



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΤΕΧΝΟΛΟΓΙΑΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΕΞΑΓΩΓΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΑΠΟ
ΗΛΕΚΤΡΟΕΓΚΕΦΑΛΟΓΡΑΦΗΜΑ ΜΕ ΧΡΗΣΗ ΜΟΝΤΕΛΩΝ
ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ

Καραγιαννίδη Βασιλική-Μαρία
Α.Μ. 1059396

Επιβλέπων: Μπερμπερίδης Κωνσταντίνος

ΠΑΤΡΑ-ΦΕΒΡΟΥΑΡΙΟΣ 2023

Πανεπιστήμιο Πατρών, Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών.

Καραγιαννίδη Βασιλική-Μαρία

Copyright © 2022 –Με την επιφύλαξη παντός δικαιώματος

Το σύνολο της εργασίας αποτελεί πρωτότυπο έργο, παραχθέν από την Καραγιαννίδη Βασιλική-Μαρία, και δεν παραβιάζει δικαιώματα τρίτων καθ' οποιονδήποτε τρόπο. Αν η εργασία περιέχει υλικό, το οποίο δεν έχει παραχθεί από την ίδια, αυτό είναι ευδιάκριτο και αναφέρεται ρητώς εντός του κειμένου της εργασίας ως προϊόν εργασίας τρίτου, σημειώνοντας με παρομοίως σαφή τρόπο τα στοιχεία ταυτοποίησής του, ενώ παράλληλα βεβαιώνει πως στην περίπτωση χρήσης αυτούσιων γραφικών αναπαραστάσεων, εικόνων, γραφημάτων κ.λ.π., έχει λάβει τη χωρίς περιορισμούς άδεια του κατόχου των πνευματικών δικαιωμάτων για την συμπερίληψη και επακόλουθη δημοσίευση του υλικού αυτού.

ΠΙΣΤΟΠΟΙΗΣΗ

Πιστοποιείται ότι η Διπλωματική Εργασία με τίτλο:

«ΕΞΑΓΩΓΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΑΠΟ ΗΛΕΚΤΡΟΕΓΚΕΦΑΛΟΓΡΑΦΗΜΑ ΜΕ ΧΡΗΣΗ ΜΟΝΤΕΛΩΝ ΒΑΘΙΑΣ ΜΑΘΗΣΗΣ»

της φοιτήτριας του Τμήματος Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών

ΒΑΣΙΛΙΚΗ-ΜΑΡΙΑ ΚΑΡΑΓΙΑΝΝΙΔΗ ΤΟΥ ΙΩΑΝΝΗ

Αριθμός Μητρώου: 1059396

Παρουσιάστηκε δημόσια στο τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας
Υπολογιστών στις

...../...../.....

και εξετάστηκε από την ακόλουθη εξεταστική επιτροπή:

Μπερμπερίδης Κωνσταντίνος, Καθηγητής, Τμήμα Μηχανικών Η/Υ & Πληροφορικής (επιβλέπων)

Μουστάκας Κωνσταντίνος, Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών & Τεχνολογίας Η/Υ
(μέλος επιτροπής)

Λυμπερόπουλος Δημήτριος, Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών & Τεχνολογίας
Η/Υ (μέλος επιτροπής)

Ο Επιβλέπων:

Ο Διευθυντής του Τομέα:

.....
Μπερμπερίδης Κωνσταντίνος
Καθηγητής

.....
Λογοθέτης Μιχαήλ
Αναπληρωτής Καθηγητής

Ευχαριστίες

Αρχικά θα ήθελα να εκφράσω την βαθύτατη ευγνωμοσύνη μου στον καθηγητή μου και επιβλέποντα της διπλωματικής, Κωνσταντίνο Μπερμπερίδη, καθώς και στον μεταδιδακτορικό ερευνητή Χρήστο Μαυροκεφαλίδη, για την στήριξη, την ενθάρρυνση και την πολύτιμη βοήθεια τους. Τα σχόλια και οι προτάσεις τους συνέβαλαν καθοριστικά στην διαμόρφωση της έρευνας μου και με βοήθησαν να εξελιχθώ ως επιστήμων και ερευνήτρια που ξεκινάει τα πρώτα της βήματα. Ένα τεράστιο ευχαριστώ, θα ήθελα να το αφιερώσω στην οικογένεια, στον σύντροφο, στους φίλους και συγγενείς μου. Η υπομονή, η ενθάρρυνση και η ανιδιοτελής υποστήριξη τους είναι πολύτιμες και με βοηθούν καθημερινά να κυνηγώ τους στόχους και τα όνειρά μου. Είμαι απίστευτα τυχερή που έχω τόσο όμορφους ανθρώπους στη ζωή μου και δεν μπορώ να τους ευχαριστήσω αρκετά.

ΠΕΡΙΛΗΨΗ

Το πεδίο της αναγνώρισης συναισθημάτων έχει αναδειχθεί σε σημαντικό ερευνητικό πεδίο τα τελευταία χρόνια, με σκοπό την ενίσχυση διαφόρων εφαρμογών που σχετίζονται με την αλληλεπίδραση ανθρώπων-μηχανών, την ανθρώπινη υγεία, την ψυχαγωγία και τη ρομποτική. Οι ερευνητές, μέσω των υπολογιστικών συστημάτων, ερευνούν και καταγράφουν όλα τα ανθρώπινα δεδομένα που μπορούν να παρέχουν πληροφορίες για τα συναισθήματα. Αυτά τα δεδομένα, μπορεί να προέρχονται από σήματα που έχουν λάβει από τις φυσιολογικές διεργασίες στο σώμα των ανθρώπων, δηλαδή, από τον εγκέφαλο, την καρδιά και τη θερμοκρασία του σώματος, ή από καταγραφές των εξωτερικών χαρακτηριστικών τους, όπως τις εκφράσεις του προσώπου, το body language και την ομιλία. Από τα δεδομένα αυτά, επιλέξαμε να μελετήσουμε και να αναλύσουμε δεδομένα που προέρχονται από τον ανθρώπινο εγκέφαλο και λαμβάνονται από συσκευές ηλεκτροεγκεφαλογραφημάτων.

Στην βιβλιογραφία, υπάρχει εκτενής ανάλυση των συστημάτων αναγνώρισης συναισθημάτων, όπου πολλές και διαφορετικές προσεγγίσεις αναλύονται και αξιολογούνται. Μία συγκεκριμένη προσέγγιση, περιλαμβάνει την χρήση μοντέλων μηχανικής μάθησης, που εντοπίζουν και επεξεργάζονται τα πολύπλοκα χαρακτηριστικά των δεδομένων που δίνουν πληροφορίες για την συναισθηματική κατάσταση των ανθρώπων. Από αυτές τις προσεγγίσεις, επιλέξαμε να επικεντρωθούμε, σε αυτές που χρησιμοποιούν μοντέλα βαθιάς μάθησης και συγκεκριμένα αξιολογήθηκε η απόδοση των συνελικτικών νευρωνικών δικτύων, των αναδρομικών νευρωνικών δικτύων μακροπρόθεσμης-βραχυπρόθεσμης μνήμης και των vision μετασχηματιστών. Στα πλαίσια της έρευνας που διεξάγαμε, επικεντρωθήκαμε περισσότερο στην μελέτη και ανάπτυξη συστημάτων αναγνώρισης συναισθημάτων που χρησιμοποιούν vision μετασχηματιστές, καθώς αποτελούν πρόσφατα αναπτυγμένη και πολλά υποσχόμενη κατηγορία μοντέλων, που έχουν ξεπεράσει κάποια από τα παραδοσιακά μοντέλα μηχανικής μάθησης σε διάφορα πεδία εργασιών.

Η αξιολόγηση των μοντέλων εξαγωγής συναισθημάτων, έγινε σε τρεις διαφορετικούς τύπους πειραμάτων, τον *subject dependent*, *subject independent* και έναν υβριδικό τύπο πειραμάτων. Τα αποτελέσματα που λάβαμε από αυτά τα τρία διαφορετικά σενάρια, κατέδειξαν ότι τα μοντέλα βαθιάς μάθησης αποδίδουν με εξαιρετική ακρίβεια και αποτελεσματικότητα στο πρόβλημα της κατηγοριοποίησης των συναισθημάτων των σημάτων EEG. Από τα τρία διαφορετικά μοντέλα που αναφέραμε στο προηγούμενο σκέλος, τα αποτελέσματα έδειξαν την υπεροχή των vision μετασχηματιστών έναντι των υπόλοιπων δικτύων.

Λέξεις Κλειδιά: EEG, αναγνώριση συναισθημάτων, βαθιά μάθηση, CNN, LSTM, Vision Transformer, επαύξηση δεδομένων.

EXTENSIVE ENGLISH SUMMARY

”EMOTION EXTRACTION FROM EEG SIGNALS USING DEEP LEARNING MODELS”

Student: Karagiannidi Vasiliki-Maria Supervisor: Mpermperidis Konstantinos

The field of emotion recognition has emerged as an important research field in recent years, aiming to enhance various applications related to human-machine interaction, human health, entertainment, and robotics. Researchers, through computer systems, investigate and record all human data that can provide information about emotions. This data may come from signals received from the physiological processes in the human body, i.e., from the brain, heart and body temperature, or from recordings of their external characteristics, such as facial expressions, body language and speech. In this thesis, our focus is directed towards the examination and analysis of data acquired through EEG devices that capture signals originating from the human brain.

In the literature, there is an extensive analysis of emotion recognition systems, where many different approaches are analyzed and evaluated. One specific approach involves the use of machine learning models, which identify and process the complex characteristics of data that provide information about people’s emotional state. Of these approaches, we chose to focus on those that use deep learning models and specifically evaluated the performance of convolutional neural networks, recurrent long-short-term memory neural networks and vision transformers. In the context of the research we conducted, we focused more on the study and development of emotion recognition systems using vision transformers, as they are a recently developed and promising class of models that have surpassed some of the traditional machine learning models in various fields of work.

The evaluation of emotion extraction models, was done in three different types of experiments, *subject dependent*, *subject independent* and a hybrid type. The results we obtained from these three different scenarios, demonstrated that deep learning models perform with excellent accuracy and efficiency in the problem of emotion categorization of EEG signals. From the three different models we mentioned in the previous section, the results showed the superiority of vision transformers over the rest of the networks.

Keywords: EEG, emotion recognition, deep learning, CNNs, LSTMs, Vision Transformers, data augmentation.

Περιεχόμενα

Κατάλογος Σχημάτων	10
1 Συναισθήματα & Ανθρώπινος Εγκέφαλος	12
1.1 Εισαγωγή	12
1.2 Ανατομία & Φυσιολογία του Εγκεφάλου	13
1.2.1 Ο Εγκέφαλος και Οι Νευρώνες	13
1.2.2 Μορφολογία των Νευρικών Κυττάρων	13
1.2.3 Εγκεφαλικός Φλοιός	14
1.2.4 Υποφλοιώδης Περιοχή	16
1.3 Περιοχές του εγκεφάλου που συνδέονται με τα συναισθήματα	16
1.4 Μοντέλα Κατηγοριοποίησης Συναισθημάτων	17
1.4.1 Διακριτά Μοντέλα	17
1.4.2 Μοντέλα Πολλών Διαστάσεων	18
2 Συστήματα Καταγραφής Εγκεφαλικών Σημάτων	19
2.1 Ορισμός Συστημάτων BCI	19
2.2 Ιστορικά Στοιχεία	19
2.3 Μέθοδοι Καταγραφής Εγκεφαλικής Δραστηριότητας	20
2.4 Τρόποι Λήψης Σημάτων	21
2.4.1 Μη Επεμβατικός Τρόπος Λήψης	21
2.4.2 Ημι-επεμβατικός Τρόπος Λήψης	21
2.4.3 Επεμβατικός Τρόπος Λήψης	22
2.4.4 Σύγκριση των Τρόπων Λήψης των Σημάτων	22
2.5 Το Ηλεκτροεγκεφαλογράφημα (EEG)	22
2.5.1 Πρότυπα τοποθέτησης ηλεκτροδίων	23
2.5.2 Χαρακτηριστικά	25
2.6 Συμβολές	27
3 Μηχανική Μάθηση	28
3.1 Εισαγωγή	28
3.2 Κατηγορίες Προβλημάτων Μηχανικής Μάθησης	29
3.2.1 Επιβλεπόμενη Μάθηση	29
3.2.2 Μη-Επιβλεπόμενη Μάθηση	32
3.2.3 Ενισχυτική Μάθηση	32
3.3 Νευρωνικά Δίκτυα	32
3.3.1 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks)	32
3.3.2 Βαθιά Μάθηση (Deep Learning)	36
3.3.3 Νευρωνικά Δίκτυα Βασισμένα σε Μηχανισμούς Αυτοπροσοχής	44
4 Συστήματα Αναγνώρισης Συναισθημάτων	51
4.0.1 Εισαγωγή	51
4.0.2 Ανασκόπηση της Βιβλιογραφίας	51
4.0.3 Σύγκριση των μεθόδων & Συμπεράσματα	58

5	Ανάλυση και Σχεδιασμός	60
5.1	Προ-επεξεργασία Δεδομένων	60
5.2	Μοντέλο	64
5.2.1	Αρχιτεκτονική	64
5.3	Εκπαίδευση	68
5.3.1	Συνάρτηση Απωλειών	68
5.3.2	Σύνολα Δεδομένων	69
5.3.3	Εξισορρόπηση Δεδομένων	70
5.3.4	Επαύξηση Δεδομένων	70
5.3.5	Υπερπαράμετροι Μοντέλων	72
5.3.6	Διαδικασία Εκπαίδευσης	73
6	Πειραματική Διαδικασία	74
6.1	Αποτελέσματα	74
6.1.1	Συγκρίσεις	77
7	Συμπεράσματα και Μελλοντικές Κατευθύνσεις	78
	Βιβλιογραφία	79

Κατάλογος Σχημάτων

1.1	Στην εικόνα φαίνονται τα βασικά δομικά στοιχεία ενός νευρικού κυττάρου, συμπεριλαμβανομένων των δενδριτών, του κυτταρικού σώματος, του άξονα και της μυελίνης Πηγή: [1]	14
1.2	Οι λοβοί του εγκεφαλικού φλοιού. Πηγή:[2]	15
1.3	Το μοντέλο <i>Circumplex</i> [60] , αναπαριστά τον διδιάστατο χώρο κατηγοριοποίησης συναισθημάτων στον άνθρωπο. Το μοντέλο προτάθηκε το 1980 από τον <i>James Russell</i> . Πηγή: [75]	18
2.1	Ένα από τα πρώτα EEG δείγματα που δημοσίευσε ο Hans Begrer το 1969 [17]. Στην εικόνα το πάνω σήμα είναι το δείγμα EEG που καταγράφηκε από ένα νεαρό αγόρι, ενώ το κάτω είναι ημιτονοειδές σήμα αναφοράς με συχνότητα 10 Hz. Πηγή: [17]	20
2.2	α) Συσκευή Emotiv EPOC με 14 ηλεκτρόδια. β) Συσκευή BrainAmp με 128 ηλεκτρόδια. Πηγή: [9]	23
2.3	Τοποθέτηση των ηλεκτροδίων με βάση το διεθνές σύστημα 10-20. Πηγή: [64] . . .	24
2.4	α) Τοποθέτηση των ηλεκτροδίων με βάση το διεθνές σύστημα 10-20. β) Τοποθέτηση των ηλεκτροδίων με βάση το διεθνές σύστημα 10-10. Τα πράσινα ηλεκτρόδια αντιστοιχούν στον μετωπιαίο λοβό, τα μπλε στον κροταφικό, τα ροζ στον βρεγματικό και τα κίτρινα στον ινιακό λοβό. Πηγή: [13]	24
2.5	Ενδεικτική τοποθέτηση των ηλεκτροδίων με βάση το διεθνές σύστημα 10-5. Πηγή: [87]	24
2.6	Στην εικόνα φαίνεται το πλάτος ενός EEG καναλιού και οι αποσυνθέσεις του στις 4 βασικές συχνοτικές ζώνες.	26
3.1	Ενδεικτικός confusion matrix που αντιστοιχεί σε πρόβλημα binary classification. .	30
3.2	Παράδειγμα προσαρμογής μοντέλου παλινδρόμησης. Τα σημεία (0, 0), (0.5, 0.25), (2, 4) είναι τα παραδείγματα εκπαίδευσης. Όταν λάβει το σημείο $x_0 = 1.414$ η έξοδος του μοντέλου προσεγγίζει την ιδανική έξοδο $y_0 = 2$	31
3.3	α) Στην αριστερή εικόνα φαίνεται ένα παράδειγμα τεχνητού νευρωνικού δικτύου τριών επιπέδων. Κάθε κόμβος του δικτύου συμβολίζει έναν «τεχνητό νευρώνα» και κάθε βέλος αντιπροσωπεύει μία σύνδεση από την έξοδο ενός νευρώνα στην είσοδο ενός άλλου. β) Στην δεξιά πλευρά απεικονίζεται το μαθηματικό μοντέλο ενός τεχνητού νευρώνα. Τα x_1, \dots, x_n είναι οι είσοδοι και τα w_1, \dots, w_n είναι τα βάρη του νευρώνα.	33
3.4	Διάγραμμα τεχνητού νευρώνα.	35
3.5	Αναπαράσταση 4 σημείων στο διδιάστατο επίπεδο.	37
3.6	Στην εικόνα βλέπουμε αναλυτικά την πράξη της ετερο-συσχέτισης μίας μήτρας μεγέθους (3 × 3) με ένα (2 × 2) φίλτρο.	38
3.7	Τα βασικά στάδια ενός <i>Convolutional Neural Network</i>	39
3.8	Αναδρομικό νευρωνικό δίκτυο <i>many to many</i> συνδέσεων. Το μήκος της ακολουθίας εισόδου Tx είναι ίδιο με το μήκος της ακολουθίας εξόδου Ty	40
3.9	Με κόκκινο φαίνεται η σειρά με την οποία θα υπολογιστούν οι μερικές παράγωγοι στην οπίσθια διάδοση.	41
3.10	Διάγραμμα ενός memory block	43
3.11	Η αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή του transformer. Πηγή:[78] . . .	45

3.12	Στην αριστερή εικόνα φαίνεται η ροή του <i>multi-head attention</i> μηχανισμού και στην δεξιά τα αναλυτικά στάδια του <i>scaled dot product</i> . Ως Q, K, V συμβολίζονται οι πίνακες που περιέχουν τα query, key, value αντίστοιχα. Πηγή:[78]	47
3.13	Επισκόπηση αρχιτεκτονικής του <i>Vision Transformer</i> . Στις εξόδους των MLP δικτύων εφαρμόζεται η συνάρτηση ενεργοποίησης <i>gaussian linear error unit</i> (GELU). Πηγή:[23]	49
4.1	Αναπαράσταση της τοποθέτησης των βιοαισθητήρων στο ανθρώπινο σώμα, για την μέτρηση διάφορων φυσιολογικών σημάτων. Πηγή:[67]	53
4.2	Σχηματική αναπαράσταση ενός ιδανικού <i>sinus</i> παλμού που διαδίδεται από την ανθρώπινη καρδιά. Στην εικόνα φαίνεται το QRS σύμπλεγμα που αποτελείται από τα τρία τμήματα Q, R, S. Πηγή:[3]	55
4.3	Στην εικόνα φαίνονται οι μύες του κεφαλιού και του αυχένα που επιλέγονται για την EMG καταγραφή, λόγω της συσχέτισης τους με συναισθήματα που αποτυπώνονται στο χαμόγελο, το συνοφρύωμα και το σκίρτημα. Πηγή:[22]	55
5.1	Οι συναρτήσεις μεταφοράς τεσσάρων μη ιδανικών φίλτρων. Στην εικόνα επισημαίνονται και οι ζώνες διέλευσης, μετάβασης και αποκοπής των φίλτρων.	61
5.2	Η απόκριση πλάτους του φίλτρου εγκοπής. Το φίλτρο υποβαθμίζει την συχνότητα των 60 Hz.	62
5.3	Το διάγραμμα ροής του <i>blind source separation</i> αλγόριθμου. Με s συμβολίζονται οι επιμέρους διαχωρισμένες πηγές, με x τα «μικτά» σήματα και με y οι εκτιμήσεις των πηγών. Πηγή:[18]	64
5.4	Επισκόπηση του deep CNN μοντέλου. Στις ενδιάμεσες εξόδους των στρωμάτων εφαρμόστηκε η συνάρτηση ενεργοποίησης ReLU.	65
5.5	Δομική αναπαράσταση της LSTM αρχιτεκτονικής. Το δίκτυο δέχεται τα τμηματικά <i>flattened</i> δεδομένα EEG.	66
5.6	Η προτεινόμενη αρχιτεκτονική των Arjun <i>et al.</i> . Το δίκτυο δέχεται στην είσοδο τα <i>flattened</i> EEG segments.	68
5.7	Γραφική αναπαράσταση της συνάρτησης Softmax: $\mathbb{R} \Rightarrow [0, 1]$. Πηγή:[4]	69
5.8	Παραλλαγή της τεχνικής CutMix που δημιουργεί νέα δεδομένα από το υπάρχον σύνολο εκπαίδευσης. Τα πρώτα δύο δείγματα μπορεί να προέρχονται από το ίδιο ή από διαφορετικά άτομα. Το νέο δείγμα που προκύπτει αποτελεί αντίγραφο του δεύτερου με την διαφορά ότι ένα επιλεγμένο υπο-πλαίσιο ανήκει στο πρώτο δείγμα EEG.	71

Κεφάλαιο 1

Συναισθήματα & Ανθρώπινος Εγκέφαλος

1.1 Εισαγωγή

Τα συναισθήματα αποτελούν μία έννοια που αν και σαν λέξη χρησιμοποιούνται πολύ συχνά, δεν μπορεί κάποιος να τα ορίσει επακριβώς, καθώς διαφέρουν από άνθρωπο σε άνθρωπο, από εποχή σε εποχή και από πολιτισμό σε πολιτισμό. Αποτελούν αναπόσπαστο κομμάτι της ύπαρξης των ανθρώπων και των ζώων και σε συνδυασμό με την συνείδηση είναι ένα μέτρο διάκρισης μεταξύ των έμβιων όντων και των μηχανών.

Το ενδιαφέρον στην διατύπωση και τον προσδιορισμό των συναισθημάτων υπάρχει από την αρχαιότητα. Ο Πλάτωνας στα έργα του κάνει λόγο για το «θυμοειδές», εννοώντας το συναίσθημα, το οποίο όταν ισορροπεί με την λογική και τις επιθυμίες συμβάλλει στην ψυχική υγεία των ανθρώπων. Σύμφωνα με τα έργα του φιλόσοφου Αριστοτέλη τα συναισθήματα συνδέονται άρρηκτα με τον συλλογισμό και την ύπαρξη. Η λέξη «συναίσθημα» ξεκίνησε να υπάρχει στις αρχές του 17ου αιώνα, μέσω του Γαλλικού *"époouvoir"* που σήμαινε μία αναταραχή, ενώ η έννοια της εδραιώθηκε δύο αιώνες μετά από τους Σκοτσέζους φιλόσοφους-γιατρούς, Thomas Brown και Charles Bel. Στις μέρες μας, ένας απλοποιημένος ορισμός του συναισθήματος έχει δοθεί από την Αμερικάνικη Ένωση Ψυχολόγων (APA)[77] ως «ένα σύνθετο πρότυπο αντίδρασης, που περιλαμβάνει βιωματικά, συμπεριφορικά και φυσιολογικά στοιχεία, με τα οποία ένα άτομο επιχειρεί να αντιμετωπίσει ένα προσωπικά σημαντικό θέμα ή γεγονός».

Τόσο βιολογικοί όσο και κοινωνικοί παράγοντες διαμορφώνουν και μεταβάλλουν τα συναισθήματα των ανθρώπων μεγάλωνοντας, καθώς και τον τρόπο που εκείνοι τα εκδηλώνουν. Παρόλο που ορισμένα βιώματα προκαλούν παρόμοια συναισθήματα για κάποιους ανθρώπους, ο τρόπος που αντιλαμβάνονται και εκφράζουν αυτά που νιώθουν διαφέρει. Επίσης, το αν και κατά πόσο εκφράζουν τα συναισθήματα τους την κάθε χρονική στιγμή, είναι στην δική τους ευχέρεια και γι' αυτό είναι δύσκολο να προσδιοριστεί εξολοκλήρου αντικειμενικά η συναισθηματική τους κατάσταση μέσα από την έκφραση. Για παράδειγμα άνθρωποι που είναι περισσότερο εσωστρεφείς, συνήθως εκδηλώνουν πιο δύσκολα τα συναισθήματα τους από τους εξωστρεφείς ανθρώπους.

Η μελέτη και έρευνα των συναισθημάτων έχει προσελκύσει πολλά επιστημονικά πεδία τα τελευταία χρόνια. Μερικά από αυτά τα πεδία είναι η επιστήμη της ψυχολογίας, της φιλοσοφίας, της ιατρικής και των υπολογιστών. Συγκεκριμένα με την ανάπτυξη του πεδίου της πληροφορικής και της τεχνολογίας γενικότερα, έχει δημιουργηθεί ένας καινούριος επιστημονικός κλάδος που ονομάζεται Συναισθηματική Πληροφορική (Affective Computing) . Για τον συγκεκριμένο κλάδο συνεργάζονται ο τομέας της πληροφορικής, της ψυχολογίας και άλλων γνωστικών επιστημών. Κύρια ιδέα του τομέα είναι η αναγνώριση, ερμηνεία και επεξεργασία των συναισθημάτων από υπολογιστικά συστήματα και συσκευές [73]. Η συγκρότηση του τομέα ξεκίνησε από την ερευνητική ομάδα της Rosalind W. Picard , με την δημοσίευση της *"Affective Computing"* [56]. Κίνητρο της έρευνας της ήταν οι μηχανές να μπορούν να εντοπίζουν, να διαχωρίζουν και να επεξεργάζονται

τα συναισθήματα των ανθρώπων και μέσω αυτών των πληροφοριών να εκτελούν συγκεκριμένες λειτουργίες.

Στην επιστημονική έρευνα των ανθρώπινων συναισθημάτων, έγινε σύντομα εμφανές ότι η μελέτη και ανάλυση της εγκεφαλικής δραστηριότητας, θα μπορούσε να αποτελέσει ένα ισχυρό εργαλείο για την αναγνώριση συναισθημάτων. Η ιδέα της ανάλυσης των εγκεφαλικών μοτίβων και η συσχέτιση τους με τα συναισθήματα αποτελεί μέχρι και σήμερα ένα διαδεδομένο και ανοιχτό θέμα στον τομέα της έρευνας της συναισθηματικής πληροφορικής και της διεπαφής εγκεφάλου-υπολογιστή (BCI). Σημαντική προϋπόθεση για κάθε ερευνητή που προσεγγίζει το πρόβλημα της αναγνώρισης συναισθημάτων μέσω των εγκεφαλικών σημάτων, είναι η πλήρης κατανόηση των βασικών δομικών και λειτουργικών περιοχών του εγκεφάλου. Η κατανόηση του εγκεφάλου και των λειτουργιών του, μπορεί να βοηθήσει σημαντικά στον τρόπο με τον οποίο αλληλεπιδρούμε με την τεχνολογία και να προσαρμόσει νέα τεχνολογικά ευρήματα στα ανθρώπινα συναισθήματα και στην γνώση.

1.2 Ανατομία & Φυσιολογία του Εγκεφάλου

Η παρούσα ενότητα, αφιερώνεται στην συνοπτική επισκόπηση των μακροσκοπικών και μικροσκοπικών περιοχών του ανθρώπινου εγκεφάλου. Αρχικά, εμβαθύνει στις λεπτομέρειες των δομικών στοιχείων των νευρικών κυττάρων και ακολούθως, επεκτείνεται στην περιγραφή του εγκεφαλικού φλοιού και της υποφλοιώδους περιοχής σε μεγαλύτερο βάθος.

1.2.1 Ο Εγκέφαλος και Οι Νευρώνες

Ο εγκέφαλος είναι το πιο περίπλοκο όργανο του σώματος, αποτελούμενο από τουλάχιστον 86 δισεκατομμύρια νευρώνες και τρισεκατομμύρια κύτταρα που ονομάζονται γάγγλια. Μαζί με τον νωτιαίο μυελό αποτελούν το *Κεντρικό Νευρικό Σύστημα-ΚΝΣ*. Βρίσκεται στο πρόσθιο άκρο του νωτιαίου μυελού και αποτελεί το μεγαλύτερο και σημαντικότερο τμήμα του ΚΝΣ [33]. Συγκεκριμένα, αποτελεί την κύρια μονάδα επεξεργασίας του νευρικού συστήματος, αφού λαμβάνει και επεξεργάζεται πληροφορίες από τα αισθητήρια όργανα όλου του σώματος και διεγείρει τις κατάλληλες σωματικές αντιδράσεις. Κάποιες από τις λειτουργίες που συντονίζει, είναι η κίνηση, η αίσθηση και τα αντανακλαστικά, αλλά ρυθμίζει και πιο ζωτικές λειτουργίες του σώματος όπως τον καρδιακό ρυθμό, την αρτηριακή πίεση και την ομοιόσταση.

Οι νευρώνες (ή αλλιώς νευρικά κύτταρα) διαμορφώνουν το *Περιφερειακό Νευρικό Σύστημα-ΠΝΣ*. Κάθε νευρώνας μπορεί να διεγείρεται από μυϊκές δράσεις, όπως το κούνημα ενός μέρους του σώματος, αλλά και μη μυϊκές, όπως για παράδειγμα από οπτικά και ηχητικά μέσα. Οι νευρώνες συνεργάζονται μεταξύ τους και μοιράζονται πληροφορίες αναγκαίες για τον έλεγχο και την λειτουργία του ανθρώπινου σώματος. Τα νευρικά κύτταρα λαμβάνουν τις πληροφορίες αυτές από τα αισθητήρια όργανα, τις μεταφέρουν στον εγκέφαλο για περαιτέρω επεξεργασία και από εκεί, δημιουργούνται οι κατάλληλες αποκρίσεις που περνούν στα υπόλοιπα όργανα του σώματος, πάλι μέσω των νευρώνων.

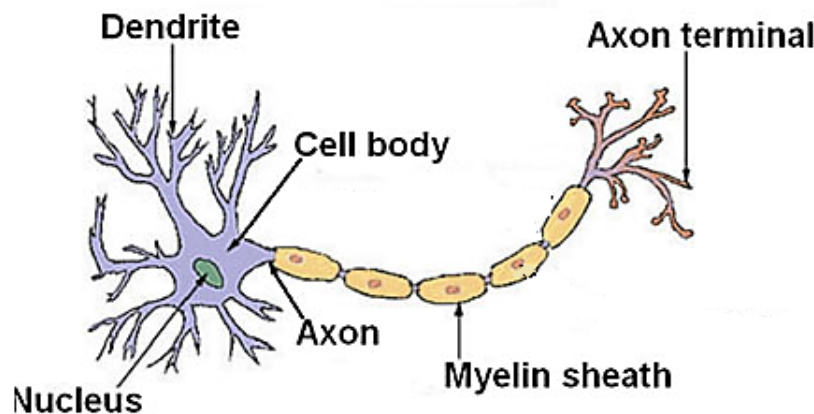
1.2.2 Μορφολογία των Νευρικών Κυττάρων

Τα νευρικά κύτταρα του Περιφερειακού Νευρικού Συστήματος, αποτελούνται από κάποιες επιμέρους δομές, τις νευρικές ίνες ή αλλιώς νευροάξονες (*Axon*), τους δενδρίτες (*Dendrite*) και το

κυτταρικό σώμα (*Cell body*) . Η δομή του νευρώνα φαίνεται στο Σχήμα 1.1. Οι λειτουργίες που επιτελούν οι συγκεκριμένες δομές είναι οι εξής:

- **Νευροάξονες:** Μεταδίδουν τα ηλεκτρικά σήματα σε νευρικά κύτταρα που μπορεί να βρίσκονται έως και 1 μέτρο μακριά από τον νευρώνα τους.
- **Δενδρίτες:** Αποτελούν παρακλάδια του κυτταρικού σώματος και έχουν τον ρόλο του υποδοχέα των ηλεκτρικών ώσεων από άλλα κύτταρα. Επίσης, έχουν βασικό ρόλο στην *συναπτική πλαστικότητα*, δηλαδή την αλλαγή των νευρικών συνδέσεων όταν οι διεργασίες το απαιτούν.
- **Κυτταρικό Σώμα:** Ο χώρος του κυττάρου που περιέχει τον πυρήνα και άλλα οργανίδια, όπως τα μιτοχόνδρια και το δίκτυο *Golgi*. Ο ρόλος του αφορά την επεξεργασία των σημάτων που λαμβάνει από τους δενδρίτες.

Στον ανθρώπινο εγκέφαλο, υπάρχουν περίπου 10.000 συνδέσεις μεταξύ των νευρικών κυτάρων, γνωστών ως *συνάψεις*. Τα ηλεκτρικά σήματα που περνάνε από αυτές τις συνάψεις, μετατρέπονται σε χημικά σήματα, όταν φτάνουν στο τελικό στάδιο του νευροάξονα. Στο στάδιο αυτό, τα χημικά σήματα που προκύπτουν ονομάζονται *νευροδιαβιβαστές* και όταν φτάσουν στον δενδρίτη ή στο σώμα του συνδεδεμένου κυττάρου μετατρέπονται ξανά σε ηλεκτρικά σήματα [61]. Οι άξονες που είναι υπεύθυνοι για την αποστολή πληροφοριών σε μεγάλες αποστάσεις, συνήθως έχουν μεγαλύτερο μήκος και επικαλύπτονται από μία ουσία, την *μυελίνη*. Στην ουσία αυτή, οφείλεται το λευκό χρώμα της λευκής ουσίας του εγκεφάλου. Ο ρόλος της μυελίνης, είναι να προστατεύει τις συνάψεις κατά την μετάδοση των ηλεκτρικών ώσεων σε μεγάλες αποστάσεις.



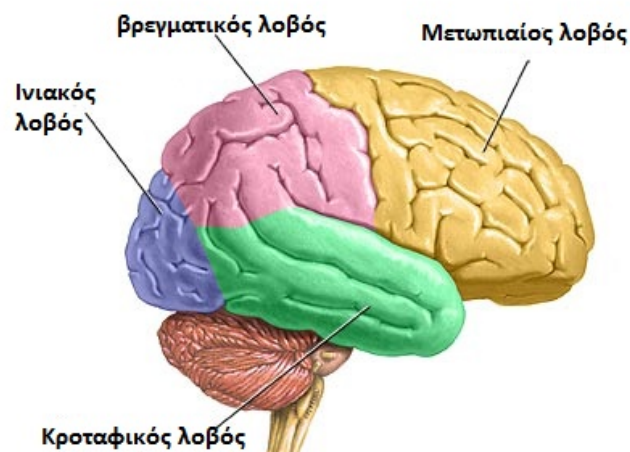
Σχήμα 1.1: Στην εικόνα φαίνονται τα βασικά δομικά στοιχεία ενός νευρικού κυττάρου, συμπεριλαμβανομένων των δενδριτών, του κυτταρικού σώματος, του άξονα και της μυελίνης Πηγή: [1]

1.2.3 Εγκεφαλικός Φλοιός

Ο εγκεφαλικός φλοιός (*Cerebral Cortex*) είναι το εξωτερικό στρώμα του εγκεφάλου, το οποίο καταλαμβάνει το μισό της συνολικής μάζας του. Η τσαλαχωμένη επιφάνεια του αποτελείται από ανυψωμένους κόλπους που ονομάζονται *έλικες* και βαθιά ανοίγματα που ονομάζονται *αύλακες*. Η δομή αυτή επιτρέπει την επεξεργασία περισσότερων πληροφοριών από περισσότερα νευρικά κύτταρα. Συγκεκριμένα, ο ανθρώπινος εγκεφαλικός φλοιός αποτελείται από 6 διαφορετικά στρώματα

νευρώνων με πλήθος 14-16 δισεκατομμυρίων νευρικών κυττάρων, που αποτελούν τον νεοφλοιό (*neocortex*). Ο νεοφλοιός περιλαμβάνει το 90% του φλοιού, ενώ το υπόλοιπο 10% ανήκει στον δευτερογενή φλοιό (*allocortex*) με 5 στρώματα νευρώνων.

Ο εγκεφαλικός φλοιός χωρίζεται περαιτέρω στο αριστερό και στο δεξί ημισφαίριο δια μέσω της διαμήχους σχισμής (*longitudinal fissure*) και τα ημισφαίρια ενώνονται κάτω από τον φλοιό μέσω του *corpus callosum*. Το αριστερό ημισφαίριο του εγκεφάλου, είναι κυρίως υπεύθυνο για την αντίληψη, κατανόηση και παραγωγή των γλωσσικών πληροφοριών, ενώ το δεξί ημισφαίριο εμπλέκεται περισσότερο στην αντίληψη και την ερμηνεία των οπτικών και χωρικών πληροφοριών. Για παράδειγμα, το αριστερό ημισφαίριο ελέγχει την ομιλία, την γραφή, τους αριθμητικούς υπολογισμούς, ενώ το δεξί παίζει σημαντικό ρόλο στην μνήμη, στην ικανότητα της προσοχής, στην παραγωγή των συναισθημάτων κ.α. Κάθε ημισφαίριο του εγκεφάλου έχει τέσσερις λοβούς, τον μετωπιαίο, βρεγματικό, κροταφικό και ινιακό. Ο καθένας από αυτούς τους λοβούς συνεργάζεται με τα υπόλοιπα τμήματα του εγκεφάλου επιτελώντας κάποιες ζωτικές λειτουργίες.



Σχήμα 1.2: Οι λοβοί του εγκεφαλικού φλοιού. Πηγή:[2]

Ο **μετωπιαίος λοβός** είναι το εμπρόσθιο τμήμα του εγκεφάλου που βρίσκεται πίσω από το μέτωπο του κεφαλιού. Βασικές λειτουργίες όπως η ομιλία, η λήψη αποφάσεων, η επίλυση προβλημάτων και άλλες, ρυθμίζονται από αυτόν. Οι κύριες υποπεριοχές του λοβού είναι ο **κινητικός φλοιός**, ο **προμετωπιαίος φλοιός** και η περιοχή **Broca**. Ο κινητικός φλοιός εντοπίζεται στην οπίσθια περιοχή του μετωπιαίου λοβού και είναι υπεύθυνος για τον έλεγχο των κινήσεων του σώματος. Ο προμετωπιαίος φλοιός βρίσκεται στο μπροστινό μέρος του λοβού και εμπλέκεται σε ανώτερες γνωστικές λειτουργίες, όπως την συγκέντρωση, την ανάκληση της προσωρινής μνήμης αλλά και τον έλεγχο άλλων περιοχών του εγκεφάλου. Η περιοχή Broca, καταλαμβάνει μία συγκεκριμένη υποπεριοχή του προμετωπιαίου φλοιού και είναι υπεύθυνη για την παραγωγή, τον έλεγχο και την κατανόηση της ομιλίας και της γλώσσας.

Ο **βρεγματικός λοβός** βρίσκεται στο μέσον του εγκεφάλου και συγκεκριμένα μεταξύ του μετωπιαίου και του ινιακού λοβού. Ελέγχει πολλές βασικές λειτουργίες, συμπεριλαμβανομένης της επεξεργασίας των αισθητηριακών πληροφοριών, όπως την αφή, την πίεση, την θερμοκρασία και τον πόνο. Επιπλέον, διαδραματίζει κρίσιμο ρόλο στην αναγνώριση αντικειμένων και ανθρώπων και είναι υπεύθυνος για την κατανόηση και των χωρικών σχέσεων, δηλαδή της αντίληψης του βάθους, της θέσης και του προσανατολισμού στον χώρο. Βλάβη στον βρεγματικό λοβό, μπορεί να οδηγήσει

τους ανθρώπους σε εσφαλμένη αντίληψη του περιβάλλοντος και έτσι να στερηθούν η ικανότητα της αποτελεσματικής πλοήγησης σε αυτό. Ο βρεγματικός λοβός, εμπλέκεται ακόμη και σε άλλες γνωστικές λειτουργίες υψηλού επιπέδου, όπως την γλωσσική επεξεργασία και την συγκέντρωση.

Ο λοβός που εντοπίζεται στο πίσω άκρο του εγκεφάλου, ονομάζεται **ινιακός**. Είναι κυρίως υπεύθυνος για την ερμηνεία και επεξεργασία των οπτικών πληροφοριών. Συγκεκριμένα, συσχετίζεται με οπτικές λειτουργίες, όπως την ανάλυση και αναγνώριση των χρωμάτων, της κίνησης και του προσανατολισμού αντικειμένων στο οπτικό πεδίο. Ο ινιακός λοβός, όπως και ο βρεγματικός, εμπλέκεται και σε λειτουργίες που αφορούν την κατανόηση των χωρικών σχέσεων και την χαρτογράφηση του οπτικού περιβάλλοντος, καθώς σχετίζεται άμεσα με τις λειτουργίες της όρασης.

