ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Πολυτροπική εκτίμηση της κατάθλιψης με χρήση βαθιάς μάθησης μέσα από εξόρυξη οπτικοακουστικής και σημασιολογικής πληροφορίας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΠΟΛΥΞΕΝΗΣ ΚΑΛΛΙΓΑ

Επιβλέπων : Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π

ΕΡΓΑΣΤΗΡΙΟ ΕΥΦΥΩΝ ΣΥΣΤΗΜΑΤΩΝ

Αθήνα, Ιούλιος 2020

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Πολυτροπική εκτίμηση της κατάθλιψης με χρήση βαθιάς μάθησης μέσα από εξόρυξη οπτικοακουστικής και σημασιολογικής πληροφορίας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΠΟΛΥΞΕΝΗΣ ΚΑΛΛΙΓΑ

Επιβλέπων : Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 9η Ιουλίου 2020.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Ανδρέας Σταφυλοπάτης	Γιώργος Στάμου	Παναγιώτης Τσανάκας
Καθηγητής Ε.Μ.Π	Καθηγητής Ε.Μ.Π	Καθηγητής Ε.Μ.Π

Αθήνα, Ιούλιος 2020

(Υπογραφή)

.....

Πολυξένη Καλλιγά

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περίληψη

Αντικείμενο αυτής της διπλωματικής εργασίας είναι η ανάπτυξη ενός βέλτιστου μοντέλου για την πολυτροπική εκτίμηση της ανθρώπινης κατάθλιψης μέσα από βιντεοσκοπημένες συνεδρίες, με τεχνικές βαθιάς μηχανικής μάθησης πάνω σε εκφράσεις του προσώπου των ασθενών και χαρακτηριστικά της ομιλίας, πέρα απο την αξιοποίηση της ανάλυσης του λόγου σε κείμενο. Σκοπός της πολυτροπικής αυτής προσέγγισης είναι η έκβαση ενός αξιόπιστου αποτελέσματος που δεν βασίζεται αποκλειστικά στην ανάλυση της φυσικής γλώσσας του ασθενή αλλα σε πιο ισχυρά φυσικά ενδεικτικά μέσα έκφρασης (στοιχεία ομιλίας και εκφράσεων προσώπου) που καταπολεμούν το πρόβλημα της διαστρέβλωσης του αποτελέσματος από τον ανθρώπινο παράγοντα. Συγκεκριμένα, θα συγκρίνουμε τα σχετικά πλεονεκτήματα των διάφορων προσεγγίσεων με στόχο να πετύχουμε σημαντικές βελτιώσεις στην εκτίμησης της κατάθλιψης μέσω των PHQ8 τιμών [9] οι οποίες και έχουν καταγραφεί πριν γίνει οποιαδήποτε αλληλεπίδραση ασθενή και εικονικού ψυχολόγου. Για να καθορίσουμε λοιπόν σε τι βαθμό η συγχώνευση διαφορετικών προσεγγίσεων πάνω στα διάφορα χαρακτηριστικά έκφρασης της κατάθλιψης είναι δυνατή και αποτελεσματική, θα αξιοποιήσουμε μια μεγάλη ποικιλία περιγραφικών χαρακτηριστικών για κάθε κατηγορία (ήχος, εικόνα, κείμενο) , ακολουθούμενα απο τεχνικές μηχανικής μάθησης καθώς επίσης και ποικίλα μοντέλα βαθιάς μάθησης.

Πιο συγκεκριμένα, τα πειράματά μας εκτελούνται πάνω στο αποθετήριο δεδομένων Distress Analysis Interview Corpus-Wizard of Oz (DAIC-WOZ) [1]. Αρχικά, οι πιο πρόσφατες και εξελιγμένες τεχνολογίες και έρευνες στην επιστήμη της οπτικοακουστικής αναγνώρισης ψυχικών διαταραχών και της ανάλυσης κειμένου βάση του περιεχομένου διερευνώνται λεπτομερώς και οι πιο σημαντικές προσεγγίσεις καταγράφονται. Ακολούθως, παρατίθενται οι βασικές θεωρητικές αρχές, πάνω στις οποίες βασίζεται η προτεινόμενη προσέγγισή μας στο πρόβλημα, καθώς και διάφορες δοκιμές που αποδείχθηκαν λιγότερο αποδοτικές.

Έπειτα, αναπτύσσεται το στάδιο της προεπεξεργασίας των δεδομένων για κάθε μια από τις τρεις κατηγορίες που έχουν εξαχθεί με σκοπό την βέλτιστη προετοιμασία τους πριν εισαχθούν σε έναν Baseline εκτιμητή για τον υπολογισμό του βαθμού της κατάθλιψης που πάσχει, αν και εφόσον πάσχει ο κάθε ασθενής. Στο συγκεκριμένο Baseline μοντέλο το τελικό PHQ8 σκορ υπολογίζεται από τον μέσο όρο των αποτελεσμάτων που προκύπτουν από τα επιμέρους μοντέλα του ήχου και του βίντεο.

Τέλος, προτείνουμε ένα υβριδικό μοντέλο ταξινόμησης και αξιολόγησης της κατάθλιψης εκμεταλλευόμενοι τους περιγραφητές ήχου, βίντεο και κειμένου που εξήγαμε προηγουμένως, με την εξής δομή: 1) Ένα Συνελικτικό Βαθύ Νευρωνικό Δίκτυο για κάθε ένα από τους περιγραφητές ήχου, εικόνας και κειμένου για την αξιολόγηση του PHQ8 score κατάθλιψης, ξεχωριστά για τα δύο φύλα (άνδρας, γυναίκα) όσων αφορά τον ήχο και το βίντεο, 2) Ένα μοντέλο ανάλυσης περιεχομένου των απομαγνητοφωνημένων συνεντεύξεων και ταξινόμησης με χρήση Random Forest για την παρουσία ή μη του φαινομένου της κατάθλιψης, 3) Ένα πολυπαραγοντικό μοντέλο παλινδρόμησης που συνδυάζει τις εκτιμήσεις των PHQ8 scores από τα μοντέλα παλινδρόμησης των 3 περιγραφητών καθώς επίσης και του ταξινομητή από την ανάλυση περιεχομένου. Τα αποτελέσματά μας υποδεικνύουν πως η προτεινόμενη προσέγγισή μας επιφέρει αξιόλογα αποτελέσματα, μάλιστα τα βέλτιστα μέχρι στιγμής, επιτυγχάνοντας τις ακόλουθες τιμές στις μετρικές που

αξιολογήθηκαν: ρίζα μέσου τετραγωνικού σφάλματος RMSE=4.543 και μέσο απόλυτο σφάλμα MAE=3.347 στο σύνολο των testing δεδομένων, αρκετά πιο βέλτιστα από τα αποτελέσματα του Baseline μοντέλου που φέρει τις μετρικές: RMSE=7.050 και MAE=5.660, γεγονός που χρήζει την προσέγγιση μας άξια σύγκρισης με σχετικές μεθοδολογίες στο πεδίο της αξιολόγησης της κατάθλιψης. Στην προσέγγιση αυτή έχουμε ως στόχο να αξιοποιήσουμε την γενικότερη συμπεριφορά των δεδομένων μας ώστε να αντιμετωπίσουμε το πρόβλημα της υπερεκπαίδευσης που φαίνεται να αντιμετωπίζουν οι παλαιότερες προσεγγίσεις.

Το υποκείμενο κίνητρο αυτής της μελέτης είναι η ανάγκη να βελτιστοποιήσουμε το πρόβλημα της αξιολόγησης της κατάθλιψης και κατ' επέκτασιν της γενικότερης αναγνώρισης συναισθήματος από πολυτροπική εξόρυξη δεδομένων, σε ένα επίπεδο όπου οι συμπεριφορές κατά την αλληλεπίδραση ανθρώπου μεταξύ ανθρώπου ή ανθρώπου με μηχανής να μπορούν να αναγνωριστούν αξιόπιστα σε πραγματικές συνθήκες από έναν εγκέφαλο υπολογιστή, χωρίς να βασίζονται αποκλειστικά στην ανάλυση της φυσικής γλώσσας αλλα σε πιο ισχυρά ενδεικτικά στοιχεία που καταπολεμούν το πρόβλημα της διαστρέβλωσης της έκβασης του αποτελέσματος από την ανθρώπινη αντίληψη. Το ευρύτερο αυτό πρότζεκτ ξεκίνησε ως μέρος μιας μεγαλύτερης προσπάθειας να δημιουργηθεί ένας πράκτορας-υπολογιστής που παίρνει συνέντευξη σε ανθρώπους και αναγνωρίζει λεκτικά και μη, στοιχεία ψυχικής διαταραχής χωρίς να υπάρχει κίνδυνος να επηρεαστεί η έκβαση του αποτελέσματος από τον υποκειμενικό παράγοντα της ανθρώπινης κρίσης.

Λέξεις Κλειδιά

Μηχανική μάθηση, βαθιά μάθηση, νευρωνικά δίκτυα, συνελικτικά δίκτυα, κατάθλιψη, φυσική επεξεργασία γλώσσας, ανάλυση περιεχομένου, πολυτροπικά μοντέλα, συγχώνευση μοντέλων, υβριδικός αλγόριθμος, επιβλεπόμενη μάθηση, ταξινόμηση, παλινδρόμηση, random forest, ακουστικά χαρακτηριστικά, οπτικά χαρακτηριστικά

Abstract

The scope of this thesis is the development of an optimal model for the multimodal assessment of human depression through recorded interviews, building Deep Learning models with the use of facial and prosodic features, besides speech to text analysis. The purpose of this multimodal approach is the outcome of a reliable conclusion that do not jeopardize of being distorted by the misdirection of the analysis of the natural language processing, but more powerful and robust features (audio and video) determine the outcome of depression diagnosis. In particular, in this work we compare the relative merits of the various approaches to advance depression estimation by means of PHQ-8 scores[1] recorded prior to every human-agent interaction. In order to establish to what extent fusion of the approaches is possible and beneficial, we deal through this with a great variety of handcrafted feature descriptors for each modality (audio, video and text) followed by machine learning techniques as well as with various deep neural network models.

