 Hz και 1 Hz αντίστοιχα, αυτή η επιλογή των 4 Hz ως συχνότητα επαναδειγματοληψίας σχετίζεται με την συχνότητα δειγματοληψίας των υπόλοιπων φυσιολογικών σημάτων ενδιαφέροντος και ενδέχεται να έχει εισάγει κάποια παραμόρφωση στα σήματα. Για το λόγο αυτό, θα μπορούσαμε να χρησιμοποιήσουμε μια στρατηγική συγχώνευσης δεδομένων ή μια αρχιτεκτονική που παραλείπει την ανάγκη να γίνει εκ νέου δειγματοληψία. Παράλληλα, μια σημαντικά υψηλότερη συχνότητα είναι πιθανό να συλλάβει περαιτέρω συναισθηματικές πληροφορίες από το φυσιολογικό σήμα. Ωστόσο, μια φορητή συσκευή που να επιτρέπει τη συλλογή δεδομένων σε σημαντικά υψηλότερη συχνότητα δεν είναι προς το παρόν διαθέσιμη.

6.3 Μελλοντικές προεκτάσεις

Στις συγκεκριμένη ενότητα παρατίθενται ένα σύνολο μελλοντικών προεκτάσεων της παρούσας διπλωματικής εργασίας. Η πλειονότητα των προτάσεων σχετίζεται με την περαιτέρω βελτίωση των αρχιτεκτονικών που παρουσιάστηκαν, με σκοπό την δημιουργία ένα καλύτερου μοντέλου για την αναγνώριση συναισθημάτων σε νατουραλιστικές συζητήσεις από φυσιολογικά σήματα καταγεγραμμένα από φορητές συσκευές.

6.3.1 Επεκτάσεις Μηχανισμού Προσοχής

Στην προσπάθεια βελτίωσης της διαδικασίας, θα ήταν χρήσιμο να πραγματοποιηθεί μια οπτική απεικόνιση του μηχανισμού προσοχής, γνωστή ως χάρτης προσοχής. Με αυτό τον τρόπο, θα μπορούσαμε να παρατηρήσουμε τις χρονικές στιγμές που προσφέρουν παραπάνω πληροφορία για την ταξινόμηση της συναισθηματικής κατάστασης, με τελικό στόχο την εξαγωγή ενός συμπεράσματος σχετικά με το λόγο που συμβαίνει αυτό, όπως για παράδειγμα η έντονη εκδήλωση κάποιας συγκεκριμένης συναισθηματικής κατάστασης ή κάποια εκφερόμενη άποψη από τον συνομιλητή.

Παράλληλα, ο μηχανισμός προσοχής θα μπορούσε να προστεθεί πριν την εισαγωγή των τιμών των φυσιολογικών σημάτων ενδιαφέροντος στην κρυφή κατάσταση του επιπέδου μακράς βραχυπρόθεσμης μνήμης, με σκοπό την ανάθεση βαρών σε αυτά, ανάλογα με την πληροφορία που προσφέρει η κάθε τιμή για την ταξινόμηση της συναισθηματικής κατάστασης. Με αυτό τον τρόπο, καθίσταται δυνατό να παρατηρήσουμε το σύνολο των φυσιολογικών σήματων που μπορούν να λειτουργήσουν ως δείκτες συναισθηματικών καταστάσεων συγκριτικά μεταξύ τους.

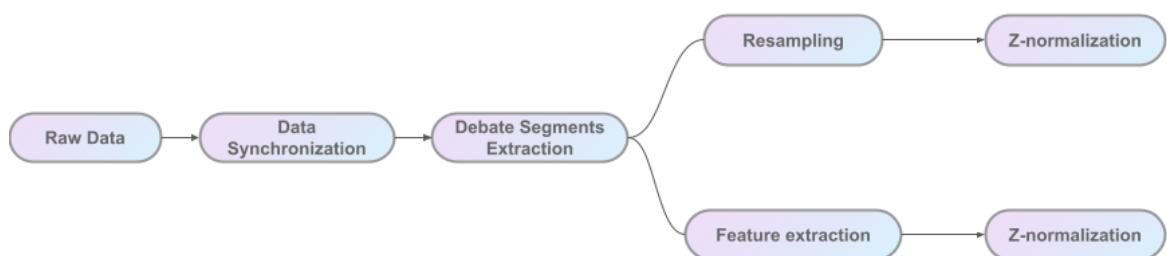
6.3.2 Δυναμικά Χαρακτηριστικά Συναισθημάτων ως επιπλέον πληροφορία