Ο **κροταφικός λοβός** αποτελεί την περιοχή κάτω από τον μετωπιαίο λοβό και βρεγματικό λοβό. Κύρια περιοχή του λοβού είναι η περιοχή *Wernick*. Η περιοχή αυτή βρίσκεται κοντά στον ακουστικό φλοιό, παίζει καθοριστικό ρόλο στην κατανόηση της γλώσσας και είναι απαραίτητη για την επεξεργασία και την ερμηνεία των ήχων και της ομιλίας. Εκτός από τη γλώσσα, ο κροταφικός λοβός εμπλέκεται επίσης στην ακουστική επεξεργασία, τον σχηματισμό και ανάκτηση της μακροπρόθεσμης μνήμης, την οπτική αναγνώριση αντικειμένων και τη χωρική πλοήγηση. Επιπλέον, είναι γνωστό ότι εμπλέκεται στη μετατροπή των αισθητηριακών εισροών σε αντιλήψεις με νόημα και έχει συνδεθεί με τη μάθηση και τη συναισθηματική επεξεργασία.

1.2.4 Υποφλοιώδης Περιοχή

Η περιοχή αυτή όπως φαίνεται και από το όνομά της, εντοπίζεται κάτω από τον εγκεφαλικό φλοιό, στο εσωτερικό μέρος του εγκεφάλου. Η υποφλοιώδης περιοχή αποτελείται από ομάδες νευρικών μετασχηματισμών, κάποιες από τις οποίες είναι η *παρεγκεφαλίδα*, η *υπόφυση*, τα *γάγγλια* και το *λιμβικό σύστημα*. Οι δραστηριότητες στις οποίες εμπλέκονται οι περιοχές αυτές είναι η μνήμη, τα συναισθήματα και η παραγωγή ορμονών. Είναι βασικοί κόμβοι πληροφοριών του νευρικού συστήματος, καθώς αναπροσαρμόζουν και μεταδίδουν πληροφορίες σε πολλές περιοχές του εγκεφάλου.

1.3 Περιοχές του εγκεφάλου που συνδέονται με τα συναισθήματα

Στις προηγούμενες υπό-ενότητες έγινε μία μικρή ανάλυση της μορφολογίας του εγκεφάλου, ο οποίος αποτελεί κομβική δομή, στην οποία δημιουργούνται και αναλύονται τα ανθρώπινα συναισθήματα. Η επεξεργασία των συναισθημάτων, είναι μία πολύπλοκη και συνεργατική διαδικασία που περιλαμβάνει τις περιοχές του αριστερού και δεξιού ημισφαιρίου. Τα συναισθήματα υφίστανται επεξεργασία στο αριστερό ημισφαίριο και ιδιαίτερα σε περιοχές όπως η *αμυγδαλή*, ο *νησιωτικός φλοιός* και η *φαιά ουσία περίξ του υδραγωγού (PAG)*. Ωστόσο, η σήμανση και η κατηγοριοποίηση των συναισθημάτων συμβαίνουν κυρίως στο αριστερό ημισφαίριο, το οποίο σε συνεργασία με το δεξιό ημισφαίριο συνδυάζουν την γλωσσική επεξεργασία και την συναισθηματική έκφραση των συναισθημάτων.

Η *αμυγδαλή* είναι μία δομή σε σχήμα αμυγδάλου που βρίσκεται κοντά στο εγκεφαλικό στέλεχος. Η λειτουργία της σχετίζεται με την συναισθηματική συμπεριφορά και τα συνειδητά κίνητρα. Σε αυτήν, αναπτύσσονται κυρίως αρνητικά συναισθήματα όπως ο φόβος, καθώς και σωματικές αι-

σθήσεις που προκύπτουν από τις συναισθηματικές εμπειρίες. Επικοινωνεί με τα υπόλοιπα μέρη του εγκεφάλου για να διεγείρει αντιδράσεις σε συγκεκριμένα ερεθίσματα. Βλάβη στην αμυγδαλή μπορεί να προκαλέσει σοβαρές επιπλοκές στον έλεγχο των συναισθημάτων και των παρορμήσεων. [10]

Η *φαιά ουσία* *πέριξ του υδραργύρου* ή αλλιώς *κεντρική φαιά ουσία* (PAG), βρίσκεται στο εγκεφαλικό στέλεχος. Το κύριο συναίσθημα που αναπτύσσεται σε αυτήν την περιοχή είναι το αίσθημα του πόνου και του στρες. Κύριος στόχος της περιοχής είναι να μειώσει την ύπαρξη πόνου υποβιάζοντας τον, όταν ανιχνεύεται στα νεύρα. Η περιοχή αυτή εμπλέκεται στην μείωση του άγχους και γενικά προστατεύει τον ανθρώπινο ψυχισμό από τα αρνητικά και επίπονα συναισθήματα. Επιπλέον, το PAG έχει εμπλακεί στη ρύθμιση της διάθεσης, της ανταμοιβής και του εθισμού και μπορεί να παίζει ρόλο στη ρύθμιση του ύπνου και των *κινκάρδιων ρυθμών*. Έτσι, αποτελεί μία πολύπλοκη περιοχή του εγκεφάλου με ευρύ φάσμα λειτουργιών που σχετίζονται με το συναίσθημα, την συμπεριφορά, αλλά και άλλες φυσιολογικές λειτουργίες που συνδέονται με το αυτόνομο νευρικό σύστημα. [44]

Ο *νησιωτικός φλοιός* βρίσκεται βαθιά στην πλάγια αύλακα του εγκεφάλου, μεταξύ του κροταφικού και του μετωπιαίου λοβού. Πολλές από τις λειτουργίες με τις οποίες συνδέεται επηρεάζουν τα συναισθήματα. Κάποιες από αυτές είναι η αντίληψη, η αυτογνωσία, η γνωσιακή λειτουργία και οι διαπροσωπικές συμπεριφορές. Επιπλέον, ένα σημαντικό συναίσθημα που αναπτύσσεται στον νησιωτικό φλοιό είναι το αίσθημα της αηδίας, δηλαδή της αρνητικής αντίδρασης που προκύπτει από την εμπλοκή δυσάρεστων ερεθισμάτων όπως από μία δυσάρεστη οσμή, γεύση κ.λ.π. Οι νευροεπιστήμονες θεωρούν ότι, όμοια με την περι-υδραγωγική φαιά ουσία, σχετίζεται με την αίσθηση του πόνου. Πολύ συχνά το αίσθημα του πόνου δημιουργείται σε αυτήν την περιοχή, ως αντίδραση στην ύπαρξη σωματικής βλάβης.

1.4 Μοντέλα Κατηγοριοποίησης Συναισθημάτων

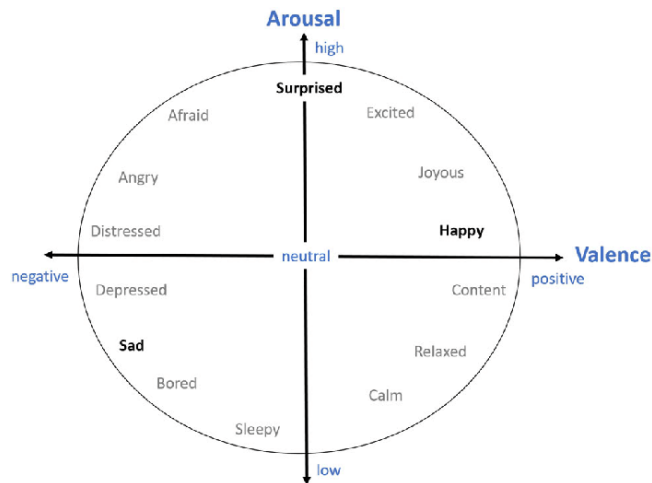
Οι ερευνητές που ασχολούνται με την συναισθηματική επιστήμη, μέχρι και σήμερα, χρησιμοποιούν διαφορετικά μοντέλα διάκρισης των συναισθημάτων, καθώς δεν έχει καθιερωθεί μία συγκεκριμένη μέθοδος διαχωρισμού τους. Οι υποθέσεις που φαίνεται να υιοθετούνται πιο συχνά είναι δύο και υποστηρίζουν ότι: i) τα συναισθήματα χωρίζονται σε έναν μικρό αριθμό διακριτών κατηγοριών, ii) τα συναισθήματα μπορούν να χαρακτηριστούν από έναν συγκεκριμένο αριθμό διαστάσεων.

1.4.1 Διακριτά Μοντέλα

Η ιδέα αυτών των μοντέλων κατηγοριοποίησης είναι ότι υπάρχουν κάποια βασικά συναισθήματα που είναι κοινώς αναγνωρίσιμα από ανθρώπους διαφορετικών πολιτισμών. Μελέτες έχουν δείξει ότι όταν τα βασικά συναισθήματα εκδηλώνονται από τις εκφράσεις του προσώπου, του σώματος και την φωνή των ατόμων, άτομα διαφορετικών εθνικοτήτων και κουλτούρας μπορούν να τα αναγνωρίσουν. Μία προσπάθεια διαχωρισμού των συναισθημάτων έγινε από τον Paul Ekman ο οποίος υποστήριξε ότι τα συναισθήματα είναι διακριτά και διαφορετικά νευροφυσιολογικά συστήματα επεξεργάζονται και ενεργοποιούν διαφορετικά συναισθήματα [65]. Μελετώντας τις εκφράσεις του προσώπου πολλών ανθρώπων από διαφορετικούς πολιτισμούς καθιέρωσε ως βασικά συναισθήματα τον θυμό, την αηδία, τον φόβο, την ευτυχία, την έκπληξη και την λύπη. Σύμφωνα με αυτήν τη θεωρία, ο συνδυασμός αυτών των συναισθημάτων δημιουργεί πιο περίπλοκα συναισθήματα.

1.4.2 Μοντέλα Πολλών Διαστάσεων

Τα μοντέλα συναισθημάτων πολλών διαστάσεων βασίζονται στην θεωρία ότι υπάρχει ένας συγκεκριμένος αριθμός μεταβλητών με πεδίο ορισμού το σύνολο των πραγματικών αριθμών που χαρακτηρίζουν τα συναισθήματα. Για παράδειγμα, αν δύο μεταβλητές χαρακτηρίζουν τον χώρο των συναισθημάτων, τότε συγκεκριμένες τιμές του χώρου πλαισιώνουν συγκεκριμένα συναισθήματα. Σε αντίθεση με τα προηγούμενα μοντέλα, τα μοντέλα πολλών διαστάσεων βασίζονται στην αρχή ότι τα συναισθήματα εκδηλώνονται από τη συνεργασία διαφορετικών νευρικών και εγκεφαλικών συστημάτων.



Σχήμα 1.3: Το μοντέλο *Circumplex* [60] , αναπαριστά τον διδιάστατο χώρο κατηγοριοποίησης συναισθημάτων στον άνθρωπο. Το μοντέλο προτάθηκε το 1980 από τον *James Russell*. Πηγή: [75]

Στο σχήμα 1.3 φαίνεται το διδιάστατο κυκλικό μοντέλο *Circumplex* που αναπαριστά τα συναισθήματα στις διαστάσεις Valence, Arousal. Η διάσταση valence (σθένος) είναι το μέτρο της θετικότητας/αρνητικότητας του συναισθήματος, ενώ η διάσταση arousal (έγερση/ξεσέχωμα) αντιστοιχεί στην ένταση του συναισθήματος, δηλαδή αν είναι υποτονικό ή διεγερτικό. Όπως φαίνεται από την εικόνα από μία συγκεκριμένη τιμή των μεταβλητών valence, arousal προκύπτει ένα συναίσθημα. Το σημείο (0, 0) είναι το απόλυτα ουδέτερο συναίσθημα.

Ένα ακόμη dimensional μοντέλο που έχει επικρατήσει στα πλαίσια της έρευνας είναι το τριδιάστατο μοντέλο PAD, την δομή του οποίου σκέφτηκαν οι Albert Mehrabian και James A. Russell. Το μοντέλο αυτό δημιουργήθηκε για να περιγράψει τις συναισθηματικές καταστάσεις στον χώρο των 3 διαστάσεων. Η ονομασία PAD προέρχεται από από τα αρχικά Pleasure (ευχαρίστηση), Arousal (έγερση) και Dominance (κυριαρχία). Η διάσταση της ευχαρίστησης δείχνει πόσο ευχάριστο ή δυσάρεστο είναι ένα συναίσθημα, για παράδειγμα ο φόβος έχει χαμηλή σε αυτήν την διάσταση. Η κλίμακα arousal υποδεικνύει πόσο παθητικό ή ενεργητικό είναι το συναίσθημα και όχι τόσο την ένταση του συναισθήματος. Για αυτό η έντονη λύπη στην κλίμακα της διέγερσης έχει χαμηλή τιμή παρ' όλο που μπορεί να θεωρηθεί ένα έντονο συναίσθημα. Η τελευταία διάσταση της κυριαρχίας δείχνει την επιβολή που έχει το συναίσθημα, για παράδειγμα ο ενθουσιασμός είναι ένα κυρίαρχο συναίσθημα, ενώ η απόγνωση ένα συναίσθημα που υποτάσσει το άτομο.

Κεφάλαιο 2

Συστήματα Καταγραφής Εγκεφαλικών Σημάτων

2.1 Ορισμός Συστημάτων BCI

Ένα σύστημα διεπαφής εγκεφάλου-υπολογιστή (BCI) ή αλλιώς σύστημα διεπαφής εγκεφάλου μη-χανής (BMI), αποτελεί ένα σύστημα παρακολούθησης της εγκεφαλικής δραστηριότητας μέσω μίας εξωτερικής συσκευής. Ως εξωτερική συσκευή συνήθως χρησιμοποιείται ένας υπολογιστής, και τα εγκεφαλικά σήματα που καταγράφονται είναι τα ηλεκτρομαγνητικά σήματα που ανταλλάσσουν οι νευρώνες (τα λεγόμενα brain waves), όταν πραγματοποιείται μία εγκεφαλική λειτουργία.

Σε αντίθεση με άλλα συστήματα, τα συστήματα BCI παρακολουθούν άμεσα τις ηλεκτρικές δραστηριότητες του εγκεφάλου και αυτό μπορούν και το κάνουν με 3 τρόπους, δηλαδή επεμβατικά, ημι-επεμβατικά και μη επεμβατικά. Όταν η καταγραφή πραγματοποιηθεί, τα ηλεκτρικά σήματα μετασχηματίζονται σε μορφή κατανοητή για το σύστημα χρησιμοποιώντας τεχνικές αναγνώρισης προτύπων και μηχανικής μάθησης [13]. Έχοντας γεφυρώσει το χάσμα της επικοινωνίας, ο ανθρώπινος εγκέφαλος μπορεί να δώσει άμεσα εντολές σε εφαρμογές των υπολογιστικών συστημάτων, χωρίς να απαιτείται η χρήση ομιλίας ή η γραφής από το άτομο που την χρησιμοποιεί.

2.2 Ιστορικά Στοιχεία

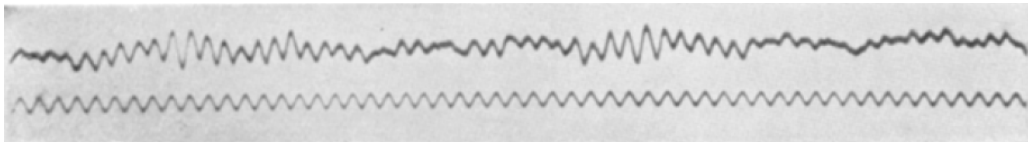
Η εξέλιξη των BCIs έχει συμβάλλει τόσο στην κατανόηση της δομής και της λειτουργικότητας του εγκεφάλου, όσο και στην γρήγορη πρόβλεψη και τον εντοπισμό σοβαρών προβλημάτων υγείας. Επιπλέον, τα συστήματα BCI έχουν χρησιμοποιηθεί για να αποκαταστήσουν χρήσιμες λειτουργίες σε άτομα με αναπηρίες από νευρομυικές διαταραχές όπως η εγκεφαλική παράλυση, το εγκεφαλικό επεισόδιο, ο τραυματισμός του νωτιαίου μυελού, η αμυοτροφική πλάγια σκλήρυνση κ.α. Οι ερευνητές μέχρι και σήμερα συνεχίζουν να χρησιμοποιούν ηλεκτροεγκεφαλικά (EEG), ενδοφλοιώδη, ηλεκτροκορτικογραφικά (ECoG) σε συνδυασμό με άλλα σήματα όπως ηλεκτρομυικά (EMG) κ.α., για την χρήση ρομποτικών βραχιόνων, αναπηρικών καρεκλών, προσθετικών μελών και άλλων συσκευών.

Η πρώτη συσκευή καταγραφής BCI πραγματοποιήθηκε το 1924 από τον Hans Begrer [76]. Η εξωτερική συσκευή που χρησιμοποίησε ήταν ένα τριχοειδές ηλεκτρόμετρο Lippmann, το οποίο συνδεόταν με ασημένια σύρματα κάτω από το κρανίο των ασθενών. Τα αποτελέσματα του πειράματος δεν ήταν καθόλου ικανοποιητικά για τον Begrer. Στη συνέχεια χρησιμοποίησε πιο εξελιγμένες συσκευές μέτρησης όπως το γαλβανόμετρο καταγραφής διπλού πηνίου Siemens, το οποίο ανίχνευε τάσεις με πολύ μικρό πλάτος. Έτσι ο Begrer κατάφερε να αναλύσει για πρώτη φορά σήματα EEG στις συχνότητες alpha και beta που κατέγραψε το σύστημα από άτομα που έπασχαν από εγκεφαλικές ασθένειες. Η ανακάλυψη των σημάτων αυτών έγινε αποδεκτή αργότερα το 1934, αφού μέχρι τότε τα αποτελέσματα των μετρήσεων αποδίδονταν σε θόρυβο, έως ότου οι

επιστήμονες Lord Adrian και Hallowell Davis επιβεβαιώσουν την εγκυρότητα τους.

Η ανάγκη της ανακάλυψης συστημάτων επικοινωνίας του ανθρώπινου εγκεφάλου και των μηχανών ξεκίνησε από τον τομέα της νευροϊατρικής, όμως σήμερα τα BCIs εφαρμόζονται και στους τομείς της ψυχαγωγίας και του gaming , στα «έξυπνα περιβάλλοντα», στον τομέα της ασφάλειας και ελέγχου ταυτότητας, στον τομέα του affective computing κ.λ.π.

Από τα πρώτα συστήματα διεπαφής εγκεφάλου-μηχανής πολύ ενδιαφέρον είναι το σύστημα που χρησιμοποιήθηκε για την σύνθεση του μουσικού έργου "*Music for Solo Performer*" , από τον συνθέτη Alvin Lucier το 1965 [47]. Ο συνθέτης συνδύασε τα σήματα EEG που μέτρησε απευθείας από τον εγκέφαλο του, και με την χρήση φίλτρων και ενισχυτών παρήγαγε ήχους από μουσικά όργανα κάθε φορά που το σύστημα εντόπιζε τα alpha waves που προέρχονταν από τον εγκέφαλο του μουσικού.



Σχήμα 2.1: Ένα από τα πρώτα EEG δείγματα που δημοσίευσε ο Hans Begrer το 1969 [17]. Στην εικόνα το πάνω σήμα είναι το δείγμα EEG που καταγράφηκε από ένα νεαρό αγόρι, ενώ το κάτω είναι ημιτονοειδές σήμα αναφοράς με συχνότητα 10 Hz. Πηγή: [17]

2.3 Μέθοδοι Καταγραφής Εγκεφαλικής Δραστηριότητας

Κάποια από τα είδη της εγκεφαλικής δραστηριότητας που καταγράφουν τα συστήματα BCI είναι τα ηλεκτρικά ρεύματα που προκύπτουν από τα σήματα που ανταλλάσσουν οι νευρώνες. Η δραστηριότητα αυτού του είδους ονομάζεται ηλεκτροφυσιολογική. Για να πραγματοποιηθούν οι νευρικές λειτουργίες, τα κύτταρα χρειάζονται ενέργεια, την οποία λαμβάνουν σε μορφή γλυκόζης. Η παροχή της γλυκόζης γίνεται μέσω των αιμοφόρων αγγείων του εγκεφάλου στις ενεργές νευρικές περιοχές. Έτσι, η παρακολούθηση της εγκεφαλικής δραστηριότητας μπορεί να γίνει με την καταγραφή των επιπέδων γλυκόζης του αίματος. Η κατηγορία αυτή ονομάζεται αιμοδυναμική. Στη συνέχεια θα ακολουθήσει μία ανάλυση πάνω στα δύο είδη εγκεφαλικής δραστηριοποίησης.

Η πρώτη κατηγορία εγκεφαλικής δραστηριότητας είναι η ηλεκτροφυσιολογική. Οι νευρώνες στέλνουν ηλεκτρικά και χημικά σήματα κατά μήκος των οδών του εγκεφάλου. Όταν ένας νευρώνας πυροδοτεί ένα ηλεκτρικό σήμα, δημιουργείται θετικό φορτίο στον άξονά του και αρνητικό φορτίο στον δένδριτη του κυττάρου που λαμβάνει αυτό το σήμα, επιτρέποντας έτσι ρεύμα να περάσει από την θετική πολικότητα προς την αρνητική. Το ηλεκτρικό σήμα μετατρέπεται σε χημικό στο τερματικό στάδιο του άξονα του κυττάρου-αποστολέα και ξαναγίνεται ηλεκτρικό όταν φτάσει στον δένδριτη του νευρώνα-παραλήπτη. Τα φορτία που ανταλλάσσουν δύο μεμονωμένοι νευρώνες είναι απειροελάχιστα και το ρεύμα που δημιουργείται δεν είναι δυνατόν να ανιχνευθεί από τις συμβατικές συσκευές. Όταν ομάδες νευρώνων πυροδοτούνται μαζί, οι ηλεκτρικές ώσεις που μεταδίδονται στις διάφορες οδούς συναθροίζονται σε μία μετρήσιμη ποσότητα που είναι γνωστή ως *πρωτεύον ρεύμα* [51]. Δευτερεύοντα ρεύματα δημιουργούνται λόγω της επαφής των πρωτογενών ρευμάτων με τα

φορτία του εγκεφαλικού ιστού και του υγρού του εγκεφάλου [14]. Οι δύο τύποι ρευμάτων, καθώς και τα μαγνητικά πεδία που παράγουν καταγράφονται από τα ηλεκτροεγκεφαλογράφημα (EEG), τα ηλεκτροκορτικογράφημα (ECoG) και τα μαγνητοεγκεφαλογράφημα (MEG). Οι τεχνικές χαρακτηρίζονται άμεσες, καθώς η ηλεκτροφυσιολογική δραστηριότητα έχει άμεση συσχέτιση με την λειτουργία των νευρώνων.

Η δεύτερη εγκεφαλική δραστηριότητα που καταγράφεται στην βιβλιογραφία είναι η αιμοδυναμική. Το αίμα απελευθερώνει γλυκόζη και οξυγόνο στους ενεργούς νευρώνες με μεγαλύτερο ρυθμό απ' ό,τι στους νευρώνες που είναι σε αδράνεια. Οι αλλαγές αυτές στην περιεκτικότητα της γλυκόζης και του οξυγόνου οδηγούν σε αλλαγές στον όγκο του αίματος ανά μονάδα εγκεφαλικού ιστού (*Cerebral Blood Volume-CBV*), στον ρυθμό της ροής του αίματος στον εγκέφαλο (*Cerebral Blood Flow-CBF*) και αλλαγές στην συγκέντρωση δεοξυαιμοσφαιρίνης και οξυαιμοσφαιρίνης [15] [39]. Για την παρακολούθηση της αιμοδυναμικής εγκεφαλικής δραστηριότητας έχουν δημιουργηθεί διάφορες μέθοδοι νευροαπεικόνισης, όπως η λειτουργική μαγνητική τομογραφία (fMRI), η φασματοσκοπία εγγύς υπέρυθρης ακτινοβολίας (NIRS) και η τομογραφία εκπομπής ποζιτρονίων (PET scan). Οι τεχνικές αυτές χαρακτηρίζονται ως έμμεσες, καθώς δεν σχετίζονται άμεσα με την λειτουργία των νευρικών κυττάρων.

2.4 Τρόποι Λήψης Σημάτων

Στα BCI συστήματα οι μέθοδοι λήψης των εγκεφαλικών σημάτων είναι 3, ανάλογα με το κατά πόσο τα ηλεκτρόδια λαμβάνουν τα εγκεφαλικά σήματα άμεσα ή έμμεσα από τον εγκεφαλικό φλοιό. Οι κατηγορίες λήψης είναι η επεμβατική, ημι-επεμβατική και μη-επεμβατική και τα σήματα που καταγράφονται σε κάθε περίπτωση είναι είτε τα ηλεκτρικά σήματα που ανταλλάσσουν οι νευρώνες του εγκεφάλου, είτε τα μαγνητικά πεδία που παράγονται από τα ρεύματα αυτά, είτε η παρατήρηση διαφορετικών διεργασιών όπως για παράδειγμα η ροή του αίματος στον εγκέφαλο.

2.4.1 Μη Επεμβατικός Τρόπος Λήψης

Σε αυτήν την κατηγορία η τοποθέτηση των ηλεκτροδίων γίνεται στο εξωτερικό του κεφαλιού χωρίς να απαιτείται η πραγματοποίηση επέμβασης για να ληφθούν τα εγκεφαλικά σήματα. Μερικές τεχνικές που μελετούν την εγκεφαλική δραστηριότητα με μη επεμβατικό τρόπο είναι το μαγνητοεγκεφαλογράφημα (MEG), η τομογραφία εκπομπής ποζιτρονίων (PET scan), η λειτουργική μαγνητική τομογραφία (fMRI), η φασματοσκοπία υπέρυθρου (fNIRS), και το ηλεκτροεγκεφαλογράφημα (EEG). Από τις παραπάνω μεθόδους η μέθοδος που θα αναλύσουμε περισσότερο στη συνέχεια είναι το ηλεκτροεγκεφαλογράφημα, καθώς είναι η πιο ευρέως χρησιμοποιούμενη τεχνική νευροαπεικόνισης, λόγω της υψηλής χρονικής ανάλυσης, του χαμηλού κόστους και της υψηλής φορητότητας.

2.4.2 Ημι-επεμβατικός Τρόπος Λήψης

Στον ημι-επεμβατικό τρόπο λήψης ανήκει το ηλεκτροκορτικογράφημα (ECoG). Τα ηλεκτρικά σήματα του φλοιού καταγράφονται από ηλεκτρόδια που διαπερνούν το σκαλπ, αλλά όχι την φαιά ουσία. Όπως και στην περίπτωση των επεμβατικών τεχνικών, απαιτείται η διεξαγωγή κρανιοτομής. Τα ηλεκτρόδια είτε διατάσσονται πιο πυκνά σε πλέγματα για να καλύπτουν μεγαλύτερες περιοχές

του εγκεφάλου, είτε μεμονωμένα πάνω από την φλοιώδη επιφάνεια.

2.4.3 Επεμβατικός Τρόπος Λήψης

Η τεχνική αυτή απαιτεί την διεξαγωγή χειρουργικής επέμβασης για την άμεση εισχώρηση μικροηλεκτροδίων, στην σκληρή μνίνη που περιβάλλει τον εγκέφαλο. Σε αντίθεση με τις προηγούμενες τεχνικές, οι εγκεφαλικές πληροφορίες που καταγράφονται έχουν μεγαλύτερη ανάλυση και βελτιωμένη ακρίβεια. Τα συστήματα αυτής της κατηγορίας χρησιμοποιούν μικροηλεκτρόδια για την παρατήρηση μίας μικρής περιοχής του εγκεφάλου (single unit BCI), ή για την ανίχνευση σημάτων από περισσότερες περιοχές (multi unit BCI). Υπάρχουν διάφορα μήκη ηλεκτροδίων, για παράδειγμα έως 1.5 mm (Utah, Blackrock Microsystems) ή 10 mm (FMA, MicroProbes) [80].

2.4.4 Σύγκριση των Τρόπων Λήψης των Σημάτων

Κάθε τεχνική λήψης δεδομένων από τον εγκέφαλο έχει κάποια θετικά και κάποια αρνητικά. Τα επεμβατικά συστήματα BCI λαμβάνουν τα δεδομένα κατευθείαν από τον εγκέφαλο και γι' αυτό είναι πιο αξιόπιστα. Όμως, η κρανιοτομή απαιτεί την συγκέντρωση μίας μεγάλης ομάδας ιατρικού προσωπικού και ιατρικών εργαλείων και γι' αυτό το κόστος είναι πολύ αυξημένο, αλλά υπάρχει και το ρίσκο αιμορραγίας, μόλυνσης ή και ανεπανόρθωτης βλάβης του ασθενή. Για τους παραπάνω λόγους, η τεχνική αυτή συνήθως εφαρμόζεται σε πολύ ειδικές περιπτώσεις, κυρίως σε τυφλούς ή παράλυτους ασθενείς.

Αντίθετα με τον επεμβατικό τρόπο λήψης εγκεφαλικών δεδομένων, ο μη επεμβατικός τρόπος είναι πολύ πιο ασφαλής και το κόστος που απαιτεί πολύ μικρότερο. Αν και τα δεδομένα μπορεί να έχουν περισσότερο θόρυβο, ο χρόνος καταγραφής τους μπορεί να είναι πολύ μεγάλος, κάτι που δεν μπορεί να συμβεί στον επεμβατικό τρόπο λήψης, καθώς ο χρόνος περιορίζεται στον χρόνο που μπορεί να είναι εκτεθειμένος ο εγκέφαλος του ατόμου στο εξωτερικό περιβάλλον.

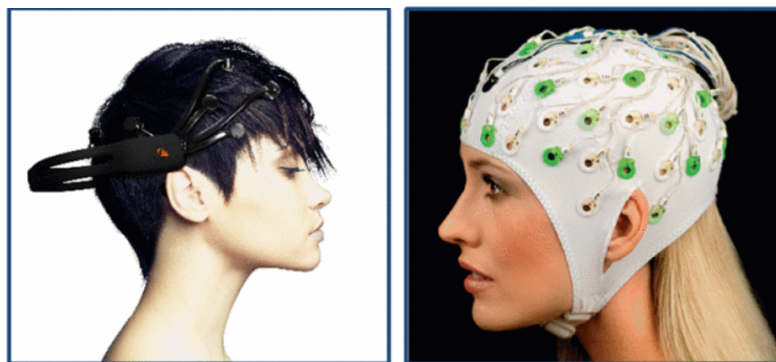
2.5 Το Ηλεκτροεγκεφαλογράφημα (EEG)

Το ηλεκτροεγκεφαλογράφημα ανήκει στην κατηγορία των μη επεμβατικών μεθόδων λήψης του τομέα BCI. Είναι ένας τρόπος παρακολούθησης των ηλεκτρικών κυματομορφών του εγκεφάλου που ενεργοποιούνται από τα νευρικά κύτταρα και ανιχνεύονται στην επιφάνεια του κεφαλιού. Η καταγραφή γίνεται με την εφαρμογή μεταλλικών ηλεκτροδίων στο скальπ και η δραστηριότητα που ανιχνεύεται προκύπτει από τα ηλεκτρικά δυναμικά των νευρικών κυττάρων στον νεοφλοιό (neocortex) και στον αλλοφλοιό (allocortex) [62]. Τα μεταλλικά ηλεκτρόδια που χρησιμοποιούνται συνήθως είναι κατασκευασμένα από κράμα χαλκού, νικελίου και ψευδαργύρου, ενώ ως επίστρωση χρησιμοποιείται το πολυτετραφθοροαιθυλένιο (Teflon).

Για τις συμβατικές συσκευές απαιτείται η χρήση ενός αγωγίμου υλικού-τζελ, το οποίο τοποθετείται στα ηλεκτρόδια της συσκευής για να μειωθεί η αντίσταση στο τριχωτό της κεφαλής που προκύπτει από τα νεκρά κύτταρα του δέρματος. Μειώνοντας την ηλεκτρική αντίσταση, αυξάνεται η αγωγιμότητα και έτσι βελτιώνεται η ανάλυση του σήματος [13]. Μία άλλη εναλλακτική είναι η επιλογή «στεγνών» συσκευών EEG που δεν απαιτούν τη χρήση τζελ στα ηλεκτρόδια. Σύμφωνα

με την βιβλιογραφία [28], οι συσκευές αυτές δεν προϋποθέτουν την τεχνική προετοιμασία που απαιτείται στην περίπτωση των συμβατικών «υγρών» συσκευών, όμως ακόμη αναμένεται να δώσουν το ίδιο καλής ποιότητας σήματα.

Υπάρχουν διάφορες κατηγορίες συσκευών που τοποθετούνται στο κεφάλι του ενδιαφερόμενου για την καταγραφή των σημάτων του εγκεφάλου του. Μερικές από αυτές έχουν τη μορφή κράνους, τιάρας, κορδέλας που δένει στο κεφάλι (headband) ή ακόμα και χρήση μεμονωμένων ηλεκτροδίων. Με την ανάπτυξη του τομέα, οι ερευνητές έχουν δημιουργήσει συσκευές που είναι άνετες για τον χρήστη, φορητές και παρέχουν ποιοτικά σήματα. Κάποια όργανα που έχουν δημιουργηθεί είναι το MindSet και το MindWave από την εταιρία NeuroSky. Οι συσκευές αυτές είναι ασύρματες και περιέχουν έναν αισθητήρα EEG. Ο χρήστης συγκεντρώνει την σκέψη του σε ένα αντικείμενο, οι αισθητήρες καταγράφουν την εγκεφαλική του δραστηριότητα και σύμφωνα με αυτήν επιτελούν συγκεκριμένες λειτουργίες. Μία ακόμη συσκευή EEG είναι η EPOC (Σχήμα 2.2, εικόνα α) της εταιρίας Emotiv. Σε αντίθεση με τις προηγούμενες συσκευές, έχει 14 EEG κανάλια και ένα γυροσκόπιο που ανιχνεύει τις κινήσεις του κεφαλιού [26]. Οι προηγούμενες συσκευές ανήκουν στην κατηγορία των dry συσκευών. Ένα σύστημα EEG που χρησιμοποιεί wet αισθητήρες είναι το BrainAmp, το οποίο είναι ασύρματο και έχει 128 ηλεκτρόδια (Σχήμα 2.2, εικόνα β)

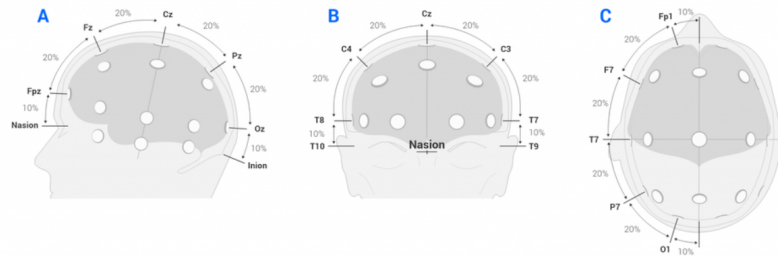


Σχήμα 2.2: α) Συσκευή Emotiv EPOC με 14 ηλεκτρόδια.
β) Συσκευή BrainAmp με 128 ηλεκτρόδια.

Πηγή: [9]

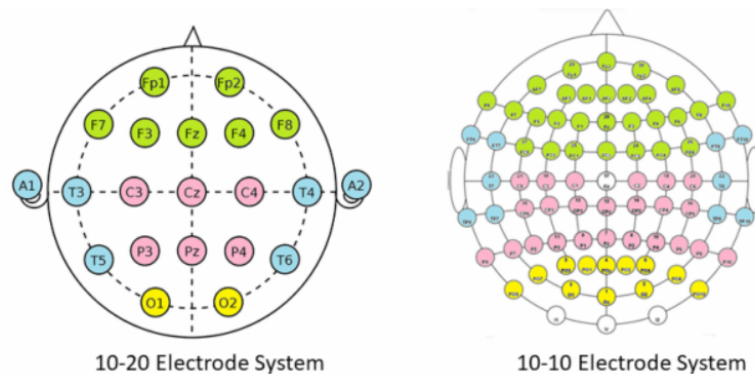
2.5.1 Πρότυπα τοποθέτησης ηλεκτροδίων

Η χωρική τοποθέτηση των ηλεκτροδίων ενός συστήματος BCI στο κεφάλι του χρήστη, συνήθως γίνεται με βάση το διεθνές σύστημα 10-20 [32]. Όπως φαίνεται στην εικόνα 2.4α, υπάρχουν 21 ηλεκτρόδια που καλύπτουν όλη την επιφάνεια του εγκεφαλικού φλοιού. Η απόσταση δύο γειτονικών ηλεκτροδίων είναι ίση με το 10% ή 20% της μέγιστης διαμέτρου του κεφαλιού, μετρούμενης κατά μήκος ή κατά πλάτος (εικόνα 2.3). Τα γράμματα που φαίνονται στην εικόνα κατονομάζουν τα ηλεκτρόδια και αντιστοιχούν στην εγκεφαλική περιοχή που τοποθετούνται. Συγκεκριμένα, τα γράμματα αντιστοιχούν στις περιοχές των ημισφαιρίων (Fr-προ-μετωπιαία, F-μετωπιαία, T-κροταφική, O-ινιακή, C-κεντρική και P-βρεγματική) και οι αριθμοί που ακολουθούν τα γράμματα αντιστοιχούν στα ημισφαίρια (οι μονοί αριθμοί ανήκουν στο αριστερό και οι ζυγοί στο δεξί ημισφαίριο). Οι αριθμοί αυξάνονται από την προ-μετωπιαία στην ινιακή περιοχή. Το γράμμα "z" αντιπροσωπεύει τη μέση των ημισφαιρίων (το "z" προέρχεται από το zero) "A" αντιπροσωπεύει τους λοβούς. [74]

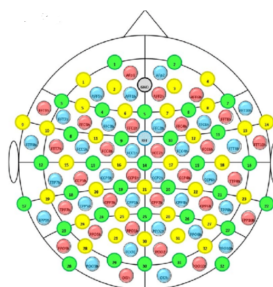


Σχήμα 2.3: Τοποθέτηση των ηλεκτροδίων με βάση το διεθνές σύστημα 10-20.
Πηγή: [64]

Για την παρακολούθηση περισσότερων περιοχών έχουν εγκαθιδρυθεί τα συστήματα 10-10 και 10-5, τα οποία αποτελούν επεκτάσεις του συστήματος 10-20. Η διαφορά τους έγκειται στις αποστάσεις των γειτονικών ηλεκτροδίων, οι οποίες είναι μικρότερες όσο αυξάνονται τα ηλεκτρόδια. Στο 10-10 σύστημα ο αριθμός των ηλεκτροδίων ποικίλει από 64 έως 85 ηλεκτρόδια, ενώ στο 10-5 από 320 έως 329. Τα συστήματα αυτά χρησιμοποιούνται κυρίως για ιατρικούς σκοπούς όπου απαιτείται η ύπαρξη πολλών ηλεκτροδίων για μεγαλύτερης ακρίβειας καταγραφή των εγκεφαλικών σημάτων. Η εικόνα 2.4 δείχνει την τοποθέτηση των ηλεκτροδίων για τα συστήματα 10-20 και 10-10 και η εικόνα 2.6 για το σύστημα 10-5.



Σχήμα 2.4: α) Τοποθέτηση των ηλεκτροδίων με βάση το διεθνές σύστημα 10-20. β) Τοποθέτηση των ηλεκτροδίων με βάση το διεθνές σύστημα 10-10. Τα πράσινα ηλεκτρόδια αντιστοιχούν στον μετωπιαίο λοβό, τα μπλε στον χροταφικό, τα ροζ στον βρεγματικό και τα κίτρινα στον ινιακό λοβό.
Πηγή: [13]



Σχήμα 2.5: Ενδεικτική τοποθέτηση των ηλεκτροδίων με βάση το διεθνές σύστημα 10-5. Πηγή: [87]

2.5.2 Χαρακτηριστικά

Οι κυματομορφές που καταγράφουν τα ηλεκτροεγκεφαλογράφημα συνήθως έχουν ημιτονοειδή μορφή και το πλάτος τους κυμαίνεται στα 0.5-10μV peak to peak (περίπου 100 φορές χαμηλότερη από τις επεμβατικές τεχνικές). Το συχνотικό περιεχόμενο των σημάτων κυμαίνεται στο εύρος 1-20Hz. Κάποια χαρακτηριστικά που καθορίζουν τις κυματομορφές EEG και μας δίνουν πληροφορίες είναι:

- Το πλάτος σε Volts.
- Η συχνότητα, δηλαδή ο ρυθμός μεταβολής του σήματος.
- Η χωρική μεταβλητότητα, δηλαδή η απόκριση του σήματος στις διάφορες περιοχές του εγκεφάλου.