In particular, experiments are carried out on the Distress Analysis Interview Corpus-Wizard of Oz (DAIC-WOZ) database [1]. Initially, the up-to-date technologies in the fields of audiovisual mental disorders recognition and context-aware analysis are being thoroughly explored through bibliographical research and the most important approaches are documented. Following this, the basic theoretical principles and technologies, on which our proposed approach is based, as well as the potential successful tries that are made, are presented.

Second, the preprocessing of extracted data is presented and examined for the most efficient data representation for each of the three modules to be used in the baseline random forest regressor to compute depression severity, where the final PHQ8 score is computed through the fusion of audio and video modalities by averaging the regression outputs of the unimodal random forest regressors.

Finally, we propose a hybrid depression classification and depression estimation framework from audio, video and text descriptors which contains three main components: 1) Deep Convolutional Neural Network (DCNN) based audio visual and text multi-modal depression recognition frameworks, trained with male and female participants (only for audio visual), respectively; 2) Content Analysis and Random Forest based depression classification framework from the interview transcripts; 3) A multivariate regression model fusing the audio visual and text PHQ-8 estimations from the DCNN models, and the depression classification result from the text information. In the DCNN based depression estimation framework, audio/video/text feature descriptors are first input into a DCNN to learn high-level features, which are then fed to the multivariate regression model to predict the final PHQ-8 score. The results show that the proposed depression recognition framework obtains very promising results, actually the best ones till now, with root mean square error (RMSE) as 4.543 and mean absolute error (MAE) as 3.347 on the test set, which are all lower than the baseline results with RMSE as 7.050 and MAE as 5.660, worthy of comparison with relevant methodologies in the challenging field of depression assessment. In this approach we aim to leverage the more generic representation of data to tackle the

[1] Personal Health Questionnaire Depression Scale (PHQ-8)

problem of overfitting that most of the other similar methodologies could not address effectively.

The underlying motivation for this work is the need to advance depression estimation and by extension general emotion recognition for multimedia retrieval to a level where behaviours expressed during human-human, or human-agent interactions, can be reliably sensed in real-life conditions, without jeopardizing of being distorted by the misdirection of the analysis of natural language processing, but instead, being defined by more powerful and robust features (audio and video). This project was implemented as part of a larger effort to create a computer agent that interviews people and identifies verbal and non-verbal indicators of mental illness without jeopardizing of being affected by the subjective factor of human intervention.

Keywords

Machine learning, deep learning, neural networks, convoluted networks, depression, natural language processing, content analysis, multi-modal frameworks, model fusion, hybrid algorithm, supervised learning, classification, regression, random forest regressor, prosodic features, visual features

Ευχαριστίες

Θα ήθελα αρχικά να ευχαριστήσω τον επιβλέποντα καθηγητή αυτής της εργασίας, κύριο Ανδρέα Σταφυλοπάτη που μου έδωσε την δυνατότητα να εκπονήσω την διπλωματική μου εργασία στο Εργαστήριο Ευφυών Συστημάτων του τομέα Τεχνολογίας Πληροφορικής και Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

Ιδιαίτερες ευχαριστίες θα ήθελα να δοσω στον κύριο Δρ. Γεώργιο Σιόλα για την πολύτιμη στήριξη και καθοδήγησή του καθ' όλη την διάρκεια της εκτέλεσης της διπλωματικής εργασίας, καθώς και την άμεση ανταπόκριση του σε όποια δυσκολία αντιμετώπιζα. Θα ήθελα επίσης να τον ευχαριστήσω θερμά για την προθυμία του να με συμβουλεύσει για την συνέχιση των σπουδών μου μετά το προπτυχιακό επίπεδο, καθώς θεωρώ ιδιαίτερα σημαντική την άποψη του.

Στη συνέχεια θα ήθελα να ευχαριστήσω τους καθηγητές Γεώργιο Στάμου και Παναγιώτη Τσανάκα που συμμετείχαν στην τριμελή επιτροπή της διπλωματικής μου εργασίας.

Κλείνοντας θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου που είναι δίπλα μου και με στηρίζουν όλα αυτά τα χρόνια.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

1. Εισαγωγή

Με την πάροδο των χρόνων οι εξελίξεις στην αλληλεπίδραση του ανθρώπου με τη μηχανή έχουν επιφέρει την έκρηξη του ενδιαφέροντος των επιστημόνων γύρω από τον κλάδο της αυτόματης αναγνώρισης της συναισθηματικής κατάστασης του ανθρώπου (Affective Computing), ένας κλάδος που φέρει πρωτεύοντα ρόλο σε οποιαδήποτε επιστήμη κοινωνικού περιεχομένου (διαφήμιση, περίθαλψη, ψυχαγωγία, πολιτικές ανακατευθύνσεις κλπ.). Το συναίσθημα έχει πρωτεύοντα ρόλο στις διαπροσωπικές σχέσεις των ανθρώπων και περιέχει σημαντική πληροφορία για την κατάσταση του ατόμου, τις προθέσεις του και τις πράξεις του στο άμεσο μέλλον. Τι είναι το συναίσθημα λοιπόν στην ψυχολογία...Είναι η εξωτερίκευση της ψυχικής κατάστασης που βιώνει το άτομο και έχει προκληθεί από κάποιο/α ερέθισμα/τα που προκάλεσε έντονες ψυχολογικές μεταβολές. Και πώς αντικατοπτρίζεται αυτή στο περιβάλλον; Τόσο ο τόνος και η ένταση της φωνής, όσο και οι εκφράσεις του προσώπου σε συνδυασμό με την λεκτική πληροφορία του ομιλητή μπορούν να επιφέρουν την πλήρη αποσαφήνιση της διανοητικής του κατάστασης. Ο Beethoven, αφού είχε γίνει κουφός, δήλωσε πως μπορούσε να κρίνει από την έκφραση του προσώπου ενός εκτελεστή, αν ερμήνευε το μουσικό του κομμάτι στον σωστό τόνο!! [2]

Οι ραγδαίες λοιπόν τεχνολογικές, οικονομικές αλλα και κοινωνικές εξελίξεις στον κόσμο μας έφεραν μεγάλες διακυμάνσεις στην ψυχολογία των ανθρώπων, δίνοντας στην ασθένεια της κατάθλιψης μια κυρίαρχη θέση στις σύγχρονες κοινωνίες. Τι είναι όμως η κατάθλιψη; Είναι μια κατάσταση αδιαθεσίας και αποτροπής από κάθε δραστηριότητα που μπορεί να επηρεάσει τις σκέψεις του ατόμου, τη συμπεριφορά, τα αισθήματα και γενικώς την αίσθηση της ευεξίας. Έτσι κατέστη αναγκαία και η εφαρμογη του Affective Computing στην αναγνώριση άμεσων αλλά και έμμεσων σημαδιών ψυχικών διαταραχών μέσα από μια σειρά αλυσιδωτών υπολογιστικών διεργασιών που ανήκουν στους κλάδους της Μηχανικής Μάθησης, της Τεχνητής Νοημοσύνης και της επεξεργασίας σημάτων συμπεριφοράς. Η αναγνώριση τους από την ανθρώπινη συνείδηση αδυνατούσε να φέρει πλήρη ακριβή συμπεράσματα είτε λόγω φυσικών εμποδίων, όπως η υποκειμενικότητα και η διαφορετική προσωπική αντίληψη τόσο του ασθενή όσο και του ακροατή, είτε λόγω μη έγκαιρης διάγνωσής τους.

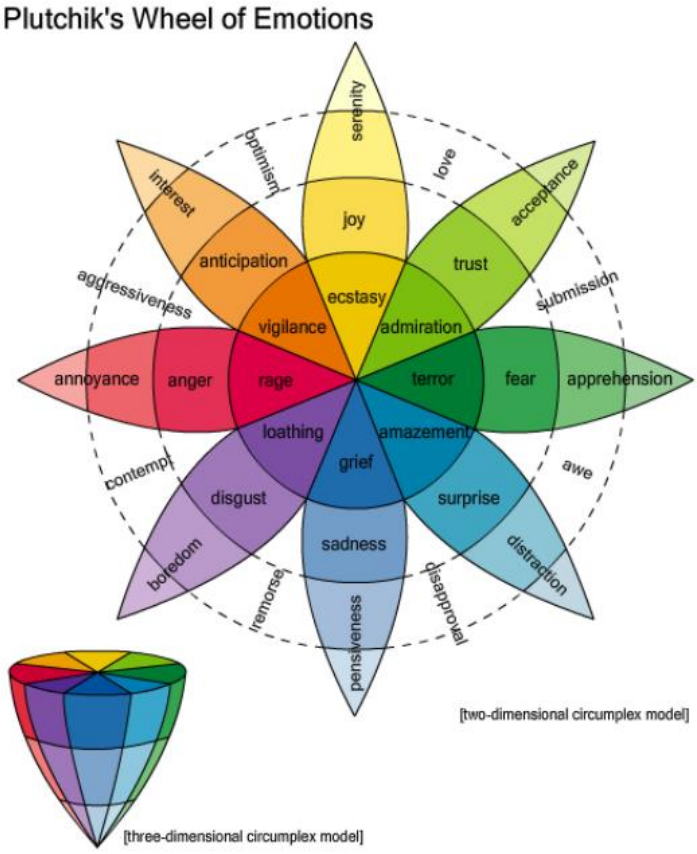
1.1 Ο ρόλος της Μηχανικής Μάθησης στην Κατάθλιψη

Τι είναι Μηχανική μάθηση; Μηχανική μάθηση είναι υποπεδίο της επιστήμης των υπολογιστών που αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη. Το 1959, ο Άρθουρ Σάμιουελ ορίζει τη μηχανική μάθηση ως "Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί". Η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μαθαίνουν από τα δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά. Τέτοιοι αλγόριθμοι λειτουργούν κατασκευάζοντας μοντέλα από πειραματικά δεδομένα, προκειμένου να κάνουν προβλέψεις βασιζόμενες στα δεδομένα ή να εξάγουν αποφάσεις που εκφράζονται ως το αποτέλεσμα.[3] Η κατηγοριοποίησης αυτών των δεδομένων, η αποτύπωση των μεταξύ τους σχέσεων και οι ιδιότητές τους αποτελούν βάση της διαδικασίας αυτής. Για αυτόν τον λόγο έχει αναπτυχθεί ένας ξεχωριστός υπό-κλάδος της τεχνητής νοημοσύνης, αυτός της μηχανικής μάθησης.