Είναι προφανές ότι τα άτομα έχουν τον δικό τους λεπτό τρόπο έκφρασης των συναισθημάτων τους. Στην προσπάθεια μοντελοποίησης της έντασης των συναισθηματικών καταστάσεων που βιώνει κάθε ατόμο ως επιπλέον πληροφορία για την περαιτέρω βελτίωση των μοντέλων ταξινόμησης, δύναται να πραγματοποιηθεί συγχώνευση των δυναμικών χαρακτηριστικών με το διάνυσμα εξοδού είτε του επίπεδου μακράς βραχυπρόθεσμης μνήμης, είτε του επιπέδου προσοχής των αρχιτεκτονικών που αναλύθηκαν παραπάνω, με σκόπιο την περαιτέρω βελτίωση των μοντέλων.

Παράρτημα Α'

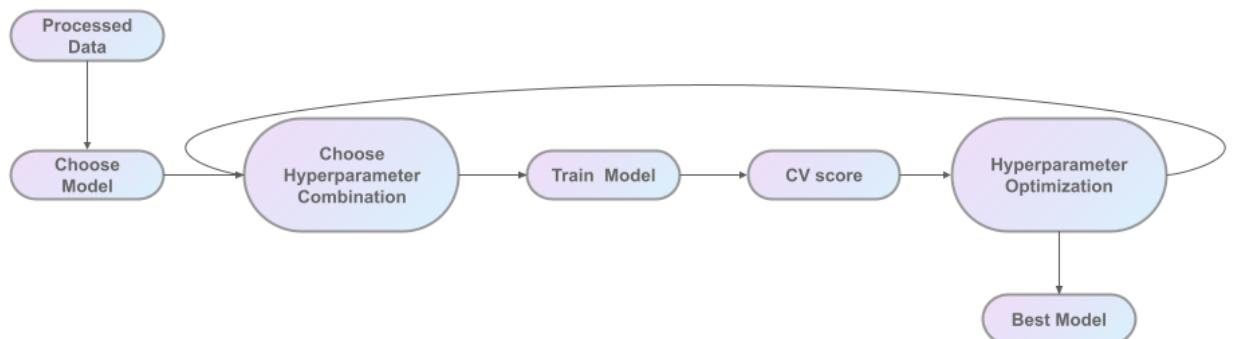
Διαγράμματα

Στάδιο προεπεξεργασίας



Σχήμα Α'.1: Διάγραμμα στάδιου προεπεξεργασίας

Στάδιο εκπαίδευσης



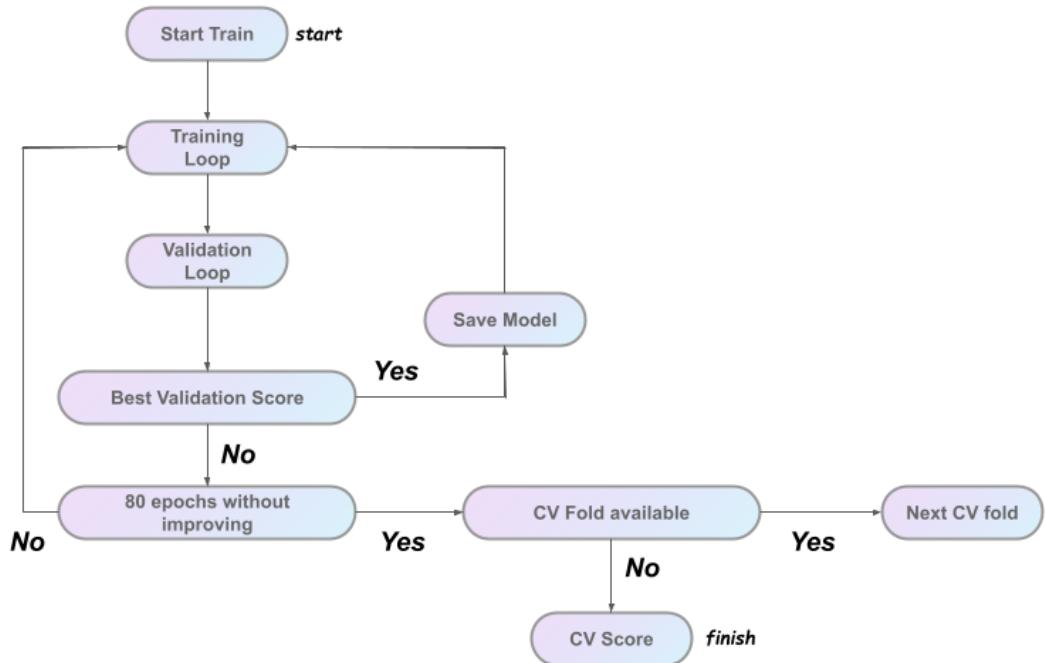
Σχήμα Α'.2: Διάγραμμα στάδιου εκπαίδευσης