Η μεταφορά των κυματομορφών στο πεδίο των συχνοτήτων κάνει εμφανή την ύπαρξη κάποιων συχνοτικών ζωνών. Αν και το φάσμα είναι συνεχής συνάρτηση, η εγκεφαλική κατάσταση του ατόμου μπορεί να κάνει κυρίαρχες 5 βασικές συχνότητες. Έτσι, μπορεί να γίνει μία κατηγοριοποίηση των κυματομορφών με βάση τις κύριες συχνοτικές ζώνες:

Κύματα Delta (δ): Οι κυματομορφές αυτού του είδους έχουν συχνότητες στο εύρος 0.5-4 Hz και υψηλά πλάτη που όμως μεταβάλλονται με αργό ρυθμό. Εμφανίζονται στον βρεγματικό και μετωπιαίο λοβό του εγκεφάλου. Οι delta κυματομορφές είναι κυρίως παρούσες στα βρέφη και κατά την διάρκεια του βαθύ ύπνου των ενηλίκων, ενώ συχνά οι νευρώνες τις διοχετεύουν σε μεγάλο βαθμό όταν έχουν προκληθεί εγκεφαλικοί τραυματισμοί ή τα άτομα παρουσιάζουν μαθησιακά προβλήματα και προβλήματα διάσπασης προσοχής. [6]

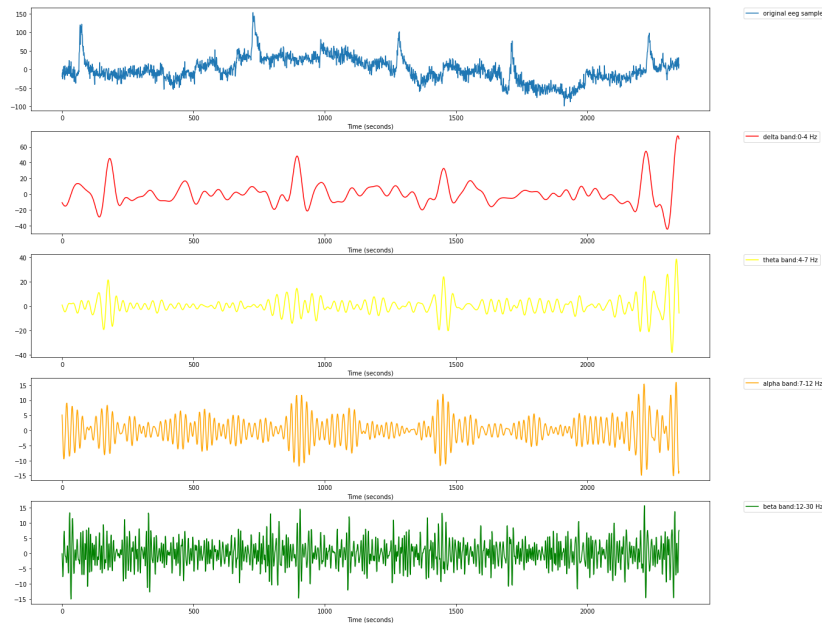
Κύματα Theta (θ): Η συχνοτική μπάντα των σημάτων αυτών καλύπτει το εύρος 4-8Hz. Η περιοχή 4-5 Hz ονομάζεται slow theta και η 6-8 Hz high theta. Είναι τύποι σημάτων που παρατηρούνται στις μετωπιαίες και στις πλευρικές περιοχές του εγκεφάλου των νέων παιδιών και όταν τα άτομα βρίσκονται υπό συνθήκες υπνηλίας. Επίσης, συνδέεται και με τον διαλογισμό. Υψηλά επίπεδα theta κυματομορφών συνδέονται με μη φυσιολογική εγκεφαλική λειτουργία, με βλάβες της υποφλοιώδους περιοχής και με μεταβολικές εγκεφαλοπάθειες. [6] [40]

Κύματα Alpha (α): Είναι οι κυματομορφές του συχνοτικού εύρους 8-13 Hz. Οι συχνότητες 8-9 Hz αντιστοιχούν στα slow alpha και οι 11-13 Hz στα high alpha waves. Η φυσιολογική λειτουργία των alpha κυματομορφών εξαπλώνεται στις κροταφικές και βρεγματικές περιοχές των ημισφαιρίων, αλλά είναι κυρίως αισθητές στον ινιακό λοβό. Εμφανίζονται όταν οι άνθρωποι χαλαρώνουν και δεν δέχονται οπτικά ερεθίσματα, ενώ εξασθενούν με το άνοιγμα των ματιών και την έντονη νοητική δραστηριότητα. Σε κατάσταση υπνηλίας τα alpha σήματα παρουσιάζονται μαζί με τις θ και β συχνότητες, ενώ συνυπάρχουν με τις δ στην κατάσταση του βαθύ ύπνου. [6] [52]

Κύματα Beta (β): Βρίσκονται στο διάστημα 13-30 Hz και χωρίζονται στα διαστήματα 13-20 Hz (low beta) και 20-30Hz (high beta). Το πλάτος τους κυμαίνεται κοντά στα 10-20 μV και σπάνια ξεπερνάει τα 30 μV. Πυροδοτούνται πιο συχνά από τις υπόλοιπες κατηγορίες κυματομορφών και οι καταγραφές έχουν δείξει ότι εντοπίζονται στις περιοχές του μετωπιαίου λοβού. Είναι συνδεδεμένα με τις κινητικές δραστηριότητες των ατόμων και είναι ο κυρίαρχος ρυθμός σε άτομα με αυξημένο άγχος. Συγκεκριμένα, τα high beta waves χαμηλού πλάτους παρουσιάζονται, όταν είναι αυξημένες οι νοητικές δραστηριότητες, ενώ όταν το πλάτος είναι απότομα αυξημένο σε κάποιες συχνότητες,

τότε πιθανώς τα άτομα να βρίσκονται υπό την επήρεια φαρμάκων, όπως βενζοδιαζεπινών, βαρβιτουρικών και χλωράλης.[6]

Κύματα Gamma (γ): Είναι οι κυματομορφές με συχνότητες στο εύρος 30-100Hz. Εμφανίζονται στον βρεγματικό, κροταφικό και προ-μετωπιαίο λοβό του εγκεφαλικού φλοιού, σχετίζονται με τις αισθήσεις και διεγείρονται όταν διάφορες γνωστικές λειτουργίες λαμβάνουν χώρα, όπως η μνήμη, η εκμάθηση και η προσοχή. Τα πλάτη των κυμάτων είναι πολύ μικρά (κοντά στα 2 μV) και μπορούν να αυξηθούν με τον διαλογισμό. [38] [48]



Σχήμα 2.6: Στην εικόνα φαίνεται το πλάτος ενός EEG καναλιού και οι αποσυνθέσεις του στις 4 βασικές συχνοτικές ζώνες.

Η ηλεκτροεγκεφαλογραφία είναι με διαφορά η πιο διαδεδομένη νευροαπεικονιστική μέθοδος. Χρησιμοποιείται ευρέως σε κλινικές και ερευνητικές μελέτες εδώ και δεκαετίες, όμως έχει επίσης κάποια μειονεκτήματα σε σύγκριση με άλλα συστήματα καταγραφής της εγκεφαλο-νευρικής λειτουργίας.

Τα συστήματα EEG έχουν εξελιχθεί πολύ και πλέον είναι διαθέσιμα για οποιονδήποτε επιθυμεί να τα έχει. Με εξαίρεση τις συσκευές EEG που χρησιμοποιούνται για εργαστηριακούς σκοπούς, τα headsets που έχουν δημιουργηθεί είναι ελαφριά, οικονομικά και σχετικά εύκολα στην χρήση. Ένα ακόμη προτέρημα σε σχέση με τα υπόλοιπα συστήματα διεπαφής εγκεφάλου-υπολογιστή, είναι η καλή χρονική ανάλυση (temporal resolution), δηλαδή ικανότητα τους να ανιχνεύουν τις αλλαγές στα ηλεκτρικά σήματα των διάφορων περιοχών του εγκεφάλου, σε μικρό χρονικό διάστημα. [27]

Οι συσκευές EEG έχουν προσφέρει πολλά στην επιστημονική και τεχνολογική κοινότητα, όμως έχουν κάποια μειονεκτήματα. Η χωρική κάλυψη του εγκεφαλικού φλοιού εξαρτάται από τον αριθμό των ηλεκτροδίων της συσκευής. Για να λάβουμε πληροφορίες από πολλές περιοχές του εγκεφαλικού φλοιού απαιτείται μεγάλος αριθμός ηλεκτροδίων. Τα συστήματα EEG είναι επιδεκτικά σε θόρυβο και παρεμβολές που μπορεί να προέρχονται από άλλες λειτουργίες που σχετίζονται με άλλα μέρη του ανθρώπινου οργανισμού. Κάποιες από αυτές είναι οι βιο-ηλεκτρικές παρεμβολές που προκαλούνται από το ανοιγόκλεισμα των οφθαλμών (ανιχνεύονται από τεχνικές ηλεκτρο-οφθαλμογραφίας EOG)

και από τους μύες (ανιχνεύονται από συστήματα ηλεκτρομυογραφίας EMG). Επίσης, η ποιότητα των μετρήσεων μπορεί να επηρεαστεί από παρεμβολές που προκαλούν εξωτερικές ηλεκτρομαγνητικές συσκευές, όπως μία κοινή γραμμή ρεύματος. Ένα ακόμη αρνητικό των συσκευών, έχει σχέση με την χρήση του gel ηλεκτρολυτών, το οποίο πρέπει να ανανεώνεται συχνά καθώς στεγνώνει και αλλοιώνεται από μικροοργανισμούς του δέρματος και του εξωτερικού περιβάλλοντος. Λύση σε αυτό δίνουν οι dry συσκευές που όμως έχουν χειρότερο λόγο σήματος προς θόρυβο (SNR) σε σχέση με τα wet EEGs . [27]

2.6 Συμβολές

Η ερευνητική και επιστημονική κοινότητα, έχει κάνει σημαντικές ανακαλύψεις που οφείλει στην ανάπτυξη των συστημάτων BCI. Στην ιατρική, η πλέον αυτοματοποιημένη μελέτη των εγκεφαλικών σημάτων και των συνδέσεων τους με τα συναισθήματα, έχει συνεισφέρει σημαντικά στην διάγνωση και παρακολούθηση διάφορων νευρολογικών και ψυχολογικών διαταραχών, όπως η σχιζοφρένεια, η κατάθλιψη και οι αγχώδεις διαταραχές. Επιπλέον, οι εξελίξεις στα συστήματα BCI και EEG , έχουν οδηγήσει στην ανάπτυξη καινοτόμων εφαρμογών, όπως βιντεοπαιχνίδια, συστήματα εικονικής και επαυξημένης πραγματικότητας και ρομπότ, που βελτιώνουν συνεχώς τις συναισθηματικές εμπειρίες των χρηστών.

Κεφάλαιο 3

Μηχανική Μάθηση

Σκοπός του κεφαλαίου είναι οι αναγνώστες να αποκτήσουν βασικές γνώσεις πάνω στους τομείς και τα εργαλεία του κλάδου της μηχανικής μάθησης. Χωρίς αυτές, το υλοποιητικό στάδιο της διπλωματικής θα ήταν ημιτελές.

3.1 Εισαγωγή

Η μηχανική μάθηση (machine learning) αποτελεί υποκλάδο της τεχνητής νοημοσύνης (artificial intelligence). Βασίζεται στην δημιουργία μοντέλων και αλγόριθμων που αντλούν πληροφορίες από διάφορες συλλογές δεδομένων, για να πραγματοποιήσουν έναν επιθυμητό στόχο. Ένας χρήσιμος τρόπος κατανόησης των αλγόριθμων μηχανικής μάθησης, είναι συγκρίνοντάς τους με τους συμβατικούς αλγόριθμους. Ο συνηθισμένος σχεδιασμός αλγορίθμων προϋποθέτει την αξιοποίηση των γνώσεων του προβλήματος και τη διαμόρφωση κατάλληλων βημάτων και κανόνων που το λύνουν. Αντίθετα, οι αλγόριθμοι μηχανικής μάθησης διατυπώνουν αυτόματα τους κανόνες από τα δεδομένα.

Για να καταλάβουμε καλύτερα την διαφορά τους αρκεί να ανατρέξουμε στο πρόβλημα του διαχωρισμού μήλων και μπανανών. Η λύση των τυπικών αλγόριθμων θα βασιζόταν στην εκ των προτέρων εύρεση των χαρακτηριστικών εκείνων που διαχωρίζουν τα δύο φρούτα. Ένας ακριβής αλγόριθμος θα πρέπει να λαμβάνει υπόψιν πολλά χαρακτηριστικά και μοτίβα, ούτως ώστε να μην κάνει λανθασμένες προβλέψεις. Παρ' όλα αυτά, η μοντελοποίηση των χαρακτηριστικών με τρόπο που να είναι κατανοητός για τα συστήματα που θα υλοποιήσουν τον αλγόριθμο, δεν είναι καθόλου εύκολη διαδικασία. Σε αντίθεση με την τυπική αλγοριθμική λογική, οι αλγόριθμοι μηχανικής μάθησης μαθαίνουν νέα μοτίβα όσο περισσότερο «παρατηρούν» τα δεδομένα, ενώ συχνά αποκαλύπτουν και χαρακτηριστικά που δεν είναι διακριτά με το γυμνό μάτι.

Ένα μοντέλο μηχανικής μάθησης χαρακτηρίζεται από έναν συγκεκριμένο αριθμό παραμέτρων, των οποίων οι τιμές μεταβάλλονται όσο το μοντέλο «μαθαίνει» πληροφορίες από τα δεδομένα. Η προσαρμογή των παραμέτρων γίνεται στο στάδιο της εκπαίδευσης (training) και τα δεδομένα που χρησιμοποιούνται σε αυτό το στάδιο αποτελούν τα δεδομένα εκπαίδευσης (training data). Το σύνολο των παραδειγμάτων εκπαίδευσης επιλέγεται, έτσι ώστε να αντιπροσωπεύει τις ιδιότητες του γενικού συνόλου δεδομένων και οι πληροφορίες που μαθαίνει το δίκτυο πρέπει να γενικεύονται όσο καλύτερα γίνεται σε όλο το πεδίο ενδιαφέροντος και όχι μόνο στα παραδείγματα που έχει δει. Όταν ολοκληρωθεί η διαδικασία της εκπαίδευσης, τα μοντέλα αποκτούν την τελική τους μορφή και σε αυτό το σημείο πρέπει να αξιολογηθεί η απόδοσή τους. Τα δίκτυα τροφοδοτούνται με νέα δεδομένα και προσπαθούν να προβλέψουν τα αποτελέσματα που αντιστοιχούν σε αυτά. Ένα μοντέλο μηχανικής μάθησης έχει καλή απόδοση, όταν μπορεί να αναπαράγει ακριβείς προβλέψεις στα νέα δεδομένα. Αυτή η διαδικασία είναι γνωστή ως inference.

Η μηχανική μάθηση έχει πολλές δυνατότητες και προοπτικές για περαιτέρω ανάπτυξη. Όσο η υπολογιστική και επεξεργαστική ισχύς αυξάνονται, τόσο περισσότερο θα εξελίσσεται. Το κυριότερο πρόβλημα που λύνει, είναι η διευθέτηση του μεγάλου όγκου δεδομένων που παράγονται

καθημερινά. Οι άνθρωποι δυσκολεύονται όλο και πιο πολύ να διαχειριστούν και να κατανοήσουν όλες τις πληροφορίες. Έτσι, η μηχανική μάθηση μπορεί να βοηθήσει στον χειρισμό των δεδομένων αποτελεσματικά, αλλά και στην λήψη των κατάλληλων αποφάσεων και δράσεων που αφορούν σε αυτά. Αρκεί μόνο η αξιοποίηση ενός μοντέλου που θα εκπαιδευτεί, μία τεχνική βελτιστοποίησης και πολλά δεδομένα.

3.2 Κατηγορίες Προβλημάτων Μηχανικής Μάθησης

Παρακάτω παρουσιάζονται εν συντομία οι τρεις βασικοί τύποι προβλημάτων μηχανικής μάθησης. Ο διαχωρισμός των προβλημάτων έγινε με βάση τον τρόπο της εκπαίδευσης των μοντέλων.

3.2.1 Επιβλεπόμενη Μάθηση

Στον τύπο της επιβλεπόμενης μάθησης (αλλιώς επιτηρούμενης μάθησης ή μάθηση με επιτήρηση) (supervised learning) έχουμε N δυάδες $\{(x_n, y_n), n = 1 \dots N\}$ που αποτελούν τα παραδείγματα εκπαίδευσης ή εποπτικά σήματα. Τα x_n είναι οι μεταβλητές εισόδου ή feature vectors και τα y_n οι ιδανικές εξοδοί που αντιστοιχούν σε κάθε είσοδο. Ο κύριος στόχος των προβλημάτων επιβλεπόμενης μάθησης είναι τα μοντέλα να αποκτήσουν γνώσεις από τα δεδομένα εκπαίδευσης που θα χρειαστούν για την σωστή πρόβλεψη των άγνωστων εξόδων y_m που αντιστοιχούν σε νέες εισόδους x_m . Τα ζευγάρια $\{(x_m, y_m), m = 1 \dots M\}$ αποτελούν τα παραδείγματα αξιολόγησης. Τα εποπτευόμενα προβλήματα μάθησης χωρίζονται περαιτέρω σε δύο τύπους ανάλογα με την φύση της επιδιωκόμενης εξόδου.

Κατηγοριοποίηση (Classification)

Στα προβλήματα κατηγοριοποίησης τα δεδομένα εισόδου αντιστοιχούν σε διακριτές εξόδους, που ονομάζονται ετικέτες κλάσης (class labels). Ένα τέτοιο παράδειγμα αποτελεί η κατηγοριοποίηση σημάτων EEG στα θετικά ή αρνητικά συναισθήματα. Η μαθηματική μορφή αυτού του προβλήματος θα ήταν η εύρεση μίας συνάρτησης $f(x) : x \in \mathbb{R} \Rightarrow y \in \{positive, negative\}$. Η συνάρτηση $f(x)$ δεν είναι δυνατόν να οριστεί μαθηματικά, όμως με χρήση μεθόδων εκπαίδευσης μπορούμε να την προσεγγίσουμε. Παραδείγματα αλγόριθμων κατηγοριοποίησης είναι ο γραμμικός ταξινομητής (linear classifier), οι μηχανές υποστήριξης διανυσμάτων (SVMs), τα δέντρα αποφάσεων, οι k -πλησιέστεροι γείτονες (k -nearest neighbours) και το random forest .

Όταν ένα μοντέλο μηχανικής μάθησης χρησιμοποιείται για την κατηγοριοποίηση δεδομένων, υπάρχουν τέσσερα πιθανά αποτελέσματα που μπορεί να συμβούν.

- **True Positives (TP):** Ο αριθμός των παραδειγμάτων που έχουν κατηγοριοποιηθεί σωστά, δηλαδή οι προβλεπόμενες κλάσεις και οι πραγματικές ταυτίζονται. Για παράδειγμα, σε ένα πρόβλημα διάγνωσης ασθενειών, ένα true positive αποτέλεσμα θα ήταν η περίπτωση όπου το μοντέλο προβλέπει σωστά ότι ένας ασθενής έχει μία συγκεκριμένη ασθένεια, δεδομένου ότι νοσεί και στην πραγματικότητα.
- **True Negatives (TN):** Στο παράδειγμα της κατηγοριοποίησης ενός ζώου στις κλάσεις σκύλος, γάτα, πουλί, true negative αποτέλεσμα θα ήταν η πρόβλεψη ότι το ζώο δεν είναι γάτα και η πρόβλεψη αυτή έρχεται σε συμφωνία με την πραγματικότητα. Εάν το μοντέλο

προβλέπει σωστά ότι το ζώο είναι σκύλος, τότε αυτό θα θεωρηθεί true negative αποτέλεσμα για τις κατηγορίες γάτα και πουλί.

- **False Positives (FP):** Στην περίπτωση αυτή το μοντέλο προβλέπει ότι το παράδειγμα ανήκει σε μία κατηγορία, ενώ στην πραγματικότητα ανήκει σε κάποια άλλη.
- **False Negatives (FN):** Ο αριθμός των παραδειγμάτων που έχουν κατηγοριοποιηθεί στις λάθος κατηγορίες, δηλαδή το μοντέλο προβλέπει ότι ένα παράδειγμα δεν ανήκει σε μία κατηγορία, ενώ στην πραγματικότητα ανήκει σε αυτήν.

Η αξιολόγηση των μοντέλων μηχανικής μάθησης είναι εξίσου σημαντική με την κατασκευή τους και πρέπει πάντα να είναι αξιόπιστη και αντικειμενική. Υπάρχουν πολλές μετρικές αξιολόγησης μοντέλων κατηγοριοποίησης στην βιβλιογραφία. Στη συνέχεια περιγράφονται οι πιο συνηθισμένες μετρικές αξιολόγησης, με βάση τους 4 τύπους αποτελεσμάτων που αναλύθηκαν:

1. **Accuracy:** Ποσοστό σωστά προβλεπόμενων παρατηρήσεων.

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

2. **Precision:** Ο λόγος των σωστά κατηγοριοποιημένων παραδειγμάτων προς τον αριθμό των παραδειγμάτων που το μοντέλο πρόβλεψε ότι ανήκουν σε κάποια κατηγορία.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

3. **Ανάκληση-Recall:** Ορίζεται ως το κλάσμα των παραδειγμάτων που έχουν κατηγοριοποιηθεί σε μία κλάση σε σχέση με όλα τα παραδείγματα που ανήκουν πραγματικά σε αυτήν την κλάση.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

4. **F1 Score:** Ορίζεται ως ο σταθμισμένος μέσος όρος της ανάκλησης (recall) και του precision.

$$\text{F1-Score} = 2 \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. **Confusion Matrix:** Ιδιαίτερος τύπος αξιολόγησης της απόδοσης των μοντέλων καθώς δεν επιστρέφει μία αριθμητική τιμή. Ωστόσο, παρέχει πολύτιμες πληροφορίες για τα προβλεπόμενα αποτελέσματα. Συγκεκριμένα, εμφανίζει αναλυτικά τις σωστές και λανθασμένες προβλέψεις για κάθε τάξη. Το μέγεθος του πίνακα ποικίλει ανάλογα με τον αριθμό των κατηγοριών. Για παράδειγμα σε ένα binary classification πρόβλημα, ο πίνακας είναι 2x2 ενώ για τρεις κατηγορίες 3x3 κ.ο.κ.

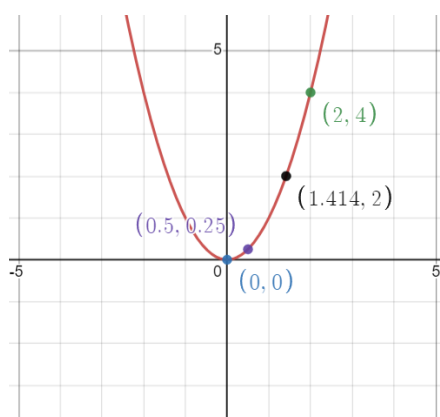
Actual Values	class A	TP	FN
	class B	FP	TN
		class A	class B
		Predicted Values	

Σχήμα 3.1: Ενδεικτικός confusion matrix που αντιστοιχεί σε πρόβλημα binary classification.

Παλινδρόμηση (*Regression*)

Τα προβλήματα παλινδρόμησης αφορούν στην εύρεση μίας σχέσης $y = f(x)$ μεταξύ της εξαρτημένης μεταβλητής y και της ανεξάρτητης x . Οι παράμετροι των μοντέλων παλινδρόμησης προσαρμόζονται με στόχο να προσεγγίσουν την συνεχή συνάρτηση $f(x)$, με όσο το δυνατόν μικρότερο σφάλμα. Έστω ότι εκπαιδεύουμε regression μοντέλο χρησιμοποιώντας τα σημεία $(0, 0)$, $(0.5, 0.25)$, $(2, 4)$ που διέρχονται από την καμπύλη $y = x^2$, όπως φαίνεται στο σχήμα 3.2. Όταν το μοντέλο τροφοδοτηθεί με ένα νέο σημείο $x_0 = 1.414$ η έξοδος του θα προσεγγίζει με κάποια ακρίβεια την ιδανική έξοδο $y_0 = 2$.

Πληροφοριακά, οι πιο δημοφιλείς μέθοδοι παλινδρόμησης είναι η γραμμική (linear regression) και η λογαριθμική (logistic regression) παλινδρόμηση. Υπάρχουν δύο τύποι παλινδρόμησης, η παρεμβολή (interpolation) και η παρέκταση (extrapolation). Στην παρεμβολή οι μεταβλητές που πρέπει να προβλεφθούν βρίσκονται ανάμεσα στο εύρος των δεδομένων εκπαίδευσης. Το παράδειγμα του σχήματος 3.1 ανήκει στην κατηγορία της παρεμβολής, καθώς η μεταβλητή $x_0 = 1.414$ ανήκει στο εύρος $[0, 2]$ των τετμημένων εκπαίδευσης. Αντίθετα, η πρόβλεψη των μεταβλητών που βρίσκονται εκτός του εύρους των μεταβλητών εκπαίδευσης που έχουν χρησιμοποιηθεί για την προσαρμογή του μοντέλου, είναι τύπος παρέκτασης. Ένα παράδειγμα extrapolation αποτελεί η πρόβλεψη των καιρικών φαινομένων της αυριανής ημέρας με βάση τα μετεωρολογικά φαινόμενα των τριών τελευταίων ημερών.



Σχήμα 3.2: Παράδειγμα προσαρμογής μοντέλου παλινδρόμησης. Τα σημεία $(0, 0)$, $(0.5, 0.25)$, $(2, 4)$ είναι τα παραδείγματα εκπαίδευσης. Όταν λάβει το σημείο $x_0 = 1.414$ η έξοδος του μοντέλου προσεγγίζει την ιδανική έξοδο $y_0 = 2$.

Συμπερασματικά, τόσο η κατηγοριοποίηση όσο και η παλινδρόμηση αποτελούν πολύτιμες στρατηγικές που λύνουν δύο βασικές κατηγορίες προβλημάτων της μηχανικής μάθησης. Στην περίπτωση της κατηγοριοποίησης, τα μοντέλα προσπαθούν να προσδιορίσουν αν τα άγνωστα παραδείγματα εμπίπτουν σε μία δεδομένη κατηγορία ή όχι. Απ' την άλλη, οι μέθοδοι παλινδρόμησης εκχωρούν συνεχείς τιμές στις μεταβλητές εισόδου. Όπως φαίνεται, το πρόβλημα της παλινδρόμησης είναι πιο περίπλοκο από αυτό της κατηγοριοποίησης, αφού στην τελευταία περίπτωση οι αποκρίσεις των μοντέλων λαμβάνουν συγκεκριμένες τιμές, ενώ στην παλινδρόμηση οι πιθανές προβλεπόμενες τιμές είναι άπειρες.

3.2.2 Μη-Επιβλεπόμενη Μάθηση

Η μη-επιβλεπόμενη μάθηση (αλλιώς μάθηση χωρίς επιτήρηση) (unsupervised learning) χαρακτηρίζει τα προβλήματα στα οποία δίνονται μόνο τα feature vectors χωρίς τις ετικέτες που αντιστοιχούν σε αυτά. Τα δεδομένα εκπαίδευσης αποτελούνται από N μεταβλητές $\{x_n, n = 1..N\}$. Αναλύοντας τα x_n , οι αλγόριθμοι είναι σε θέση να ανακαλύπτουν μοτίβα και να τα οργανώνουν σε ομάδες με βάση την ομοιότητά τους. Αυτή η διαδικασία είναι γνωστή ως ομαδοποίηση (clustering). Μερικές από τις βασικές εφαρμογές της ομαδοποίησης σχετίζονται με τους τομείς της μείωσης διαστατικότητας (dimensionality reduction), της εξαγωγής χαρακτηριστικών (feature extraction) και της εκμάθησης χαρακτηριστικών (feature learning). Οι εφαρμογές αυτές έχουν ως κοινό στόχο την αναπαράσταση των δεδομένων σε ένα χώρο μικρότερων διαστάσεων.

Σε αντίθεση με την εποπτευόμενη μάθηση, οι αλγόριθμοι μάθησης χωρίς επίβλεψη επιτρέπουν στους χρήστες να αντιμετωπίσουν πιο σύνθετα προβλήματα επεξεργασίας. Επιπλέον, εξοικονομείται χρόνος και κόστος, καθώς δεν απαιτείται η εκ των προτέρων επισήμανση των δεδομένων. Στην περίπτωση που κάποια εργασία περιλαμβάνει δεδομένα με ετικέτες και δεδομένα χωρίς ετικέτες, κάνουμε λόγο για ημι-επιβλεπόμενη μέθοδο μάθησης (semi-supervised learning).

3.2.3 Ενισχυτική Μάθηση

Οι αλγόριθμοι της ενισχυτικής μάθησης μοντελοποιούνται από έναν «έξυπνο πράκτορα» (ένα υπολογιστικό σύστημα), τις δράσεις που λαμβάνει μέσα στο περιβάλλον του και ένα σύστημα ποινής/ανταμοιβής. Για κάθε καλή ενέργεια που αφορά το περιβάλλον του, ο πράκτορας λαμβάνει θετική ανατροφοδότηση (ανταμοιβή) και για κάθε κακή δράση λαμβάνει ποινή. Η ανατροφοδότηση που έχει λάβει από προηγούμενες δράσεις, τον βοηθούν να αποκτήσει όλο και περισσότερες γνώσεις για το περιβάλλον του και για τις ενέργειες που πρέπει να λάβει στη συνέχεια. Ένα παράδειγμα ενισχυτικής μάθησης είναι η εκπαίδευση ενός οχήματος σε ένα δεδομένο περιβάλλον με εμπόδια. Οι δράσεις που καταλήγουν σε συγκρούσεις του οχήματος με τα εμπόδια τιμωρούνται με ποινές.

Η ενισχυτική μάθηση διαφέρει από την μάθηση με επίβλεψη, καθώς δεν παρέχονται οι βέλτιστες δράσεις (ετικέτες y) που πρέπει να ληφθούν από τους πράκτορες για κάθε δεδομένη κατάσταση του περιβάλλοντος (είσοδος x). Επίσης, θα ήταν λάθος να πούμε ότι είναι τύπος μάθησης χωρίς επίβλεψη, καθώς οι τρέχουσες ενέργειες που λαμβάνουν οι πράκτορες επηρεάζονται από προηγούμενες δράσεις και ανταμοιβές.

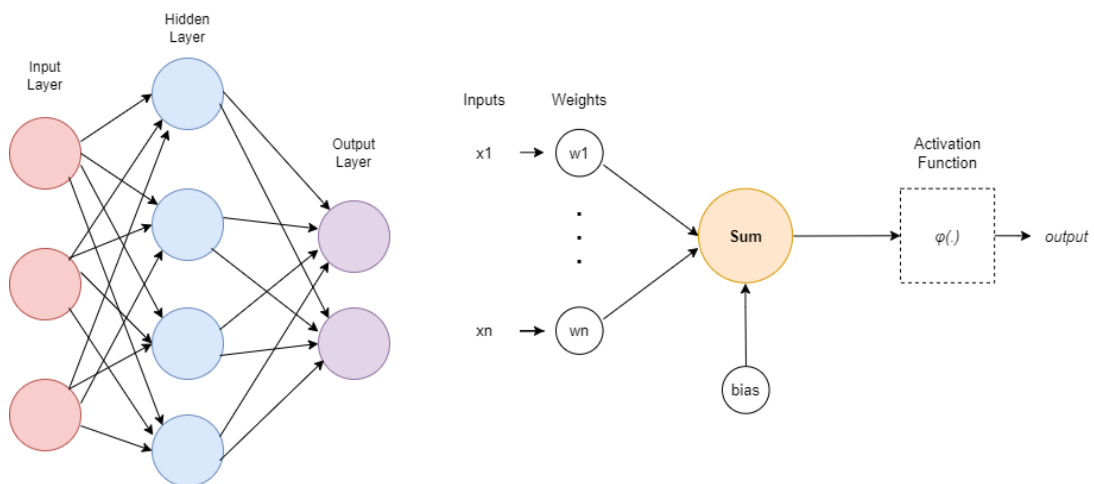
3.3 Νευρωνικά Δίκτυα

3.3.1 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks)

Τα πρώτα τεχνητά νευρωνικά δίκτυα (ή εν συντομία νευρωνικά δίκτυα) (artificial neural networks) δημιουργήθηκαν στις αρχές του 1940 από τους McCulloch και Pitts [35], δύο ερευνητές που στην προσπάθειά τους να κατανοήσουν τον ανθρώπινο εγκέφαλο, δημιούργησαν υπολογιστικά μοντέλα που προσομοιώνουν τις βασικές του λειτουργίες. Η βασική μονάδα των McCulloch-Pitt δικτύων είναι οι νευρώνες, οι οποίοι μέσω συνάψεων ανταλλάσσουν και επεξεργάζονται δεδομένα. Κάποιοι ακόμη ερευνητές που έχουν συμβάλει σημαντικά στην εξέλιξη των νευρωνικών δικτύων είναι ο Frank Rosenblatt που διεξήγαγε το πρώτο δίκτυο Perceptron γύρω στο 1960 [58] [59], και ο Bernard Widrow ο οποίος είναι γνωστός για την δημιουργία της Widrow-Hoff μεθόδου εκμάθη-

σης που εφαρμόζεται μέχρι και σήμερα στον τομέα της αναγνώρισης προτύπων [82].

Ένα νευρωνικό δίκτυο αποτελεί ένα δίκτυο διασυνδεδεμένων κόμβων, εμπνευσμένο από το δίκτυο των νευρώνων του εγκεφάλου. Στην εικόνα 3.3α. φαίνεται ένα νευρωνικό δίκτυο ενός κρυφού επιπέδου (*hidden layer*). Οι κόμβοι των τεχνητών νευρωνικών δικτύων ονομάζονται «τεχνητοί νευρώνες» (*artificial neurons*) και κάθε στρώμα του δικτύου μπορεί να έχει έναν ή περισσότερους κόμβους. Ένας τεχνητός νευρώνας επεξεργάζεται, στέλνει και λαμβάνει σήματα από τους κόμβους με τους οποίους είναι συνδεδεμένος. Κάθε νευρώνας περιέχει μία ή περισσότερες εισόδους. Μία είσοδος μπορεί να δεχτεί δεδομένα από εξωτερικές πηγές (όπως εικόνες, λέξεις κ.α.), είτε δεδομένα που προκύπτουν από τις εξόδους άλλων νευρώνων. Μόλις ο κόμβος ολοκληρώσει την επεξεργασία των δεδομένων που έχει λάβει, τροφοδοτεί τα δεδομένα στις εισόδους άλλων νευρώνων ή στην έξοδο του δικτύου.



Σχήμα 3.3: α) Στην αριστερή εικόνα φαίνεται ένα παράδειγμα τεχνητού νευρωνικού δικτύου τριών επιπέδων. Κάθε κόμβος του δικτύου συμβολίζει έναν «τεχνητό νευρώνα» και κάθε βέλος αντιπροσωπεύει μία σύνδεση από την έξοδο ενός νευρώνα στην είσοδο ενός άλλου. β) Στην δεξιά πλευρά απεικονίζεται το μαθηματικό μοντέλο ενός τεχνητού νευρώνα. Τα x_1, \dots, x_n είναι οι εισόδοι και τα w_1, \dots, w_n είναι τα βάρη του νευρώνα.

Στην εικόνα 3.3β φαίνονται αναλυτικά οι μαθηματικοί συμβολισμοί που αντιπροσωπεύουν τις λειτουργίες ενός τεχνητού νευρώνα. Η έξοδος του κάθε νευρώνα προκύπτει από το σταθμισμένο άθροισμα των εισόδων του. Μερικές φορές, οι σταθμισμένες εισόδοι προστίθενται με μία σταθερή τιμή που καλείται *bias*. Στη συνέχεια, το αποτέλεσμα του αθροίσματος περνάει από μία βαθμωτή συνάρτηση που καλείται συνάρτηση ενεργοποίησης $\varphi(\cdot)$ (*activation function*) (συνήθως μη γραμμική). Εάν w_i είναι τα βάρη του i -οστού νευρώνα, η σχέση εισόδου-εξόδου ενός επιπέδου γενικεύεται για όλους τους νευρώνες ενός επιπέδου ως εξής:

$$\mathbf{Y} = \varphi(\mathbf{W}\mathbf{X} + \mathbf{B}), \quad (3.1)$$

$$\mathbf{W} = (\mathbf{w}_1 \mathbf{w}_2 \cdots \mathbf{w}_n)^T = \begin{pmatrix} w_1^1 & w_1^2 & \cdots & w_1^n \\ w_2^1 & w_2^2 & \cdots & w_2^n \\ \vdots & \vdots & \ddots & \vdots \\ w_n^1 & w_n^2 & \cdots & w_n^n \end{pmatrix} \quad (3.2)$$

$$\mathbf{X} = (x_1 \ x_2 \ \cdots \ x_n)^T, \quad (3.3)$$

$$\mathbf{B} = (b_1 \ b_2 \ \cdots \ b_n)^T \quad (3.4)$$

Από την σχέση αυτή φαίνεται πως οι είσοδοι μετασχηματίζονται με γραμμικές σχέσεις, το αποτέλεσμα τροφοδοτείται σε απλές μη γραμμικές συναρτήσεις και η τελική έξοδος που προκύπτει είναι μη γραμμική. Σύμφωνα με το θεώρημα του G. Cybenko-1989 ([21]), «οποιαδήποτε συνεχής συνάρτηση $g(x)$, κάτω από κατάλληλες συνθήκες ομαλότητας, μπορεί να προσεγγιστεί οσοδήποτε κοντά από ένα Νευρωνικό Δίκτυο με ένα κρυφό επίπεδο και πεπερασμένο αριθμό νευρώνων». Με άλλα λόγια, ένα ANN αυθαίρετου αριθμού νευρώνων, μπορεί να προσεγγίσει οποιαδήποτε καλώς ορισμένη συνεχή συνάρτηση με ένα αυθαίρετο σφάλμα ϵ . Χάρη σε αυτό και άλλα θεωρήματα, τα νευρωνικά δίκτυα μπόρεσαν να λύσουν μεγάλο αριθμό μη γραμμικών προβλημάτων.

Στα ANN τα κρυφά επίπεδα (*hidden layers*) είναι τα στρώματα νευρώνων που βρίσκονται μεταξύ των επιπέδων εισόδου και εξόδου. Διαδραματίζουν πολύ σημαντικό ρόλο, καθώς επεξεργάζονται και μετασχηματίζουν τα δεδομένα εισόδου και στη συνέχεια τα μεταβιβάζουν στο επίπεδο εξόδου. Οι λεπτομέρειες της επεξεργασίας που συμβαίνει στα κρυφά επίπεδα είναι άγνωστη, όμως επιτρέπουν την αναπαράσταση και την εκμάθηση σύνθετων μοτίβων από τα δεδομένα. Αυτό το κάνουν προσαρμόζοντας τις άγνωστες παραμέτρους του δικτύου (δηλαδή τα βάρη και τις πολώσεις των νευρώνων). Υπό αυτήν την έννοια, τα κρυφά επίπεδα μπορούν να θεωρηθούν ως οι κύριοι παράγοντες που αξιολογούν τα πιο σημαντικά χαρακτηριστικά των δεδομένων εισόδου και χρησιμοποιούν αυτές τις πληροφορίες για να παράγουν τις εξόδους των δικτύων.

Καθώς τα τεχνητά νευρωνικά δίκτυα συνεχίζουν να εξελίσσονται, η αρχική απλή εικόνα που είχαμε για αυτά έχει αντικατασταθεί από πιο προηγμένες, βαθύτερες και πιο σύνθετες αρχιτεκτονικές. Η παρουσία και συνεισφορά τους είναι σημαντική σε πολλούς τομείς, συμπεριλαμβανομένων των πεδίων του image και speech recognition, της ρομποτικής, της βιοϊατρικής και της αναγνώρισης συναισθημάτων, του υπερκλάδου που καλύπτει την ερευνητική μελέτη της παρούσας διπλωματικής.