Όπως προαναφέραμε ο υποκλάδος στον οποίο συναντιούνται οι δυο κλάδοι της μηχανικής μάθησης και της ανάλυσης ανθρωπίνων συναισθημάτων και κατ' επέκτασιν των ψυχικών διαταραχών καλείται Affective Computing. Πριν την παρέμβαση του υπολογιστή στην ανάλυση συναισθημάτων, η διάγνωση της κατάθλιψης υστερούσε καθώς βασιζόταν στις εκτιμήσεις των αυτο-αναφορών των ασθενών και μόνο, κινδυνεύοντας η διάγνωση να διαστρεβλωθεί λόγω του υποκειμενικού της χαρακτήρα. [4] Με τις πρόσφατες όμως εξελίξεις του Affective Computing, μια μεγάλη ποικιλία νέων χαρακτηριστικών εισήχθη με σκοπό την καλύτερη εκτίμηση της κατάθλιψης και την συνεισφορά στην ακριβέστερη εκτίμηση της σοβαρότητας της κατάστασης του ασθενή από τον ψυχολόγο. Εκπαιδεύοντας επομένως τα κατάλληλα μοντέλα για την εξόρυξη κατάλληλων σχέσεων μεταξύ χαρακτηριστικών που έχουν αντληθεί τόσο από την ομιλία, όσο και από τις εκφράσεις του προσώπου, χωρίς να λείπει η εξόρυξη πληροφορίας από τον λόγο, η αυτοματοποίηση της ανίχνευσης της κατάθλιψης, καθώς και άλλων ψυχικών ασθενειών, εισέρχεται επιτυχώς στη διαδικασία διάγνωσης ψυχικών διαταραχών ευρύτερα.

1.2 Αναγνώριση Συναισθήματος στην Τεχνολογία

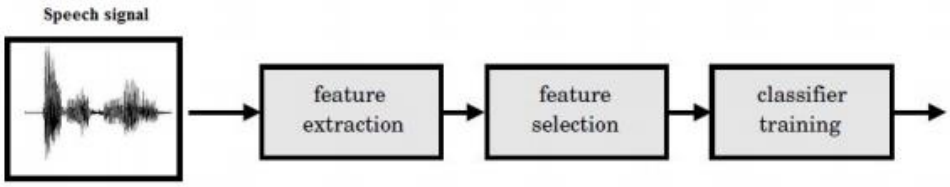
Για την ερμηνεία της μεγάλης ποικιλίας των διαφορετικών συναισθημάτων που υπάρχουν, παρόμοια και με άλλους υποστηρικτές, ο Plutchik [5] υποστήριξε την θεωρία πως τα βασικά συναισθήματα είναι τετριμμένα και μπορούν να συγχωνευτούν για να προκύψουν παράγωγα συναισθήματα. Η προσέγγιση του περιγράφει τις σχέσεις μεταξύ των συναισθημάτων μέσω της έννοιας του χρώματος και των αναμίξεών του, σύμφωνα με έναν τροχό χρωμάτων (Σχ. 1.1). Τα 8 βασικα συναισθήματα, σύμφωνα με τη θεωρία αυτή, απεικονίζονται στον κύκλο, στο κέντρο του δισδιάστατου μοντέλου. Τα συναισθήματα στις άσπρες περιοχές προκύπτουν από την μίξη δύο βασικών συναισθημάτων. Στο τρισδιάστατο μοντέλο, η κατακόρυφη διάσταση του κώνου αναπαριστά την ένταση και ο κύκλος αναπαριστά το βαθμό ομοιότητας των συναισθημάτων. Διακρίνονται 8 τομείς, οι οποίοι αντιστοιχούν στις βασικές συναισθηματικές διαστάσεις, ως τέσσερα ζεύγη αντιθέτων.



Σχήμα 1.1: Plutchik's Wheel of Emotions

1.2.1 Αναγνώριση Συναισθήματος μέσω Φωνής

Οι άνθρωποι χρησιμοποιώντας ως μέσο μετάδοσης τη φωνή εκφράζουν την συναισθηματική τους κατάσταση. Τι είναι η φωνή όμως; Είναι κύματα πίεσης του αέρα που μπορούν να αποτυπωθούν στο χαρτί ως ένα ηλεκτρικό σήμα με τη διαδικασία της ηχογράφησης. Στη συνέχεια το ηλεκτρικό αυτό σήμα υφίσταται τις διαδικασίες της δειγματοληψίας και της διακριτοποίησης, όπου και προκύπτει το τελικό ψηφιακό σήμα φωνής με το οποίο εργαζόμαστε. Για να αναγνωρίσουμε λοιπόν συναισθήματα από ηλεκτρικά σήματα εργαζόμαστε σε ένα πρόβλημα αναγνώρισης προτύπων, στο οποίο στόχος είναι η εκπαίδευση ενός συστήματος μηχανικής μάθησης πάνω σε ένα σύνολο ηχογραφημένων στιγμιότυπων με γνωστό συναισθηματικό περιεχόμενο (Επιβλεπόμενη μάθηση [3]). Τα βασικά βήματα του pipeline απεικονίζονται στο Σχήμα 1.2. Το είδος των χαρακτηριστικών του ηχητικού σήματος που εξάγουμε αναφέρεται στην ενότητα 3.2.1.1.



Σχήμα 1.2: Στάδια αναγνώρισης προτύπων

Γενικά, η ομιλία κάθε ατόμου διαφέρει από κάποιου άλλου. Στοιχεία που διαφοροποιούν τις διάφορες φωνές είναι βιολογικά, όπως το φύλο και η ηλικία, κοινωνικό-πολιτισμικά, όπως η γλώσσα, αλλά σημαντική επιρροή έχει και η έκφραση συναισθήματος. Ωστόσο, η έκφραση αυτή δεν είναι η ίδια μεταξύ των ομιλητών, αφού κάθε ένας από αυτούς χαρακτηρίζεται από τον προσωπικό του χαρακτήρα, την κουλτούρα του και άλλα ιδιαίτερα στοιχεία. Όλα τα παραπάνω συμβάλλουν στη διαφοροποίηση του σήματος φωνής κάθε ατόμου, και άρα στα ακουστικά χαρακτηριστικά που εξάγονται από κάθε ηχητικό σήμα. Ετσι, η οποιαδήποτε διαφοροποίηση μεταξύ των ομιλητών θα οδηγήσει πιθανότητα σε χαμηλή απόδοση του συστήματος, αφού αυξάνεται η πιθανότητα λανθασμένης αναγνώρισης ή κοινώς του θορύβου. Για παράδειγμα, δύο άνθρωποι που βιώνουν το ίδιο συναίσθημα, όπως η λύπη, θα το εκφράσουν πιθανότατα με διαφορετικό τρόπο, ο οποίος εξαρτάται από το βαθμό της λύπης τους, το χαρακτήρα τους, αλλά και το κοινωνικό-πολιτισμικό περιβάλλον, μέσα στο οποίο έχουν μάθει να εκφράζονται. Επίσης, ασάφεια μπορεί να δημιουργηθεί και από την ομοιότητα των ακουστικών χαρακτηριστικών δύο διαφορετικών συναισθημάτων, όπως λύπη και πλήξη. Επιπλέον, ως στοιχείο διαφοροποίησης μπορεί να θεωρηθεί και το περιβάλλον ηχογράφησης κάθε ομιλητή, καθώς επιφέρει μεταβολές στο τελικό σήμα φωνής. Έτσι, σημαντικό ρόλο μπορεί να παίξει η αλλαγή της στάθμης της ενέργειας του σήματος φωνής εξαιτίας διαφορετικών συνθηκών ηχογράφησης, όπως η αύξηση της ενέργειας η οποία είναι αυτή που χαρακτηρίζει κάποια είδη συναισθήματος, όπως θυμό ή ενθουσιασμό. Όλα τα παραπάνω παραδείγματα δείχνουν μεταβολές που μπορεί να παρατηρηθούν σε ένα σήμα φωνής και αντίστοιχα να επηρεάσουν τα εξαγόμενα ακουστικά χαρακτηριστικά. Τέτοιες μεταβολές είναι γενικά ανεκτές από το ανθρώπινο σύστημα αναγνώρισης, καθώς εκπαιδεύεται διαρκώς σε ποικιλία ομιλητών στις καθημερινές αλληλεπιδράσεις. Επίσης, λόγω της πολυπλοκότητάς του, ο ανθρώπινος εγκέφαλος μπορεί να αντιλαμβάνεται πλήθος καινούργιων ερεθισμάτων. Όμως, δεν ισχύει το ίδιο για ένα αυτόματο σύστημα αναγνώρισης, ανεπτυγμένο σε έναν υπολογιστή. Αυτό, εκπαιδεύεται με βάση κάποιον περιορισμένο αριθμό δεδομένων και στη συνέχεια, καλείται να αναγνωρίσει το συναίσθημα ενός δεδομένου εισόδου, το οποίο συχνά θα προέρχεται από διαφορετικό ομιλητή ή περιβάλλον ηχογράφησης, συγκριτικά με τα δεδομένα εκπαίδευσης. Με αυτόν τον τρόπο, γίνεται κατανοητό ότι η διαφοροποίηση της φωνής μεταξύ των ομιλητών αποτελεί κρίσιμο στοιχείο της απόδοσης ενός συστήματος, ειδικά σε πραγματικές εφαρμογές όπου τα ηχητικά δεδομένα προέρχονται από διαφορετικούς ομιλητές ή ακουστικά περιβάλλοντα. Ωστόσο έχει αναπτυχθεί μια πληθώρα εργαλείων που προσπαθούν να εξαλείψουν αυτούς τους εξωτερικούς παράγοντες και να γενικεύσουν το πρόβλημα της αναγνώρισης συναισθήματος με ικανοποιητική ακρίβεια. Εκτενέστερη ανάλυση των ηχητικών σημάτων που θα αξιοποιήσουμε στην εργασία μας παρατίθεται στην παράγραφο 3.2.1.