Στάδιο πρόβλεψης



Σχήμα Α'.3: Διάγραμμα στάδιου πρόβλεψης

Εκπαίδευση αρχιτεκτονικών νευρωνικών δικτύων με διασταυρωμένη επικύρωση Κ-Πτυχών



Σχήμα A.4: Διάγραμμα εκπαίδευσης αρχιτεκτονικών νευρωνικών με διασταυρωμένη επικύρωση Κ-Πτυχών

Παράτημα Β'

Ακρωνύμια και συντομογραφίες

UTC Coordinated Universal Time

HMI Human Machine Interaction

SVM Support Vector Machine

MLP Multi Layer Perceptron

CNN Convolutional Neural Network

RNN Recurrent Neural Network

LSTM Long Short-Term Memory

GRU Gated Recurrent UnitRadial basis function

AUROC Area Under Receiver Operating Characteristic

PSD Power Spectral Density

HR Heart Rate

BVP Blood Volume Pressure

EDA Electrodermal Activity

T Temperature

ECG Electrocardiogram

EEG Electroencephalogram

GSR Galvanic Skin Response

EMG Electromyogram

RES Respiratory

PPG Photoplethysmogram

IBI Inter Beat Interval

HRV Heart Rate Variability

BPM Beats Per Minute

MSE Multi Scale Entropy

RBF Radial Basis Function

LOSO Leave One Subject Out

H High

L Low

PA Positive Affect

NA Negative Affect

MSSD Mean Squared Successive Difference

AR Auto Regression

Παράρτημα Γ'

Ονοματολογία

x μια βαθμωτή τιμή.

\mathbf{x} ένα διάνυσμα.

\mathbf{X} ένας πίνακας.

$x^{(t)}$ τιμή μιας ακολουθίας την χρονική στιγμή t .

x_i i -οστή θέση του διανύσματος \mathbf{x} .

$f(x; \theta)$ συνάρτηση του x παραμετροποιημένη με θ .

$\frac{df}{dx}$ παράγωγος της f ως προς το x .

$\frac{\partial f}{\partial x}$ μερική παράγωγος της f ως προς το x .

$\arg_x \min f(x)$ τιμή x που ελαχιστοποιεί τη συνάρτηση f .

\tilde{p}_{data} η εμπειρική κατανομή που ορίζεται από το σύνολο εκπαίδευσης.

p_{data} η κατανομή ενός συνόλου δεδομένων.

$\mathbf{E}_{x \sim P}[f(x)]$ προσδοκία της $f(x)$ σε σχέση με το $P(x)$.

Βιβλιογραφία

- [1] M. Thelwall and G. Paltoglou., “Seeing stars of valence and arousal in blog posts.” *Frontiers in Psychology.*, vol. 4, pp. 116–123, 2012.
- [2] C. Y. Park, N. Cha, S. Kang, A. Kim, Ahsan, H. Khandoker, L. Hadjileontiadis, A. Oh, Y. Jeong, and U. Lee., “K-emocon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations.” *Scientific Data*, vol. 7, no. 1, p. 293, 2020.
- [3] J. Mayer, D. Caruso, and P. Salovey, “Emotional intelligence meets traditional standards for an intelligence.” *Intelligence*, vol. 27, pp. 267–298, 1999.
- [4] R.-E. Mastoras, D. Iakovakis, S. Hadjidimitriou, V. Charisis, S. Kassie, T. Al-saadi, A. Khandoker, and L. J. Hadjileontiadis, “Touchscreen typing pattern analysis for remote detection of the depressive tendency.” *Scientific reports*, vol. 9, pp. 1–12, 2019.
- [5] B. Kidwai, “A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation.” *18th IEEE International Conference on Mobile Data Management (MDM)*, IEEE, pp. 371–375, 2017.
- [6] C. Breazeal, “Emotion and sociable humanoid robots.” *International journal of human-computer studies*, vol. 59, pp. 119–155, 2003.
- [7] R. Picard, *Affective Computing*. MIT Press, 2000.
- [8] A. Mollahosseini, B. Hasani, and M. Mahoor, “Affectnet: A database for facial expression, valence, and arousal computing in the wild.” *IEEE Transactions on Affective Computing*, vol. 10, pp. 18–31, 2017.
- [9] D. G. et al., “Technique for automatic emotion recognition by body gesture analysis.” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, IEEE, pp. 1–6, 2008.
- [10] C. O. Alm, D. Roth, and R. Sproat, “Emotions from text: Machine learning for text-based emotion prediction.” *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pp. 579–586, 2005.
- [11] C. B. et al., “Analysis of emotion recognition using facial expressions, speech and multimodal information.” *Proceedings of the 6th international conference on Multimodal interfaces*, pp. 205–211, 2004.