Gradient Descent

Παρατηρώντας την σχέση 3.1, εύκολα καταλαβαίνουμε ότι η τελική έξοδος του δικτύου έχει τη μορφή $f(\mathbf{X}, \theta)$, δηλαδή εξαρτάται από την μήτρα εισόδου \mathbf{X} και τις άγνωστες παραμέτρους $\theta = \{\mathbf{W}, \mathbf{B}\}$. Ως εκ τούτου, για να ολοκληρωθεί η ανάλυση των ANNs, απομένει να εξηγήσουμε την σχέση που καθορίζει τις άγνωστες παραμέτρους θ , που αποτελεί και σημαντικό στοιχείο της εκπαίδευσης του δικτύου. Το πρόβλημα αυτό, συνοψίζεται στην εύρεση των κατάλληλων παραμέτρων που ελαχιστοποιούν μία συνάρτηση σφάλματος-απωλειών (*Loss Function* ή *Error Function*).

Οι τιμές των παραμέτρων καθορίζονται μέσω μίας επαναληπτικής διαδικασίας, κατά την οποία τα βάρη προσαρμόζονται κατά ένα μικρό ποσοστό με βάση την τιμή της συνάρτησης απωλειών σε

κάθε βήμα. Έστω $f_{\vartheta}(x)$ η έξοδος ενός νευρωνικού δικτύου με παραμέτρους ϑ , η οποία προσεγγίζει με κάποιο σφάλμα μία συνάρτηση $u(x)$. Τότε, η συνάρτηση απωλειών ορίζεται ως $\mathcal{L}(f_{\vartheta}(x), u(x))$. Ο κανόνας για την αναπροσαρμογή των παραμέτρων στο χρονικό βήμα t είναι ο εξής:

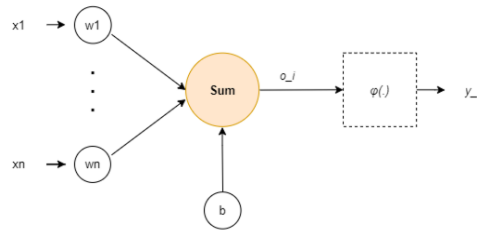
$$\vartheta^{t+1} = \vartheta^t - \lambda \frac{\partial \mathcal{L}}{\partial \vartheta^t} \quad (3.5)$$

Η βαθμωτή μεταβλητή λ αντιπροσωπεύει τον ρυθμό εκμάθησης (*learning rate*) και καθορίζει το μέγεθος του βήματος που πρέπει να ληφθεί κατά την ενημέρωση του ϑ . Η συνάρτηση κόστους $\mathcal{L}(f(\vartheta, x), u(x))$ είναι μη αρνητική συνάρτηση που μετράει την απόσταση της εκτιμώμενης εξόδου $f_{\vartheta}(x)$ από την πραγματική $u(x)$. Μία συνάρτηση κόστους που χρησιμοποιείται συχνά είναι η $\mathcal{L}_q(f(\vartheta, x), u(x)) = |f - u|^q$. Ειδική περίπτωση της \mathcal{L}_q είναι η τετραγωνική συνάρτηση απωλειών $\mathcal{L}_2(f, u) = (f - u)^2$.

Error Backpropagation

Η προς τα πίσω διάδοση σφαλμάτων (*Error Back Propagation*) προσδιορίζει την κλήση της συνάρτησης σφάλματος-απωλειών σε σχέση με τις παραμέτρους του δικτύου. Ο αλγόριθμος που υλοποιεί την ιδέα, διασχίζει το δίκτυο με αντίστροφη σειρά, από το επίπεδο εξόδου στο επίπεδο εισόδου, σύμφωνα με τον κανόνα της αλυσίδας (*chain rule*).

Ως παράδειγμα, θα εξετάσουμε την περίπτωση του νευρωνικού δικτύου του σχήματος..... όμως, η ανάλυση θα ξεκινήσει από την οπίσθια διάδοση σφάλματος στο επίπεδο ενός i νευρώνα (εικόνα 3.4).



Σχήμα 3.4: Διάγραμμα τεχνητού νευρώνα.

Έστω o_i η είσοδος της συνάρτησης ενεργοποίησης $\phi()$ και y_i η τελική έξοδος του νευρώνα που αντιστοιχεί στο διάνυσμα εισόδου \mathbf{x}_i . Αν \hat{y}_i είναι η ιδανική έξοδος που αντιστοιχεί στην είσοδο \mathbf{x}_i και $\mathcal{L}(y, \hat{y}) = 0.5 \cdot (y - \hat{y})^2$ είναι η συνάρτηση απωλειών, προκειμένου να ενημερώσουμε τα βάρη w , πρέπει να υπολογίσουμε τις μερικές παραγώγους $\{\frac{\partial \mathcal{L}}{\partial w_j} | j = 1 \dots n\}$. Διασχίζοντας το δίκτυο από την έξοδο προς την είσοδο και χρησιμοποιώντας τον κανόνα της αλυσίδας, υπολογίζουμε:

$$\frac{\partial \mathcal{L}}{\partial w_j} = \frac{\partial \mathcal{L}}{\partial y_i} \cdot \frac{\partial y_i}{\partial o_i} \cdot \frac{\partial o_i}{\partial w_j} = (y_i - \hat{y}_i) \cdot \phi'(o_i) \cdot x_j$$

Η παραπάνω εξίσωση αφορά τον υπολογισμό της μερικής παραγώγου $\frac{\partial \mathcal{L}}{\partial w_j}$ για κάθε βάρος w_j του νευρώνα. Μία ακόμα χρήσιμη αναπαράσταση της μεθόδου βασχ προπαγατιον, είναι η χρήση του κανόνα της αλυσίδας για την αναπαράσταση πινάκων.

Αν θεωρήσουμε ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο 2 επιπέδων, είσοδο X και έξοδο Y , οι εξισώσεις που το αναπαριστούν είναι:

$$Z_1 = \mathbf{W}_1 \cdot X + B_1$$

$$Z_1 = f_1(Z_1)$$

$$Z_2 = \mathbf{W}_2 \cdot Z_1 + B_2$$

$$Y = f_2(Z_2)$$

Με χρήση του κανόνα της αλυσίδας, οι εξισώσεις που επιλύουν τις μερικές παραγώγους ως προς τις παραμέτρους είναι:

$$\nabla_{[\mathbf{W}_2, B_2]} \mathcal{L} = (\nabla_Y \mathcal{L} \odot f'_2(Z_2)) \cdot [Z_1^T 1]$$

$$\nabla_{[\mathbf{W}_1, B_1]} \mathcal{L} = (\nabla_{Z_1} \mathcal{L} \odot f'_1(W_1)) \cdot [X^T 1]$$

$$\nabla_{Z_1} \mathcal{L} = W_2^T \cdot (\nabla_Y \mathcal{L} \odot f'_2(Z_2))$$

Η παραπάνω σχέση μπορεί να γενικευτεί για πλήρως συνδεδεμένο νευρωνικό δίκτυο με $k - 1$ κρυφά επίπεδα:

Στο τελευταίο k επίπεδο

:

$$\nabla_{[\mathbf{W}_k, B_k]} \mathcal{L} = (\nabla_Y \mathcal{L} \odot f'_k(Y)) \cdot [Z_{k-1}^T 1]$$

Στα ενδιάμεσα επίπεδα $i = 1..(k - 1)$:

$$\nabla_{[W_i, B_i]} L = (\nabla_{Z_i} \mathcal{L} \odot f'_i(Z_i)) \cdot [Z_{i-1}^T 1]$$

$$\nabla_{Z_i} \mathcal{L} = W_{i+1}^T \cdot (\nabla_{Z_{i+1}} \mathcal{L} \odot f'_{i+1}(Z_i))$$

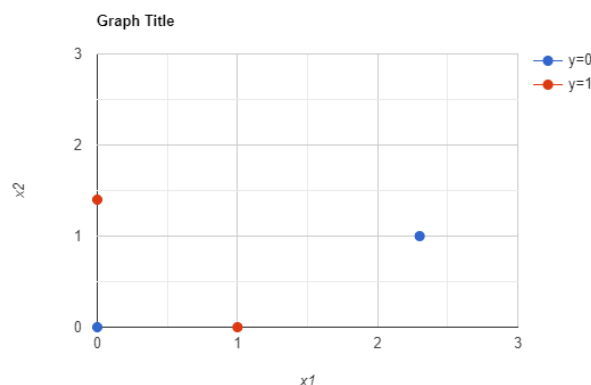
Ο συμβολισμός \odot είναι ο πολλαπλασιασμός στοιχείο προς στοιχείο. Ο όρος $\nabla_X f$ συμβολίζει την *Ιακωβιανή μήτρα*, δηλαδή τον πίνακα που περιέχει τις μερικές παραγώγους μίας παραγωγίσιμης συνάρτησης f ως προς τις ανεξάρτητες μεταβλητές x . Να σημειωθεί πως για να χρησιμοποιηθεί ο αλγόριθμος *backpropagation* πρέπει η συνάρτηση κόστους \mathcal{L} και οι συναρτήσεις ενεργοποίησης του νευρωνικού δικτύου να είναι παραγωγίσιμες στις περιοχές ενδιαφέροντος.

3.3.2 Βαθιά Μάθηση (Deep Learning)

Τα νευρωνικά δίκτυα έχουν εξελιχθεί με την πάροδο του χρόνου και παρουσιάζονται σε διάφορες μορφές και τύπους. Ανάλογα με το πλήθος των *hidden layers*, τα νευρωνικά δίκτυα χωρίζονται σε *shallow neural networks* και *deep neural networks*. Στις ρηχές (*shallow*) αρχιτεκτονικές κατατάσσονται οι πρώιμες μορφές νευρωνικών δικτύων, όπως το δίκτυο *Perceptron* του *Rosenblatt*. Συγκεκριμένα, ένα ρηχό νευρωνικό δίκτυο συνήθως έχει ένα επίπεδο εισόδου, 1 ή 2 κρυφά επίπεδα και ένα επίπεδο εξόδου, ενώ ένα deep νευρωνικό δίκτυο μπορεί να έχει παραπάνω από 2 κρυφά επίπεδα. Η απόδοση των ρηχών και βαθιών νευρωνικών δικτύων ποικίλλει ανάλογα με την

εφαρμογή και η επιλογή του κατάλληλου «βάθους» είναι πολύ σημαντική για την δημιουργία ενός επιτυχημένου μοντέλου. Σε αυτήν την ενότητα θα αναλυθούν οι λόγοι για την ανάπτυξη deep αρχιτεκτονικών.

Ας πάρουμε για παράδειγμα το πρόβλημα της δυαδικής κατηγοριοποίησης τεσσάρων σημείων x_1, x_2, x_3, x_4 . Τα σημεία αναπαριστώνται στο $2D$ επίπεδο που φαίνεται στην εικόνα 3.5. Το μπλε χρώμα αντιστοιχεί στην ετικέτα 0 και το κόκκινο στην ετικέτα 1. Η έξοδος του μοντέλου πρέπει να εξάγει σωστά τις τιμές 0 και 1 για καθένα από τα 4 σημεία.



Σχήμα 3.5: Αναπαράσταση 4 σημείων στο διδιάστατο επίπεδο.

Ένα ρηχό MLP δίκτυο, δεν μπορεί να καθορίσει το κατάλληλο διαχωριστικό όριο των σημείων αυτών, καθώς δεν υπάρχει κάποια ευθεία γραμμή που να τα διαχωρίζει. Μία λύση σε αυτό θα ήταν ο μετασχηματισμός των δεδομένων σε κάποιο υπερεπίπεδο, στο οποίο ίσως και να υπάρχει ένα γραμμικό διαχωριστικό όριο. Ωστόσο, αυτή η προσέγγιση μπορεί να αυξήσει την πολυπλοκότητα των δεδομένων και να επιβαρύνει τη διαδικασία εκπαίδευσης, καθιστώντας το δίκτυο πιο επιρρεπές στην υπερπροσαρμογή. Επιπλέον, μπορεί να οδηγήσει σε αύξηση του υπολογιστικού κόστους. Μία εναλλακτική λύση θα ήταν η χρήση *handcrafted* χαρακτηριστικών, δηλαδή η εξαγωγή των κατάλληλων μοτίβων-χαρακτηριστικών από τα δεδομένα που σχετίζονται με το γνωστικό πεδίο. Όμως, η εύρεση και η εξαγωγή αυτών των χαρακτηριστικών απαιτεί σημαντική χρονική δέσμευση και τα νέα χαρακτηριστικά ενδέχεται να μην αντιπροσωπεύουν όλες τις χρήσιμες πληροφορίες στα δεδομένα ή να μην γενικεύονται σε άλλες βάσεις δεδομένων.

Οι τεχνικές αυτές έχουν χρησιμοποιηθεί εκτενώς στο παρελθόν επειδή μπορούν να παρέχουν μια καλή αναπαράσταση των δεδομένων, αλλά δεν είναι πάντα εύκολες και πρακτικές. Από την άλλη πλευρά, η χρήση περισσότερων κρυφών στρωμάτων συνήθως είναι αρκετή για την απευθείας εκμάθηση κρυφών χαρακτηριστικών από τα δεδομένα. Για αυτόν τον λόγο, συχνά τα deep neural networks αποτελούν μία πιο ευέλικτη και ισχυρή επιλογή.

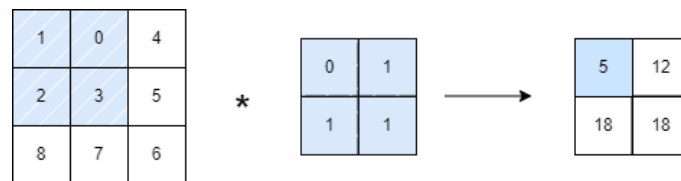
Αναλογιζόμενοι την δομή των νευρωνικών δικτύων, εύλογα αναμένεται ότι οι βαθιές αρχιτεκτονικές θα έχουν μεγαλύτερο υπολογιστικό κόστος από τις ρηχές. Η προσθήκη περισσότερων επιπέδων νευρώνων, αυξάνει τα βάρη και άρα τους υπολογισμούς που πρέπει να γίνουν, όμως η εκπαίδευση των *deep neural networks* παραμένει σχετικά γρήγορη. Σε αυτό έχουν συνεισφέρει πολλοί παράγοντες. Ένας τρόπος μείωσης του χρόνου εκπαίδευσης είναι η χρήση της *mini-batch gradient descent* τεχνικής. Η τεχνική αυτή επιτρέπει στο μοντέλο να ενημερώνει τα βάρη και τις

υπερ-παραμέτρους χρησιμοποιώντας μικρές «παρτίδες» δεδομένων αντί για ολόκληρο το σύνολο δεδομένων, μειώνοντας έτσι τον χρόνο εκπαίδευσης.

Οι αρχιτεκτονικές βαθειάς μάθησης έχουν εξελιχθεί σημαντικά και στο προσκήνιο έχουν εμφανιστεί πιο εξελιγμένες αρχιτεκτονικές από τα βασικά Multi Layer Perceptrons . Κάποιες από αυτές είναι τα Συνελικτικά Νευρωνικά Δίκτυα (*Convolutional Neural Networks*), τα Αναδρομικά Νευρωνικά Δίκτυα (*Recurrent Neural Networks*) και οι *Transformers* . Αυτές οι αρχιτεκτονικές σχεδιάστηκαν για να αντιμετωπίσουν συγκεκριμένες προκλήσεις και χρησιμοποιούνται εκτενώς σε ένα μεγάλο φάσμα πεδίων της επιστήμης και της τεχνολογίας.

Συνελικτικά Νευρωνικά Δίκτυα (CNN)

Τα συνελικτικά νευρωνικά δίκτυα είναι τύπος νευρωνικών δικτύων που εφαρμόζεται κυρίως σε δεδομένα με 2D μορφή, όπως εικόνες. Διαφέρει από την αρχιτεκτονική των συμβατικών fully connected δικτύων από την οπτική ότι ένας νευρώνας επεξεργάζεται τις τιμές ενός μικρού «παράθυρου» της εισόδου που λαμβάνει. Το παράθυρο ολισθαίνει από αριστερά προς δεξιά και από πάνω έως κάτω, κατά μήκος της εισόδου και παράγει μία μοναδική τιμή σε κάθε βήμα, έως ότου έχει διατρέξει όλη την είσοδο. Το ολισθαίνον παράθυρο ονομάζεται πυρήνας (*kernel*) και η διαδικασία της ολίσθησης *stride* . Οι δημιουργοί αυτής της αρχιτεκτονικής, δανείστηκαν το όνομα από την πράξη της συνέλιξης, καθώς η κύρια λειτουργία των νευρωνικών δικτύων αυτού του είδους, έχει μεγάλη ομοιότητα με αυτήν. Η λειτουργία αυτή καλείται *ετερο-συσχέτιση* (*cross correlation*). Το αποτέλεσμα της ετερο-συσχέτισης της εισόδου με το φίλτρο αναφέρεται ως *χάρτης χαρακτηριστικών* (*feature map*) [85].

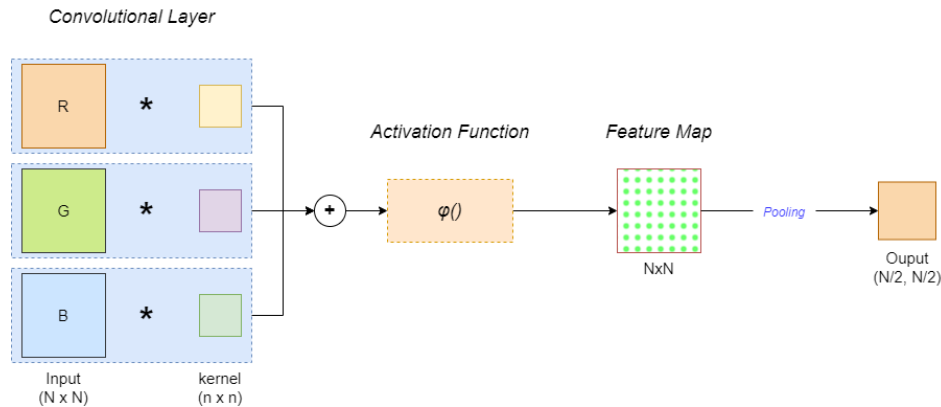


Σχήμα 3.6: Στην εικόνα βλέπουμε αναλυτικά την πράξη της ετερο-συσχέτισης μίας μήτρας μεγέθους (3×3) με ένα (2×2) φίλτρο.

Η διαδικασία της ετεροσυσχέτισης φαίνεται στην εικόνα 3.6. Το σχηματισμένο πλαίσιο εξόδου προέκυψε από τον υπολογισμό $1 \cdot 0 + 0 \cdot 1 + 2 \cdot 1 + 3 \cdot 1 = 5$. Με παρόμοιο τρόπο προκύπτουν και τα υπόλοιπα κελιά της μήτρας. Αν το μέγεθος της αρχικής μήτρας είναι (H, W) και το μέγεθος του φίλτρου (h, w) , η μήτρα που θα προκύψει θα έχει μέγεθος $(H - h + 1, W - w + 1)$. Με βάση αυτούς τους υπολογισμούς, στο παράδειγμα η τελική μήτρα έχει μήκος (2×2) . Για να παραμείνει το μέγεθος της τελικής μήτρας ίδιο με αυτό της μήτρας εισόδου, συνήθως εφαρμόζεται η τεχνική του γεμίσματος **padding** . Η μήτρα εισόδου γεμίζει με μηδενικά γύρω από τα όρια της, έτσι ώστε να υπάρχει αρκετός χώρος για να μετακινείται ο πυρήνας χωρίς να επικαλύπτει τα κελιά της.

Στο παραπάνω σκέλος έγινε η ανάλυση του *συνελικτικού επιπέδου* (*convolutional layer*) που χρησιμοποιείται για την εξαγωγή διάφορων χαρακτηριστικών από την είσοδο. Συνήθως μετά από αυτό το στρώμα, ακολουθεί το *στρώμα συγκέντρωσης* (*pooling*). Εκτελείται η υποδειγματοληψία του χάρτη χαρακτηριστικών, δηλαδή μειώνονται οι περιττές πληροφορίες και διατηρούνται τα πιο σημαντικά χαρακτηριστικά. Όπως και στο επίπεδο συνέλιξης, ένα παράθυρο μετακινείται κατά μήκος του χάρτη και εφαρμόζεται μία συνάρτηση για την δημιουργία μίας μοναδικής εξόδου

από όλες τις τιμές μέσα στο παράθυρο του χάρτη χαρακτηριστικών. Οι πιο συνηθισμένες τεχνικές *pooling* είναι το *max* και *average pooling*. Στην πρώτη τεχνική, το παράθυρο που εκτείνεται στον χάρτη διατηρεί την μέγιστη τιμή του πλαισίου, ενώ στην δεύτερη, την μέση τιμή των τιμών του.



Σχήμα 3.7: Τα βασικά στάδια ενός *Convolutional Neural Network*.

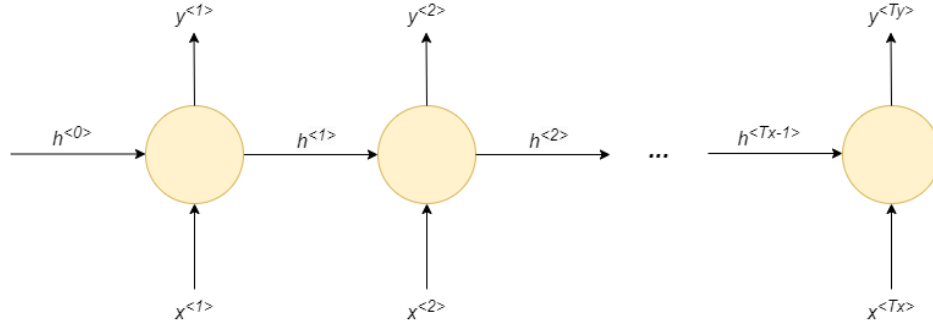
Στην εικόνα 3.7 φαίνεται η βασική επεξεργασία των συνελικτικών νευρωνικών δικτύων, όταν τροφοδοτηθούν με μία εικόνα τριών καναλιών (R, G, B). Σε κάθε κανάλι, επιλέγεται ένα φίλτρο διαστάσεων $(n \times n)$ και εκτελείται *2D* ετεροσυσχέτιση του φίλτρου με την αντίστοιχη συνιστώσα της εικόνας. Έτσι, προκύπτουν 3 εικόνες, οι οποίες προστίθενται και περνούν μέσα από μη γραμμική συνάρτηση ενεργοποίησης. Η τελική εικόνα είναι ο χάρτης χαρακτηριστικών, ο οποίος περνάει από το στάδιο *pooling*, δηλαδή η διάσταση του μειώνεται ανάλογα με την διάσταση του παραθύρου υποδειγματοληψίας (στην εικόνα η διάσταση του παραθύρου είναι (2×2)). Η διαδικασία αυτή επαναλαμβάνεται έως ότου η διάσταση της τελικής εξόδου να αποκτήσει την επιθυμητή τιμή.

Ένα βασικό πλεονέκτημα των συνελικτικών νευρωνικών δικτύων, έναντι των τυπικών αρχιτεκτονικών, είναι ότι απαιτούν λιγότερες παραμέτρους. Για να τροφοδοτήσουμε μία είσοδο διάστασης $N \times N$ σε ένα τυπικό πλήρως συνδεδεμένο νευρωνικό δίκτυο, τα στοιχεία πρέπει πρώτα να αναδιαμορφωθούν σε μονοδιάστατο διάνυσμα και τότε κάθε στοιχείο της εισόδου τροφοδοτείται σε κάθε κόμβο στο επίπεδο εισόδου του δικτύου. Συνεπώς, κάθε κόμβος του επιπέδου εισόδου θα έχει $(N^2 \times \text{πλήθος των κόμβων εισόδου})$ βάρη. Επιπλέον, η διαδικασία της αναδιαμόρφωσης της *2D* εισόδου σε διάνυσμα, προκαλεί την απώλεια των χωρικών πληροφοριών των στοιχείων της. Αντίθετα, τα CNNs εξάγουν τα καλύτερα χαρακτηριστικά με πολύ λιγότερες παραμέτρους για εκπαίδευση, διατηρώντας τις χωρικές πληροφορίες της εισόδου, γεγονός που τα καθιστά πιο υπολογιστικά αποδοτικά και αποτελεσματικά για εργασίες επεξεργασίας εικόνας.

Αναδρομικά Νευρωνικά Δίκτυα (RNN)

Τα αναδρομικά νευρωνικά δίκτυα (*recurrent neural networks*), εξειδικεύονται στην επεξεργασία ακολουθιακών δεδομένων της μορφής $x^{(1)}, \dots, x^{(\tau)}$. Οι δείκτες $1, \dots, \tau$ μπορεί να αναφέρονται σε χρονικά βήματα, ή στην θέση των δειγμάτων στην ακολουθία, ανάλογα με το είδος των δεδομένων. Τα δίκτυα αυτά, βασίζονται στην ιδέα του *parameter sharing*. Σε κάθε στρώμα του δικτύου, χρησιμοποιούν το ίδιο σύνολο παραμέτρων για την επεξεργασία των τμημάτων της ακολουθίας εισόδου. Αυτή η ιδιότητα, επιτρέπει στα RNN να διατηρούν πληροφορίες από προηγούμενα βήματα και να μοντελοποιούν τις χρονικές ή χωρικές εξαρτήσεις των δεδομένων.

Η ιδέα του *parameter sharing* είναι εξαιρετικά χρήσιμη όταν σημαντικές πληροφορίες βρίσκονται σε διαφορετικές θέσεις των ακολουθιών. Για παράδειγμα, ένα εκπαιδευμένο αναδρομικό νευρωνικό δίκτυο, με εισόδους τις προτάσεις «Η γάτα σκότωσε το ποντίκι» και «Το ποντίκι σκοτώθηκε από την γάτα», μπορεί να καταλάβει ότι και οι δύο προτάσεις μεταφέρουν τις ίδιες πληροφορίες, ανεξάρτητα από την σειρά των λέξεων. Επιπλέον, μπορούν και επεξεργάζονται αποτελεσματικά πολύ μεγάλα μήκη ακολουθιών, σε αντίθεση με τα τυπικά *fnns* όπου ο αριθμός των παραμέτρων θα αυξανόταν εκθετικά. Το επόμενο σχήμα δείχνει την οργάνωση ενός rnn δικτύου που παράγει μία έξοδο σε κάθε βήμα και κάθε κρυφό επίπεδο συνδέεται με το κρυφό επίπεδο του επόμενου βήματος.



Σχήμα 3.8: Αναδρομικό νευρωνικό δίκτυο *many to many* συνδέσεων. Το μήκος της ακολουθίας εισόδου T_x είναι ίδιο με το μήκος της ακολουθίας εξόδου T_y .

Στο σχήμα 3.8, ως h συμβολίζεται η κρυφή κατάσταση (*hidden state* ή *hidden state variable*). Στο βήμα t η κρυφή κατάσταση είναι συνάρτηση της εισόδου και της χρονικής κατάστασης του βήματος $t - 1$, δηλαδή $h(t) = f(x(t), h(t - 1))$. Οι εξισώσεις που συμβολίζουν την εμπρόσθια διάδοση (*forward propagation*) είναι οι ακόλουθες:

$$\mathbf{h}^{(t)} = f(\mathbf{b} + \mathbf{W}_h \mathbf{h}^{(t-1)} + \mathbf{U}_x \mathbf{x}^{(t)}), \quad (3.6)$$

$$\mathbf{y}^{(t)} = g(\mathbf{c} + \mathbf{V}_h \mathbf{h}^{(t)}) \quad (3.7)$$

Όπου, τα b, c, V_h, W_h, U_x είναι οι παράμετροι του δικτύου και ϕ συμβολίζεται η συνάρτηση ενεργοποίησης. Για την εκτέλεση του αλγόριθμου, απαιτείται η αρχικοποίηση της αρχικής κατάστασης $h^{(0)}$ και στη συνέχεια ο αλγόριθμος εφαρμόζεται για $t = 1 \dots T$. Παρατηρούμε ότι οι τιμές των παραμέτρων μένουν ίδιες για όλα τα βήματα της ακολουθίας.

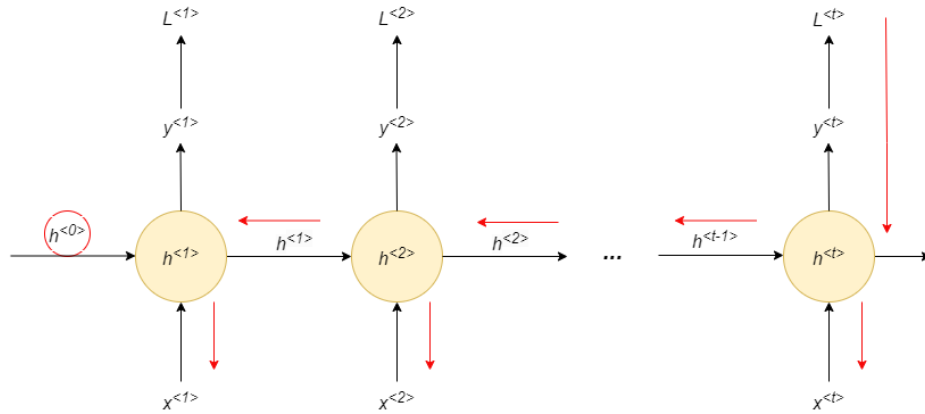
Στο σχήμα 3.8 φαίνεται ο δημοφιλής τύπος αναδρομικού νευρωνικού δικτύου **Many To Many**. Υπάρχουν πολλά ακόμη σχεδιαστικά πρότυπα αναδρομικών neural networks μερικά από τα οποία είναι:

- **One To Many:** Αναφέρεται στο σενάριο όπου το rnn δίκτυο έχει σχεδιαστεί για να παράγει πολλαπλές εξόδους από μία είσοδο. Χρησιμοποιείται σε εφαρμογές όπως *music generation*, όπου εισάγεται μία νότα και το δίκτυο προβλέπει τις επόμενες νότες της ακολουθίας.
- **Many To One:** Τα δίκτυα αυτής της αρχιτεκτονικής δημιουργούν μία μόνο έξοδο από πολλαπλές εισόδους. Συνήθως χρησιμοποιούνται σε εφαρμογές *sentiment analysis*, όπου η είσοδος είναι ακολουθίες λέξεων (π.χ. κείμενα, ομιλία κ.α.) και η έξοδος υποδεικνύει το συναίσθημα που κυριαρχεί στην είσοδο. Αυτή η αρχιτεκτονική είναι γνωστή και ως *fully recurrent architecture*, καθώς η κρυφή κατάσταση που προκύπτει στο βήμα t τροφοδοτείται ξανά στο βήμα $t + 1$.

- **Many To Many:** Οι αρχιτεκτονικές αυτής της κατηγορίας, περιέχουν δύο διαφορετικά είδη δικτύων. Στον ένα τύπο δικτύου, η ακολουθία εξόδου έχει ίδιο μήκος με την ακολουθία εισόδου και εφαρμόζεται σε εργασίες *name entity recognition*, όπου το δίκτυο παίρνει μία ακολουθία εισόδου και δημιουργεί ετικέτες για κάθε οντότητα της ακολουθίας στην έξοδο. Στον άλλο τύπο, οι ακολουθίες εισόδου και εξόδου μπορούν να έχουν διαφορετικά μήκη. Εφαρμόζεται σε εργασίες αυτόματης μετάφρασης προτάσεων, όπου οι προτάσεις εισόδου και εξόδου μπορούν να έχουν διαφορετικό αριθμό λέξεων.

Το πρόβλημα των Μακροπρόθεσμων Εξαρτήσεων

Στο προηγούμενο σκέλος έγινε η ανάλυση της εμπρόσθιας διάδοσης των αναδρομικών νευρωνικών δικτύων, όπου οι εξόδοι κάθε χρονικού βήματος προκύπτουν από τις εξισώσεις 3.6, 3.7 για τα δίκτυα *many to many*. Η οπίσθια διάδοση των recurrent neural networks είναι απαραίτητη για την ανανέωση των τιμών των παραμέτρων μέσω του αλγόριθμου *gradient descent* και φαίνεται να διασχίζει το δίκτυο στο πεδίο του χρόνου από το τελευταίο χρονικό βήμα προς το αρχικό. Γι' αυτό, αναφερόμαστε στην οπίσθια διάδοση των αναδρομικών νευρωνικών δικτύων ως **backpropagation through time**. Στην εικόνα 3.9 απεικονίζεται η κατεύθυνση με την οποία υπολογίζονται οι μερικές παράγωγοι της συνάρτησης απωλειών σε σχέση με τις παραμέτρους του δικτύου σε κάθε χρονικό βήμα.



Σχήμα 3.9: Με κόκκινο φαίνεται η σειρά με την οποία θα υπολογιστούν οι μερικές παράγωγοι στην οπίσθια διάδοση.

Παρακάτω φαίνονται οι εξισώσεις της οπίσθιας διάδοσης στον χρόνο, ενός απλού αναδρομικού νευρωνικού δικτύου:

$$\mathbf{h}^t = \mathbf{W}_h \mathbf{h}^{t-1} + \mathbf{U}_x \mathbf{x}^t, \mathbf{o}^t = \mathbf{V}_h \mathbf{h}^t \quad (3.8)$$

Η ολική συνάρτηση απωλειών προκύπτει ως $\mathcal{L}(\mathbf{x}, \mathbf{y}, \theta) = \sum_{t=1}^T l(\mathbf{o}^t, \mathbf{y}^t)$, όπου $l(\cdot)$ είναι η συνάρτηση απωλειών και \mathbf{y}^t η ιδανική έξοδος. Στη συνέχεια φαίνονται οι ιαχωριανές μήτρες των μερικών παραγώγων τις \mathcal{L} με τους πίνακες των παραμέτρων.

$$\nabla_{V_h} \mathcal{L} = \sum_{t=1}^T \nabla_{\mathbf{o}^t} l(\mathbf{o}^t, \mathbf{y}^t) \cdot (\mathbf{h}^t)^T$$

,

$$\nabla_{W_h} \mathcal{L} = \sum_{t=1}^T \nabla_{\mathbf{o}^t} l(\mathbf{o}^t, \mathbf{y}^t) \cdot \mathbf{V}_h^T \cdot \nabla_{W_h} \mathbf{h}^t$$

$$\nabla_{U_x} \mathcal{L} = \sum_{t=1}^{\tau} \nabla_{\mathbf{o}^t} l(\mathbf{o}^t, \mathbf{y}^t) \cdot \mathbf{V}_h^T \cdot \nabla_{U_x} \mathbf{h}^t$$

Λόγω του αναδρομικού στοιχείου οι προηγούμενες κρυφές καταστάσεις επηρεάζονται από τις επόμενες: $\nabla_{\mathbf{h}^t} \mathbf{h}^{t+1} = \mathbf{V}_h^T$, επομένως $\nabla_{\mathbf{h}^t} \mathbf{h}^{\tau} = (\mathbf{V}_h^T)^{\tau-t}$. Άρα:

$$\nabla_{\mathbf{w}_h} \mathbf{h}^t = \sum_{j=1}^t (\mathbf{V}_h^T)^{t-j} \cdot \mathbf{h}_j$$

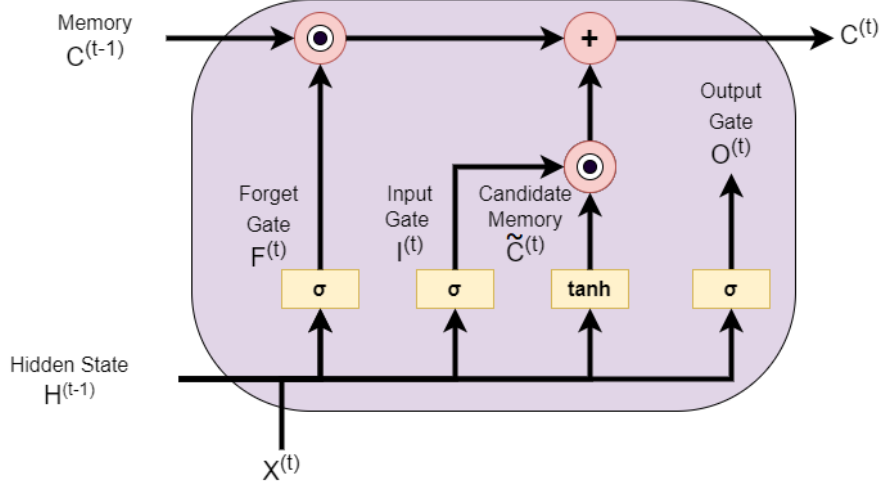
$$\nabla_{\mathbf{v}_x} \mathbf{h}^t = \sum_{j=1}^t (\mathbf{V}_h^T)^{t-j} \cdot \mathbf{x}_j$$

Ο πίνακας των παραμέτρων \mathbf{V}_h είναι τετραγωνικός πίνακας διάστασης $n \times n$ με n ανεξάρτητα ιδιοδιανύσματα $q_i, i = 1 \dots n$ και μπορεί να γραφτεί στη μορφή $\mathbf{V}_h = \mathbf{Q} \Lambda \mathbf{Q}^{-1}$, (όπου Λ ο πίνακας των ιδιοτιμών και \mathbf{Q} ο πίνακας των ιδιοδιανυσμάτων του πίνακα \mathbf{V}_h). Αν αντικαταστήσουμε το \mathbf{V}_h με την ιδιο-αποσύνθεση του (*eigendecomposition*) στις εξισώσεις των μερικών παραγώγων, θα παρατηρήσουμε ότι οι ιδιοτιμές του πίνακα θα υψωθούν στην δύναμη του χρόνου Λ^t . Όταν το χρονικό μήκος t είναι πολύ μεγάλο, οι ιδιοτιμές με πολύ μικρό μέγεθος θα τείνουν στο μηδέν και οι ιδιοτιμές με μέγεθος μεγαλύτερο της μονάδας αναμένεται να γίνουν πολύ μεγάλες. Αυτό δημιουργεί πρόβλημα στην εκμάθηση μακροπρόθεσμων εξαρτήσεων, καθώς τα βάρη που σχετίζονται με μακροπρόθεσμες αλληλεπιδράσεις θα έχουν εκθετικά μικρότερες τιμές από τα βάρη που σχετίζονται με τις βραχυπρόθεσμες.

Το παραπάνω πρόβλημα υφίσταται κυρίως στα αναδρομικά νευρωνικά δίκτυα και μπορεί να επιβραδύνει σε μεγάλο βαθμό την εκπαίδευσή τους, καθώς το σήμα που σχετίζεται με τις βραχυπρόθεσμες εξαρτήσεις θα τείνει να «κρύβεται». Ο ισχυρισμός αυτός αποδεικνύεται και μέσω πειραμάτων που έχουν διεξαχθεί [16]. Σύμφωνα με αυτά, για ακολουθίες μεγέθους μόνο 10 ή 20 μονάδων, η πιθανότητα της επιτυχημένης εκπαίδευσης ενός τυπικού rnn δικτύου, μέσω του *stochastic gradient descent* τείνει πολύ γρήγορα στο μηδέν. Για την επίλυσή του, έχουν προταθεί διάφορες τεχνικές, όπως το *gradient clipping*, η κανονικοποίηση των βαρών σε ένα καθορισμένο εύρος και η χρήση εναλλακτικών ακολουθιακών μοντέλων, όπως τα *LSTM*, *GRU* δίκτυα και πιο μοντέρνες αρχιτεκτονικές που βασίζονται σε *Transformers* για να μετριάσουν το πρόβλημα.

LSTM δίκτυα

Τα LSTM δίκτυα (*long short term memory*) αποτελούν επέκταση των αναδρομικών νευρωνικών δικτύων. Σχεδιάστηκαν για να λύσουν το πρόβλημα της έλλειψης μακροπρόθεσμης διατήρησης πληροφοριών. Όπως και τα παραδοσιακά αναδρομικά δίκτυα, περιέχουν επαναλαμβανόμενες μονάδες που συνδέονται μεταξύ τους και χρησιμοποιούν κοινές παραμέτρους. Όμως, οι μονάδες των δικτύων αυτών περιέχουν ορισμένα πρόσθετα στοιχεία που μιμούνται τις βασικές λειτουργίες μνήμης και γι' αυτό αποκαλούνται μονάδες μνήμης (*memory blocks*). Συγκεκριμένα, σχεδιάστηκαν για να αναγνωρίζουν και να αποκλείουν πληροφορίες που δεν είναι χρήσιμες. Σημαντικά στοιχεία κάθε μονάδας είναι τρεις πύλες (*gates*): η πύλη εισόδου, εξόδου και η πύλη λήθης (*forget gate*), οι οποίες είναι υπεύθυνες για τις λειτουργίες εγγραφής, ανάγνωσης και αποκλεισμού των περιεχομένων των μπλοκ μνήμης.



Σχήμα 3.10: Διάγραμμα ενός memory block .