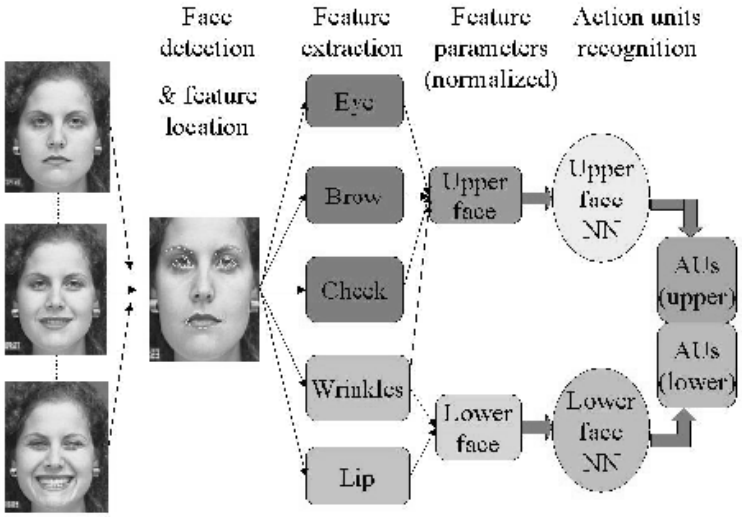
1.2.2 Αναγνώριση Συναισθήματος μέσω Έκφρασης Προσώπου

Οι άνθρωποι για να εκφράσουν την ψυχολογική τους κατάσταση εκτός της φωνής, χρησιμοποιούν ως μέσο και τις εκφράσεις των προσώπων τους. Ωστόσο τα περισσότερα συστήματα αυτόματης ανάλυσης έκφρασης προσπαθούν να αναγνωρίσουν ένα μικρό σύνολο συγκεκριμένων εκφράσεων όπως χαρά, θυμός, που ωστόσο εμφανίζονται σπάνια αυτούσιες σε συγκεκριμένα στιγμιότυπα και εντοπίζονται δύσκολα. Τα ανθρώπινα συναισθήματα εκφράζονται συνήθως μέσα από μετακινήσεις συγκεκριμένων διακριτών μυϊκών σημείων του προσώπου [7] . Ωστόσο σύμφωνα με την παραπομπή [6], ο Carl-Herman Hjortsjö ανέπτυξε ένα αυτόματο σύστημα (που στη συνέχεια βελτιστοποιήθηκε από επόμενους ερευνητές) για τον εντοπισμό «λεπτών» αλλαγών στις εκφράσεις του

προσώπου με βάση τα σταθερά μυϊκά και σκελετικά χαρακτηριστικά που διαθέτουν όλοι οι άνθρωποι (φρύδια, μάτια, στόμα), καθώς και παροδικά χαρακτηριστικά του προσώπου, όπως το βάθος των αυλακώσεων του προσώπου σε μετωπική πρόσοψη του προσώπου. Έτσι λοιπόν, σε αντίθεση με τα συμβατικά συστήματα, κατάφεραν το σύστημά τους να αναγνωρίζει λεπτομερείς αλλαγές στις εκφράσεις του προσώπου βασισμένο στο σύστημα κωδικοποίησης δράσης του προσώπου (FACS) [7], αντί για τις 6 (θυμός, φόβος, χαρά, λύπη, έκπληξη, αηδία) βασικές εκφράσεις των συμβατικών συστημάτων οι οποίες και ταυτοποιούνται με την εξαγωγή τοπικών περιγραφητών (local descriptors), γεγονός που σημαίνει μεγαλύτερη χρονική και τοπική πολυπλοκότητα, καθώς και μικρότερη ακρίβεια στα αποτελέσματα. Αυτά τα χαρακτηριστικά λοιπόν ,σε σύνολο 64, ονομάστηκαν Action Units (AU) και όπως παρουσιάζεται στην παραπομπή [7] , συνδυασμοί αυτών (πάνω από 7000 συνδυασμοί έχουν παρατηρηθεί) αλλα και ατομικά AU δίνουν μια πληθώρα από συναισθήματα και αντιδράσεις. Άλλα σημεία του προσώπου που δημιουργούν τα AU είναι τα χείλη, τα ζυγωματικά, οι ρυτίδες, τα λακκάκια και πολλά ακόμη. Οι τιμές που παίρνουν τα χαρακτηριστικά αυτά ανήκουν σε μια κλίμακα 0-5 και αντιπροσωπεύουν τον βαθμό σύσπασης του μυ που εξετάζεται.

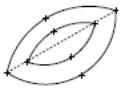
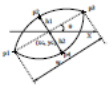
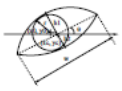
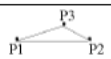
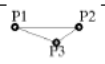
Επομένως αφού εξαχθούν τα απαιτούμενα AUs, μέσω της διαδικασίας feature extraction , προκύπτουν πρότυπα (patterns) τα οποία θα τροφοδοτηθούν σε κατάλληλο σύστημα μηχανικής μάθησης για την ανάλυση του συναισθήματος.

Επιπροσθέτως, ένα ακόμη πλεονέκτημα της τεχνικής αυτής είναι ότι καθορίζει το συναίσθημα του ατόμου μέσω των εκφράσεων του σε πραγματικό χρόνο (real-time), σε αντίθεση με τις περισσότερες τεχνικές αναγνώρισης συναισθήματος μέσω έκφρασης οι οποίες για να αναγνωρίσουν το συναίσθημα που παράγεται από κάποιο ερέθισμα πρεπει να προσθέσουν μια καθυστέρηση ανάμεσα σε αυτό και στο επόμενο ερέθισμα που προσδίδει μια ολική καθυστέρηση σε όλο το σύστημα [7].



Σχήμα 1.3: Περιγραφή Συστήματος Αναγνώρισης των AUs

Table 1. Multi-state facial component models of a front face

Component	State	Description/Feature
Lip	Opened	
	Closed	
	Tightly closed	
Eye	Open	
	Closed	
Brow	Present	
Cheek	Present	
Furrow	Present	
	Absent	

Σχήμα 1.4:Χαρακτηριστικά εντοπισμού των σημείων του προσώπου

Table 4. Basic upper face action units or AU combinations

AU 1	AU 2	AU 4
		
Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together
AU 5	AU 6	AU 7
		
Upper eyelids are raised.	Cheeks are raised.	Lower eyelids are raised.
AU 1+4	AU 4+5	AU 1+2
		
Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.	Inner and outer portions of the brows are raised.
AU 1+2+4	AU1+2+5+6+7	AU0(neutral)
		
Brows are pulled together and upward.	Brow, eyelids, and cheek are raised.	Eyes, brow, and cheek are relaxed.

Σχήμα 1.5: AUs και συνδυασμοί τους για την περιγραφή του πάνω μέρους του προσώπου

1.3 Αντικείμενο της Διπλωματικής Εργασίας

Στην παρούσα διπλωματική εργασία μελετούμε όλα τα στάδια μιας διαδικασίας pipeline για την δημιουργία ενός αποδοτικού συστήματος μηχανικής μάθησης που αξιολογεί την ανθρώπινη κατάθλιψη με τεχνικές βαθιάς μάθησης εκμεταλλευόμενο όχι ένα αλλά τρία διαφορετικά είδη πληροφορίας από τον άνθρωπο: τις εκφράσεις του προσώπου του, τους τόνους και την ένταση της ομιλίας του, καθώς και την ανάλυση του λόγου του σε κείμενο. Το πρόβλημα αυτό αποτελεί ένα ανοικτό πρόβλημα στο ερευνητικό πεδίο της Συναισθηματικής Ανάλυσης (Sentiment Analysis) λόγω της δυσκολίας να αντληθεί χρήσιμη πληροφορία από μια πληθώρα λεκτικών και μη, περιγραφικών χαρακτηριστικών με σκοπό την δημιουργία ενός ισχυρού μοντέλου που θα αξιολογεί με όσο το δυνατόν καλύτερη προσέγγιση τον βαθμό της κατάθλιψης (PHQ8) που πάσχει, εάν πάσχει ο ασθενής, απαλλαγμένο τόσο από το πρόβλημα της υπερεκπαίδευσης όσο και της υποεκπαίδευσης. Για πρώτη φορά το εισήγαγε ο οργανισμός The Audio/Visual Emotion Challenge and Workshop (AVEC2017) [10] διοργανώνοντας τον διαγωνισμό Real-life Depression and Affect καλώντας τους ερευνητές των επιστημών της αναγνώρισης συναισθήματος αλλά και της οπτικοακουστικής επεξεργασίας σήματος να συνεργαστούν για την προσέγγιση ενός ιδανικού μοντέλου αξιολόγησης της κατάθλιψης πάνω σε πραγματικά δεδομένα.