- [12] M. Frank and E. Svetieva, “Microexpressions and deception..” *Understanding facial expressions in communication*, Springer, pp. 227–242, 2015.
- [13] J. Gross and R. Levenson, “Emotion elicitation using films.” *Cognition and emotion*, vol. 9, pp. 87–108, 1995.
- [14] S. K. et al., “Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices.” *IEEE Journal of Biomedical and Health Informatics*, vol. 22, pp. 98–107, 2017.
- [15] T. Hastie, R. Tibshirani, and J. Friedman, “The elements of statistical learning: Data mining, inference, and prediction.” *Springer Science & Business Media*, vol. 2, 2009.
- [16] J. Ocumpaugh, “Machine learning: Trends, perspectives, and prospects.” *Science*, vol. 349, p. 255–260, 2015.
- [17] M. Mohammed, M. B. Khan, and E. B. M. Bashier, *Machine Learning: Algorithms and Applications*, 1st ed. CRC Press, 2017.
- [18] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Elsevier, 2011.
- [19] I. S. et al., “Cybersecurity data science: an overview from machine learning perspective,” *Journal of Big Data*, vol. 7, no. 41, p. 1–29, 2020.
- [20] V. Francois-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, “An introduction to deep reinforcement learning.” *Foundations and Trends in Machine Learning*, vol. 11, no. 3-4, 2018.
- [21] J. Juan and G. Russell, “An introduction to machine learning approaches for biomedical research.” *Frontiers in Medicine*, vol. 8, 2011.
- [22] R. S. Sutton and A. G. Barto, *Reinforcement Learning, Second Edition*, 2nd ed. MIT Press, 2018.
- [23] C. Cortes and V. Vapnik, “Support-vector networks.” *Machine Learning*, vol. 20, p. 273–297, 1995.
- [24] S. S. et al., “A survey on supervised and unsupervised learning techniques.” *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*, pp. 627–644, 2020.
- [25] L.-Y. Chuang, J.-H. Tsai, and C.-H. Yang, “Operon prediction using particle swarm optimization and reinforcement learning.” *2010 International Conference on Technologies and Applications of Artificial Intelligence*, pp. 366–372, 2010.
- [26] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. O’Reilly Media, Inc., 2019.
- [27] T. Chen and C. Guestrin, *XGBoost*. ACM, 2016.
- [28] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators.” *Neural networks*., vol. 2, pp. 359–366, 1989.
- [29] J. B. Diederik Kingma, “Adam: A method for stochastic optimization.” *3rd International Conference for Learning Representations*., 2015.

- [30] M. Zeiler, “Adadelta: An adaptive learning rate method.” *ArXiv*, 2012.
- [31] J. Duchi, E. Hazan, and Y. Singer, “Adaptive subgradient methods for online learning and stochastic optimization,” *Journal of Machine Learning Research*, 2011.
- [32] S. Ruder, “An overview of gradient descent optimization algorithms,” *ArXiv*, 2016.
- [33] S. Hochreiter and J. Schmidhuber, “Long short-term memory.” *Neural Computation.*, vol. 9, p. 1735–1780, 1997.
- [34] T. Lei, Y. Zhang, S. Wang, H. Dai, and Y. Artzi, “Training rnns as fast as cnns.” *ArXiv*, 2017.
- [35] J. Su, Z. Tan, D. Xiong, R. Ji, X. Shi, and Y. Liu, “Lattice-based recurrent neural network encoders for neural machine translation.” *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017.
- [36] T. Fawcett, “Roc graphs: Notes and practical considerations for researchers.” *Machine learning*, vol. 31, pp. 1–38, 2017.
- [37] E. Dejonckheere, M. Mestdagh, M. Houben, I. Rutten, L. Sels, P. Kuppens, and F. Tuerlinckx, “Complex affect dynamics add limited information to the prediction of psychological well-being.” *Nature Human Behaviour*, p. 478–491, 2019.
- [38] N. Clifford and B. Scott, “Emotion in human-computer interaction.” *Human-computer interaction fundamentals*, vol. 20094635, pp. 53–68, 2009.
- [39] F. N. et al., “Varieties of affect: Emotions and episodes, moods, and sentiments.” 2009.
- [40] R. Davidson, “On emotion, mood, and related affective constructs.” *The nature of emotion: Fundamental questions.*, pp. 51–55, 1994.
- [41] P. Ekman, “An argument for basic emotions.” *Cognition and emotion.*, vol. 6, pp. 169–200, 1992.
- [42] P. E. et al., “Unmasking the face: A guide to recognizing emotions from facial clues.” 1992.
- [43] Russell, “A circumplex model of affect,” *Journal of personality and social psychology.*, vol. 39, p. 1161, 1980.
- [44] R. Plutchik, “Emotions: A general psychoevolutionary theory,” *Approaches to emotion.*, pp. 197–219, 1984.
- [45] E. Cambria, A. Livingstone, and A. Hussain., “The hourglass of emotions.” *Cognitive behavioural systems.*, pp. 144–157, 2012.
- [46] S. Du, Y. Tao, and A. Martinez., “Compound facial expressions of emotion.” *Proceedings of the National Academy of Sciences.*, vol. 111, pp. 1454–1462, 2014.