Στο σχήμα 3.10 φαίνεται ένα *memory block* (ή αλλιώς *lstm block*) που δέχεται δεδομένα σε κάποιο βήμα t . Στο block τροφοδοτείται η είσοδος X^t και η κρυφή κατάσταση του προηγούμενου βήματος H^{t-1} , οι οποίες διέρχονται μέσα από τις τρεις πύλες, δηλαδή μέσα από fully connected επίπεδα όπως φαίνεται στις παρακάτω εξισώσεις:

$$\text{Input Gate: } \mathbf{I}^{(t)} = \sigma(\mathbf{X}^{(t)}\mathbf{W}_{xi} + \mathbf{H}^{(t-1)}\mathbf{W}_{hi} + \mathbf{bi}) \quad (3.9)$$

$$\text{Forget Gate: } \mathbf{F}^{(t)} = \sigma(\mathbf{X}^{(t)}\mathbf{W}_{xf} + \mathbf{H}^{(t-1)}\mathbf{W}_{hf} + \mathbf{bf}) \quad (3.10)$$

$$\text{Output Gate: } \mathbf{O}^{(t)} = \sigma(\mathbf{X}^{(t)}\mathbf{W}_{xo} + \mathbf{H}^{(t-1)}\mathbf{W}_{ho} + \mathbf{bo}) \quad (3.11)$$

Όπου W_x, W_h είναι τα βάρη που πολλαπλασιάζονται με την είσοδο και την κρυφή κατάσταση του προηγούμενου βήματος και b είναι οι πολώσεις των τριών πυλών. Η συνάρτηση $\sigma(\cdot)$ είναι η σιγμοειδής συνάρτηση ενεργοποίησης οι οποία μετασχηματίζει τις τελικές τιμές των πυλών στο εύρος $[0, 1]$.

Το υποψήφιο κελί μνήμης (*candidate memory cell*) υπολογίζεται με παρόμοιο τρόπο από την έκφραση:

$$\tilde{\mathbf{C}}^{(t)} = \tanh(\mathbf{X}^{(t)}\mathbf{W}_{xc} + \mathbf{H}^{(t-1)}\mathbf{W}_{hc} + \mathbf{bc}). \quad (3.12)$$

Συνήθως ως συνάρτηση ενεργοποίησης του υποψήφιου κελιού μνήμης επιλέγεται η υπερβολική εφαπτομένη, γι' αυτό και οι τιμές του κελιού κυμαίνονται στο εύρος $[-1, 1]$. Στα lstm δίκτυα, η πύλη $\mathbf{I}^{(t)}$ καθορίζει το ποσοστό της εισόδου που θα διατηρηθεί στην μνήμη μέσω της $\tilde{\mathbf{C}}^{(t)}$. Ο μηχανισμός της $\mathbf{F}^{(t)}$ απ' την άλλη, ελέγχει την ποσότητα του περιεχομένου του προηγούμενου κελιού μνήμης $\mathbf{C}^{(t-1)}$ που θα ληφθεί υπόψιν στο παρόν χρονικό βήμα. Δηλαδή, όταν η πύλη λήθης είναι πολύ μικρή (πλησιάζει το 0) και η πύλη εισόδου κοντά στο 1, η προηγούμενη μνήμη θα αποκλειστεί και το κελί μνήμης του τωρινού βήματος θα γίνει ίσο με το υποψήφιο κελί μνήμης.

Συνεπώς, το νέο κελί μνήμης δομείται ως εξής:

$$\mathbf{C}^{(t)} = \mathbf{F}^{(t)} \odot \mathbf{C}^{(t-1)} + \mathbf{I}^{(t)} \odot \tilde{\mathbf{C}}^{(t)} \quad (3.13)$$

Η κρυφή κατάσταση $\mathbf{H}^{(t)}$ (εξ. 3.14) καθορίζεται από την πύλη εξόδου. Όταν η τιμή της $\mathbf{O}^{(t)}$ προσεγγίζει την μονάδα, η κρυφή κατάσταση διοχετεύεται με όλες τις πληροφορίες της μνήμης

$C^{(t)}$, ενώ όταν πλησιάζει το μηδέν, οι πληροφορίες της τωρινής μνήμης διατηρούνται μόνο στην τρέχουσα χρονική μονάδα και δεν περνάνε στην επόμενη.

$$H^{(t)} = O^{(t)} \odot \tanh(C^{(t)}) \quad (3.14)$$

3.3.3 Νευρωνικά Δίκτυα Βασισμένα σε Μηχανισμούς Αυτοπροσοχής

Στις προηγούμενες ενότητες καλύφθηκαν εκτενώς κάποιες από τις πιο σημαντικές αρχιτεκτονικές της βαθιάς μάθησης, τα *cnh* και *rnn* δίκτυα. Συγκεκριμένα, τα *rnn* δίκτυα κατάφεραν να διευρύνουν το εύρος των δεδομένων που αξιοποιούνται στο πεδίο μελέτης της βαθιάς μάθησης, εισάγοντας τον τύπο των ακολουθιακών δεδομένων. Το πρόβλημα της εκμάθησης των μακροπρόθεσμων εξαρτήσεων, λύθηκε σε ένα βαθμό από τα *lstm* δίκτυα, με την χρήση του νέου *memory block*, όμως ορισμένες αδυναμίες αυτών των δικτύων παρέμειναν στο προσκήνιο. Από τις εξισώσεις 3.13 και 3.6 φαίνεται ότι τα *lstm* και *rnn* νευρωνικά δίκτυα, δεν μπορούν να υποστηρίξουν την παράλληλη υλοποίηση των επιπέδων κατά την εκπαίδευση, γιατί για να εκτελεστεί ένα *layer* απαιτείται η εκ των προτέρων υλοποίηση των προηγούμενων. Ως εκ τούτου, η εκπαίδευση των δικτύων αυτών με μεγάλο όγκο δεδομένων και με ακολουθίες μεγάλου μήκους απαιτεί σημαντική χρονική και υπολογιστική δαπάνη.

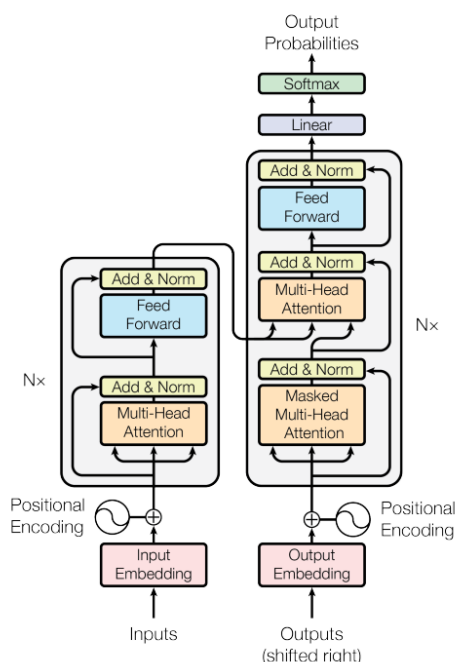
Τα συνελικτικά νευρωνικά δίκτυα, σε αντίθεση με τα προηγούμενα, υποστηρίζουν την ταυτόχρονη εκτέλεση των διαφορετικών επιπέδων και γι' αυτό η εκπαίδευση και η εξαγωγή συμπερασμάτων μπορεί να γίνει αποτελεσματικά και σε μικρότερο χρόνο. Αδυνατούν όμως, να εντοπίσουν τους συσχετισμούς των ακολουθιακών δεδομένων, όταν τα μήκη τους μεταβάλλονται. Η κύρια μονάδα των δικτύων αυτών είναι οι πυρήνες, οι οποίοι ολισθαίνουν κατά μήκος των δεδομένων εισόδου και επεξεργάζονται τμήματα της εισόδου που έχουν πάντα ίδια μεγέθη. Όταν κληθεί να χειριστεί ακολουθίες διαφορετικού μεγέθους, ο πυρήνας δεν θα μπορεί να λάβει πάντα υπόψη τις εξαρτήσεις των διαφορετικών τμημάτων των ακολουθιών. Λόγω αυτού, τα *cnh* δεν μπορούν να μαθαίνουν όλες τις μακροπρόθεσμες εξαρτήσεις των δεδομένων.

Ένα νέο μοντέλο έχει φέρει την επανάσταση στον τομέα της μηχανικής μάθησης και κυρίως στον τομέα της επεξεργασίας φυσικής γλώσσας, ο ***transformer***. Οι εφαρμογές του επεκτείνονται πέρα από την επεξεργασία φυσικής γλώσσας, στην αναγνώριση ομιλίας, στην επεξεργασία βίντεο, εικόνας και ήχου και περαιτέρω. Ο λόγος που το μοντέλο αυτό έχει γίνει τόσο δημοφιλές, οφείλεται στην ικανότητα του να επεξεργάζεται ακολουθιακά δεδομένα με υψηλή αξιοπιστία και ακρίβεια. Σε προβλήματα στα οποία διαθέτουμε μεγάλο όγκο δεδομένων, το μοντέλο του *transformer* είναι συχνά η προτιμότερη επιλογή έναντι των αναδρομικών νευρωνικών δικτύων και των παραλλαγών τους. Κύριο πλεονέκτημα της λειτουργίας του είναι ότι μπορεί και χειρίζεται όλα τα δεδομένα ταυτόχρονα, σε αντίθεση με τα αναδρομικά νευρωνικά δίκτυα, τα οποία επεξεργάζονται ένα δείγμα σε κάθε αλγοριθμικό βήμα.

Το δίκτυο *transformer* έχει καταφέρει να λύσει με επιτυχία το πρόβλημα των μακροπρόθεσμων εξαρτήσεων. Αυτό οφείλεται στον ειδικό μηχανισμό *attention*. Για παράδειγμα, αν χρησιμοποιήσουμε ένα δίκτυο *transformer* για να διαγνώσει ψυχικές ασθένειες, από δεδομένα EEG, το δίκτυο θα μπορεί να παρατηρήσει και να αναλύσει τις αλλαγές στα δυναμικά της εγκεφαλικής δραστηριότητας σε όλο το χρονικό δείγμα ταυτόχρονα και θα αποθηκεύσει τις χρονικές συσχετίσεις τους. Σε αντίθεση, ένα *rnn* δίκτυο, μπορεί να δυσκολευτεί να εντοπίσει τις συσχετίσεις μεταξύ δυναμικών που απέχουν πολύ μεταξύ τους χρονικά. (Σημείωση: Το συγκεκριμένο παράδειγμα δόθηκε για να

γίνει κατανοητό το πλεονέκτημα των μοντέλων transformer . Απαιτείται περισσότερη έρευνα για την επιβεβαίωση της υπεροχής των transformers έναντι των rnns για την ανίχνευση ασθενειών από σήματα EEG .)

Η πρωτοποριακή αρχιτεκτονική του μοντέλου transformer παρουσιάστηκε και δημοσιεύθηκε για πρώτη φορά το 2017 μέσα από το paper με τίτλο "*Attention is all you need*" [78] . Η δομή της απεικονίζεται στο σχήμα 3.11. Η αρχιτεκτονική του transformer χωρίζεται σε δύο επιμέρους τμήματα, την κωδικοποίηση και την αποκωδικοποίηση. Αρχικά, η ακολουθία που αναπαριστά την είσοδο $\mathbf{x} = (x_1, \dots, x_n)$, μετασχηματίζεται σε μία ακολουθία συνεχών τιμών $\mathbf{z} = (z_1, \dots, z_n)$ που ονομάζεται *μνήμη* ή *πλαίσιο* (*context*). Η αρχική αναπαράσταση της εισόδου συμβαίνει στο τμήμα της κωδικοποίησης, το οποίο αποτελείται από μονάδες *αυτο-προσοχής* και πλήρως συνδεδεμένα νευρωνικά δίκτυα. Το τμήμα της αποκωδικοποίησης, λαμβάνει στην είσοδο την ακολουθία \mathbf{z} , μαζί με τις προηγούμενες εξόδους και χρησιμοποιώντας τις μονάδες της κωδικοποίησης με ορισμένες τροποποιήσεις, δημιουργεί την ακολουθία εξόδου $\mathbf{y} = (y_1, \dots, y_m)$.



Σχήμα 3.11: Η αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή του transformer. Πηγή:[78]

Για να γίνει πλήρως κατανοητή η αρχιτεκτονική του transformer στους αναγνώστες, στη συνέχεια θα ακολουθήσει αναλυτική περιγραφή των δομικών στοιχείων του σχεδιασμού του μαζί με τις μαθηματικές σχέσεις που τα περιγράφουν. Η ανάλυση θα διεξαχθεί με την σειρά των βημάτων που ακολουθεί η αρχιτεκτονική, από την τροφοδότηση της εισόδου στον κωδικοποιητή μέχρι την έξοδο του αποκωδικοποιητή.

Attention

Ο τίτλος της δημοσίευσης [78] μας εισάγει σε μία από τις πιο βασικές δομές του transformer τον μηχανισμό «προσοχής», μία λειτουργία που αντιπροσωπεύει την συνειδητή προσοχή στους έμβιους οργανισμούς. Συγκεκριμένα, ο μηχανισμός αυτός, επιτρέπει στο δίκτυο να δίνει περισσότερη αξία

στα πιο σημαντικά τμήματα των δεδομένων, και λιγότερη σε αυτά που δεν παρέχουν σημαντικές πληροφορίες στο μοντέλο. Στην πιο απλή μορφή της, η λειτουργία αυτή μοντελοποιείται με το σταθμισμένο άθροισμα των τιμών εισόδου, όμως στην περίπτωση του transformer ο μηχανισμός προσοχής είναι πιο σύνθετος, αφού περιλαμβάνει περισσότερους μετασχηματισμούς για τον υπολογισμό των βαρών προσοχής μεταξύ διαφορετικών τμημάτων της εισόδου. Αν θεωρήσουμε ως είσοδο του δικτύου τον πίνακα $\mathbf{x} \in \mathbb{R}^{d_m \times d}$ (n το πλήθος των αντικειμένων εισόδου και d η διάσταση τους), οι σχέσεις που εκφράζουν τον μηχανισμό προσοχής είναι οι εξής:

$$Q(\mathbf{x}_i) = \mathbf{x}_i W_q, K(\mathbf{x}_i) = \mathbf{x}_i W_k, V(\mathbf{x}_i) = \mathbf{x}_i W_v \quad (3.15)$$

$$\text{όπου } W_q, W_k \in \mathbb{R}^{d \times d_k} \text{ και } W_v \in \mathbb{R}^{d \times d_v}$$

$$attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.16)$$

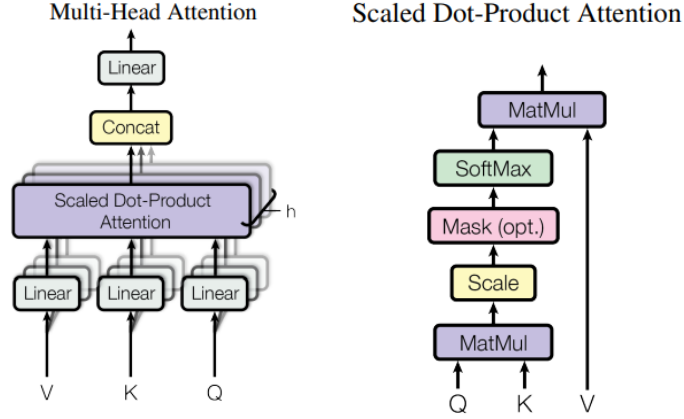
$$\text{όπου } Q = \begin{bmatrix} Q(\mathbf{x}_1) \\ \vdots \\ Q(\mathbf{x}_n) \end{bmatrix} \in \mathbb{R}^{n \times d_k}, K = \begin{bmatrix} K(\mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}_n) \end{bmatrix} \in \mathbb{R}^{n \times d_k}, V = \begin{bmatrix} V(\mathbf{x}_1) \\ \vdots \\ V(\mathbf{x}_n) \end{bmatrix} \in \mathbb{R}^{n \times d_v}$$

Οι πίνακες Q, K, V περιγράφονται ως «ερωτήματα» (*queries*), «κλειδιά» (*keys*) και «τιμές» (*values*), οι καταχωρήσεις τους προκύπτουν από τον πολλαπλασιασμό των πινάκων W_q, W_k, W_v με όλα τα $x_i, i = 1 \dots n$. Οι τιμές των πινάκων βάρους W_q, W_k, W_v προσδιορίζονται κατά την διάρκεια της εκπαίδευσης και οι αναπαραστάσεις x_i , αποτελούν τα διανύσματα των δεδομένων εισόδου που καλούνται *input embeddings*. Κάθε είσοδος \mathbf{x}_i αντιστοιχίζεται σε έναν πίνακα $Q(\mathbf{x}_i)$. Στον μηχανισμό προσοχής, η συμβατότητα όλων των πινάκων $Q(\mathbf{x}_i)$ με όλους τους πίνακες-κλειδιά $K(\mathbf{x}_j)$ (όπου το j κυμαίνεται από 1 έως n και $j \neq i$), υπολογίζεται από το εσωτερικό γινόμενο QK^T της σχέσης 3.16. Στις θέσεις που το γινόμενο QK^T έχει μεγαλύτερη τιμή, τόσο μεγαλύτερη θα είναι η προσοχή του δικτύου.

Η σχέση 3.16 δίνει τις κανονικοποιημένες τιμές των γινομένων QK^T για μία μοναδική αναπαράσταση των Q, K, V . Η κανονικοποίηση $\sqrt{d_k}$ γίνεται για να διασφαλιστεί ότι το εσωτερικό γινόμενο QK^T δεν θα λάβει ακραία μεγάλες τιμές που θα προκαλέσουν αστάθειες κατά την διάρκεια της εκπαίδευσης. Για να μπορέσει το μοντέλο να εξάγει περισσότερες πληροφορίες από την είσοδο, παράγει διαφορετικές αναπαραστάσεις των πινάκων Q, K, V , πολλαπλασιάζοντας τους με h διαφορετικούς πίνακες βαρών. Όλες οι έξοδοι που προκύπτουν συνενώνονται και διέρχονται από ένα τελευταίο γραμμικό επίπεδο. Η διαδικασία αυτή φαίνεται στην εικόνα 3.12 και εκφράζεται από τις σχέσεις 3.17, 3.18. Στην ερευνητική κοινότητα συνήθως αναφέρεται ως **multi-head attention** μηχανισμός. Σε κάθε βήμα εκμάθησης, το μοντέλο μαθαίνει τις νέες και βελτιωμένες εκδοχές των query, key και value .

$$head_i = attention(W_i^Q Q^T, W_i^K K^T, W_i^V V^T) \quad (3.17)$$

$$MLH = Concat(head_1, \dots, head_h) W^o \quad (3.18)$$



Σχήμα 3.12: Στην αριστερή εικόνα φαίνεται η ροή του *multi-head attention* μηχανισμού και στην δεξιά τα αναλυτικά στάδια του *scaled dot product*. Ως Q, K, V συμβολίζονται οι πίνακες που περιέχουν τα query, key, value αντίστοιχα. Πηγή:[78]

Στην σχέση 3.17 τα βάρη $W_i^Q, W_i^K \in \mathbb{R}^{d_m \times d_k}$, $W_i^V \in \mathbb{R}^{d_m \times d_v}$ και $W^o \in \mathbb{R}^{hd_v \times d_m}$.

Positional Encoding

Στη γενική περίπτωση των αναδρομικών και συνελικτικών νευρωνικών δικτύων τα μοντέλα μπορούν και συγκεντρώνουν πληροφορίες που σχετίζονται με την χωρική οργάνωση των δεδομένων στην είσοδο. Στα μεν πρώτα, τα μοντέλα επεξεργάζονται τα δεδομένα με την σειρά με την οποία τροφοδοτούνται σε αυτά, ενσωματώνοντας πληροφορίες των προηγούμενων βημάτων στην επεξεργασία του τρέχοντος βήματος. Στα δε συνελικτικά νευρωνικά δίκτυα, ο πυρήνας ολισθαίνει κατά μήκος των δεδομένων και το δίκτυο επεξεργάζεται τμήματα της εισόδου με καθορισμένη σειρά. Αυτό επιτρέπει στα CNN να καταγράφουν τις θέσεις των χαρακτηριστικών της εισόδου. Τα εργαλεία που αναλύσαμε και περιγράφουν την αρχιτεκτονική του transformer δεν παρέχουν την γνώση της ακριβούς θέσης των x_i στην είσοδο.

Γι' αυτόν τον λόγο, στις εισόδους των τμημάτων του κωδικοποιητή και αποκωδικοποιητή, διοχετεύονται οι πληροφορίες των θέσεων των input embeddings σε κωδικοποιημένη μορφή. Αν X είναι οι ενσωματώσεις εισόδου διάστασης $n \times d$, τότε αν ένας πίνακας P , περιέχει τις κωδικοποιημένες θέσεις των δεδομένων εισόδου θα έχει ίδιες διαστάσεις με τις ενσωματώσεις εισόδου $n \times d$ (όπου n το πλήθος των δεδομένων και d η διάσταση των input embeddings). Οι πιο συνηθισμένες κωδικοποιήσεις που χρησιμοποιούνται έχουν ημιτονοειδή μορφή όπως περιγράφεται παρακάτω.

$$P_{i,2j} = \sin\left(\frac{i}{10000^{2j/d}}\right)$$

$$P_{i,2j+1} = \cos\left(\frac{i}{10000^{2j/d}}\right)$$

Όπου, $i = 0, \dots, n-1, j = 0, \dots, \lfloor \frac{d-1}{2} \rfloor$. Αν η είσοδος που θα επεξεργαστεί το δίκτυο έχει μορφή ακολουθίας, ο δείκτης i αντιπροσωπεύει την σχετική θέση των δεδομένων στην ακολουθία και ο δείκτης d την διάσταση των input embeddings. Εκτός από τον παραπάνω τρόπο κωδικοποίησης, οι πληροφορίες της χωρικής τοποθέτησης των δεδομένων στην είσοδο μπορούν να διαμορφωθούν και με learning τρόπο, για παράδειγμα, εκπαιδεύοντας ένα δίκτυο με είσοδο τις σχετικές θέσεις των δεδομένων στην ακολουθία και δημιουργώντας μία αναπαράσταση των θέσεων διάστασης d . Σε αυτήν την περίπτωση κάνουμε λόγο για *positional embeddings*.

encoder

Έχοντας αναλύσει τα νέα εργαλεία στα οποία βασίζεται η αρχιτεκτονική του transformer, μπορούμε να προχωρήσουμε στην διάρθρωση της δομής του κωδικοποιητή και του αποκωδικοποιητή. Στο σχήμα 3.11 ο κωδικοποιητής αναπαρίστανται στο αριστερό τμήμα. Στην είσοδο του διέρχονται τα input embeddings προστιθέμενα με τα positional encodings (ή embeddings). Ο encoder περιλαμβάνει ένα επαναλαμβανόμενο μπλοκ με multi-head attention μηχανισμούς και ένα τυπικό **position-wise FFN** δίκτυο. Το FFN δίκτυο αποτελείται από γραμμικά στάδια που στην έξοδο τους εφαρμόζεται η συνάρτηση ενεργοποίησης ReLU. Μετά την έξοδο κάθε υπο-επιπέδου ακολουθούν **residual** στάδια, δηλαδή οι έξοδοι προστίθενται με την είσοδο των προηγούμενων επιπέδων. Με τις residual συνδέσεις, υπάρχει ροή πληροφοριών από τα αρχικά στα τελικά στάδια και έτσι το δίκτυο αποκτάει «μνήμη». Οι έξοδοι των residual επιπέδων διέρχονται από το στάδιο του *layer normalization* [12].

decoder

Ο ρόλος του αποκωδικοποιητή, έχει να κάνει με την παραγωγή σειριακών ακολουθιών, δηλαδή μία προς μία. Όμοια με τον κωδικοποιητή, περιέχει μηχανισμούς multi-head attention και point-wise ffn δίκτυα, ακολουθούμενα από residual συνδέσεις και *layer normalization* στάδια. Η είσοδος του αποκωδικοποιητή τροφοδοτείται με τις προηγούμενες παραγόμενες εξόδους και τις ενσωματωμένες κωδικοποιήσεις των θέσεων. Στο μεταγενέστερο multi-head attention στάδιο, τροφοδοτείται με τις εξόδους του κωδικοποιητή που περιέχουν τις πληροφορίες του attention μηχανισμού της εισόδου. Αξίζει να σημειωθεί ότι το πρώτο multi-head attention στάδιο, υλοποιείται με μία τροποποιημένη δομή. Όπως φαίνεται στο σχήμα 3.12, ένα στάδιο *masking* εφαρμόζεται πριν την softmax συνάρτηση και μετά το κανονικοποιημένο εσωτερικό γινόμενο των πινάκων Q, K . Η μάσκα εμποδίζει το δίκτυο απ' το να «προσέξει» δεδομένα που δημιουργήθηκαν σε επόμενα βήματα και καθιστά το μοντέλο κατάλληλο για εργασίες πρόβλεψης ακολουθίας.

Vision Transformer

Τρία χρόνια μετά την εμφάνιση της δημοσίευσης που συστήνει το μοντέλο του τυπικού transformer, οι Dosovitskiy *et al.*[23] παρουσίασαν μία παραλλαγή αυτού, η οποία χρησιμοποιεί ως δομικό στοιχείο τον κωδικοποιητή. Η νέα αρχιτεκτονική δημιουργήθηκε για να διευρύνει τα πεδία εφαρμογών του μετασχηματιστή πέραν του τομέα του NLP, στα πεδία της υπολογιστικής όρασης (computer vision). Έτσι, ο vision μετασχηματιστής προοριζόταν για να χειριστεί $2D$ δεδομένα εικόνων της μορφής $\mathbf{x} \in \mathbb{R}^{H \times W \times C}$. Τα στάδια της αρχιτεκτονικής φαίνονται στην εικόνα 3.13. Αρχικά, τα στοιχεία των εικόνων αναδιατάσσονται σε $2D$ patches διάστασης $\mathbf{x}_p \in \mathbb{R}^{N \times (P^2 \cdot C)}$ με $N = \frac{HW}{P^2}$ το πλήθος των patches. H, W είναι οι διαστάσεις της κάθε εικόνας, C είναι ο αριθμός των καναλιών (π.χ. σε RGB εικόνες το $C = 3$) και (P, P) είναι οι διαστάσεις των δισδιάστατων patches. Οι σχέσεις 3.19 έως 3.22 εκφράζουν τις εξισώσεις του δικτύου.

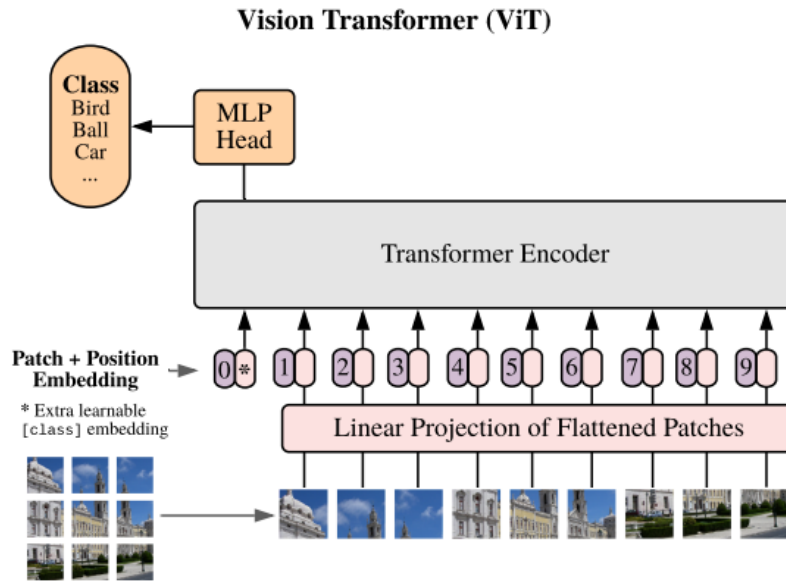
$$\mathbf{z}_0 = [\mathbf{x}_{class}; \mathbf{x}_p^1 \mathbf{E}; \mathbf{x}_p^2 \mathbf{E}; \dots; \mathbf{x}_p^N \mathbf{E}] + \mathbf{E}_{pos}, \quad \mathbf{E} \in \mathbb{R}^{(P^2 \cdot C) \times D}, \mathbf{E}_{pos} \in \mathbb{R}^{(N+1) \times D} \quad (3.19)$$

$$\mathbf{z}_l' = \text{MSA}(\text{LN}(\mathbf{z}_{l-1})) + \mathbf{z}_{l-1}, \quad l = 1 \dots L \quad (3.20)$$

$$\mathbf{z}_l = \text{MLP}(\text{LN}(\mathbf{z}_l')) + \mathbf{z}_l', \quad l = 1 \dots L \quad (3.21)$$

$$\mathbf{y} = LN(\mathbf{z}_L^0) \quad (3.22)$$

Τα N στο πλήθος \mathbf{x}_p patches διέρχονται από ένα εκπαιδύσιμο γραμμικό επίπεδο που στην έξοδο μετασχηματίζει την διάσταση τους σε διάσταση D , όπως φαίνεται στην εξίσωση 3.19. Κατά την διάρκεια της εκπαίδευσης, ένα *class token* συνενώνεται με την ακολουθία των patches και ο ρόλος του είναι να συνοψίζει τα περιεχόμενα των δεδομένων εισόδου βοηθώντας το μοντέλο να κάνει καλύτερες προβλέψεις. Η τιμή του class token καθορίζεται πριν την εκπαίδευση μέσω ενός MLP δικτύου. Οι επόμενες εξισώσεις εκφράζουν το δίκτυο του encoder, που χρησιμοποιεί μηχανισμούς αυτοπροσοχής και MLP δίκτυα. Η έξοδος του κωδικοποιητή παράγει μία διανυσματική αναπαράσταση της εικόνας εισόδου που χρησιμεύει στην λήψη των αποφάσεων ταξινόμησης. Οι τιμές της εξόδου y αντιπροσωπεύουν τις κλάσεις κατηγοριοποίησης της εισόδου.



Σχήμα 3.13: Επισκόπηση αρχιτεκτονικής του *Vision Transformer*. Στις εξόδους των MLP δικτύων εφαρμόζεται η συνάρτηση ενεργοποίησης *gaussian linear error unit* (GELU). Πηγή:[23]

Η αρχιτεκτονική του Vision Transformer δεν διαφέρει και πολύ από αυτήν του τυπικού Transformer. Σε προβλήματα κατηγοριοποίησης εικόνων έχει πετύχει *state-of-the-art* επιδόσεις και συγκριτικά με τα CNN δίκτυα που μέχρι τότε χρησιμοποιούνταν κατά κόρον σε προβλήματα συσχετισμένα με τον τομέα της υπολογιστικής όρασης, έχει καταφέρει να καταγράψει επιτυχώς τις *global* εξαρτήσεις των δεδομένων και να εντοπίζει σύνθετα μοτίβα και αντικείμενα. Επίσης, χειρίζεται με μεγάλη ακρίβεια δεδομένα εισόδου με μεταβλητό μέγεθος, αντιμετωπίζοντας τα ως ακολουθίες από patches. Αυτό μπορεί να διευκολύνει την χρήση του μοντέλου σε ένα ευρύ φάσμα μεγεθών.

Πέραν των παραπάνω πλεονεκτημάτων, το μοντέλο του Vision Transformer έχει και κάποια πιθανά μειονεκτήματα. Όπως και οι συμβατικοί μετασχηματιστές, έτσι και ο vision μετασχηματιστής απαιτεί μεγάλο όγκο δεδομένων προκειμένου να ολοκληρωθούν οι εργασίες που επιτελεί με επιτυχία και τα αποτελέσματα που θα δοθούν να έχουν ικανοποιητική ακρίβεια. Επιπλέον, τα μοντέλα αυτά απαιτούν πολλούς υπολογισμούς τόσο κατά την διάρκεια της εκπαίδευσης, όσο και κατά την διάρκεια του *inference*. Γι' αυτό οι χρήστες των μοντέλων πρέπει να διαθέτουν προηγμένο υπολογιστικό εξοπλισμό.

Εν κατακλείδι, το μοντέλο Vit αποτελεί πανίσχυρο εργαλείο που μπορεί να επεκταθεί από το πρόβλημα του *image classification* και σε προβλήματα κατηγοριοποίησης άλλων δεδομένων που μπορούν να τμηματοποιηθούν σε *patches*. Η χρήση του σε διάφορα προβλήματα και η αποτελεσματική επίλυση τους, έχουν δείξει ότι συχνά ξεπερνά την απόδοση ορισμένων παραδοσιακών μοντέλων, όπως τα LSTM και CNN δίκτυα.

Κεφάλαιο 4

Συστήματα Αναγνώρισης Συναισθημάτων

4.0.1 Εισαγωγή

Τα ανθρώπινα συναισθήματα μπορούν να βγουν στην επιφάνεια και να γίνουν διακριτά με πολλούς τρόπους. Κάποιοι από αυτούς είναι μέσα από τις εκφράσεις του προσώπου και του σώματος, της έντασης και του ύφους της φωνής, των λέξεων που χρησιμοποιούνται κατά την ομιλία, αλλά και μέσα από την ανάλυση των βιολογικών αντιδράσεων. Βασισμένοι σε αυτήν την υπόθεση, οι ερευνητές έχουν μελετήσει μεθόδους εντοπισμού συναισθημάτων χρησιμοποιώντας τις πληροφορίες που δίνει κάθε τύπος δεδομένων ξεχωριστά αλλά και συνδυαστικά. Έτσι, ένας νέος τρόπος διαχωρισμού των τεχνικών αναγνώρισης μπορεί να γίνει με βάση τα δεδομένα που αξιοποιούνται. Στις επόμενες ενότητες του κεφαλαίου θα ακολουθήσει περιγραφή των συστημάτων αναγνώρισης συναισθημάτων και παρουσίαση των προσεγγίσεων που έχουν δημοσιευτεί στην βιβλιογραφία.

4.0.2 Ανασκόπηση της Βιβλιογραφίας

Συστήματα βασισμένα στις εκφράσεις προσώπου

Σε αυτό το πεδίο έρευνας χρησιμοποιούνται ως δεδομένα εικόνες με τα πρόσωπα ανθρώπων ή βίντεο που καταγράφουν τις εκφράσεις τους με την πάροδο του χρόνου. Στην περίπτωση που τα δεδομένα που διατίθενται είναι εικόνες, κάνουμε λόγο για **στατική** αναγνώριση συναισθημάτων προσώπου, ενώ όταν έχουμε καταγραφές σε βίντεο, η αναγνώριση των συναισθημάτων είναι **δυναμική**. Υπάρχουν πολλά σύνολα δεδομένων διαθέσιμα, που περιέχουν εικόνες ή βίντεο συνοδευόμενα με τις σημάνσεις τους, οι οποίες βασίζονται σε κατηγορικά ή πολυδιάστατα μοντέλα συναισθημάτων. Μερικά από αυτά είναι τα AffectNet, EMOTIC, Google Facial Expression Dataset κ.α. Στην βιβλιογραφία, η έρευνα για την αυτόματη αναγνώριση συναισθημάτων προσώπου (*automatic facial emotion recognition-FER*), χωρίζεται σε δύο υποκλάδους.

Ο πρώτος κλάδος περιλαμβάνει την ανάλυση των εικόνων (ή των βίντεο) που γίνεται με μη αυτοματοποιημένο τρόπο, ξεκινώντας από την επιλογή και εξαγωγή των χαρακτηριστικών που συλλαμβάνουν τις πιο σημαντικές πτυχές των δεδομένων (*handcrafted features*). Αυτή η διαδικασία συνήθως αναλύεται σε τρία στάδια: α) Στην ανίχνευση του προσώπου β) στην εξαγωγή και επιλογή των χαρακτηριστικών που σχετίζονται με τις εκφράσεις του προσώπου και γ) στην κατηγοριοποίηση των εκφράσεων σε συναισθήματα. Οι N. Sebe *et al.* [63] χρησιμοποίησαν ένα τρισδιάστατο μοντέλο παρακολούθησης του προσώπου και της παραμόρφωσης του κατά την πάροδο του χρόνου, το οποίο προσαρμόζει 16 επιφάνειες Bézier, χρησιμοποιώντας ως σημεία ελέγχου τα χαρακτηριστικά της έκφρασης του προσώπου, δηλαδή τις γωνίες των ματιών, του στόματος κ.α. Η έξοδος του μοντέλου, εξάγει ένα διάνυσμα χαρακτηριστικών κίνησης, το οποίο στη συνέχεια τροφοδοτείται στις εισόδους ταξινομητών, όπως SVMs, kNNs και δέντρα αποφάσεων, για την κατηγοριοποίηση της έκφρασης σε 7 συναισθήματα.

Στον δεύτερο κλάδο, οι ερευνητές επιλέγουν μοντέλα βαθιάς μάθησης, τα οποία τροφοδοτούνται με τις καταγραφές των προσώπων και κατηγοριοποιούν τα δεδομένα σε συναισθήματα, χωρίς

να εξάγουν εκ των προτέρων τα *handcrafted features*. Η διαδικασία της αξιολόγησης και της εκπαίδευσης γίνεται από «άκρο σε άκρο» (*end to end*) ξεκινώντας με την εισαγωγή των δεδομένων στο μοντέλο. Οι N. Jain *et al.* [31] πρότειναν την δημιουργία ενός μοντέλου που συνδυάζει CNN και RNN νευρωνικά δίκτυα. Αρχικά κάθε εικόνα περνάει από το CNN δίκτυο, το οποίο παράγει ένα διάνυσμα χαρακτηριστικών. Στη συνέχεια, κάθε διάνυσμα διέρχεται από το αναδρομικό νευρωνικό δίκτυο, το οποίο επιστρέφει την τιμή των *emotion labels* της εικόνας. Οι Deepak Kumar Jain *et al.* (2019) [30] πρότειναν την εκπαίδευση ενός *deep convolutional neural network* με 6 συνελκτικά επιπέδα και *residual* συνδέσεις που εφαρμόζονται μετά το δεύτερο και τέταρτο στρώμα συνέλιξης. Για το στάδιο της κατηγοριοποίησης, ενέταξαν 2 πλήρως συνδεδεμένα επίπεδα. Οι ερευνητές σύγκριναν την απόδοση του μοντέλου τους με το μοντέλο των N. Jain *et al.* [31], αξιοποιώντας τα δεδομένα της βάσης JAFFE και απέδειξαν ότι η ακρίβεια της υλοποίησης τους είναι καλύτερη κατά 0.32%

Συστήματα βασισμένα σε κείμενο

Τα δεδομένα κειμένου χρησιμοποιούνται αρκετά συχνά για την ανάλυση λέξεων ή φράσεων που μπορεί να υποδηλώνουν την ύπαρξη συναισθημάτων, καθώς είναι ευρέως προσβάσιμα. Τα δεδομένα αυτά μπορεί να προέρχονται από ηχογραφημένες ομιλίες που μετατράπηκαν σε κείμενο, από γραπτά μηνύματα, άρθρα και άλλων ειδών αρχεία σε μορφή κειμένων [66]. Ο αποθηκευτικός χώρος των αρχείων αυτών είναι πολύ μικρότερος σε σύγκριση με τον χώρο που απαιτούν δεδομένα άλλου είδους και μπορεί να μειωθεί οσοδήποτε χρειάζεται για να διατηρηθούν οι λέξεις/φράσεις που υποδεικνύουν συναισθηματική φόρτιση. Τα τελευταία χρόνια έχει δημιουργηθεί ένας νέος τομέας με τίτλο "*Sentiment Analysis*", ο οποίος εντοπίζει και αναλύει την ύπαρξη ή τις υπόνοιες συναισθηματικής φόρτισης από τα δεδομένα κειμένων.

Συχνά το ερευνητικό πεδίο του τομέα *sentiment analysis* συγχέεται με τον τομέα του *emotion detection*, καθώς οι έννοιες *sentiment* και *emotion* είναι παραπλήσιες. Στην πραγματικότητα, *sentimental analysis* αποτελεί οποιαδήποτε εργασία επεξεργασίας φυσικής γλώσσας (NLP) που στοχεύει στον εντοπισμό της γνώμης (*opinion mining*) ή του συναισθήματος που αντανακλούν τα δεδομένα. Στο ερευνητικό άρθρο τους, οι W. Medhat *et al.* (2014) [49], παρουσίασαν και ταξινόμησαν 54 άρθρα που λύνουν ειδικά προβλήματα του τομέα *sentiment analysis*. Η ταξινόμηση τους έγινε με βάση την μέθοδο επιλογής και εξαγωγής χαρακτηριστικών κειμένου (π.χ. εύρεση της χ^2 στατιστικής συνάρτησης των λέξεων, μέθοδοι μετασχηματισμού χαρακτηριστικών όπως η PCA κ.α.), των τεχνικών κατηγοριοποίησης συναισθημάτων (χρήση μοντέλων μηχανικής μάθησης ή χρήση *lexicon based* μοντέλων) και του προβλήματος που επιλύουν (*emotion detection*, *transfer learning* κ.α.).