Το υποκείμενο κίνητρο αυτής της μελέτης είναι η ανάγκη να βελτιστοποιήσουμε το πρόβλημα της αξιολόγησης της κατάθλιψης και κατ' επέκτασιν της γενικότερης αναγνώρισης συναισθήματος από πολυτροπική εξόρυξη δεδομένων, σε ένα επίπεδο όπου οι συμπεριφορές κατά την αλληλεπίδραση ανθρώπου μεταξύ ανθρώπου ή ανθρώπου με μηχανής να μπορούν να αναγνωριστούν αξιόπιστα από πραγματικά δεδομένα σε real-life συνθήκες από έναν υπολογιστή. Είναι αντιληπτό επομένως ότι μια τέτοια προσέγγιση θα αντιμετωπίζει το πρόβλημα της μεγάλης απόκλισης των δεδομένων καθώς εξωτερικοί παράγοντες όπως φύλο, θόρυβος περιβάλλοντος, γλώσσα, πολιτισμός μπορούν να αμβλύνουν σημαντικά τα δεδομένα μας. Στη συγκεκριμένη περίπτωση λοιπόν, το αποθετήριο των δεδομένων μας αποτελείται από συνεντεύξεις ανθρώπων που έχουν μαγνητοσκοπηθεί κάτω από όσο το δυνατόν όμοιες συνθήκες περιβάλλοντος, μειώνοντας τον θόρυβο που θα κληθούμε να εξαλείψουμε από τα δεδομένα μας. Ωστόσο λόγω της υψηλής κατανάλωσης μνήμης που έχουν τα εξαγόμενα χαρακτηριστικά, το αποθετήριο δεδομένων μας αποτελείται από έναν ανεπαρκή αριθμό δειγμάτων, στο σύνολο 189, γεγονός που θέλει προσεκτική διαχείριση για την αποφυγή υπερεκπαίδευσης του μοντέλου μας στα δείγματα εκπαίδευσης.

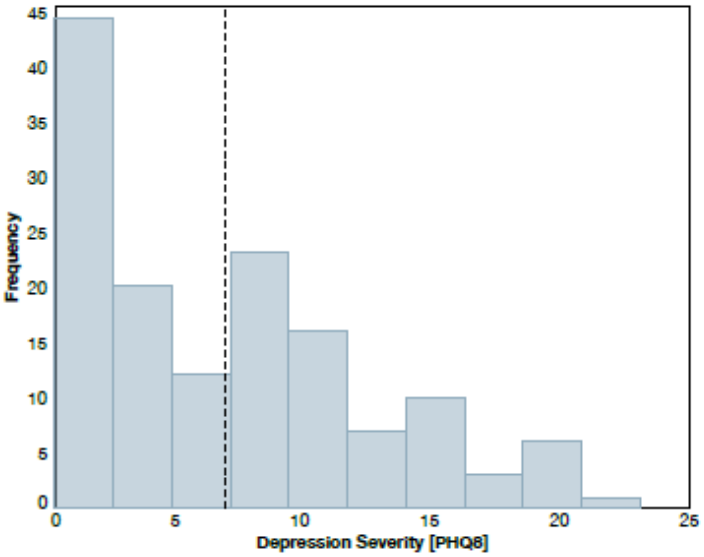
Συνολικά λοιπόν, στόχος της διπλωματικής αυτής εργασίας είναι η επίτευξη ενός ανταγωνιστικού μοντέλου αξιολόγησης της κατάθλιψης, απαλλαγμένου από το πρόβλημα της υπερεκπαίδευσης αλλα και με πλήρη εκμετάλλευση της πληροφορίας με χρήσιμο περιεχόμενο από τα τρία είδη μετάδοσης συναισθήματος του ανθρώπου, επιδιώκοντας μια χαμηλής πολυπλοκότητας προσέγγιση που θα μπορούσε να εφαρμοστεί σε μεγαλύτερο πλήθος δειγμάτων με την ίδια απόδοση. Έτσι θα κληθούμε να ερευνήσουμε κατά πόσο ο συνδυασμός περισσοτέρων από ενός είδους χαρακτηριστικών συνεισφέρει στην απόδοση του συστήματος και κυρίως θα επικεντρωθούμε στην απόδοση των οπτικοακουστικών στοιχείων καθώς αυτά είναι που αδυνατούν να ερμηνεύσουν οι ψυχολόγοι λόγω του έμμεσου χαρακτήρα τους, παρά τα λεκτικά χαρακτηριστικά που δίνουν άμεση και ξεκάθαρη ερμηνεία της κατάστασης του ασθενή αλλά ωστόσο μπορούν να διαστρεβλώσουν εύκολα την έκβαση της διάγνωσης .

1.4 Αποθετήριο Συνεντεύξεων Ανάλυσης Ψυχικών Διαταραχών-DAICWOZ

Στην παρούσα διπλωματική, η βάση δεδομένων πάνω στην οποία θα εκτελέσουμε τα πειράματα μας είναι η Distress Analysis Interview Corpus – Wizard of Oz (DAICWOZ) [1], το οποίο είναι μέρος ενός μεγαλύτερου αποθετηρίου δεδομένων, του Distress Analysis Interview Corpus (DAIC) το οποίο περιέχει κλινικές συνεντεύξεις κατάλληλα διευθετημένες ώστε να υποστηρίζουν την διάγνωση καταστάσεων ψυχολογικών διαταραχών, όπως το άγχος, η κατάθλιψη και μετατραυματικές διαταραχές άγχους. Οι συνεντεύξεις αυτές συλλέχθηκαν ως μέρος ενός ευρύτερου πρότζεκτ με απώτερο σκοπό να δημιουργηθεί ένας πράκτορας-υπολογιστής που παίρνει συνέντευξη σε ανθρώπους και αναγνωρίζει αυτόματα σε πραγματικές συνθήκες λεκτικά και μη, στοιχεία ψυχικής διαταραχής χωρίς να υπάρχει κίνδυνος να επηρεαστεί η έκβαση του αποτελέσματος από τον υποκειμενικό παράγοντα της ανθρώπινης κρίσης [8]. Το κάθε δείγμα συνέντευξης αποτελείται από τις καταγραφές ήχου και βιντεο, αλλά και από τις απομαγνητοφωνημένες ερωτοαπαντήσεις τα οποία προέρχονται απο τις συνεδρίες ψυχανάλυσης του κάθε ασθενή με έναν εικονικό πράκτορα, την Έλλη, ο οποίος βέβαια καθοδηγούνταν από άνθρωπο εκτός δωματίου. Επιπλέον το κάθε δείγμα αποτελείται από κάποια λεκτικά και μη, χαρακτηριστικά τα οποία έχουν εξαχθεί βάσει κάποιων Toolkits που θα αναφέρουμε στη συνέχεια. Αντιμετωπίζοντας ένα πρόβλημα επιβλεπόμενης μάθησης, παρατίθενται για κάθε συμμετέχοντα-δείγμα το PHQ8 score [9] του το οποίο είναι και η μετρική που θα αξιολογεί το σύστημα μας με τιμές στο διάστημα [0,24] (διατεταγμένο σύνολο τιμών) όπου για τιμές <10, ο ασθενής δεν πάσχει από κατάθλιψη, ενώ με >10 πάσχει. Τέλος, δίνονται και τα 8 διαφορετικα PHQ8 score (Σχ.1.6) των οποίων το άθροισμα δίνει το τελικό PHQ8 score και το καθένα αντιπροσωπεύει μια διαφορετική διαταραχή (έλλειψη ύπνου, ενδιαφέροντος κλπ.) με τιμές που κυμαίνονται απο 0-3 (όσο μεγαλύτερη τιμή τόσο πιο έντονο το πρόβλημα), καθώς και το φύλο του ασθενή.

How often during the past 2 weeks were you bothered by...	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4. Feeling tired or having little energy	0	1	2	3
5. Poor appetite or overeating	0	1	2	3
6. Feeling bad about yourself, or that you are a failure, or have let yourself or your family down	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed. Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3

Σχήμα 1.6: Παρουσίαση των 8 διαφορετικών PHQ8 score



Σχήμα 1.7: Κατανομή των PHQ8 score

1.5 Δομή της Διπλωματικής Εργασίας

Στην παρούσα διπλωματική εργασία αντιμετωπίζουμε το πρόβλημα της εκτίμησης της κατάθλιψης με ανάλυση του ηχητικού σήματος, των εκφράσεων του προσώπου και της πληροφορίας του λόγου του ασθενή ως ένα πρόβλημα που μπορεί να επιλυθεί αποτελεσματικά αξιοποιώντας χαρακτηριστικά ήχου, εικόνας και κειμένου. Αρχικά, στο κεφάλαιο 2 θα αναπτυχθούν έννοιες και ορισμοί της Μηχανικής Μάθησης ως θεωρητικό υπόβαθρο πάνω στο οποίο θα βασιστεί η δόμηση των συστημάτων Βαθιάς Μηχανικής Μάθησης της εργασίας. Στο κεφάλαιο 3 θα γίνει μια περιγραφή της δημιουργίας του συνόλου των δεδομένων εισόδου μέσα από την επεξεργασία ηχητικού σήματος, εικόνας και κειμένου, που προηγήθηκαν. Έπειτα, στο κεφάλαιο 4 θα αναλυθούν οι προτεινόμενες αρχιτεκτονικές και οι προτεινόμενες ενσωματώσεις δεδομένων που συνθέτουν τα συστήματα εκτίμησης και κατηγοριοποίησης της κατάθλιψης που υλοποιήσαμε. Ενώ, στο κεφάλαιο 5 θα εξηγηθεί λεπτομερώς η τελική υβριδική πολυτροπική προσέγγιση που εφαρμόστηκε για την συνένωση των επιμέρους προτεινόμενων συστημάτων. Τέλος, στο κεφάλαιο 6 καταλήγουμε στα βασικά συμπεράσματα της διπλωματικής εργασίας και δίνουμε προτάσεις που θα μπορούσαν να εφαρμοστούν στο μέλλον για την βελτίωση της επίδοσης των συστημάτων μας αλλά και για την χρήση τους σε άλλες εφαρμογές.