- [47] G. Y. et al., “The ordinal nature of emotions.” *Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, IEEE, pp. 248–255, 2017.
- [48] C. Daniel, S. Rui, K. D. Kamble, S. Huddar, Niranjan, and M. Galen, “Universals and cultural variations in 22 emotional expressions across five cultures.” *Emotion*, vol. 18, p. 75, 2018.
- [49] A. Cowen and D. Keltner, “Self-report captures 27 distinct categories of emotion bridged by continuous gradients.” *Proceedings of the National Academy of Sciences*, vol. 114, pp. E7900–E7909, 2017.
- [50] A. C. et al., “Mapping the passions: Toward a high dimensional taxonomy of emotional experience and expression.” *Psychological Science in the Public Interest.*, vol. 20, pp. 69–90, 2019.
- [51] D. Keltner, “Toward a consensual taxonomy of emotions.” *Cognition and Emotion.*, vol. 33, pp. 14–19, 2019.
- [52] R. Jack, C. Crivelli, and T. Wheatley, “Data-driven methods to diversify knowledge of human psychology.” *Trends in cognitive sciences.*, vol. 22, pp. 1–5, 2018.
- [53] D. Keltner, J. Tracy, D. Sauter, and A. Cowen, “What basic emotion theory really says for the twenty-first century study of emotion.” *Journal of nonverbal behavior.*, vol. 43, pp. 195–201, 2019.
- [54] G. van Kleef, A. Cheshin, A. Fischer, and I. Schneider, “The social nature of emotions.” *Frontiers in Psychology.*, vol. 7, p. 896, 2016.
- [55] G. van Kleef., *Machine Learning: Algorithms and Applications*. Cambridge University Press, 2016.
- [56] J. A. Russell, “A circumplex model of affect.” *Journal of Personality and Social Psychology*, vol. 39, no. 6, p. 1161–1178, 1980.
- [57] K. e. a. Plarre, “Continuous inference of psychological stress from sensory measurements collected in the natural environment.” In *Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks*, p. 97–108, IEEE, 2011.
- [58] J. Ocumpaugh, “Baker rodrigo ocumpaugh monitoring protocol (bromp) 2.0 technical and training manual.” *New York, NY Manila, Philipp. Teach. Coll. Columbia Univ. Ateneo Lab. for Learn*, 2015.
- [59] C. Y. Park, “Pyteap v0.1.2.” *Zenodo*, 2020.
- [60] M. Soleymani, F. Villaro-Dixon, T. Pun, and G. Chanel, “Toolbox for emotional feature extraction from physiological signals (teap).” *Frontiers in ICT*, p. 1, 2017.
- [61] A. P. et al., “Pytorch: An imperative style, high-performance deep learning library.” *Advances in Neural Information Processing Systems*, pp. 8024–8035, 2019.
- [62] M. Benedek and C. Kaernbach, “A continuous measure of phasic electrodermal activity.” *Journal of neuroscience methods*, pp. 80–91, 2010.

- [63] A. Haag, S. Goronzy, P. Schaich, and J. Williams, “Emotion recognition using bio-sensors: First steps towards an automatic system.” *Affective Dialogue Systems*, 2004.
- [64] V. P. et al., “Scipy 1.0: Fundamental algorithms for scientific computing in python.” *Nature Methods*, vol. 17, pp. 261–272, 2020.
- [65] J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, “Attention-based models for speech recognition.” *Advances in Neural Information Processing Systems*, pp. 577–585, 2015.
- [66] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate.” *arXiv*, 2014.
- [67] L. Smith., “Cyclical learning rates for training neural networks.” *arXiv*, 2015.