Συστήματα βασισμένα σε ομιλία

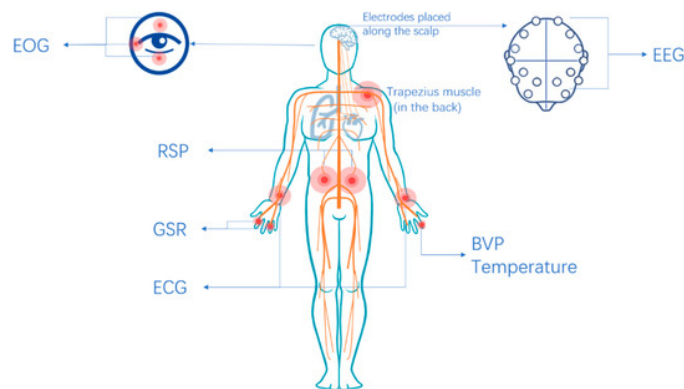
Ένας ακόμη τύπος δεδομένων που μπορεί να αξιοποιηθεί από τα υπολογιστικά μοντέλα για τον εντοπισμό συναισθημάτων, είναι η ομιλία. Οι πληροφορίες που μεταδίδονται και χρησιμεύουν στις εφαρμογές αναγνώρισης της συναισθηματικής κατάστασης του ομιλητή είναι δύο ειδών, η λεκτική και η μη-λεκτική. Οι λεκτικές πληροφορίες περιλαμβάνουν τις λέξεις και τις προφορικές προτάσεις. Οι μη λεκτικές πληροφορίες είναι εξίσου σημαντικές και έχουν έρθει στο επίκεντρο της έρευνας τα τελευταία χρόνια. Συγκεκριμένα, στις μη-λεκτικές πληροφορίες συμπεριλαμβάνονται μεταξύ άλλων, ο τόνος (η ποιότητα της φωνής, για παράδειγμα όταν κάποιο άτομο αισθάνεται την θλίψη μπορεί να υπάρχει αναστεναγμός στην φωνή), η συχνότητα (*pitch*) (η φωνή σε χαμηλές συχνότη-

τες μπορεί να υποδεικνύει θυμό, ενώ σε υψηλές χαρά, ενθουσιασμό ή φόβο), ο ρυθμός των λέξεων (συνήθως το συναίσθημα του θυμού προκαλεί γρήγορη παραγωγή λέξεων) και η ένταση της φωνής (για παράδειγμα η θλίψη συνοδεύεται από φωνή χαμηλής έντασης).

Οι Stuhlsatz *et al.* [71] πρότειναν ένα *Generalized Discriminant Analysis* (GerDA) δίκτυο που αποτελεί επέκταση της παραδοσιακής *Linear Discriminant Analysis* μεθόδου, για την εκμάθηση των σημαντικών χαρακτηριστικών των δεδομένων από πολλαπλές βάσεις. Κύριος στόχος του δικτύου είναι η εύρεση της βέλτιστης συνάρτησης από ένα σύνολο μη γραμμικών συναρτήσεων \mathcal{F} που συμβολίζεται με $f_{optimal} \in \mathcal{F} | \{f_{optimal} : \mathbb{R}^d \rightarrow \mathbb{R}^r\}$ και μετασχηματίζει τα δεδομένα πολλών διαστάσεων σε έναν υπο-χώρο μικρότερων διαστάσεων. Η βέλτιστη συνάρτηση $f_{optimal}$ προκύπτει μέσα από την εκπαίδευση ενός *Deep Neural Network* που μεγιστοποιεί το κριτήριο *Fisher* $Q_h(F) := \text{trace}\{\mathbf{S}_T^{-1}\mathbf{S}_B\}$ με \mathbf{S}_T η συνολική μήτρα διασποράς και \mathbf{S}_B η μήτρα διασποράς εντός των κλάσεων. Όταν ολοκληρώθηκε η εκπαίδευση του δικτύου, τα προκύπτοντα χαρακτηριστικά χρησιμοποιήθηκαν ως είσοδοι σε έναν ταξινομητή *minimum distance*. Η μέση ακρίβεια της μεθόδου για τις κλάσεις *valence*, *arousal* προέκυψε ίση με 79.5%.

Συστήματα βασισμένα σε φυσιολογικά σήματα

Τα προηγούμενα συστήματα εντοπισμού και ανάλυσης συναισθημάτων έχουν έναν περιορισμό, στο ότι η εκδήλωση των συναισθημάτων μπορεί να είναι παραπλανητική, δηλαδή οι άνθρωποι συχνά μεταποιούν εκούσια ή ακούσια την έκφραση του προσώπου, τις λέξεις και τα χαρακτηριστικά της φωνής τους και εξωτερικεύουν συναισθήματα που δεν αντικατοπτρίζουν τα πραγματικά. Επιπλέον, οι πολιτισμικοί και κοινωνικοί παράγοντες μπορεί να επηρεάσουν τον τρόπο έκφρασης των κοινών συναισθημάτων, πράγμα που σημαίνει ότι η φυσική εκδήλωση τους δεν είναι συνεπής μεταξύ των πολιτισμικών ομάδων. Γι' αυτούς του λόγους, πολλά συστήματα αναγνώρισης συναισθημάτων βασίζονται στις μετρήσεις φυσιολογικών σημάτων. Η σχέση μεταξύ των συναισθημάτων και των φυσιολογικών σημάτων, έχει ερευνηθεί εκτενώς από τον τομέα της ψυχοφυσιολογίας, αποκαλύπτοντας ότι τα συναισθήματα επηρεάζουν το κεντρικό και αυτόνομο νευρικό σύστημα των ανθρώπων. Επιπλέον, ερευνητές και ειδικοί του τομέα *human-computer interaction* έχουν σχεδιάσει και κατασκευάσει καινοτόμες συσκευές που χρησιμοποιούν ειδικούς αισθητήρες για την μέτρηση των φυσιολογικών αποκρίσεων και τοποθετούνται σε συγκεκριμένα μέρη του σώματος. Στην επόμενη εικόνα φαίνεται η πλαισίωση των αισθητήρων στο ανθρώπινο σώμα που μετρούν κάποια βασικά φυσιολογικά σήματα.



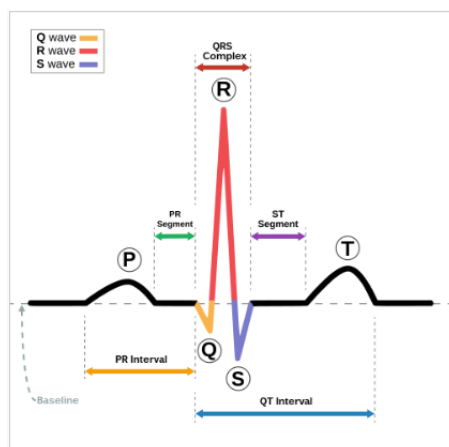
Σχήμα 4.1: Αναπαράσταση της τοποθέτησης των βιοαισθητήρων στο ανθρώπινο σώμα, για την μέτρηση διάφορων φυσιολογικών σημάτων. Πηγή:[67]

Το αυτόνομο νευρικό σύστημα αποτελεί το τμήμα του περιφερειακού νευρικού συστήματος που επηρεάζει διάφορες σωματικές παραμέτρους, όπως τον καρδιακό και αναπνευστικό ρυθμό, την αρτηριακή πίεση, την θερμοκρασία, κ.α. Οι παράμετροι αυτές, επηρεάζονται και μεταβάλλονται με τις αλλαγές των συναισθημάτων που βιώνουν οι οργανισμοί. Γι' αυτό, οι ερευνητές συχνά επιλέγουν παραμέτρους, όπως η μεταβλητότητα του καρδιακού ρυθμού (HRV), η ηλεκτροδυναμική δραστηριότητα (EDA), η θερμοκρασία και τα πρότυπα αναπνοής [24]. Όσον αφορά το κεντρικό νευρικό σύστημα, τα τμήματα του που επηρεάζονται από τα συναισθήματα, βρίσκονται στον εγκέφαλο (βλ. ενότητα 1.3) και στην ερευνητική κοινότητα έχουν διεξαχθεί εκτενείς μελέτες που ερευνούν τις συσχετίσεις της εγκεφαλικής δραστηριότητας με τα συναισθήματα. Στο υπόλοιπο κεφάλαιο, θα εμβαθύνουμε στις τεχνικές μέτρησης της φυσιολογικής απόκρισης του κεντρικού και του αυτόνομου νευρικού συστήματος, επισημαίνοντας σημαντικές μελέτες που έχουν διεξαχθεί στον τομέα ενδιαφέροντος.

ECG

Το ηλεκτροκαρδιογράφημα (ECG) είναι η διαδικασία καταγραφής της ηλεκτρικής δραστηριότητας που παράγεται από τους καρδιακούς παλμούς, όταν η καρδιά συστέλλεται και χαλαρώνει. Η καταγραφή γίνεται με την τοποθέτηση ηλεκτροδίων στο δέρμα και σε συγκεκριμένα σημεία του σώματος, όπως το στήθος, τα χέρια και τα πόδια. Το πρότυπο τοποθέτησης των ηλεκτροδίων που χρησιμοποιείται πιο συχνά, είναι γνωστό ως "12-lead". Κατά το πρότυπο αυτό, 10 ηλεκτρόδια τοποθετούνται με μη επεμβατικό τρόπο ένα σε κάθε χέρι, ένα στο δεξί πόδι και τα υπόλοιπα έξι στην επιφάνεια του θώρακα. Τα τρία ηλεκτρόδια των άκρων του σώματος συνδέονται μεταξύ τους δημιουργώντας 3 επιπλέον σήματα. Τα δώδεκα σήματα που προκύπτουν είναι γνωστά ως Lead I, Lead II, Lead III, aVR, aVL, aVF, V1, V2, V3, V4, V5, and V6 [57]. Τα σήματα ECG συχνά καταγράφονται και με συσκευές λιγότερων ηλεκτροδίων που είναι εύχρηστες και φορητές, όπως «έξυπνα ρολόγια» και οθόνες *Holter* τριών ηλεκτροδίων.

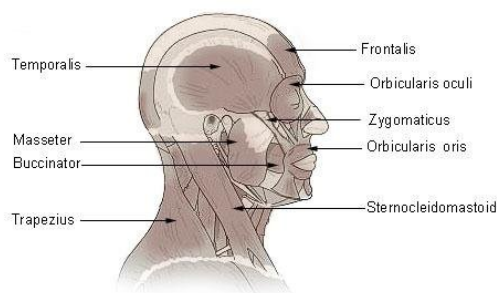
Στην δημοσίευση τους οι Y. L. Hsu *et al.* (2020) [29] συγκέντρωσαν 34 χαρακτηριστικά για καθένα από τα προ-επεξεργασμένα ECG δεδομένα διάρκειας ενός λεπτού. Τα χαρακτηριστικά αυτά προέρχονται από γραμμικές και μη γραμμικές αναλύσεις στον χρόνο και στη συχνότητα. Εκτός από το στάδιο της προεπεξεργασίας, το οποίο είναι πολύ σημαντικό για τα μετέπειτα στάδια, οι συγγραφείς αξιοποίησαν τον αλγόριθμο ανίχνευσης του συμπλέγματος QRS των ECG κυματομορφών, που προτάθηκε από τους Pan-Tompkins (1985)[54]. Ο αλγόριθμος αυτός εντοπίζει τις κορυφές R και τις αιχμές Q και S (βλ. σχήμα 4.2) κατά τον σχηματισμό του καρδιακού παλμού. Μέσα από αυτόν τον αλγόριθμο μπορούν να εξαχθούν οι παράμετροι RR intervals και HRV (μεταβλητότητα καρδιακού ρυθμού) των ECG δεδομένων, που παρέχουν μία καθαρή εικόνα σχετικά με τις αλλαγές του καρδιακού παλμού κατά την συναισθηματική διέγερση.



Σχήμα 4.2: Σχηματική αναπαράσταση ενός ιδανικού *sinus* παλμού που διαδίδεται από την ανθρώπινη καρδιά. Στην εικόνα φαίνεται το QRS σύμπλεγμα που αποτελείται από τα τρία τμήματα Q, R, S. Πηγή:[3]

EMG

Η ηλεκτρομυογραφία (EMG) είναι η τεχνική ανάλυσης των ηλεκτρικών διακυμάνσεων που παράγονται στους σκελετικούς μύες [68]. Η ηλεκτρική δραστηριότητα τους προέρχεται από τις εκούσιες ή ακούσιες μυϊκές συσπάσεις. Οι EMG καταγραφές χρησιμοποιούνται εκτενώς στους τομείς της κλινικής ιατρικής και της βιοϊατρικής για την ανίχνευση μη κανονικών λειτουργιών των μυών, συχνά όμως, αξιοποιούνται στον τομέα της αναγνώρισης συναισθημάτων, για την εύρεση συσχετίσεων μεταξύ της συναισθηματικής διέγερσης και των μυϊκών αντιδράσεων. Τα πιο συνήθη μυϊκά μέρη που καταγράφονται για τους σκοπούς της αναγνώρισης συναισθημάτων εντοπίζονται στο κεφάλι και στον αυχένα και είναι: οι κύριοι ζυγωματικοί μύες (*Zygomaticus Major*), ο βυκανήτης μύς (*buccinator*), ο σφιγκτήρας των φρυδιών (*Corrugator supercilii*), ο μετωπιαίος μύς (*Frontalis*), ο σφιγκτήρας των βλεφάρων (*Orbicularis oculi*), ο μασητήρας (*Masseter*), ο μύς του ώμου και του λαιμού (*Trapezius*) και ο στερνοκλειδομαστοειδής (*sternocleidomastoid*) (βλ. σχήμα 4.3). Η έρευνα που περιγράφει τις εξαρτήσεις μεταξύ συναισθημάτων και μυών προσώπου έγινε για πρώτη φορά από τους Ekman και Friesen το 1980 [25].



Σχήμα 4.3: Στην εικόνα φαίνονται οι μύες του κεφαλιού και του αυχένα που επιλέγονται για την EMG καταγραφή, λόγω της συσχέτισης τους με συναισθήματα που αποτυπώνονται στο χαμόγελο, το συνοφρύωμα και το σκίρτημα. Πηγή:[22]

Οι έρευνες που διεξάγονται συνήθως δεν βασίζονται ποτέ μόνο στα EMG δεδομένα, αλλά χρησιμοποιούν συνδυαστικά και κάποιες άλλες τεχνικές μέτρησης φυσιολογικών σημάτων, για τα

προβλήματα αναγνώρισης συναισθημάτων. Στην πράξη το ηλεκτρομυογράφημα ανιχνεύει αποτελεσματικά τα ηλεκτρικά σήματα που οφείλονται στα έντονα συναισθήματα, ενώ οι αλλαγές στην διακύμανση των EMG σημάτων είναι πολύ μικρές όταν ο παράγοντας valence είναι ασθενής. Στην δημοσίευση [50], οι συγγραφείς Nakasone *et al.* (2005) υλοποίησαν ένα Bayesian δίκτυο που ανιχνεύει και αναλύει τα ανθρώπινα συναισθήματα σε πραγματικό χρόνο. Το δίκτυο αυτό τροφοδοτείται από την σύντηξη δεδομένων ηλεκτρομυογραφίας και γαλβανικής απόκρισης δέρματος (GSR) ως είσοδο. Σύμφωνα με τα αποτελέσματα των πειραμάτων τους, το μοντέλο τους ανιχνεύει με επιτυχία τα διάφορα επίπεδα διέγερσης (arousal) που εντοπίζονται μέσω των μετρήσεων GSR, ενώ λιγότερο συχνά ανιχνεύει τα αρνητικά συναισθήματα σθένους (valence) που σχετίζονται με την μυϊκή δραστηριότητα.

Οι συγγραφείς J. Chen *et al.* (2022) [19] δημιούργησαν μία καινούρια βάση δεδομένων, την PME_4 , με σκοπό την ταξινόμηση των ανθρώπινων συναισθημάτων έντεκα υποκειμένων. Συγκεκριμένα, βιντεοσκοπήσαν τα πρόσωπα των συμμετεχόντων και κατέγραψαν την ηλεκτρική δραστηριότητα του εγκεφάλου (καταγραφές EEG) και των μυών του προσώπου (καταγραφές EMG). Για την δημιουργία της βάσης, ζητήθηκε από τους συμμετέχοντες να εκφέρουν μία τυχαία πρόταση «Ο ουρανός είναι πράσινος» και να προσποιηθούν παράλληλα ότι αισθάνονται επτά βασικά συναισθήματα. Εκτός από την δημιουργία της βάσης, ερεύνησαν και σύγκριναν διάφορες μεθόδους προ-επεξεργασίας δεδομένων, εξαγωγής χαρακτηριστικών και κατηγοριοποιητών. Αξίζει να αναφερθεί, ότι η χρήση της μεθόδου *bootstrap* για την μείωση του θορύβου στα δεδομένα EMG και EEG, συνέβαλλε στην βελτίωση της ακρίβειας κατά 20% για τα δεδομένα EEG και κατά 10% για τα EMG. Εν συντομία, η μέθοδος δημιουργεί νέα δεδομένα από τα ήδη υπάρχοντα, υπολογίζοντας τον μέσο όρο πολλαπλών δοκιμών του ίδιου συναισθήματος για κάθε άτομο.

EEG

Οι προηγούμενοι τύποι βιοσημάτων αποτελούν διεγέρσεις του αυτόνομου νευρικού συστήματος. Τα σήματα EMG και ECG συνδέονται περισσότερο με τις αντιδράσεις που προκαλούν τα συναισθηματικά ερεθίσματα, παρά με τις άμεσες πληροφορίες των συναισθημάτων. Σύμφωνα με τον τομέα της γνωστικής νευροεπιστήμης έχει αποδειχθεί ότι τα δεδομένα που προέρχονται από το κεντρικό νευρικό σύστημα και συγκεκριμένα από τον εγκέφαλο, συνδέονται στενά με την ρύθμιση και την επεξεργασία των συναισθημάτων. Ως εκ τούτου, εδώ και πολλές δεκαετίες, οι ερευνητές του τομέα της συναισθηματικής αναγνώρισης, συχνά επικεντρώνονται στην μελέτη των δεδομένων EEG και στην σύνδεση τους με τα συναισθήματα.

Οι Y. P. Lin *et al.* [43] κατέγραψαν τα δεδομένα EEG 26 συμμετεχόντων ενώ άκουγαν μουσική. Αρχικά πραγματοποιήθηκε οπτική επιθεώρηση των δεδομένων για τον εντοπισμό και την αφαίρεση του θορύβου κίνησης. Μη επικαλυπτόμενα *Hanning* παράθυρα ενός δευτερολέπτου εφαρμόστηκαν σε κάθε κανάλι πριν τον μετασχηματισμό των δεδομένων στο πεδίο της συχνότητας μέσω του *Short Time Fourier Transform* 512 σημείων. Από τις προκύπτουσες χρονικές σειρές του φάσματος, για την μελέτη επιλέχθηκαν τέσσερις τύποι χαρακτηριστικών, οι PSD_{30} , $DASM_{12}$, $RASM_{12}$, PSD_{24} . Τα χαρακτηριστικά αυτά επιλέχθηκαν με βάση τα 30 κανάλια EEG και τους μετασχηματισμούς τους, στο πεδίο συχνοτήτων. Οι συγγραφείς θέλησαν να συγκρίνουν και να εκτιμήσουν την απόδοση των τεσσάρων κατηγοριών χαρακτηριστικών, χρησιμοποιώντας MLP και SVM κατηγοριοποιητές. Η μελέτη έδειξε ότι τα χαρακτηριστικά $DASM_{12}$ είχαν κατά μέσο όρο την υψηλότερη ακρίβεια κατηγοριοποίησης στις 5 συχνοτικές ζώνες ($\alpha, \beta, \gamma, \delta, \theta$). Επιπλέον το δίκτυο SVM απέδωσε καλύτερα στην κατηγοριοποίηση των δεδομένων, σε σύγκριση με το δίκτυο MLP.

Ο W. Zheng [86] πρότεινε μία νέα στατιστική μέθοδο βασισμένη στην *Group Sparse Canonical Correlation Analysis (GSCCA)* μέθοδο, για την επιλογή των καταλληλότερων καναλιών EEG. Εάν $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_n]$ και $\mathbf{L} = [\mathbf{I}_1, \dots, \mathbf{I}_n]$ είναι οι πίνακες που περιέχουν όλα τα δοκιμαστικά *raw feature vectors* \mathbf{f}_t και τα διανύσματα ετικετών συναισθημάτων \mathbf{I}_t που αντιστοιχούν στα \mathbf{f}_t , η GSCCA μέθοδος θα βρει τους γραμμικούς συνδυασμούς των \mathbf{F}, \mathbf{L} που έχουν μέγιστη συσχέτιση. Εκτός από την αποτελεσματική συσχέτιση των συναισθημάτων και των EEG χαρακτηριστικών, η μέθοδος τους λαμβάνει υπόψιν και τις συνεισφορές διαφορετικών περιοχών του εγκεφάλου. Επιπλέον, σε αντίθεση με τις παραδοσιακές τεχνικές CCA, GSCCA, η μέθοδος που προτείνεται από τον W. Zheng εκτελείται χωρίς κάποιο στάδιο προ-επεξεργασίας.

Οι T. Song *et al.* [70] πρότειναν την εκπαίδευση ενός νέου δικτύου που ονομάζεται *Dynamical Graph Convolution Neural Network*. Το δίκτυο αυτό αποτελεί προέκταση των Convolutional Neural Networks και βασίζεται στην αναπαράσταση των EEG καναλιών στο πεδίο των γράφων. Σε αυτήν την αναπαράσταση κάθε κανάλι EEG αντιστοιχεί σε έναν κόμβο του γράφου και οι ακμές του γράφου αντιπροσωπεύουν τις χωρικές σχέσεις των καναλιών, δηλαδή τις φυσικές θέσεις στο σκαλπ του κεφαλιού. Οι συγγραφείς θεώρησαν απαραίτητη την μοντελοποίηση των χρονικών και δυναμικών αλληλεπιδράσεων μεταξύ των διαφορετικών περιοχών του εγκεφάλου (λειτουργικές σχέσεις), οι οποίες δεν μπορούν να αποτυπωθούν πλήρως από τις στατικές χωρικές πληροφορίες. Η παραπάνω λογική, διαμορφώνεται με μία μήτρα γειτνίασης, της οποίας οι τιμές προσαρμόζονται κατά την εκπαίδευση του δικτύου για την εκμάθηση των λειτουργικών σχέσεων των περιοχών του εγκεφάλου.

Οι Y. Li *et al.* [42] πρότειναν μία νέα μέθοδο ονόματι *R2G-STNN*. Η μέθοδος βασίζεται σε δι-κατευθυντήρια *long-short term memory* δίκτυα, που μαθαίνουν τις συσχετίσεις των εγκεφαλικών σημάτων και τις μεταβολές τους στον χρόνο, στα διάφορα τμήματα του εγκεφάλου. Η χωρική εκμάθηση των χαρακτηριστικών, διαιρείται ιεραρχικά σε δύο μέρη: αρχικά εκτελείται η χωρική εκμάθηση των σχέσεων των καναλιών που ανήκουν στην ίδια περιοχή του εγκεφάλου και στη συνέχεια, το δίκτυο εκμεταλλεύεται τις πληροφορίες αυτές για να μάθει τις σχέσεις των διαφορετικών περιοχών του εγκεφάλου. Επιπλέον, οι συγγραφείς εισήγαγαν και ένα επίπεδο *regional-attention*, για να μπορεί το δίκτυο να μαθαίνει τις διαφορετικές συνεισφορές των εγκεφαλικών περιοχών στην διάπλαση των συναισθημάτων.

Η παρούσα διπλωματική εργασία βασίστηκε σε μεγάλο βαθμό στην εργασία των Arjun *et al.* [11]. Οι συγγραφείς, παρουσίασαν μία νέα προσέγγιση στο πρόβλημα της κατηγοριοποίησης των συναισθημάτων από σήματα EEG, βασισμένη στην αρχιτεκτονική του *Vision Transformer*. Δημιούργησαν δύο παραλλαγές των δικτύων αυτών, μία που να δέχεται τα πρωτότυπα δεδομένα EEG και μία που τροφοδοτείται με τις δισδιάστατες φασματικές εικόνες τους ως είσοδο. Οι φασματικές εικόνες των δεδομένων EEG, δημιουργήθηκαν μέσω του *Continuous Wavelet Transform* με παράμετρο κλίμακας ίση με 48 και *mother wavelets* τις *db4*, *coif5*. Τα αποτελέσματα των πειραμάτων έδειξαν ότι το μοντέλο αποδίδει καλύτερα όταν τροφοδοτηθεί με τα ανεπεξέργαστα EEG δεδομένα. Συγκεκριμένα, η ακρίβεια του δικτύου ήταν βελτιωμένη κατά 2%.

Multimodal Systems

Πολλές έρευνες αξιοποιούν περισσότερες από μία πηγές σημάτων για βελτίωση της ακρίβειας και της αξιοπιστίας των συστημάτων αναγνώρισης συναισθημάτων. Τα συστήματα που χρησιμοποιούνται σε αυτές τις περιπτώσεις είναι «πολυτροπικά» (*multimodal*). Οι Soleymani *et al.* [69], παρουσίασαν μία προσέγγιση βασισμένη στον συνδυασμό EEG καταγραφών και εκφράσεων προσώπου. Σύμφωνα με το πείραμα τους, οι συναισθηματικές πληροφορίες των EEG σημάτων προέρχονται σε

μεγάλο βαθμό από τις μυϊκές συσπάσεις του προσώπου των ατόμων, αλλά εξακολουθούν να δίνουν περαιτέρω πληροφορίες για την συναισθηματική τους κατάσταση. Οι N. Jia *et al.* [5] αξιοποιούν δεδομένα ήχου, βίντεο και σύλληψης κίνησης (MoCAP) από τα οποία εκχαιεύουν χαρακτηριστικά συναισθημάτων μέσω νευρωνικών δικτύων με *attention* μηχανισμούς.

Οι K. H. Kim *et al.* [34] εκμεταλλεύτηκαν 3 μεθόδους καταγραφής φυσιολογικών σημάτων του αυτόνομου νευρικού συστήματος: το ηλεκτροκαρδιογράφημα (ECG), την διακύμανση της θερμοκρασίας του δέρματος (SKT) και την ηλεκτροδυναμική δραστηριότητα (EDA). Η ακρίβεια της κατηγοριοποίησης των σημάτων αυτών σε συστήματα τριών και τεσσάρων κατηγοριών ήταν ίση με 78.4% και 61.8%. Ενδιαφέρουσα εργασία ήταν αυτή των J. S. Leventont *et al.* [41], όπου εξέτασαν την σύνδεση των συναισθηματικών αποκρίσεων με το φαινόμενο της επακόλουθης μνήμης σε παιδιά σχολικής ηλικίας 5-8 χρονών. Για την μελέτη χρησιμοποίησαν EEG, Respiration και ECG καταγραφές. Τα ευρήματα τους κατέδειξαν ότι η μνήμη των μικρότερων παιδιών επηρεάζεται λιγότερο από τις συναισθηματικές εμπειρίες.

4.0.3 Σύγκριση των μεθόδων & Συμπεράσματα

Έχοντας ερευνήσει την υπάρχουσα βιβλιογραφία στον τομέα της αναγνώρισης συναισθημάτων, θα καταλάβει κανείς ότι η επιλογή των κατάλληλων σημάτων ποικίλει ανάλογα με την εφαρμογή και τον στόχο της υλοποίησης. Η χρήση των οπτικό-ακουστικών σημάτων έχει το πλεονέκτημα της εύκολης, ασφαλούς και λιγότερο δαπανηρής συλλογής δεδομένων, καθιστώντας την μακροχρόνια καταγραφή των δεδομένων λιγότερο κουραστική. Όμως, οι πληροφορίες των μη φυσιολογικών σημάτων είναι υποκειμενικές και συχνά επηρεάζονται από παράγοντες όπως τον πολιτισμό, την κουλτούρα και το εξωτερικό πλαίσιο των ανθρώπων. Παράλληλα με την ανάπτυξη των οπτικό-ακουστικών μέσων καταγραφής, η ραγδαία ανάπτυξη της τεχνολογίας, έχει συνεισφέρει σημαντικά στην δημιουργία και βελτίωση των συσκευών καταγραφής βιολογικών σημάτων. Κύριο πλεονέκτημα των συστημάτων που αξιοποιούν τα σήματα αυτού του είδους, είναι ότι παρέχουν ακριβείς πληροφορίες που δεν επηρεάζονται από εξωτερικούς παράγοντες. Επιπλέον, έχει αποδειχθεί ότι τα φυσιολογικά σήματα παρέχουν καλή απόδοση και σε μερικές περιπτώσεις ξεπερνάνε την απόδοση των γλωσσικών ή συμπεριφορικών σημάτων. Ωστόσο, παρά τα πλεονεκτήματά τους, έχουν ακόμη και πολλά μειονεκτήματα που πρέπει να ληφθούν υπόψιν και να αντιμετωπιστούν κατάλληλα. Μερικά από αυτά αναλύονται στη συνέχεια:

- **Αξιοποίηση εξειδικευμένου εξοπλισμού:** Η καταγραφή των φυσιολογικών σημάτων μπορεί να γίνει μόνο από εξειδικευμένες συσκευές, που συχνά απαιτούν μεγάλο κόστος και χρόνο προετοιμασίας.
- **Επίδραση εξωτερικών παρεμβολών:** Τα φυσιολογικά σήματα συχνά επηρεάζονται από άλλες πηγές όπως σωματικές κινήσεις, μυϊκές συσπάσεις κ.α. Επιπλέον, οι μη επεμβατικές μέθοδοι συχνά επηρεάζονται από εξωτερικό θόρυβο.
- **Πολυπλοκότητα των δεδομένων:** Η ερμηνεία των φυσιολογικών δεδομένων είναι δύσκολη και απαιτείται μεγάλος όγκος δεδομένων για να γίνει αξιόπιστη εκχαιεύση πληροφοριών από τα υπολογιστικά συστήματα

- **Δυσκολία γενίκευσης των αποτελεσμάτων:** Οι βιολογικές διεργασίες ποικίλουν από άνθρωπο σε άνθρωπο και γι' αυτό η γενίκευση των ευρημάτων και η σύγκριση τους είναι δύσκολη.

Κεφάλαιο 5

Ανάλυση και Σχεδιασμός

Το πρόβλημα που καλείται να επιλύσει η παρούσα διπλωματική αφορά την ανάλυση προ-καταγεγραμμένων σημάτων EEG και τον εντοπισμό των μοτίβων των δεδομένων που σχετίζονται με την συναισθηματική κατάσταση των ανθρώπων το χρονικό διάστημα της καταγραφής. Η προσέγγιση, στην οποία βασιστήκαμε για να αντιμετωπίσουμε το πρόβλημα, περιλαμβάνει *end to end* μεθόδους βαθιάς μάθησης, οι οποίες λειτουργούν σαν *black boxes* και αυτοματοποιούν τις διαδικασίες του εντοπισμού των σχετικών μοτίβων από τα δεδομένα. Απ' όλες τις μελέτες που υπάρχουν στην ερευνητική κοινότητα, στα πλαίσια της συγκεκριμένης διπλωματικής αξιολογήσαμε την απόδοση των CNN, LSTM και Transformer δικτύων, με παραπάνω εστίαση στο μοντέλο του Vision Transformer. στη συνέχεια, θα ακολουθήσει ανάλυση πάνω στα κομβικά στάδια διεξαγωγής των πειραμάτων, τα οποία είναι απαραίτητα για την επιτυχή διεξαγωγή της κατηγοριοποίησης των σημάτων EEG σε συναισθήματα.

5.1 Προ-επεξεργασία Δεδομένων

Τα σήματα που καταγράφουν οι συσκευές EEG είναι κατά μεγάλο βαθμό επηρεασμένα από θόρυβο και παρεμβολές άλλων σημάτων. Οι πηγές θορύβου μπορεί να προέρχονται είτε από το εξωτερικό περιβάλλον, είτε από εσωτερικά σήματα που παρεμβάλλονται κατά την καταγραφή. Έτσι, τα σήματα EEG συνήθως περιέχουν επιπλέον πληροφορίες που επικαλύπτουν την δυναμική των εγκεφαλικών σημάτων. Για την αφαίρεση του θορύβου και των παρεμβολών, υπάρχουν απλές τεχνικές, όπως τα βαθυπερατά, υψιπερατά, ζωνοδιαβατά και απόρριψης ζώνης φίλτρα. Κάποιες από αυτές τις τεχνικές εφαρμόζονται στα πεδία του χρόνου και των συχνοτήτων, ενώ κάποιες άλλες ασχολούνται με την χωρική ανάλυση των δεδομένων. Το κεφάλαιο αυτό θα αφιερωθεί στην ανάλυση κάποιων μεθόδων προεπεξεργασίας και αποθορυβοποίησης που χρησιμοποιούνται στις εργασίες αναγνώρισης συναισθημάτων από σήματα EEG.

Υπό-δειγματοληψία

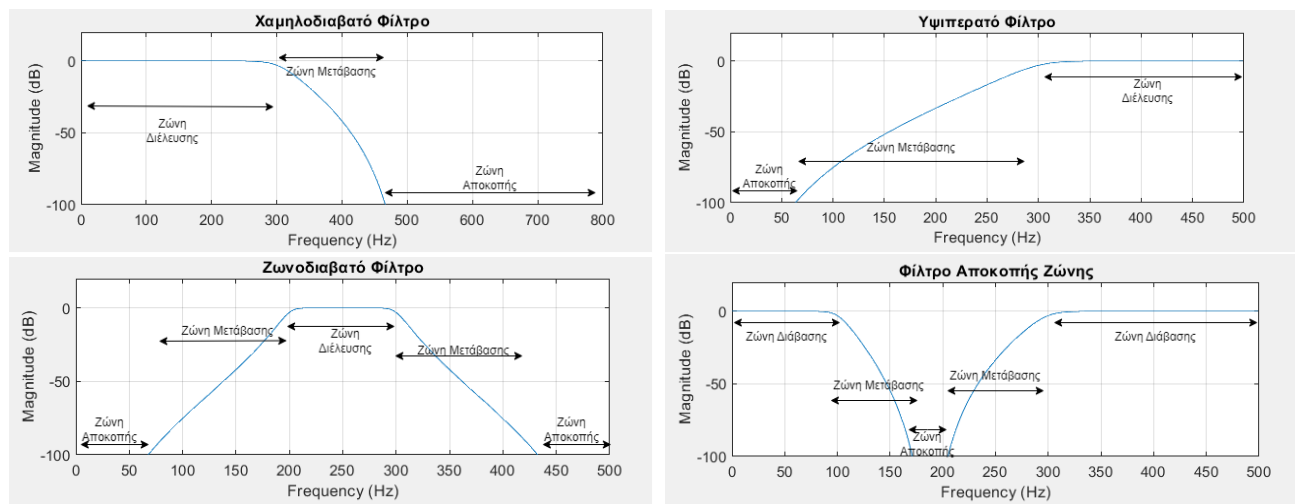
Το στάδιο της υποδειγματοληψίας συνήθως είναι το πρώτο στάδιο της προετοιμασίας των δεδομένων EEG και περιλαμβάνει την προσεκτική μείωση του μεγέθους των δεδομένων, έτσι ώστε να μην χανθούν πολύτιμες πληροφορίες. Στον τομέα που μας ενδιαφέρει, η τεχνική μείωσης των δεδομένων που χρησιμοποιείται πιο συχνά ονομάζεται *αποδεκατισμός* (*decimation*). Το στάδιο του αποδεκατισμού εφαρμόζεται διατηρώντας κάθε n -οστό δείγμα του καταγεγραμμένου σήματος. Ωστόσο, είναι σημαντικό να σημειωθεί ότι πρέπει να ληφθεί μέριμνα ώστε να αποφευχθεί το φαινόμενο *αναδίπλωσης της συχνότητας* (*frequency aliasing*) κατά την εφαρμογή του αποδεκατισμού.

Για να καταλάβουμε καλύτερα το φαινόμενο αυτό, θα συμβολίσουμε την καινούρια συχνότητα δειγματοληψίας του σήματος ως f_s . Εάν το σήμα έχει στοιχεία με συχνότητες μεγαλύτερες από $\frac{f_s}{2}$ (ρυθμός *Nyquist*), στο νέο σήμα, οι υψηλές συχνότητες αναδιπλώνονται και εμφανίζονται ως χαμηλές. Για να αποφευχθεί το πρόβλημα, ο αποδεκατισμός συχνά συνοδεύεται από την χρήση

φίλτρων κατά της παραμόρφωσης αυτής. Τα φίλτρα αυτά είναι χαμηλοδιαβατά και καταστέλλουν το πρόβλημα της αναδίπλωσης σε αποδεκτά επίπεδα. Το χαμηλοδιαβατό φίλτρο, πρέπει να εφαρμοστεί πολύ προσεκτικά, καθώς χρήσιμες πληροφορίες των συναισθημάτων αποτυπώνονται στις υψηλές συχνότητες των δεδομένων και λανθασμένη επιλογή φίλτρου, μπορεί να συμβάλλει στην απώλεια των πληροφοριών αυτών.

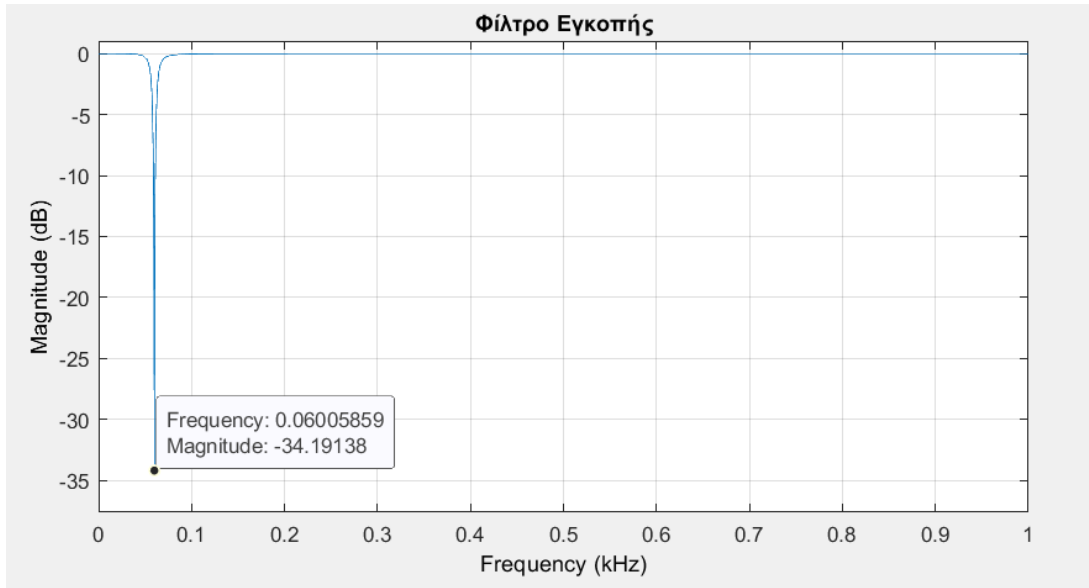
Φιλτράρισμα στο πεδίο συχνοτήτων

Η επεξεργασία των σημάτων EEG στο πεδίο των συχνοτήτων, εφαρμόζεται συχνά και στοχεύει στην διατήρηση και ενίσχυση των χρήσιμων συχνοτικών περιοχών του εγκεφάλου. Η επιλογή του φίλτρου γίνεται με βάση το συχνοτικό εύρος των σημάτων EEG, το οποίο μπορεί να διαφέρει ανάλογα με την συσκευή που έχει χρησιμοποιηθεί. Τα πιο συνηθισμένα φίλτρα είναι το: χαμηλοπερατό, υψιπερατό, ζωνοδιαβατό και αποκοπής ζώνης. Οι συναρτήσεις μεταφοράς των φίλτρων αυτών φαίνονται στην εικόνα 5.1. Το χαμηλοπερατό φίλτρο επιτρέπει την διέλευση των χαμηλών συχνοτήτων, ενώ αποκόπει τις υψηλές συχνότητες που βρίσκονται πάνω από την συχνότητα αποκοπής. Το υψιπερατό φίλτρο εκτελεί την αντίστροφη διαδικασία του χαμηλοπερατού. Το ζωνοδιαβατό φίλτρο, επιτρέπει την διέλευση μίας συγκεκριμένης μπάντας συχνοτήτων και το φίλτρο απόρριψης ζώνης εμποδίζει την διέλευση κάποιων συχνοτικών ζωνών.



Σχήμα 5.1: Οι συναρτήσεις μεταφοράς τεσσάρων μη ιδανικών φίλτρων. Στην εικόνα επισημαίνονται και οι ζώνες διέλευσης, μετάβασης και αποκοπής των φίλτρων.