1.6 Συγγενείς Εργασίες

Στην παρούσα ενότητα θα παρουσιάσουμε παρόμοιες και συγγενείς έρευνες στον τομέα της Αναγνώρισης και Εκτίμησης της Κατάθλιψης, του οποίου η σημαντική θέση που καταλαμβάνει στα βασικά προβλήματα των σύγχρονων κοινωνιών έχει διεγείρει το ενδιαφέρον πολλών ερευνητών παγκοσμίως. Τα τελευταία χρόνια λοιπόν έχουν μελετηθεί ποικίλες μέθοδοι για το πρόβλημα της αναγνώρισης (classification) και εκτίμησης (regression) της κατάθλιψης. Ο Cohn κ.α. [11] επιχείρησαν την κλινική διάγνωση της

κατάθλιψης μέσω εκφράσεων του προσώπου και χαρακτηριστικών της φωνής εφαρμόζοντας στη συνέχεια την μέθοδο SVM (Μηχανές Διανυσμάτων Υποστήριξης) και της λογιστικής παλινδρόμησης (logistic regression) για το θετικό ή αρνητικό πόρισμα. Ο Nicholas κ.α. [12] επικεντρώθηκαν στην μελέτη και συνεισφορά διαφόρων χαρακτηριστικών της ομιλίας, αποδεικνύοντας πως ο συνδυασμός των Mel-frequency cepstral coefficient (MFCC) και των Formant χαρακτηριστικών πέτυχαν ένα ποσοστό 80% ακρίβειας στο πρόβλημα της ταξινόμησης της κατάθλιψης (depression classification). Πριν τον διαγωνισμό του AVEC2017, διεξήχθη ο AVEC2016 [13] ο οποίος εστίασε στο πρόβλημα ταξινόμησης της κατάθλιψης (depression classification), ενώ ο AVEC2017 εισήγαγε το πρόβλημα της εκτίμησης της κατάθλιψης (depression regression). Ο Ma κ.α. [14] πρότειναν ένα αντίστοιχα αποδοτικό μοντέλο βαθιάς μηχανικής μάθησης, γνωστό ως DeepAudioNet, το οποίο εκπαιδεύτηκε πάνω στα σχετικά με την κατάθλιψη χαρακτηριστικά για την έκβαση του αποτελέσματος της ταξινόμησης της κατάθλιψης. Η Pampouchidou κ.α. [15] πραγματοποίησαν το πρόβλημα του depression classification συγχωνεύοντας τα υψηλού επιπέδου και χαμηλού επιπέδου χαρακτηριστικά από τον ήχο, το βίντεο και το κείμενο. Στον διαγωνισμό του AVEC2014, όπου το ζητούμενο ήταν η εκτίμηση των BDI-II scores, οι Jain κ.α. [16] εστίασαν στους περιγραφητές των οπτικών χαρακτηριστικών και χρησιμοποίησαν την τεχνική Fisher Vector για να εκτιμήσουν τα επίπεδα κατάθλιψης.

Στη συνέχεια θα παρουσιάσουμε κάποιες αξιόλογες προσεγγίσεις του προβλήματος αυτού που συμμετείχαν στον διαγωνισμό του AVEC2017. Κατακτώντας την πρώτη θέση, ο Sun κ.α. [21] με την τεχνική της επιλεκτικής ανάλυσης περιεχομένου κειμένου (Selected-Text feature) πάνω στον αλγόριθμο εκτίμησης Random Forest Regression επιτύγχαναν την βέλτιστη μέχρι τότε τιμή μετρικής απόδοσης RMSE=4.98 και MAE=3.87 στα testing δεδομένα. Ακολούθως, στην δεύτερη θέση οι Gong κ.α. [22] προσέγγισαν μια μέθοδο επιλογής Topic (Topic Modeling) ταξινομώντας τα δεδομένα ήχου και βίντεο σε διαφορετικές ενότητες βασισμένες στο θέμα ανάλυσης του ασθενή, εξαγόμενο από τα transcripts πριν τροφοδοτηθούν σε έναν SGD Regressor, πετυχαίνοντας τελική απόδοση RMSE=4.99 και MAE=3.96,

Είναι αξιοσημείωτο όμως πως στις παραπάνω μεθόδους μελετήθηκε είτε το πρόβλημα του classification, είτε του regression της κατάθλιψης μεμονωμένα. Ωστόσο, οι Le Yang κ.α. [17], που κατέκτησαν την τρίτη θέση στον AVEC2017, μελέτησαν τον συνδυασμό των 2 προβλημάτων του classification και depression αξιοποιώντας και τα τρία είδη δεδομένων (ήχος, βίντεο, κείμενο) μέσω προσεγγίσεων βαθιάς μάθησης, δείχνοντας τελικά μέσω πειραματικών αποτελεσμάτων μια αξιόλογη προσέγγιση με τελικές μετρικές απόδοσης RMSE=5.40 και MAE=4.36.

Έτσι λοιπόν διερευνώντας διεξοδικά τις παραπάνω έρευνες και αποκομίζοντας σημαντικές πληροφορίες τις οποίες θα αναλύσουμε στην ενότητα των πειραμάτων, προτείνουμε ένα νέο υβριδικό μοντέλο ταξινόμησης και αξιολόγησης της κατάθλιψης εκμεταλλευόμενοι τους περιγραφητές ήχου και βίντεο όσο και λόγου.

Κεφάλαιο 2

2. Θεωρητικό Υπόβαθρο

2.1 Μηχανική Μάθηση

Η μηχανική μάθηση (machine learning-ML) είναι μια περιοχή της τεχνητής νοημοσύνης (artificial intelligence-AI) η οποία αφορά αλγορίθμους και μεθόδους που επιτρέπουν στους υπολογιστές να «μαθαίνουν». Με τη μηχανική μάθηση καθίσταται εφικτή η κατασκευή προσαρμόσιμων (adaptable) προγραμμάτων υπολογιστών τα οποία λειτουργούν με βάση την αυτοματοποιημένη ανάλυση συνόλων δεδομένων και όχι τη διαίσθηση των μηχανικών που τα προγραμμάτισαν. Σκοπός της Μηχανικής Μάθησης είναι η εκπαίδευση του υπολογιστή, με συγκεκριμένες μεθόδους και αλγορίθμους, έτσι ώστε κατά την λήψη αποφάσεων να επιτυγχάνει κάποιο συγκεκριμένο αποτέλεσμα, το οποίο ορίζεται ως και η απόδοση του συστήματος αυτού, βάση κάποιας μετρικής που κρίνει την σωστή λειτουργία του. Η μηχανική μάθηση επικαλύπτεται σημαντικά με τη στατιστική, αφού και τα δύο πεδία μελετούν την ανάλυση δεδομένων, από τα οποία αντλούν γνώση. Τα δεδομένα αυτά, οποιασδήποτε μορφής και να είναι απαιτείται να μετατραπούν σε αριθμητικά και να καταχωρηθούν σε διανύσματά χαρακτηριστικών (Feature Vectors), πριν τροφοδοτηθούν στον αλγόριθμο μηχανικής μάθησης, καθώς αυτή είναι η μόνο μορφή δεδομένων που μπορεί να «κατανοήσει» ο υπολογιστής.

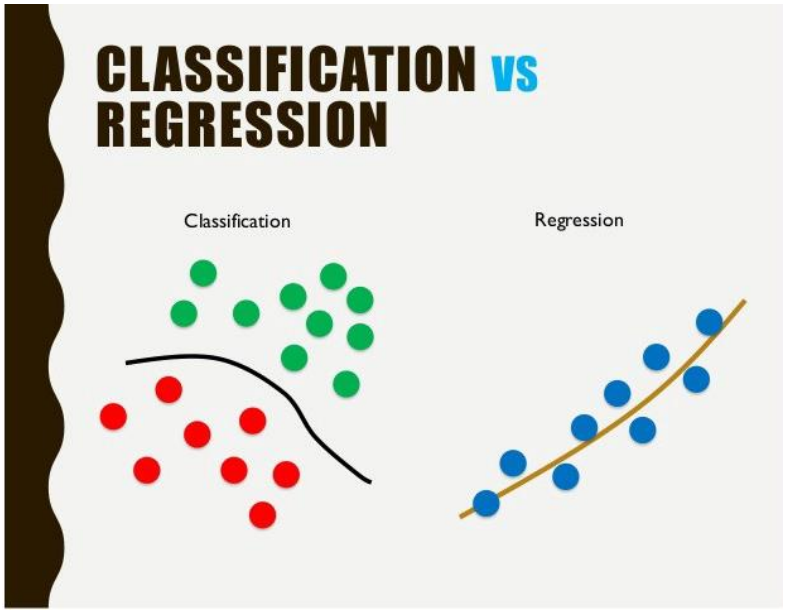
Οι μέθοδοι ML χωρίζονται σε τρεις κατηγορίες: Επιβλεπόμενη Μάθηση (Supervised Learning), Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning) και Ενισχυμένη Μάθηση (Reinforcement Learning).