Συχνά, οι συσκευές που καταγράφουν τα δεδομένα EEG, διατηρούν και πληροφορίες που βρίσκονται έξω από τις ζώνες ενδιαφέροντος και δεν συνδέονται με τις συναισθηματικές πληροφορίες. Σύμφωνα με την δημοσίευση [7] των S. M. Alarcão *et al.*, περίπου το 84% των ερευνών που μελέτησαν χρησιμοποιούν ζωνοδιαβατά φίλτρα. Ένα ακόμη φίλτρο που χρησιμοποιείται συχνά, είναι το φίλτρο εγκοπής (*notch filter*). Το φίλτρο εγκοπής χρησιμοποιείται για την αφαίρεση θορύβου που δημιουργείται από καλώδια εναλλασσόμενου ρεύματος. Συγκεκριμένα, υποβαθμίζουν μεμονωμένες συχνότητες (κοντά στα 50 ή 60 Hz) στις οποίες βρίσκεται ο θόρυβος αυτού του τύπου. Ενδεικτικά, η απόκριση συχνότητας ενός φίλτρου εγκοπής φαίνεται στην εικόνα 5.2



Σχήμα 5.2: Η απόκριση πλάτους του φίλτρου εγκοπής. Το φίλτρο υποβαθμίζει την συχνότητα των 60 Hz.

Αναφορά Κοινού Μέσου Όρου (Common Average Referencing-CAR)

Η μέθοδος CAR είναι απ' τις πιο συνηθισμένες μεθόδους χωρικού φιλτραρίσματος στα συστήματα BCI. Η μέθοδος αυτή υλοποιείται μέσω της εξίσωσης 5.1, όπου ο μέσος όρος των σημάτων όλων των ηλεκτροδίων αφαιρείται από τα σήματα καθενός ηλεκτροδίου ξεχωριστά [72]. Με την μέθοδο CAR ακυρώνεται ο εξωτερικός θόρυβος και ο θόρυβος από τα ηλεκτρικά δυναμικά του τριχωτού της κεφαλής, ο οποίος διεισδύει σε όλα τα ηλεκτρόδια.

$$x_i^{CAR} = x_i^{ER} - \frac{1}{n} \sum_{j=1}^n x_j^{ER}, i = 1 \dots n, j \neq i \quad (5.1)$$

Ως x_i^{ER} συμβολίζουμε την διαφορά δυναμικού μεταξύ του i ηλεκτροδίου και του ηλεκτροδίου αναφοράς και n είναι ο συνολικός αριθμός των καναλιών. Παραλλαγή της μεθόδου, αποτελεί η μέθοδος *mean average referencing* (MAR), όπου ο συνολικός μέσος όρος αντικαθίσταται απ' τον μέσο όρο των δυναμικών μίας ομάδας ηλεκτροδίων.

Στην ιδανική περίπτωση που τα ηλεκτρόδια της συσκευής EEG καλύπτουν όλο το μήκος του κεφαλιού και οι ηλεκτρικές πηγές του εγκεφάλου είναι σημειακές, η συνολική χωρική τάση που θα προκύψει μετά την εφαρμογή του φίλτρου θα έχει μέση τιμή μηδέν. Σε αυτήν την περίπτωση, τυχόν θόρυβος που θα βρισκόταν σε όλα τα κανάλια θα παραλείπονταν εξ' ολοκλήρου. Αυτό όμως δεν συμβαίνει στην πράξη και ως αποτέλεσμα μπορεί να υπάρχει θόρυβος σε ορισμένα κανάλια, ακόμη και μετά την εφαρμογή του φίλτρου CAR. Το φίλτρο μπορεί επίσης να εισάγει πρόσθετο θόρυβο σε ορισμένα κανάλια, εάν ο θόρυβος δεν κατανέμεται εξίσου σε όλες τις πηγές σήματος.

Laplacian Φίλτρο (LAP)

Το φίλτρο LAP, λειτουργεί ως χωρικό φίλτρο και φιλτράρει τις χαμηλές συχνότητες που περιέχουν απομακρυσμένο θόρυβο από εξωτερικές πηγές, δίνοντας έμφαση στις υψηλές συχνότητες του σήματος, δηλαδή στις κοντινές πηγές. Επιπλέον, χρησιμεύει στην μείωση του θορύβου που διαχέεται από τα γειτονικά ηλεκτρόδια και φτάνει στα ηλεκτρόδια ενδιαφέροντος. Για κάθε κανάλι,

υπολογίζει την δεύτερη παράγωγο της στιγμιαίας χωρικής κατανομής τάσης και τονίζει την δραστηριότητα των πηγών που βρίσκονται ακτινικά κάτω από το κάθε ηλεκτρόδιο [72]. Παρακάτω, φαίνεται η σχέση που υλοποιεί το συγκεκριμένο φίλτρο στην πράξη:

$$x_i^{LAP} = x_i^{ER} - \sum_{j \in S_i}^n x_j^{ER} g_{ij} \quad (5.2)$$

Όπου $g_{i,j} = \frac{1}{\sum_{j \in S_i}^n \frac{1}{d_{ij}}}$, d_{ij} είναι η απόσταση του i και j καναλιού και S_i είναι η ομάδα των γειτονικών καναλιού που συνορεύουν με το κανάλι i . Η τιμή LAP κάθε καναλιού προκύπτει ως η διαφορά μεταξύ της τιμής του καναλιού και του σταθμισμένου μέσου όρου των πιο κοντινών καναλιών. Όσο η παράμετρος d_{ij} της απόστασης των καναλιών μειώνεται, το LAP φίλτρο γίνεται όλο και πιο ευαίσθητο στον θόρυβο που προέρχεται από κοντινά κανάλια.

Τυφλός Διαχωρισμός Πηγής

Τα πρωτότυπα σήματα EEG συνήθως αποτελούνται από πολλά, διάχυτα σήματα που παράγονται από διάφορες πηγές, όπως από τις κινήσεις και το ανοιγόκλεισμα των ματιών, τις μυϊκές δραστηριότητες, τον καρδιακό ρυθμό κ.α. Ο ακριβής προσδιορισμός των σημάτων που αντιστοιχούν σε συγκεκριμένες πηγές δραστηριότητας, είναι πολύ δύσκολος και μπορεί να οδηγήσει σε σφάλματα, επηρεάζοντας αρνητικά την ερμηνεία των αποτελεσμάτων κατά την ανάλυση των δεδομένων EEG. Για να ξεπεραστεί αυτό το πρόβλημα, τα πρωτότυπα σήματα EEG υφίστανται μία ειδική επεξεργασία, που ονομάζεται *blind source separation* (BSS). Η τεχνική BSS περιλαμβάνει τον διαχωρισμό των αρχικών σημάτων από το σύνολο των διάχυτων, χωρίς να απαιτείται εκτενής γνώση σχετικά με τις πηγές ή τον τρόπο συνδυασμού τους. Στη συνέχεια ακολουθεί η μαθηματική αναπαράσταση του προβλήματος και της τεχνικής BSS.

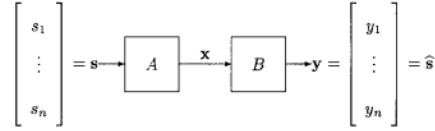
Έστω $\mathbf{s}(t) = [s_1(t) \dots s_n(t)]^T \in \mathbb{R}^{n \times 1}$ τα ανεξάρτητα σήματα διάφορων πηγών και $\mathbf{A} \in \mathbb{R}^{n \times n}$ ένας πίνακας που προκαλεί το «μπέρδεμα» των επιμέρους πηγών και δημιουργεί τα καινούρια σήματα $\mathbf{x}(t) = (x_1(t), \dots, x_n(t))^T \in \mathbb{R}^{n \times n}$. Δηλαδή ισχύει η σχέση:

$$\mathbf{x}(t) = \mathbf{A} \cdot \mathbf{s}(t) \quad (5.3)$$

Η τεχνική BSS στοχεύει στην ανάκτηση των αρχικών σημάτων των $\mathbf{s}(t)$, με βάση τα παρατηρούμενα δεδομένα $\mathbf{x}(t)$, την υπόθεση της ανεξαρτησίας των επιμέρους $s_i(t)$ και ίσως κάποιες πληροφορίες σχετικά με την κατανομή πιθανοτήτων των $s_i(t)$. Η σχέση στην οποία βασίζεται η μέθοδος του τυφλού διαχωρισμού πηγής είναι:

$$\mathbf{y}(t) = \mathbf{B} \cdot \mathbf{x}(t) \quad (5.4)$$

Όπου $\mathbf{B} \in \mathbb{R}^{n \times n}$ ο πίνακας διαχωρισμού των πηγών και $\mathbf{y}(t)$ μια εκτίμηση των ανεξάρτητων πηγών $\mathbf{s}(t)$. Στο σχήμα 5.3 παρουσιάζεται το διάγραμμα ροής του αλγόριθμου *blind source separation*.



Σχήμα 5.3: Το διάγραμμα ροής του *blind source separation* αλγόριθμου. Με \mathbf{s} συμβολίζονται οι επιμέρους διαχωρισμένες πηγές, με \mathbf{x} τα «μικτά» σήματα και με \mathbf{y} οι εκτιμήσεις των πηγών. Πηγή:[18]

Η προσαρμογή του πίνακα \mathbf{B} μπορεί να γίνει μέσα από πολλές μεθόδους. Η πιο δημοφιλής μέθοδος είναι η ανάλυση των ανεξάρτητων συνιστωσών (*Independent Component Analysis*), η οποία στοχεύει στην αποσύνθεση του μικτού σήματος σε ένα άθροισμα ανεξάρτητων σημάτων.

5.2 Μοντέλο

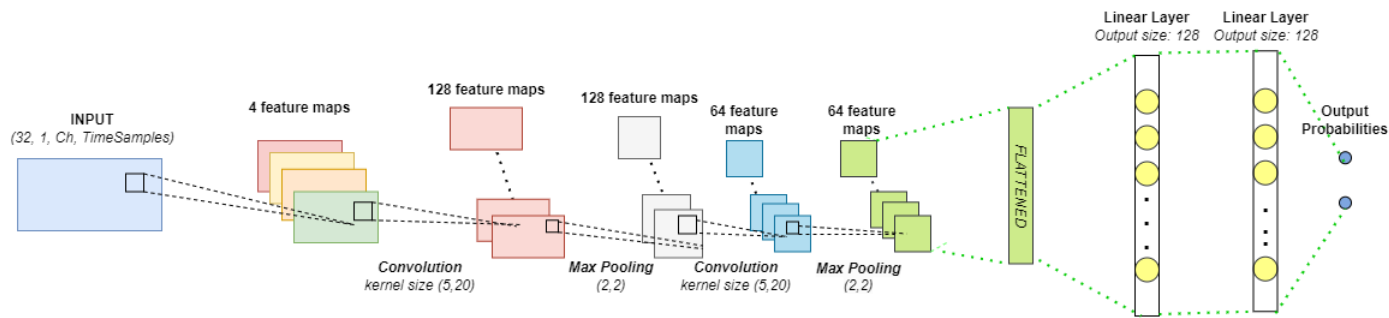
Κατά την διάρκεια της διπλωματικής, ερευνήσαμε και αναπτύξαμε τρία διαφορετικά μοντέλα βαθιών νευρωνικών δικτύων: ένα συνελικτικό νευρωνικό δίκτυο (CNN), ένα αναδρομικό δίκτυο μακροπρόθεσμης-βραχυπρόθεσμης μνήμης (LSTM) και ένα δίκτυο Vision Transformer. Μεταξύ αυτών των τριών δικτύων, δόθηκε μεγαλύτερη έμφαση στο μοντέλο του μετασχηματιστή, τόσο λόγω της πολύπλοκης και ενδιαφέρουσας δομής του, όσο και της μεγάλης προσοχής που έχει λάβει από την ερευνητική κοινότητα τα τελευταία χρόνια. Στη συνέχεια ακολουθεί ανάλυση πάνω στην αρχιτεκτονική των μοντέλων που διεξάχθηκε προκειμένου να συγκριθεί η απόδοση και η ικανότητα τους να κατηγοριοποιούν ορθά τα δεδομένα EEG σε συναισθήματα.

5.2.1 Αρχιτεκτονική

Τα αποτελέσματα της βιβλιογραφικής ανασκόπησης, έδειξαν ότι η αρχιτεκτονική των δικτύων που επιλέγει να υλοποιήσει ο κάθε συγγραφέας των μελετώμενων δημοσιεύσεων, εξαρτάται στενά από τη μορφή των δεδομένων με τα οποία εργάζεται. Συγκεκριμένα, ο τύπος του δικτύου που εφαρμόζεται, καθορίζεται από τον στόχο της έρευνας, αλλά και τα χαρακτηριστικά των δεδομένων που μοντελοποιούνται από τα δίκτυα. Επιπλέον, λαμβάνοντας υπόψη τη μορφή των δεδομένων EEG, εναπόκειται σε κάθε ερευνητή η επιλογή του κατάλληλου πεδίου που θα αναπαραστήσει τα δεδομένα (δηλαδή το πεδίο του χρόνου, του χώρου ή/και της συχνότητας) και στη συνέχεια η κατάλληλη προσαρμογή της αρχιτεκτονικής των δικτύων.

Συνελικτικό Νευρωνικό Δίκτυο

Για την υλοποίηση του δικτύου CNN, βασίστηκε στην αρχιτεκτονική που προτείνουν οι ερευνητές J. X. Chen *et al.* στην δημοσίευση [20]. Το μοντέλο παίρνει ως είσοδο ένα υποσύνολο σημάτων EEG με τη μορφή ενός τετραδιάστατου τανιστή με διαστάσεις $B \times C \times H \times N$. Η διάσταση B αντιπροσωπεύει το *batch size*, η διάσταση C τον αριθμό των καναλιών εισόδου, η διάσταση H τα ηλεκτρόδια EEG και η διάσταση N τα χρονικά δείγματα. Στην συγκεκριμένη υλοποίηση, η παράμετρος εισόδου C θα είναι ίση με 1, ενώ οι τιμές των υπόλοιπων παραμέτρων εξαρτώνται από τα δεδομένα. Η έξοδος του δικτύου, θα δημιουργήσει τις πιθανότητες κλάσεων των δεδομένων, οι οποίες σε αυτήν την περίπτωση είναι 2, επειδή τα δεδομένα ταξινομούνται σε χαμηλό/υψηλό σθένος και χαμηλή/υψηλή διέγερση. Στην εικόνα 5.4 φαίνεται το διάγραμμα του δικτύου με όλα τα *layers* που υλοποιεί.

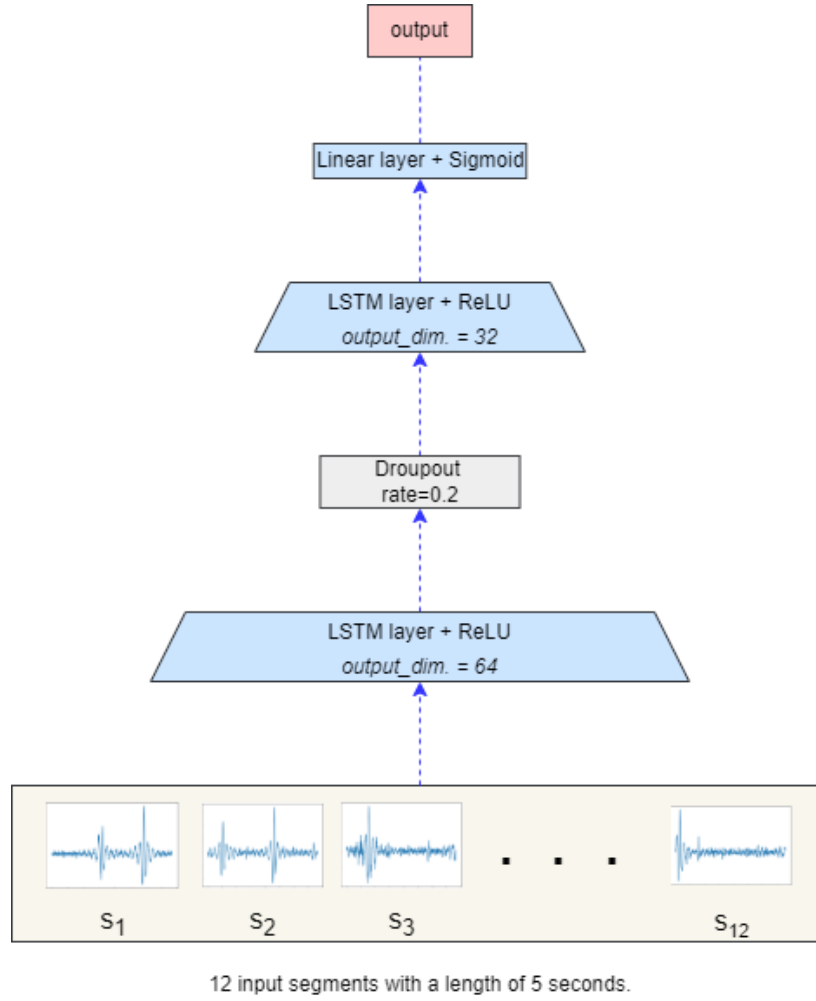


Σχήμα 5.4: Επισκόπηση του deep CNN μοντέλου. Στις ενδιάμεσες εξόδους των στρώματων εφαρμόστηκε η συνάρτηση ενεργοποίησης ReLU.

Το δίκτυο έχει τρία συνελκτικά στρώματα και δύο στρώματα μέγιστης συγκέντρωσης για την εξαγωγή χαρακτηριστικών. Αυτά τα χαρακτηριστικά στη συνέχεια τροφοδοτούνται σε δύο πλήρως συνδεδεμένα επίπεδα για κατηγοριοποίηση. Μετά την έξοδο του δεύτερου πλήρως συνδεδεμένου γραμμικού επιπέδου, εφαρμόζεται ένα *dropout* επίπεδο με *dropout rate* ίσο με 85% για την αποφυγή του *overfitting*. Η τελική έξοδος κανονικοποιείται μέσω μιας συνάρτησης ενεργοποίησης Softmax, η οποία παράγει τις πιθανότητες των κλάσεων και επιτρέπει στο μοντέλο να εκχωρήσει μία κλάση ως έξοδο.

Δίκτυο LSTM

Η αρχιτεκτονική του δικτύου LSTM που υλοποιήσαμε, φαίνεται αναλυτικά στην εικόνα 5.5. Τα δεδομένα EEG που χρησιμοποιήθηκαν σε αυτήν την υλοποίηση ελήφθησαν από την βάση δεδομένων DEAP. Τα δεδομένα αυτά, αποτελούνται από 32 κανάλια και 7068 data samples που αντιστοιχούν σε χρονικό διάστημα 60sec.. Προκειμένου να διοχετευθούν στην είσοδο του LSTM δικτύου, σύμφωνα με τους S. Alhagry *et al.* [8] έπρεπε να αναπροσαρμοστούν οι διαστάσεις των δεδομένων σε 1D αναπαράσταση. Έτσι, η διάσταση της εισόδου αναπροσαρμόστηκε σε $B \times L \times D$, όπου L είναι το sequence length που επιλέχθηκε να είναι ίσο με 12 και $D = \frac{Channels \times TimeSamples}{L} = \frac{32 \times 7068}{12}$ το μέγεθος των δειγμάτων.



Σχήμα 5.5: Δομική αναπαράσταση της LSTM αρχιτεκτονικής. Το δίκτυο δέχεται τα τμηματικά *flattened* δεδομένα EEG.

Δίκτυο Vision Transformer

Για την υλοποίηση του vision transformer, στηριχθήκαμε στις αρχές των Arjun *et al.* που περιγράφουν στην δημοσίευσή τους [11]. Τα βήματα που ακολουθούν οι συγγραφείς, από την αναδιαμόρφωση των δεδομένων έως την έξοδο του δικτύου, θα περιγραφούν στην επόμενη παράγραφο.

Υπό μαθηματικούς όρους, τα δεδομένα EEG εισόδου συμβολίζονται με $x \in \mathbb{R}^{H \times C \times W}$. Αρχικά τα δεδομένα αυτά, χωρίζονται σε *patches* και λαμβάνουν τη μορφή $x \in \mathbb{R}^{N \times (P^2 \times C)}$, όπου $N = \frac{H \times W}{P^2}$ και P το μέγεθος του κάθε patch. Τα patches μετατρέπονται σε τελικές ενσωματώσεις z_0 μέσω ενός εκπαιδευσιμου στρώματος γραμμικής προβολής, πριν εισαχθούν στον μετασχηματιστή. Για να ενισχυθεί ακόμη περισσότερο το πληροφοριακό υπόβαθρο του δικτύου, τα *patch embeddings* εμπλουτίζονται περαιτέρω από ένα «εκπαιδευσιμο» *class token* και τα *positional embeddings* που υποδεικνύουν την θέση των *tokens* εντός της ακολουθίας. Η αρχιτεκτονική του μετασχηματιστή φαίνεται στην εικόνα 5.6. Αποτελείται από επαναλαμβανόμενα μπλοκ *Multiheaded Self-Attention* (MSA) και δίκτυα MLP με συνάρτηση ενεργοποίησης την *Gaussian Error Linear Unit* (GELU). Η έξοδος κάθε μπλοκ κανονικοποιείται μέσω του *layer normalization* και συνδέεται με *residual connections*.

Οι Arjun *et al.* χρησιμοποιούν *feed forward* στάδια με δύο γραμμικά επίπεδα, που παράγουν

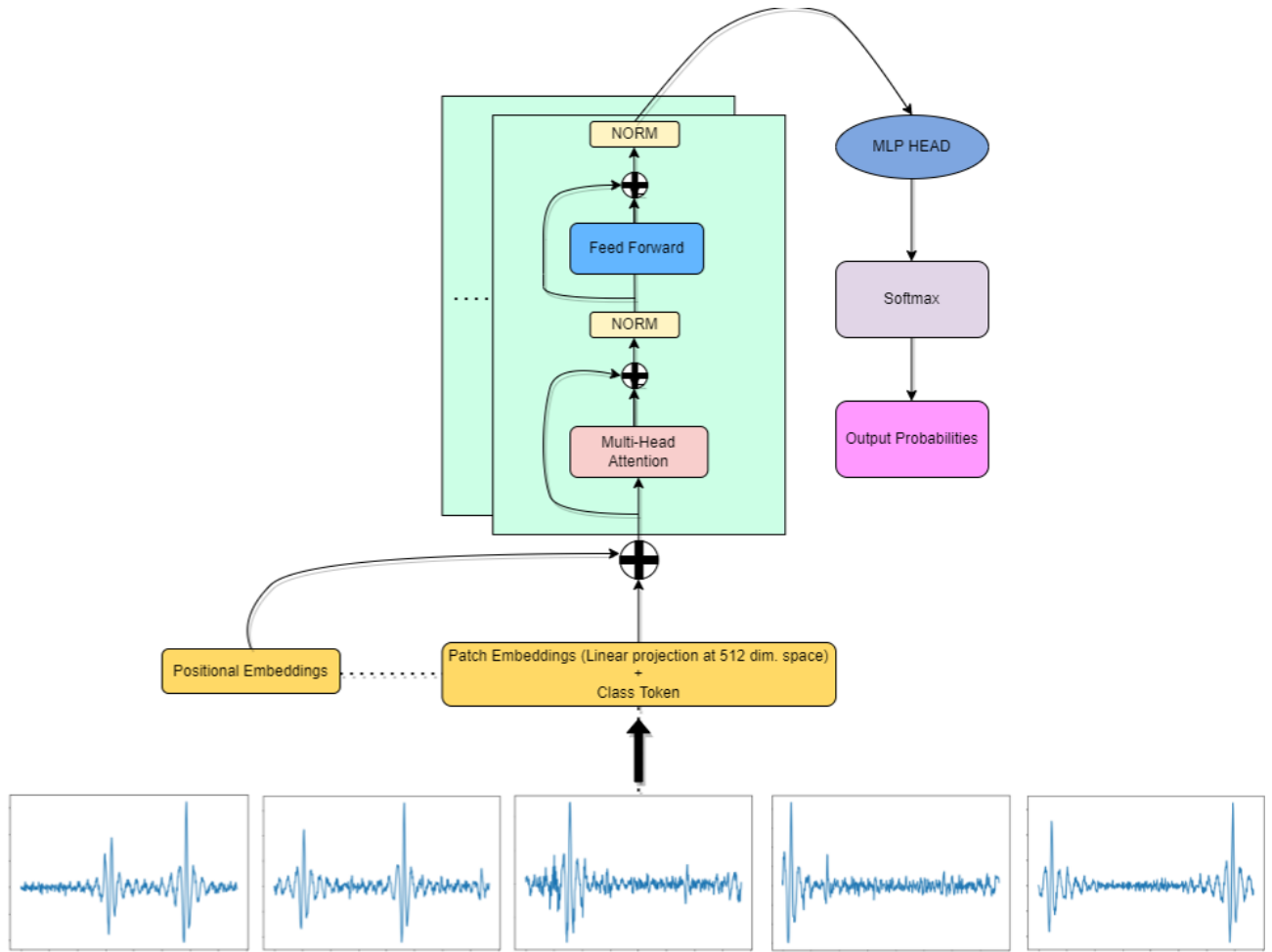
εξόδους με μεγέθη αυτά που φαίνονται στον πίνακα 5.1 Κατά την διεξαγωγή πειραμάτων με διαφορετικούς συνδυασμούς γραμμικών επιπέδων, παρατηρήθηκε ότι αύξηση στον αριθμό των γραμμικών στρωμάτων, είχε ως αποτέλεσμα την βελτίωση της ακρίβειας κατηγοριοποίησης. Κάποιοι πιθανοί λόγοι στους οποίους ευθύνεται αυτή η βελτίωση, είναι ότι η προσθήκη περισσότερων γραμμικών επιπέδων, προκαλεί αύξηση της χωρητικότητας του δικτύου και ενισχύεται η ικανότητα του μοντέλου να μαθαίνει πιο σύνθετες πληροφορίες για τα δεδομένα εκπαίδευσης. Έτσι, το μοντέλο συλλαμβάνει περισσότερες πληροφορίες για τα δεδομένα εισόδου και βελτιώνεται η ικανότητα πρόβλεψης των κλάσεων. Ωστόσο, είναι σημαντικό να αποφευχθεί η υπερβολική προσαρμογή του μοντέλου, επομένως η αύξηση βάθους στο στάδιο FFN έπρεπε να γίνει με ιδιαίτερη προσοχή.

Επίπεδα	Μέγεθος Εισόδου	Μέγεθος Εξόδου	Συνάρτηση Ενεργοποίησης
1 ^ο Linear επίπεδο	embed_dim	2× embed_dim	GELU('tanh')
2 ^ο Linear επίπεδο	2× embed_dim	embed_dim	-

Πίνακας 5.1: Επισκόπηση των επιπέδων που χρησιμοποιούν οι Arjun *et al.* στο δίκτυο που υλοποιούν και συγκεκριμένα στο FFN στάδιο.

Επίπεδα	Μέγεθος Εισόδου	Μέγεθος Εξόδου	Συνάρτηση Ενεργοποίησης
1 ^ο Linear επίπεδο	embed_dim	2× embed_dim	GELU('tanh')
2 ^ο Linear επίπεδο	2× embed_dim	4× embed_dim	GELU('tanh')
3 ^ο Linear επίπεδο	4× embed_dim	2× embed_dim	GELU('tanh')
4 ^ο Linear επίπεδο	2× embed_dim	embed_dim	-

Πίνακας 5.2: Αύξηση των επιπέδων στο στάδιο FFN του transformer για περαιτέρω βελτίωση της ακρίβειας.



Σχήμα 5.6: Η προτεινόμενη αρχιτεκτονική των Arjun *et al.*. Το δίκτυο δέχεται στην είσοδο τα *flattened* EEG segments.

5.3 Εκπαίδευση

5.3.1 Συνάρτηση Απωλειών

Η συνάρτηση απωλειών που εφαρμόστηκε στο πείραμα είναι η *Cross Entropy*. Η συνάρτηση αυτή, εκτιμά την διαφορά μεταξύ της προβλεπόμενης πιθανότητας η είσοδος να ανήκει σε μία κλάση και της επιθυμητής κλάσης εξόδου. Όταν η επιθυμητή έξοδος λαμβάνει πολλαπλές τιμές, τότε η cross entropy loss χρησιμοποιείται για προβλήματα κατηγοριοποίησης πολλαπλών κλάσεων, όπου στόχος είναι η πρόβλεψη μίας ετικέτας κλάσης μεταξύ πολλών πιθανών. Η απόσταση της προβλεπόμενης εξόδου και της επιθυμητής, υπολογίζεται σε λογαριθμική κλίμακα σύμφωνα με την παρακάτω σχέση.

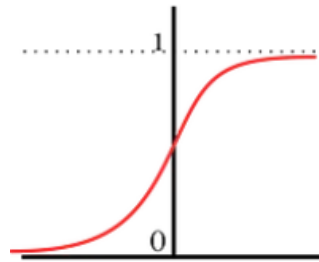
$$\mathcal{L} = - \sum_{i=1}^n (y_i \log(p_i)) \quad (5.5)$$

Όπου y_i είναι η επιθυμητή έξοδος που αντιστοιχεί στο i δείγμα εισόδου, p_i είναι η προβλεπόμενη πιθανότητα του i -οστού δείγματος που ανήκει στην κλάση y_i και n είναι ο αριθμός των κλάσεων. Η πιθανότητα p_i υπολογίζεται για κάθε τιμή της εξόδου μέσω της συνάρτησης **Softmax**. Συγκεκριμένα, η έξοδος του τελικού επιπέδου του κατηγοριοποιητή, διέρχεται μέσα από την Softmax, η

οποία μετασχηματίζει τις τιμές του στο πεδίο των κατανομών πιθανότητας κλάσεων. Δεδομένου του διανύσματος εξόδου z , η συνάρτηση υπολογίζει την πιθανότητα αυτή μέσω της σχέσης:

$$p_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (5.6)$$

Όπου z_j τα στοιχεία του διανύσματος εξόδου z_i . Η γραφική παράσταση της συνάρτησης Softmax φαίνεται στο σχήμα 5.7. Η συνάρτηση διασφαλίζει ότι το άθροισμα πιθανοτήτων των προβλεπόμενων κλάσεων θα είναι 1 και οι τιμές τους θα κυμαίνονται στο διάστημα τιμών $[0, 1]$. Η τιμή p_i αντιπροσωπεύει τον βαθμό εμπιστοσύνης του κατηγοριοποιητή στην σωστή πρόβλεψη της εξόδου.



Σχήμα 5.7: Γραφική αναπαράσταση της συνάρτησης Softmax: $\mathbb{R} \Rightarrow [0, 1]$. Πηγή: [4]

5.3.2 Σύνολα Δεδομένων

Η **DEAP** (*Dataset for Emotion Recognition using Physiological Signals*) [37], δημιουργήθηκε από τους Koelstra *et al.* και είναι μία δημοσίως διαθέσιμη βάση δεδομένων που χρησιμοποιείται ελεύθερα από ερευνητές που μελετούν τις διακυμάνσεις των συναισθημάτων στα φυσιολογικά σήματα. Η βάση DEAP, περιέχει μεγάλη πληθώρα δεδομένων και χρησιμοποιείται ευρέως σε πολλές μελέτες του τομέα *affective computing*.

Τα δεδομένα της βάσης συλλέχθηκαν από 32 υγιείς συμμετέχοντες (ηλικίας 19 έως 37) στο χρονικό διάστημα που ήταν εκτεθειμένοι σε οπτικό-ακουστικά ερεθίσματα. Τα ερεθίσματα στην συγκεκριμένη περίπτωση, ήταν μουσικά βίντεο και επιλέχθηκαν με σκοπό να προκαλούν συγκεκριμένα συναισθήματα στους συμμετέχοντες. Τα φυσιολογικά σήματα που καταγράφηκαν αποτελούνταν από καταγραφές EEG, MEG και ECG. Η βάση συνοδεύεται επίσης με συμπεριφορικά δεδομένα που περιλαμβάνουν αξιολογήσεις των συμμετεχόντων για τις συνιστώσες *valence*, *arousal*, *dominance*, *liking* των συναισθημάτων, καθώς και βίντεο που αποθανατίζουν τις εκφράσεις των προσώπων των 22 από τους 32 συμμετέχοντες.

Η χρονική διάρκεια των δεδομένων EEG ανέρχεται στα 63 sec., όπου τα πρώτα 3 δευτερόλεπτα αντιστοιχούν στο χρονικό διάστημα της προετοιμασίας των συμμετεχόντων πριν την έναρξη των βίντεο. Γι' αυτό στην υλοποίηση τα πρώτα 3 δευτερόλεπτα παραλείφθηκαν. Τα EEG και τα υπόλοιπα σήματα καταγράφηκαν με ρυθμό δειγματοληψίας 512 Hz και μέσα από χειροκίνητη επιλογή των δειγμάτων, η συχνότητα των σημάτων EEG κατέβηκε στα 128 Hz. Τα δεδομένα υποβλήθηκαν σε περαιτέρω επεξεργασία μέσω του *common average referencing*, την *blind source separation* τεχνική και την υποβολή τους σε ζωνοδιαβατό φίλτρο. Το σήμα που προέκυψε, ανήκει στη συχνοτική ζώνη 4 – 45 Hz. Τα 32 αρχεία που αντιστοιχούν σε κάθε άτομο, περιέχουν τους πίνακες *data* και *labels*. Τα περιεχόμενα των αρχείων φαίνονται παρακάτω.

Πίνακας	Μέγεθος	Περιεχόμενα
data	$32 \times 40 \times 7680$	video/trial \times channel \times data
labels	$32 \times 40 \times 4$	video/trial \times label

Ο πίνακας *labels* περιέχει τις αξιολογήσεις των συμμετεχόντων, που είναι σχετικές με το συναίσθημα που ένιωσαν κατά την παρακολούθηση των μουσικών βίντεο. Κάθε βίντεο, βαθμολογήθηκε με βάση τις 4 συνιστώσες valence, arousal, dominance, liking στο εύρος 1-9. Για το πειραματικό στάδιο της διπλωματικής, υιοθετήθηκαν μόνο τα labels valence, arousal. Επειδή παρατηρήθηκε ότι η προσέγγιση που ακολουθούν οι περισσότερες μελέτες στην βιβλιογραφία, αφορά τις κλάσεις high valence, low valence και high arousal, low arousal, θεωρήσαμε ότι τα labels με τιμές μεγαλύτερες ή ίσες του 5 αντιστοιχούν στον παράγοντα *high* – 1 και τα labels των δεδομένων με τιμές μικρότερες από 5 στον παράγοντα *low* – 0.

5.3.3 Εξισορρόπηση Δεδομένων

Η εξισορρόπηση των δεδομένων (*data balancing*) αποτελεί μία ειδική τεχνική προεπεξεργασίας, που χρησιμοποιείται για την αντιμετώπιση του προβλήματος της «ανισορρόπησης» των κλάσεων. Η ανισορροπία, μπορεί να προκαλέσει κακή απόδοση των μοντέλων μηχανικής εκμάθησης, όταν κλίνονται να χειριστούν δεδομένα που ανήκουν στην κατηγορία μειοψηφίας. Στο πειραματικό στάδιο της διπλωματικής, χρησιμοποιήσαμε την τεχνική της υπερδειγματοληψίας για να αντιμετωπίσουμε τις ανισορροπίες στις τιμές των ετικετών και να βελτιώσουμε την απόδοση του μοντέλου. Συγκεκριμένα, εφαρμόστηκε τυχαία επιλογή και αντιγραφή κάποιων υπάρχοντων δειγμάτων εκπαίδευσης που ανήκουν στην κλάση μειοψηφίας.

5.3.4 Επαύξηση Δεδομένων

Η διαδικασία της καταγραφής των δεδομένων EEG είναι χρονικά και χρηματικά δαπανηρή και προϋποθέτει ότι όλοι οι συμμετέχοντες που λαμβάνουν μέρος στο πείραμα είναι ψυχολογικά υγιείς. Οι βάσεις δεδομένων που διατίθενται από τους επιμελητές των βάσεων, έχουν ικανοποιητικό αριθμό δεδομένων, αλλά η πολυπλοκότητα των μοντέλων μηχανικής μάθησης καθιστά αναγκαία την δημόσια διάθεση και χρήση πιο πολλών παραδειγμάτων. Εν τούτοις, η περιορισμένη διάθεση δεδομένων μπορεί να οδηγήσει στο πρόβλημα της υπερβολικής προσαρμογής και στη γενικότερα κακή απόδοση των πειραμάτων.

Η επαύξηση δεδομένων (*data augmentation*) είναι μία τεχνική που χρησιμοποιείται συχνά στον τομέα της μηχανικής μάθησης και συγκεκριμένα στα πεδία της υπολογιστικής όρασης, της επεξεργασίας φυσικής γλώσσας και σε άλλα. Κατά την τεχνική αυτή, νέα δεδομένα εκπαίδευσης δημιουργούνται με τεχνητό τρόπο μέσα από τον μετασχηματισμό των αρχικών παραδειγμάτων. Στην ερευνητική κοινότητα έχουν αναπτυχθεί πολλές τεχνικές επαύξησης του συνόλου δεδομένων, όμως για το πειραματικό στάδιο αξιοποιήθηκε μία τεχνική που βασίζεται στην λειτουργία της μεθόδου **CutMix** [84]. Η ακριβής ιδέα θα περιγραφεί στη συνέχεια αναλυτικά.

Έστω $x_1, x_2 \in \mathbb{R}^{C \times L}$ δύο δεδομένα EEG εκπαίδευσης που μπορεί να προέρχονται από διαφορετικά trials του ίδιου ατόμου ή από trials διαφορετικών ατόμων. Για να μην υπάρξει σύγχυση, τα παραδείγματα εκπαίδευσης x_1, x_2 έχουν ίδιες ετικέτες. Ένα νέο ζευγάρι εκπαίδευσης (\hat{x}, \hat{y}) , προκύπτει μέσα από έναν ειδικό συνδυασμό των x_1, x_2 σύμφωνα με την σχέση:

$$\hat{x} = \mathcal{M} \odot x_1 + (\mathbf{I} - \mathcal{M}) \odot x_2 \quad (5.7)$$

$$\hat{y} = \text{label}(x_1) = \text{label}(x_2) \quad (5.8)$$

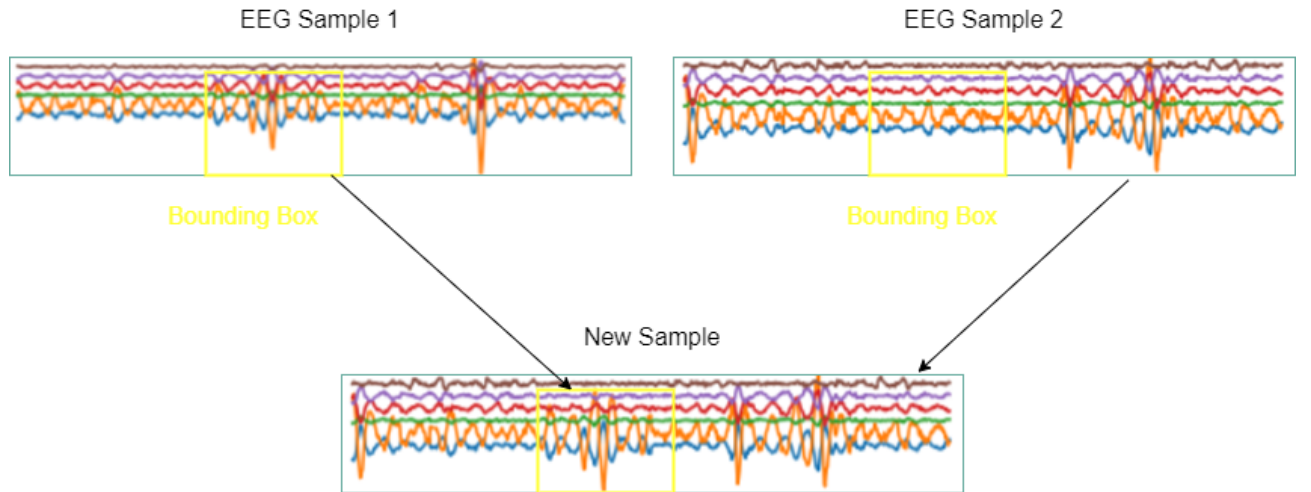
Ο πίνακας $\mathcal{M} \in \mathbb{R}^{C \times L}$ αντιπροσωπεύει μία δυαδική τετράγωνη μάσκα, δηλαδή λαμβάνει τις τιμές $\{0, 1\}$ και υποδεικνύει το πλαίσιο του δεδομένου x_1 που θα αντικατασταθεί από το δεδομένο x_2 και ο πίνακας \mathbf{I} είναι ο μοναδιαίος πίνακας με τιμές $\{0, 1\}^{C \times L}$. Αρχικά, το μέγεθος και οι συντεταγμένες του πλαισίου οριοθέτησης \mathbf{B} , προσδιορίζονται με τυχαία δειγματοληψία σύμφωνα με τις σχέσεις της CutMix δημοσίευσης:

$$r_x \sim \text{UniformDist}(C, L), \quad r_c = C\sqrt{1 - \lambda} \quad (5.9)$$

$$r_y \sim \text{UniformDist}(C, L), \quad r_l = L\sqrt{1 - \lambda} \quad (5.10)$$

Όπου $\frac{r_c r_l}{CL} = 1 - \lambda$ η σχέση αναλογίας των καναλιών και των χρονικών δειγμάτων C, L των δεδομένων EEG και των καναλιών και των χρονικών δειγμάτων r_c, r_l της υποπεριοχής.