2.1.1 Επιβλεπόμενη Μάθηση

Οι αλγόριθμοι επιβλεπόμενης μάθησης χρησιμοποιούνται όταν υπάρχει μια “σωστή απάντηση” ή αλλιώς ετικέτα (label) για κάθε παράδειγμα, όπως η κατάταξη κειμένου σε κατηγορίες με ετικέτες. Για την εκπαίδευση του συστήματος παρέχεται, μαζί με τα δεδομένα εισόδου και η επιθυμητή έξοδος για το καθένα από αυτά. Με τον τρόπο αυτό, για κάθε είσοδο που δέχεται το πρόγραμμα προβλέπει μια έξοδο και συγκρίνει το αποτέλεσμα που παράγει με το σωστό αποτέλεσμα που του παρέχεται. Με βάση τις λάθος προβλέψεις που κάνει τροποποιεί αναλόγως το μοντέλο. Ο αλγόριθμος θα σταματήσει να κάνει προβλέψεις όταν επιτύχει ένα αποδεκτό επίπεδο απόδοσης. Στους αλγόριθμους επιβλεπόμενης μάθησης το πρόγραμμα εφαρμόζει τις σχέσεις που έχει μάθει στο παρελθόν, από το σύνολο εκπαίδευσης (Training Dataset), σε νέα δεδομένα (Test Dataset) για να προβλέψει τα μελλοντικά γεγονότα. Οι αλγόριθμοι επιβλεπόμενης μάθησης χωρίζονται σε δύο κατηγορίες:

- *Αλγόριθμοι Ταξινόμησης (Classification):* Οι αλγόριθμοι ταξινόμησης προσπαθούν να αναγνωρίσουν σε ποια από όλες τις διαθέσιμες διακριτές κατηγορίες ανήκει κάθε καινούργιο δεδομένο που παρέχεται στο σύστημα χρησιμοποιώντας ως βάση το σύνολο δεδομένων εκπαίδευσης. Χαρακτηριστικό παράδειγμα ο διαχωρισμός σκύλου και γάτας σε φωτογραφίες

- *Αλγόριθμοι Παλινδρόμησης (Regression):* Οι αλγόριθμοι παλινδρόμησης, με τους οποίους ασχολούμαστε στα πλαίσια αυτής της εργασίας, προσπαθούν να προβλέψουν την τιμή(συνήθως συνεχής) από την οποία χαρακτηρίζεται κάθε νέο δεδομένο εισόδου. Στην ουσία τα προβλήματα αυτά εστιάζουν στη πρόβλεψη της εξόδου για κάθε πρότυπο εισόδου. Χαρακτηριστικό παράδειγμα η πρόβλεψη μετοχών του χρηματιστηρίου.



2.1.2 Μη Επιβλεπόμενη Μάθηση

Οι αλγόριθμοι μη-επιβλεπόμενης μάθησης σε αντίθεση με τα παραπάνω δεν διαθέτουν κάποια επισημειωμένη τιμή/ετικέτα ώστε ο αλγόριθμος να εκπαιδευτεί από αυτή. Στους αλγόριθμους μη επιβλεπόμενης μάθησης παρέχονται μόνο δεδομένα εισόδου χωρίς κάποια έξοδο. Οι αλγόριθμοι αυτοί λοιπόν, αναλύουν τα δεδομένα που τους δίνονται και προσπαθούν να αντλήσουν συμπεράσματα από αυτά, ψάχνοντας να βρουν σχέσεις και κρυμμένες δομές μέσα σε αυτά. Έτσι, τα προβλήματα χαρακτηρίζονται ως προβλήματα ομαδοποίησης (Clustering) κατά τα οποία ο αλγόριθμος επιδιώκει να δημιουργήσει ομάδες ή συστάδες (Clusters) και να τοποθετεί σε αυτές τα δεδομένα που παρουσιάζουν τις μεγαλύτερες, μεταξύ τους, ομοιότητες. Τις περισσότερες φορές το πρόβλημα αυτό δεν είναι σαφές αφού δεν γνωρίζουμε εκ των προτέρων το πλήθος των ομάδων που θέλουμε να δημιουργηθούν. Στη μέθοδο αυτή ανήκουν ένα σύνολο ιδιαίτερα σημαντικών αλγορίθμων, που εστιάζουν τη λειτουργία τους στην παραγωγή δεδομένων, όπως οι γενετικοί αλγόριθμοι.

2.1.3 Ενισχυμένη Μάθηση

Η ενισχυμένη μάθηση (reinforcement learning) αποτελεί έναν συνδυασμό των δυο παραπάνω μεθόδων αφού παρότι ο αλγόριθμος τροφοδοτείται με δεδομένα τα οποία δεν χαρακτηρίζονται από κάποιο label, διατίθεται ένα σύστημα τιμωρίας και ανταμοιβής (Punish and Reward Method) ώστε ο αλγόριθμος να εκπαιδεύεται στη σωστή κατεύθυνση. Η μέθοδος αυτή μια εξελίξιμη επιστημονική περιοχή και γίνεται ιδιαίτερα χρήσιμη διότι ο αλγόριθμος είναι σε θέση να αλληλοεπιδρά με το περιβάλλον του. Η μέθοδος ενισχυμένης μάθησης χρησιμοποιείται ιδιαίτερα σε ηλεκτρονικά παιχνίδια όπως το σκάκι, με σχετικά υψηλές αποδόσεις.

2.2 Μηχανές Διανυσματικής Υποστήριξης

Πριν την επεξήγηση των Μηχανών Διανυσματικής Υποστήριξης (Support Vector Machine-SVM), για την βαθύτερη κατανόηση τους θα ήταν χρήσιμο να διατυπώσουμε το πρόβλημα των γραμμικά διαχωρίσιμων δεδομένων. Ως γραμμικά διαχωρίσιμα δεδομένα ονομάζουμε τα δεδομένα για τα οποία υπάρχει ένα τουλάχιστον υπερεπίπεδο που διαχωρίζει πλήρως τις κατηγορίες που ανήκουν. Το υπερεπίπεδο αυτό δηλαδή, πρέπει να ‘αφήνει’ τα πρότυπα της μιας κατηγορίας στο θετικό ημιχώρο και της άλλης στον αρνητικό. Αντίθετα, τα σύνολα δεδομένων για τα οποία δεν υπάρχει διαχωριστικό επίπεδο ονομάζονται μη διαχωρίσιμα.

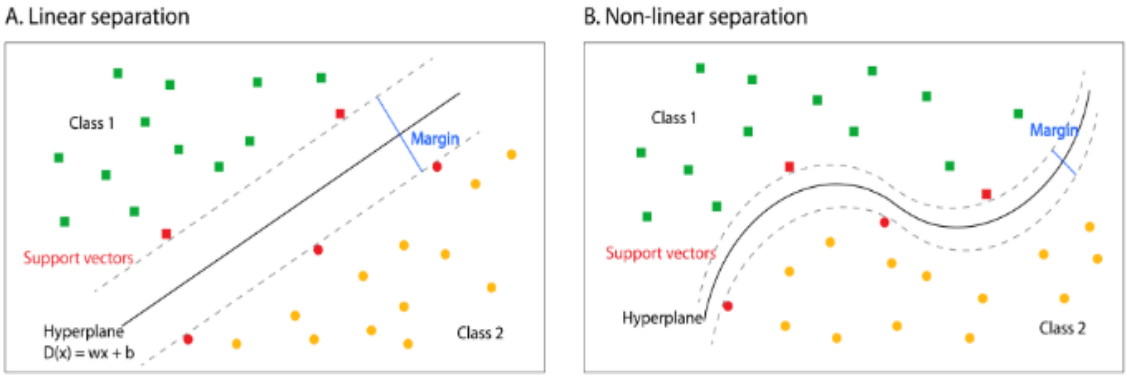
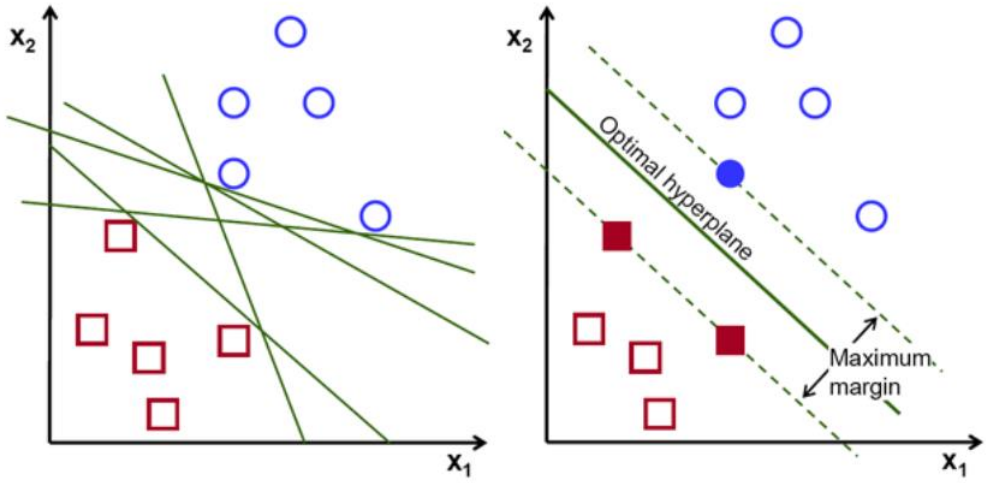


Illustration of separating two classes using SVMs. Linear (A.) and non-linear (B.) perfect separation of two classes (green and orange) with a hyperplane (black) and maximal margin (blue and dotted gray lines). Support vectors defining the hyperplane are in red. No misclassifiactions or margin violations are included.

Στα περισσότερα προβλήματα ταξινόμησης τα διανύσματα χαρακτηριστικών μεταξύ των κλάσεων δεν είναι γραμμικά διαχωρίσιμα δημιουργώντας προβλήματα στην απόδοση του ταξινομητή. Έτσι λοιπόν, τα SVM αποσκοπούν στην εύρεση του βέλτιστου υπερεπιπέδου διαχωρισμού με αποτέλεσμα να μπορούν να αποδίδουν αποτελεσματικά τόσο σε γραμμικά όσο και σε μη γραμμικά διαχωρίσιμα δεδομένα. Πιο συγκεκριμένα, τα SVM αναζητούν ένα υπερεπίπεδο μεγίστου περιθωρίου διαχωρισμού των διανυσμάτων που ανήκουν σε διαφορετικές κλάσεις. Έτσι το διαχωριστικό υπερεπίπεδο διέρχεται από τα σημεία που μεγιστοποιούν το περιθώριο (margin) μεταξύ αυτών και της κάθε κλάσης, όπως φαίνεται στο Σχήμα 2.1. Ουσιαστικά, το βέλτιστο υπερεπίπεδο θα έχει την ίδια απόσταση από τα αντίστοιχα πλησιέστερα δείγματα των 2 κλάσεων.