Η διαδικασία αυτή εφαρμόστηκε με τυχαία επιλογή των παραδειγμάτων εκπαίδευσης που διαθέτουν ίδιες ετικέτες και τα νέα δείγματα προέκυψαν μέσω της σχέσης 5.7. Οι ετικέτες των νέων δειγμάτων είναι ίδιες με αυτές των παλιών. Στην εικόνα 5.8 φαίνεται ένα παράδειγμα παραγωγής ενός νέου δείγματος EEG. Με την παραλλαγή της μεθόδου CutMix που αναλύσαμε, μπορούμε να αυξήσουμε το υποσύνολο εκπαίδευσης οσοδήποτε θέλουμε για να αυξηθεί η απόδοση του δικτύου όταν τα δεδομένα είναι ανεπαρκή.



Σχήμα 5.8: Παραλλαγή της τεχνικής CutMix που δημιουργεί νέα δεδομένα από το υπάρχον σύνολο εκπαίδευσης. Τα πρώτα δύο δείγματα μπορεί να προέρχονται από το ίδιο ή από διαφορετικά άτομα. Το νέο δείγμα που προκύπτει αποτελεί αντίγραφο του δεύτερου με την διαφορά ότι ένα επιλεγμένο υπο-πλαίσιο ανήκει στο πρώτο δείγμα EEG.

Βελτίωση της μεθόδου

Στην μέθοδο CutMix, η θέση του πλαισίου που θα αποκοπεί και θα αντικατασταθεί, είναι απολύτως τυχαία και κατά την επιλογή των συντεταγμένων του πλαισίου στα δείγματα, δεν λαμβάνονται υπόψιν οι κατανομές των τιμών της περιοχής. Ως εκ τούτου, το νέο τεχνητό δείγμα μπορεί να περιέχει σοβαρές ασυνέχειες στα άκρα του bounding box, οι οποίες μπορεί να επηρεάσουν αρνητικά τη διαδικασία της εκπαίδευσης και τη μορφή της συνάρτησης απωλειών. Ένα ακόμη πιθανό

ζήτημα, αφορά την περιοχή τιμών του πλαισίου που αντικαθίστανται. Πιο συγκεκριμένα, οι νέες τιμές μπορεί να διαφέρουν κατά πολύ από τις τιμές του υπόλοιπου δείγματος. Υπάρχει κίνδυνος, λοιπόν, το μοντέλο κατά την εκπαίδευση να δώσει περισσότερη σημασία σε χαρακτηριστικά με μεγάλες, ακραίες τιμές και έτσι να κατηγοριοποιεί λανθασμένα τα νέα δεδομένα.

Για να μετριαστούν τα παραπάνω προβλήματα, σ' αυτήν την διπλωματική προτείνουμε να περιοριστεί η τυχαιότητα της μεθόδου, αναζητώντας στα δύο δείγματα, τους υποχώρους που έχουν παρόμοιες στατιστικές ιδιότητες. Συγκεκριμένα, οι υποχώροι που διατηρήσαμε, είναι εκείνοι που χαρακτηρίζονται από την πιο κοντινή μέση τιμή. Τα αλγοριθμικά βήματα της διαδικασίας αυτής, περιγράφονται στη συνέχεια:

1. Τυχαία επιλογή μεγεθών c, l .
2. Δημιουργία ενός πυρήνα k μεγέθους $(c \times l)$ που υπολογίστηκε στο προηγούμενο βήμα.
3. Εφαρμογή 2D συνέλιξης ανάμεσα στον πυρήνα και στα δύο δείγματα x_1, x_2 : $y_1(i, j) = \sum_{m=0}^L \sum_{n=0}^C k(i-m, j-n) * x_1(m, n)$, $y_2(i, j) = \sum_{m=0}^L \sum_{n=0}^C k(i-m, j-n) * x_2(m, n)$.
4. Από τα y_1, y_2 υπολογίζουμε τους μέσους όρους όλων των υποχώρων
5. Υπολογίζουμε τις αποστάσεις των μέσων όρων που αντιστοιχούν στα δύο δείγματα EEG και διατηρούμε τον υποχώρο που αντιστοιχεί στην μικρότερη απόσταση.

5.3.5 Υπερπαραμέτροι Μοντέλων

Πολλές φορές, ένα μοντέλο κακής απόδοσης, μπορεί να πετύχει την *state of the art*, μόνο μέσα από την επιλογή των κατάλληλων υπερπαραμέτρων. Η προσαρμογή του learning rate, του momentum, του weight decay, του batch size, του αριθμού των επιπέδων και άλλων παραμέτρων που καθορίζονται πριν την εκπαίδευση, σχετίζονται στενά με την αρχιτεκτονική του μοντέλου. Επειδή το πρόβλημα της διπλωματικής αυτής είναι αρκετά βαθύ και τα δεδομένα που έχουμε στην διάθεση μας αρκετά σύνθετα, δεν αρκεί μόνο μία αρχική εκτίμηση των παραμέτρων αυτών και απαιτούνται πολλοί πειραματισμοί, προκειμένου να επιτευχθεί μία επιτυχημένη εκτέλεση. Όπως φαίνεται λοιπόν, αυτή η διαδικασία απαιτεί χρόνο και πολλές προσπάθειες, χωρίς να είναι πάντα εγγυημένη η επιτυχία.

Οι I. Loshchilov και F. Hutter [46], πρότειναν μία εναλλακτική υλοποίηση του ρυθμού εκμάθησης, ονόματι *linear warmup with cosine annealing*. Στην εκδοχή αυτή, το learning rate ακολουθεί ένα αρχικό στάδιο, κατά το οποίο αυξάνεται γραμμικά, μέχρι να φτάσει σε μία μέγιστη τιμή και στη συνέχεια μειώνεται ακολουθώντας την καμπύλη του συνημιτόνου, έως ότου φτάσει μία ελάχιστη τιμή. Το στάδιο του cosine annealing επαναλαμβάνεται κυκλικά. Η προσέγγιση είναι ιδιαίτερα χρήσιμη και έχει αποδειχτεί ότι βελτιώνει την απόδοση πολλών μοντέλων βαθιάς μάθησης. Συγκεκριμένα, η ουσία της μεθόδου είναι η εκπαίδευση του μοντέλου να γίνει σε ένα ευρύ φάσμα ρυθμών εκμάθησης, που θα βελτιώσουν την εκμάθηση του και θα το αποτρέψουν απ' το να εγκλωβιστεί σε τοπικά ελάχιστα.

Σε όλα τα μοντέλα χρησιμοποιούμε την τεχνική βελτιστοποίησης Adam με *weight decay* = 0.05. Ο ρυθμός εκμάθησης αρχικοποιείται στην τιμή 10^{-6} , αυξάνεται μέχρι την τιμή 0.005 και

ακολουθώντας μπαίνει στο στάδιο cosine annealing με ελάχιστη τιμή 10^{-6} . Οι υπόλοιπες παράμετροι είναι οι προεπιλεγμένες, $\text{betas} = (0.9, 0.999)$ και $\text{epsilon} = 10^{-8}$. Ο αριθμός των εποχών, το βάθος, τα βάρη του δικτύου ρυθμίζονται αναλόγως με το είδος της αρχιτεκτονικής, τα πειράματα και τον τρόπο εισαγωγής των δεδομένων στα μοντέλα.

5.3.6 Διαδικασία Εκπαίδευσης

Οι εγγραφές μήκους 60 δευτερολέπτων, χωρίστηκαν σε μη επικαλυπτόμενα δείγματα διάφορων μεγεθών και τροφοδοτήθηκαν στο δίκτυο εφαρμόζοντας εκ των προτέρων όλες τις τεχνικές που αναλύσαμε. Το σύνολο δεδομένων, που αποτελείται από τα διαιρεμένα δείγματα και τα 32 κανάλια, τροφοδοτήθηκε στη συνέχεια στα μοντέλα ViT, CNN και LSTM για εκπαίδευση. Στην επόμενη ενότητα θα περιγραφεί η ακριβής διαδικασία χωρισμού του συνόλου δεδομένων σε παραδείγματα εκπαίδευσης και αξιολόγησης. Ο χωρισμός αυτός εξαρτάται στενά από την πειραματική μέθοδο, η οποία θα αναλυθεί επίσης στη συνέχεια.

Κεφάλαιο 6

Πειραματική Διαδικασία

Τα μοντέλα αξιολογούνται πάνω σε τρεις διαφορετικούς τύπους πειραμάτων. Στον **subject dependent** τύπο, τα δεδομένα του κάθε subject χωρίζονται σε σύνολα εκπαίδευσης και αξιολόγησης και το δίκτυο εκπαιδεύεται και αξιολογείται με τα δεδομένα EEG ενός ατόμου κάθε φορά. Σε αυτόν τον τύπο μάθησης, το μοντέλο μαθαίνει τις πληροφορίες και τα patterns των δεδομένων EEG ενός ατόμου, έτσι ώστε όταν τροφοδοτηθεί με νέα δεδομένα (που προέρχονται από τον συγκεκριμένο άνθρωπο), να προβλέπει όσο τον δυνατόν καλύτερα τις τιμές διέγερσης και σθένους.

Στο **subject independent** πείραμα, το δίκτυο εκπαιδεύεται με τα δεδομένα EEG που προέρχονται από πολλά άτομα του συνόλου και αξιολογείται με δεδομένα που προέρχονται από καταγραφές διαφορετικών ατόμων. Αυτός ο τύπος πειραμάτων, είναι ιδιαίτερα χρήσιμος, καθώς τα μοντέλα που δημιουργούνται μπορούν και γενικεύουν την λειτουργία τους μαθαίνοντας global χαρακτηριστικά από τα EEG δεδομένα πολλών ανθρώπων.

Τέλος, αξιολογείται και ένας τρίτος τύπος πειραμάτων, κατά τον οποίο στο training set περιλαμβάνονται τα δεδομένα που αντιστοιχούν στα trials μίας ομάδας ατόμων και τα δεδομένα του test set προέρχονται από διαφορετικά trials της ομάδας αυτής. Αυτός ο τύπος πειράματος εφαρμόστηκε από τους Arjun *et al.* [11].

Δοκιμάστηκαν διάφορες στρατηγικές χωρισμού του συνόλου δεδομένων σε train και test sets, με επικρατέστερη την στρατηγική **k-fold cross validation**, όπου τα σύνολα δεδομένων χωρίζονται με k διαφορετικούς συνδυασμούς. Η ακρίβεια της μεθόδου προέκυψε ως η καλύτερη ακρίβεια ανάμεσα στις k διαφορετικές δοκιμές. Στη συνέχεια θα περιγραφεί εν συντομία η διαδικασία διεξαγωγής των πειραμάτων.

6.1 Αποτελέσματα

Ως μέθοδος αξιολόγησης της κατηγοριοποίησης των δικτύων χρησιμοποιήθηκε η μετρική *accuracy*, τόσο στα αυτούσια δεδομένα της βάσης DEAP, όσο και στον συνδυασμό των δεδομένων αυτών με δεδομένα που δημιουργήθηκαν τεχνητά χρησιμοποιώντας την τροποποιημένη μέθοδο CutMix. Τα πειράματα εφαρμόστηκαν σε δεδομένα διαφόρων χρονικών διαρκειών, όμως στη συνέχεια παρουσιάζονται οι ακρίβειες κατηγοριοποίησης για δεδομένα EEG 6 δευτερολέπτων.

Subject-independent Πειράματα

Για τα subject independent πειράματα υιοθετήθηκε η στρατηγική *5-fold cross validation* για την αξιολόγηση της απόδοσης των EEG emotion recognition πειραμάτων. Συγκεκριμένα, τα δεδομένα EEG των 25 ατόμων χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου και τα δεδομένα των υπόλοιπων 7 ατόμων χρησιμοποιήθηκαν αργότερα για την αξιολόγηση της ακρίβειας πρόβλεψης.

	Low Valence\High Valence	Low Arousal\High Arousal
CNN	53.196%	52.8409%
LSTM	54.13%	53.707%
ViT	66.28%	65.23%

Πίνακας 6.1: Η ακρίβεια των **subject independent** πειραμάτων για τις binary κλάσεις των συναισθηματικών καταστάσεων της διέγερσης και τους σθένους. Το χρονικό μήκος των δεδομένων εισόδου είναι ίσο με 6 *seconds*. Τα αποτελέσματα αυτά προέκυψαν χωρίς την χρήση αλγόριθμου επαύξησης των δεδομένων εκπαίδευσης.

	Low Valence\High Valence	Low Arousal\High Arousal
CNN	52.478%	51.362%
LSTM	54.4%	53.9%
ViT	67.07%	65.93%

Πίνακας 6.2: Εδώ βλέπουμε την ακρίβεια κατηγοριοποίησης των **subject independent** πειραμάτων, όταν έχει εφαρμοστεί ο βελτιωμένος αλγόριθμος επαύξησης των δεδομένων εκπαίδευσης.

Subject-dependent Πειράματα

Για τα subject dependent πειράματα, μετά από τον χωρισμό του συνόλου δεδομένων με πολλές στρατηγικές, ως καλύτερη στρατηγική ξεχώρισε η *4-fold cross validation* στρατηγική για την αξιολόγηση της απόδοσης των EEG emotion recognition πειραμάτων που αντιστοιχούν σε κάθε άτομο ξεχωριστά. Η εκπαίδευση και αξιολόγηση του πειράματος έγινε επαναληπτικά για κάθε ένα από τα 32 άτομα της βάσης. Τα δεδομένα που αντιστοιχούν στα 30 trials του ατόμου, επιλέχθηκαν ως το σύνολο εκπαίδευσης και τα υπόλοιπα 10 ως το σύνολο αξιολόγησης. Στους πίνακες φαίνονται οι ακρίβειες των μοντέλων για τα 10 πρώτα subjects και η μέση ακρίβεια κατηγοριοποίησης όλων των συμμετεχόντων.

δίκτυο CNN (χωρίς augmentation)

Labels	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10	Average Accuracy (over the 32 subjects)
High/Low Valence	80.17%	56.17%	86.5%	61.5%	84%	72%	56.5%	70%	64.6%	77.17%	67.26%
High/Low Arousal	55%	80.67%	60.66%	54.7%	57.5%	60.17%	61.17%	63.833%	53%	63%	59.549%

Πίνακας 6.3: Η ακρίβεια κατηγοριοποίησης των δεδομένων EEG 10 συμμετεχόντων και η μέση ακρίβεια των δεδομένων για όλους τους συμμετέχοντες. Το δίκτυο που χρησιμοποιήθηκε σε αυτήν την περίπτωση είναι το CNN που περιγράφηκε στην ενότητα 5.2.1

δίκτυο LSTM (χωρίς augmentation)

Labels	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10	Average Accuracy (over the 32 subjects)
High/Low Valence	82%	60.33%	85.76%	62.2%	85.65%	70%	58%	70%	56%	80.5%	74.67%
High/Low Arousal	56%	70%	65.6%	60%	65%	67.6%	70%	70.56%	55%	67.5%	67.81%

Πίνακας 6.4: Οι κατά προσέγγιση ακρίβειες κατηγοριοποίησης των δεδομένων EEG 5 συμμετεχόντων και η μέση ακρίβεια των δεδομένων για όλους τους συμμετέχοντες. Το δίκτυο που χρησιμοποιήθηκε σε αυτήν την περίπτωση είναι το LSTM που περιγράφηκε στην ενότητα 5.2.1

δίκτυο ViT (χωρίς augmentation)

Labels	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10	Average Accuracy (over the 32 subjects)
High/Low Valence	84.27%	78.12%	85.88%	65.9%	90%	71.7%	56.3%	80.19%	56.5%	81.2%	75.79%
High/Low Arousal	79.12%	85.78%	70.32%	66.83%	67.81%	75.9%	78.45%	70%	64.3%	78.8%	68.31%

Πίνακας 6.5: Οι κατά προσέγγιση ακρίβειες κατηγοριοποίησης των δεδομένων EEG 10 συμμετεχόντων και η μέση ακρίβεια των δεδομένων για όλους τους συμμετέχοντες. Το δίκτυο που χρησιμοποιήθηκε σε αυτήν την περίπτωση είναι το ViT που περιγράφηκε στην ενότητα 5.2.1

δίκτυο ViT (με augmentation)

Labels	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10	Average Accuracy (over the 32 subjects)
High/Low Valence	84.38%	79.31%	89.06%	70.31%	92.19%	78.12%	78.12%	84.38%	87.5%	81.2%	82.88%
High/Low Arousal	80.15%	89.68%	72.28%	66.389%	67.21%	75.49%	88.96%	77.98%	77.98%	85.13%	75.33%

Πίνακας 6.6: Οι κατά προσέγγιση ακρίβειες κατηγοριοποίησης των δεδομένων EEG 10 συμμετεχόντων και η μέση ακρίβεια των δεδομένων για όλους τους συμμετέχοντες. Το δίκτυο που χρησιμοποιήθηκε σε αυτήν την περίπτωση είναι το ViT που περιγράφηκε στην ενότητα 5.2.1

Υβριδικά Πειράματα

	Low Valence\High Valence	Low Arousal\High Arousal
CNN	77.829%	76.78%
LSTM	80.87%	78.65%
ViT	88.61%	83.8%

Πίνακας 6.7: Η ακρίβεια των υβριδικών πειραμάτων για τις binary κλάσεις των συναισθηματικών καταστάσεων της διέγερσης και τους σθένους. Τα αποτελέσματα που φαίνονται προέκυψαν δίχως κάποιον αλγόριθμο επαύξησης των δεδομένων.

	Low Valence\High Valence	Low Arousal\High Arousal
ViT	90.6%	85.39%

Πίνακας 6.8: Στον πίνακα φαίνεται η ακρίβεια κατηγοριοποίησης των δεδομένων EEG για το δίκτυο ViT, όταν εφαρμόζεται η τεχνική data augmenation στο σύνολο εκπαίδευσης, που περιγράφηκε.

6.1.1 Συγκρίσεις

Με βάση τα ευρήματα που παρουσιάστηκαν προηγουμένως, είναι προφανές ότι η αρχιτεκτονική Vision Transformer, με κατάλληλη επιλογή των υπερπαραμέτρων, ξεπερνά κατά πολύ την απόδοση των CNN και LSTM δικτύων σε όλα τα σενάρια των πειραμάτων που πραγματοποιούνται. Αυτό υποδηλώνει ότι η νέα έννοια των μηχανισμών αυτο-προσοχής, που αποτελεί τη βάση αυτών των δικτύων, υπόσχεται σημαντικές εξελίξεις στην ακρίβεια και την απόδοση της EEG αναγνώρισης συναισθημάτων.

Με την ενσωμάτωση της data augmentation τεχνικής στο δίκτυο του vision μετασχηματιστή (ViT) και την προσεκτική αναπροσαρμογή των υπερπαραμέτρων, παρατηρήσαμε σημαντική βελτίωση στην ακρίβεια του δικτύου. Είναι γνωστό ότι οι μετασχηματιστές απαιτούν μεγάλο όγκο δεδομένων εκπαίδευσης για να μάθουν αποτελεσματικά τα υποκείμενα χαρακτηριστικά. Ωστόσο, στον τομέα της αναγνώρισης συναισθημάτων από σήματα EEG, τα διαθέσιμα δεδομένα είναι περιορισμένα, γεγονός που καθιστά τα συνθετικά δείγματα πολύτιμα για την αύξηση του συνόλου δεδομένων εκπαίδευσης.

	Valence			Arousal		
	6sec.	15sec.	30sec.	6sec.	15sec.	30sec.
Arjun et al. [11] (2022)	99.4%	99.2%	97.5%	99.1%	99.2%	90.5%
Προτεινόμενη Μέθοδος	90.8%	88.6%	86.63%	90.6%	87.9%	85.39%

Πίνακας 6.9: Πίνακας που συγκρίνει την ακρίβεια των **υβριδικών πειραμάτων** της βάσης DEAP, όπως αναφέρεται στην δημοσίευση των Arjun *et al.*, με την υλοποίηση που πραγματοποιήθηκε στην παρούσα διπλωματική.

Σχετικές Έρευνες	Δίκτυο	Μέθοδος Αξιολόγησης	Ακρίβεια (%)
Pandey et al. [55] (2019)	Deep Neural Network	Leave-one-subject-out	Valence: 62.5 Arousal: 61.25
Yin et al. [83] (2020)	SVM	Leave-one-subject-out	Valence: 67.97 Arousal: 65.1
Wang et al. [81] (2022)	Hierarchical Spatial Learning Transformer (HSLT)	Leave-one-subject-out	Valence: 66.51 Arousal: 65.75
Προτεινόμενη Μέθοδος	Vision Transformer	5 fold cross validation	Valence: 67.07 Arousal: 65.93

Πίνακας 6.10: Σύγκριση της μέσης ακρίβειας της προτεινόμενης μεθόδου με πρόσφατες έρευνες, για το **subject independent** πείραμα στην βάση δεδομένων DEAP.

Στο υβριδικό πείραμα μπορούμε να κάνουμε συγκρίσεις μόνο με την υλοποίηση των Arjun *et al.*. Όπως βλέπουμε, η μέθοδος που προτείνουμε δεν κατάφερε να ξεπεράσει την απόδοση του μοντέλου τους. Τα αποτελέσματα έδειξαν ωστόσο, ότι το μοντέλο που χρησιμοποιήσαμε σ' αυτήν την διπλωματική, αποδίδει πολύ ικανοποιητικά. Στον πίνακα 6.10, παρατηρούμε ότι η απόδοση του μοντέλου στα *subject independent* πειράματα, φτάνει και σε ορισμένες περιπτώσεις, ξεπερνάει την απόδοση κάποιων πρόσφατων πειραμάτων που έχουν διεξαχθεί πάνω στο πρόβλημα της αναγνώρισης συναισθημάτων με χρήση σημάτων EEG. Επιπλέον, τα πολλά υποσχόμενα αποτελέσματα που λαμβάνονται με την αρχιτεκτονική του vision transformer, σε συνδυασμό με τις αποτελεσματικές τεχνικές επαύξησης δεδομένων που χρησιμοποιούνται, υποδηλώνουν ότι αυτή η προσέγγιση έχει μεγάλες δυνατότητες για μελλοντική έρευνα σε αυτόν τον τομέα.

Κεφάλαιο 7

Συμπεράσματα και Μελλοντικές Κατευθύνσεις

Έχοντας σχεδόν ολοκληρώσει τη συγγραφή του κειμένου της διπλωματικής και έχοντας πλέον μια πλήρη και καθαρή εικόνα αυτού του τομέα, καταλήξαμε σε κάποια βασικά συμπεράσματα, τα οποία πιστεύουμε ότι θα φανούν ιδιαίτερος χρήσιμα και θα δώσουν μία πιθανή κατεύθυνση σε νέους ερευνητές που επιθυμούν να κάνουν τα πρώτα τους βήματα στον τομέα της αναγνώρισης συναισθημάτων με χρήση σημάτων EEG .

Στην βιβλιογραφία, υπερτερούν σημαντικά οι έρευνες που ασχολούνται με την δυαδική κατηγοριοποίηση των EEGs στις κλάσεις high-low valence και high-low arousal. Αν θέλουμε τα δίκτυα που δημιουργούμε, να μας δίνουν ακριβείς και ολοκληρωμένες πληροφορίες για την συναισθηματική ποικιλομορφία που χαρακτηρίζει τους ανθρώπους, όταν εκτίθενται σε διάφορα ερεθίσματα, δεν αρκεί μόνο η περιγραφή των συναισθημάτων στις διακριτές κλάσεις. Μία αποτελεσματική προσέγγιση για την αντιμετώπιση του ζητήματος αυτού, είναι ο μετασχηματισμός του προβλήματος κατηγοριοποίησης σε πρόβλημα παλινδρόμησης. Αυτό μπορεί να επιτευχθεί διαμορφώνοντας τα δίκτυα μηχανικής μάθησης, ώστε να εξάγουν συνεχείς τιμές εντός ενός καθορισμένου εύρους αντί για διακριτές ετικέτες συναισθηματικών κλάσεων. Φυσικά, αυτό ενέχει περισσότερες απαιτήσεις, όπως την ανάγκη για περισσότερα δεδομένα, που αντιπροσωπεύουν με ακρίβεια την ποικιλομορφία των συναισθημάτων. Επιπλέον, οι συναισθηματικές ετικέτες που προκύπτουν από την αξιολόγηση των συμμετεχόντων μετά την παρακολούθηση των μουσικών βίντεο, ενέχουν έναν βαθμό υποκειμενικότητας. Συνεπώς, μία λύση σε αυτό, θα ήταν η εκπαίδευση να γίνει με μη-επιβλεπόμενο τρόπο. Όλα αυτά θα συνεισφέρουν στο ορθότερο χαρακτηρισμό του προβλήματος και θα βελτιώσουν την ακρίβεια και την ικανότητα γενίκευσης των μοντέλων.

Βιβλιογραφία

- [1] [online link]. <https://training.seer.cancer.gov/brain/tumors/anatomy/neurons.html>.
- [2] [online link]. <https://nasost.com/vasikes-leitourgies-egkefalou/>.
- [3] [online link]. <https://en.wikipedia.org/wiki/File:SinusRhythmLabels.png>.
- [4] [online link]. <https://towardsdatascience.com/sigmoid-and-softmax-functions-in-5-minutes>.
- [5] Sharmeen M Saleem Abdullah Abdullah, Siddeeq Y Ameen Ameen, Mohammed AM Sadeeq, and Subhi Zeebaree. Multimodal emotion recognition using deep learning. *Journal of Applied Science and Technology Trends*, 2(02):52–58, 2021.
- [6] Priyanka A Abhang, Bharti Gawali, and Suresh C Mehrotra. *Introduction to EEG-and speech-based emotion recognition*. Academic Press, 2016.
- [7] Soraia M. Alarcão and Manuel J. Fonseca. Emotions recognition using eeg signals: A survey. *IEEE Transactions on Affective Computing*, 10(3):374–393, 2019.
- [8] Salma Alhagry, Aly Aly Fahmy, and Reda A El-Khoribi. Emotion recognition based on eeg using lstm recurrent neural network. *International Journal of Advanced Computer Science and Applications*, 8(10), 2017.
- [9] Hamzah S. AlZu’bi, Nayel S. Al-Zubi, and Waleed Al-Nuaimy. Toward inexpensive and practical brain computer interface. In *2011 Developments in E-systems Engineering*, pages 98–101, 2011.
- [10] Katrin Amunts, O Kedo, M Kindler, P Pieperhoff, H Mohlberg, NJ Shah, U Habel, F Schneider, and K Zilles. Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anatomy and embryology*, 210(5):343–352, 2005.
- [11] Arjun Arjun, Aniket Singh Rajpoot, and Mahesh Raveendranatha Panicker. Introducing attention mechanism for eeg signals: Emotion recognition with vision transformers. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 5723–5726. IEEE, 2021.
- [12] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.
- [13] Annushree Bablani, Damodar Reddy Edla, Diwakar Tripathi, and Ramalingaswamy Cheruku. Survey on brain-computer interface: An emerging computational intelligence paradigm. *ACM computing surveys (CSUR)*, 52(1):1–32, 2019.
- [14] Sylvain Baillet, John C Mosher, and Richard M Leahy. Electromagnetic brain mapping. *IEEE Signal processing magazine*, 18(6):14–30, 2001.
- [15] Kurt Barbé, Wendy Van Moer, and Guy Nagels. Fractional-order time series models for extracting the haemodynamic response from functional magnetic resonance imaging data. *IEEE transactions on biomedical engineering*, 59(8):2264–2272, 2012.

- [16] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.
- [17] Hans Berger. On the electroencephalogram of man. *Electroencephalography and clinical neurophysiology*, pages Suppl–28, 1969.
- [18] J-F Cardoso. Blind signal separation: statistical principles. *Proceedings of the IEEE*, 86(10):2009–2025, 1998.
- [19] Jin Chen, Tony Ro, and Zhigang Zhu. Emotion recognition with audio, video, eeg, and emg: A dataset and baseline approaches. *IEEE Access*, 10:13229–13242, 2022.
- [20] JX Chen, PW Zhang, ZJ Mao, YF Huang, DM Jiang, and YN Zhang. Accurate eeg-based emotion recognition on combined features using deep convolutional neural networks. *IEEE Access*, 7:44317–44328, 2019.
- [21] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.
- [22] Renuka S Deshmukh and Vandana Jagtap. A survey: Software api and database for emotion recognition. In *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 284–289. IEEE, 2017.
- [23] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [24] Maria Egger, Matthias Ley, and Sten Hanke. Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science*, 343:35–55, 2019.
- [25] Paul Ekman, Wallace V Freisen, and Sonia Ancoli. Facial signs of emotional experience. *Journal of personality and social psychology*, 39(6):1125, 1980.
- [26] Alessandro Luiz Stamatto Ferreira, Leonardo Cunha de Miranda, Erica Esteves Cunha de Miranda, and Sarah Gomes Sakamoto. A survey of interactive systems based on brain-computer interfaces. *Journal on Interactive Systems*, 4(1), 2013.
- [27] Bernhard Graimann, Brendan Allison, and Gert Pfurtscheller. Brain–computer interfaces: A gentle introduction. In *Brain-computer interfaces*, pages 1–27. Springer, 2009.
- [28] Christoph Guger, Gunther Krausz, Brendan Z Allison, and Guenter Edlinger. Comparison of dry and gel based electrodes for p300 brain–computer interfaces. *Frontiers in neuroscience*, 6:60, 2012.
- [29] Yu-Liang Hsu, Jeen-Shing Wang, Wei-Chun Chiang, and Chien-Han Hung. Automatic ecg-based emotion recognition in music listening. *IEEE Transactions on Affective Computing*, 11(1):85–99, 2020.
- [30] Deepak Kumar Jain, Pourya Shamsolmoali, and Paramjit Sehdev. Extended deep neural network for facial emotion recognition. *Pattern Recognition Letters*, 120:69–74, 2019.

- [31] Neha Jain, Shishir Kumar, Amit Kumar, Pourya Shamsolmoali, and Masoumeh Zareapoor. Hybrid deep neural networks for face emotion recognition. *Pattern Recognition Letters*, 115:101–106, 2018.
- [32] Valer Jurcak, Daisuke Tsuzuki, and Ippeita Dan. 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage*, 34(4):1600–1611, 2007.
- [33] Eric R Kandel, James Harris Schwartz, and Thomas M Jessell. *Neurobiology of Sensation and Perception (from Principles of Neural Science by Eric Kandel): PSYC 452*. McGraw-Hill Education Create, 2014.
- [34] Kyung Hwan Kim, Seok Won Bang, and Sang Ryong Kim. Emotion recognition system using short-term monitoring of physiological signals. *Medical and biological engineering and computing*, 42:419–427, 2004.
- [35] Stephen C Kleene et al. Representation of events in nerve nets and finite automata. *Automata studies*, 34:3–41, 1956.
- [36] Byoung Chul Ko. A brief review of facial emotion recognition based on visual information. *sensors*, 18(2):401, 2018.
- [37] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.
- [38] Naomi S Kort, Pablo Cuesta, John F Houde, and Srikantan S Nagarajan. Bihemispheric network dynamics coordinating vocal feedback control. *Human brain mapping*, 37(4):1474–1485, 2016.
- [39] KL Leenders, D Perani, AA Lammertsma, JD Heather, P Buckingham, T Jones, MJR Healy, JM Gibbs, RJS Wise, J Hatazawa, et al. Cerebral blood flow, blood volume and oxygen utilization: normal values and effect of age. *Brain*, 113(1):27–47, 1990.
- [40] Dietrich Lehmann, PL Faber, Peter Achermann, Daniel Jeanmonod, Lorena RR Gianotti, and Diego Pizzagalli. Brain sources of eeg gamma frequency during volitionally meditation-induced, altered states of consciousness, and experience of the self. *Psychiatry Research: Neuroimaging*, 108(2):111–121, 2001.
- [41] Jacqueline S Leventon, Jennifer S Stevens, and Patricia J Bauer. Development in the neurophysiology of emotion processing and memory in school-age children. *Developmental cognitive neuroscience*, 10:21–33, 2014.
- [42] Yang Li, Wenming Zheng, Lei Wang, Yuan Zong, and Zhen Cui. From regional to global brain: A novel hierarchical spatial-temporal neural network model for eeg emotion recognition. *IEEE Transactions on Affective Computing*, 13(2):568–578, 2019.
- [43] Yuan-Pin Lin, Chi-Hong Wang, Tzyy-Ping Jung, Tien-Lin Wu, Shyh-Kang Jeng, Jeng-Ren Duann, and Jyh-Horng Chen. Eeg-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, 57(7):1798–1806, 2010.

- [44] Clas Linnman, Eric A Moulton, Gabi Barmettler, Lino Becerra, and David Borsook. Neuroimaging of the periaqueductal gray: state of the field. *Neuroimage*, 60(1):505–522, 2012.
- [45] Yong-Jin Liu, Minjing Yu, Guozhen Zhao, Jinjing Song, Yan Ge, and Yuanchun Shi. Real-time movie-induced discrete emotion recognition from eeg signals. *IEEE Transactions on Affective Computing*, 9(4):550–562, 2017.
- [46] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.
- [47] A Lucier. Music for solo performer (1965), for enormously amplified brain waves and percussion. *Lovely Music, Ltd*, 1982.
- [48] Antoine Lutz, Lawrence L Greischar, Nancy B Rawlings, Matthieu Ricard, and Richard J Davidson. Long-term meditators self-induce high-amplitude gamma synchrony during mental practice. *Proceedings of the national Academy of Sciences*, 101(46):16369–16373, 2004.
- [49] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113, 2014.
- [50] Arturo Nakasone, Helmut Prendinger, and Mitsuru Ishizuka. Emotion recognition from electromyography and skin conductance. In *Proc. of the 5th international workshop on biosignal interpretation*, pages 219–222. Citeseer, 2005.
- [51] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil. Brain computer interfaces, a review. *sensors*, 12(2):1211–1279, 2012.
- [52] Ernst Niedermeyer. Alpha rhythms as physiological and abnormal phenomena. *International journal of psychophysiology*, 26(1-3):31–49, 1997.
- [53] Lauri Nummenmaa, Enrico Glerean, Riitta Hari, and Jari K Hietanen. Bodily maps of emotions. *Proceedings of the National Academy of Sciences*, 111(2):646–651, 2014.
- [54] Jiapu Pan and Willis J. Tompkins. A real-time qrs detection algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3):230–236, 1985.
- [55] Pallavi Pandey and KR Seeja. Subject independent emotion recognition from eeg using vmd and deep learning. *Journal of King Saud University-Computer and Information Sciences*, 34(5):1730–1738, 2022.
- [56] RW Picard. Affective computing cambridge. *MA: MIT Press [Google Scholar]*, 1997.
- [57] Mario A. Quiroz-Juárez, Omar Jiménez-Ramírez, Rubén Vázquez-Medina, Elena Ryzhii, Maxim Ryzhii, and José L. Aragón. Cardiac conduction model for generating 12 lead ecg signals with realistic heart rate dynamics. *IEEE Transactions on NanoBioscience*, 17(4):525–532, 2018.
- [58] Frank Rosenblatt. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.
- [59] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

- [60] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.
- [61] Saeid Sanei. *Adaptive processing of brain signals*. John Wiley & Sons, 2013.
- [62] Donald L Schomer and Fernando Lopes Da Silva. *Niedermeyer’s electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2012.
- [63] Nicu Sebe, Michael S Lew, Yafei Sun, Ira Cohen, Theo Gevers, and Thomas S Huang. Authentic facial expression analysis. *Image and Vision Computing*, 25(12):1856–1863, 2007.
- [64] Margitta Seeck, Laurent Koessler, Thomas Bast, Frans Leijten, Christoph Michel, Christoph Baumgartner, Bin He, and Sándor Beniczky. The standardized eeg electrode array of the ifcn. *Clinical neurophysiology*, 128(10):2070–2077, 2017.
- [65] Michelle N Shiota. Ekman’s theory of basic emotions. *The Sage encyclopedia of theory in psychology*, pages 248–250, 2016.
- [66] Shiv Naresh Shivhare and Saritha Khethawat. Emotion detection from text. *arXiv preprint arXiv:1205.4944*, 2012.
- [67] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinyi Yang. A review of emotion recognition using physiological signals. *Sensors*, 18(7):2074, 2018.
- [68] Denise M Sloan. Emotion regulation in action: Emotional reactivity in experiential avoidance. *Behaviour research and therapy*, 42(11):1257–1270, 2004.
- [69] Mohammad Soleymani, Sadjad Asghari-Esfeden, Yun Fu, and Maja Pantic. Analysis of eeg signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing*, 7(1):17–28, 2015.
- [70] Tengfei Song, Wenming Zheng, Peng Song, and Zhen Cui. Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, 11(3):532–541, 2018.
- [71] André Stuhlsatz, Christine Meyer, Florian Eyben, Thomas Zielke, Günter Meier, and Björn Schuller. Deep neural networks for acoustic emotion recognition: Raising the benchmarks. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5688–5691, 2011.
- [72] Syahrull Hi Fi Syam, Heba Lakany, RB Ahmad, and Bernard A Conway. Comparing common average referencing to laplacian referencing in detecting imagination and intention of movement for brain computer interface. In *MATEC Web of Conferences*, volume 140, 2017.
- [73] Jianhua Tao, Tieniu Tan, and Rosalind W Picard. *Affective Computing and Intelligent Interaction: First International Conference, ACII 2005, Beijing, China, October 22-24, 2005, Proceedings*, volume 3784. Springer, 2005.

- [74] William O Tatum IV. *Handbook of EEG interpretation*. Springer Publishing Company, 2021.
- [75] Christiana Tsiourti, Astrid Weiss, Katarzyna Wac, and Markus Vincze. Multimodal integration of emotional signals from voice, body, and context: Effects of (in)congruence on emotion recognition and attitudes towards robots. *International Journal of Social Robotics*, 11, 08 2019.
- [76] Mario Tudor, Lorainne Tudor, and Katarina Ivana Tudor. Hans berger (1873-1941)—the history of electroencephalography. *Acta medica Croatica: casopis Hrvatske akademije medicinskih znanosti*, 59(4):307–313, 2005.
- [77] Gary R VandenBos. *APA dictionary of psychology*. American Psychological Association, 2007.
- [78] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [79] Guy Vingerhoets, Celine Berckmoes, and Nathalie Stroobant. Cerebral hemodynamics during discrimination of prosodic and semantic emotion in speech studied by transcranial doppler ultrasonography. *Neuropsychology*, 17(1):93, 2003.
- [80] Stephan Waldert. Invasive vs. non-invasive neuronal signals for brain-machine interfaces: will one prevail? *Frontiers in neuroscience*, 10:295, 2016.
- [81] Zhe Wang, Yongxiong Wang, Chuanfei Hu, Zhong Yin, and Yu Song. Transformers for eeg-based emotion recognition: A hierarchical spatial information learning model. *IEEE Sensors Journal*, 22(5):4359–4368, 2022.
- [82] Bernard Widrow and Marcian E Hoff. An adaptive "adaline" neuron using chemical "memistors". *Proceedings of the Institute of Radio Engineers*, 48(12):1415–1432, 1960.
- [83] Zhong Yin, Lei Liu, Jianing Chen, Boxi Zhao, and Yongxiong Wang. Locally robust eeg feature selection for individual-independent emotion recognition. *Expert Systems with Applications*, 162:113768, 2020.
- [84] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6023–6032, 2019.
- [85] Aston Zhang, Zachary C Lipton, Mu Li, and Alexander J Smola. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021.
- [86] Wenming Zheng. Multichannel eeg-based emotion recognition via group sparse canonical correlation analysis. *IEEE Transactions on Cognitive and Developmental Systems*, 9(3):281–290, 2017.
- [87] Guilherme Zimeo Morais, Joana Balardin, and João Sato. Fnirs optodes' location decider (fold): A toolbox for probe arrangement guided by brain regions-of-interest. *Scientific Reports*, 8, 02 2018.

«««