Σχήμα 2.1: Παράδειγμα προβλήματος δύο γραμμικά διαχωρίσιμων κλάσεων.

Στόχος του SVM είναι η εύρεση του βέλτιστου υπερεπίπεδου, που δίνει το μέγιστο δυνατό περιθώριο (margin) [18]

Στο σχήμα 2.1, στην αριστερή γραφική παράσταση είναι προφανές πως όλες οι διαχωριστικές γραμμές διαχωρίζουν πλήρως τα δεόμενα, ωστόσο καμία δεν έχει την αξιοπιστία της διαχωριστικής γραμμής της δεξιάς γραφικής παράστασης και αυτό διότι το περιθώριο που αφήνει από τις δύο κλάσεις είναι σαφώς μεγαλύτερο. Το βέλτιστο αυτό υπερεπίπεδο υπολογίζουν τα SVM αναζητώντας συντελεστές w,b που επιλύουν το παρακάτω πρόβλημα

$$\vec{w} \cdot \vec{x} - b = 0$$

διατηρώντας ταυτόχρονα τις συνθήκες:

$$\vec{w} \cdot \vec{x} - b = \begin{cases} 1, & if\ x \in Class1 \\ -1, & if\ x \in Class2 \end{cases}$$

οι οποίες ικανοποιούν τα διανύσματα αυτά που ονομάζονται διανύσματα υποστήριξης (support vectors) και βρίσκονται πάνω στα περιθώρια της διαχωριστικής γραμμής (όπως φαινεται στο δεξιά σχήμα της εικόνας 2.1, ενώ η μεταξύ τους απόσταση είναι ίση με $\rho = \frac{2}{\|w\|}$.

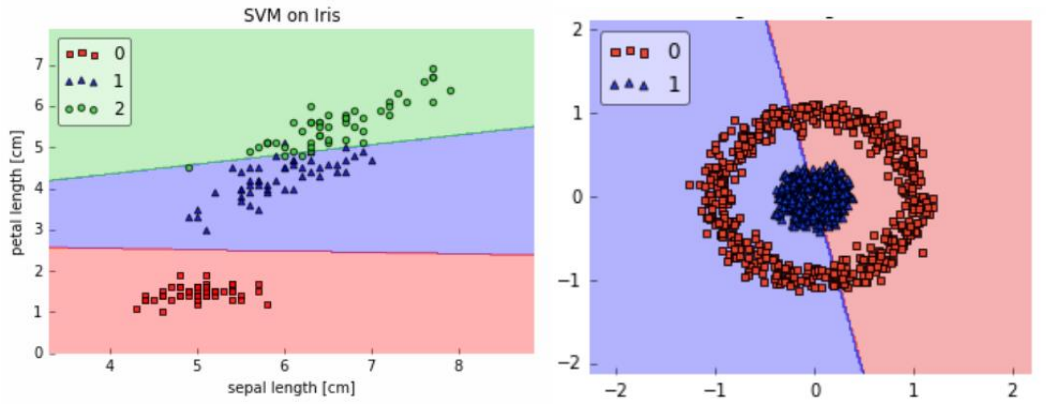
Ωστόσο για να δώσουμε μια μεγαλύτερη αποτελεσματικότητα στους αλγορίθμους SVM είτε απευθύνεται σε γραμμικά είτε σε μη γραμμικά διαχωρίσιμα πρότυπα εισόδου, εισάγουμε την έννοια της συνάρτησης του πυρήνα (kernels). Έτσι λοιπόν πλέον μια μηχανή διανυσμάτων υποστήριξης για την ταξινόμηση προτύπων εκτελεί δύο βασικά βήματα: 1) μη γραμμική αντιστοίχιση φ(.) από το χώρο εισόδου στο χώρο χαρακτηριστικών μεγαλύτερης διαστατικότητας δημιουργώντας νέες σχέσεις μεταξύ των δεδομένων και 2) γραμμική αντιστοίχιση από το χώρο χαρακτηριστικών στο χώρο εξόδου. Η σχέση πυρήνα που χρησιμοποιείται για να μετασχηματίσει τα δεδομένα σ ένα μη γραμμικό χώρο που η εύρεση βέλτιστου υπερεπιπέδου είναι πιο απλή ειναι η εξής:

$$k(\overrightarrow{x_i}, \overrightarrow{x_j}) = \varphi(\overrightarrow{x_i}) \cdot \varphi(\overrightarrow{x_j})$$

Μερικές από τις πιο διαδεδομένες συναρτήσεις πυρήνα είναι:

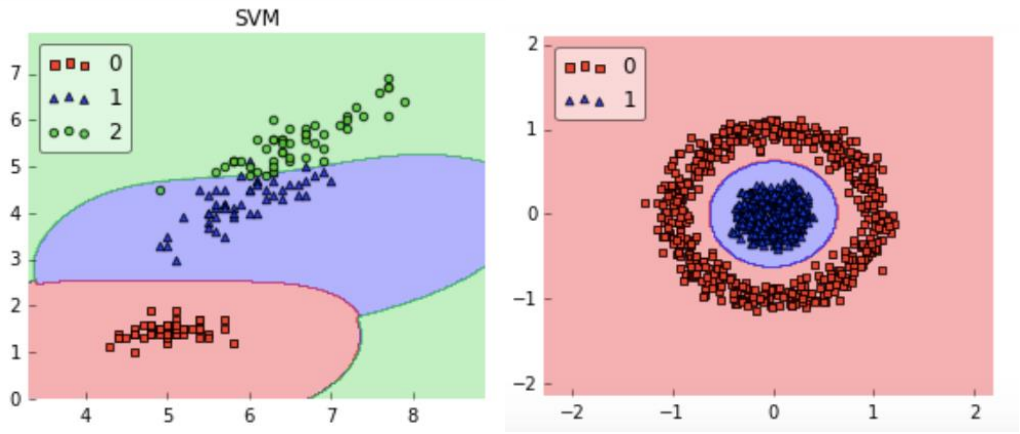
- Ο γραμμικός πυρήνα: $k(\overrightarrow{x_i}, \overrightarrow{x_j}) = \overrightarrow{x_i} \cdot \overrightarrow{x_j}$ *(για γραμμικά διαχωρίσιμα δεδομένα)*
- Ο πολυωνυμικός πυρήνας βαθμού d: $k(\overrightarrow{x_i}, \overrightarrow{x_j}) = (\overrightarrow{x_i} \cdot \overrightarrow{x_j} + 1)^d$ *(για μη-γραμμικά διαχωρίσιμα δεδομένα, όπως και οι παρακάτω συναρτήσεις)*
- Ο Γκαουσιανός πυρήνας: $k(\overrightarrow{x_i}, \overrightarrow{x_j}) = \exp\left(-\frac{\|\overrightarrow{x_i} - \overrightarrow{x_j}\|^2}{2\sigma^2}\right)$
- Ο Πυρήνας Ακτινικής Βάσης (RBF) με παράμετρο γ:
- Ο Πυρήνας Υπερβολικής Εφαπτομένης με παράμετρο γ:

Παρακάτω παραθέτουμε δύο παραδείγματα που εφαρμόζονται ένα SVM με linear kernel (Σχ. 2.2) και ένα SVM με RBF kernel (Σχ. 2.3):



Σχήμα 2.2: Γραμμικό SVM

Αριστερά εφαρμόζουμε linear SVM σε γραμμικά διαχωρίσιμα δεδομένα και δεξιά σε μη γραμμικά διαχωρίσιμα



Σχήμα 2.3: SVM με RBF kernel

Αριστερά εφαρμόζουμε SVM με RBF kernel σε γραμμικά διαχωρίσιμα δεδομένα και δεξιά σε μη γραμμικά διαχωρίσιμα

Ένα βασικό μειονέκτημα των SVM είναι η αδυναμία τους να αντιμετωπίσουν προβλήματα πολλών κλάσεων. Η βασική τεχνική που χρησιμοποιείται για να επεκταθεί η

λειτουργία τους σε παραπάνω από δύο κατηγορίες ταξινόμησης ονομάζεται ‘ένα εναντίον όλων’ και δημιουργεί τόσα υπερεπίπεδα όσες και οι κατηγορίες διαχωρισμού. Παρ ’όλα αυτά η μέθοδος αυτή, όπως γίνεται αντιληπτό, αντιμετωπίζει τόσα προβλήματα δυαδικής ταξινόμησης όσα και οι κατηγορίες του προβλήματος με αποτέλεσμα να αυξάνεται σε μεγάλο βαθμό η υπολογιστική πολυπλοκότητα της ταξινόμησης.

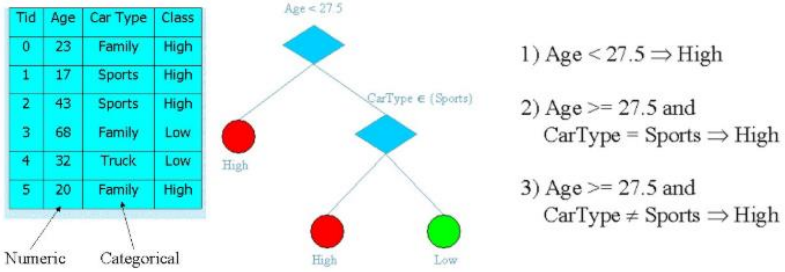
2.3 Αλγόριθμος Random Forest

2.3.1 Δέντρα Απόφασης (Decision Trees)

Τα δένδρα απόφασης (decision trees) είναι από τις πιο απλές και διαδεδομένες τεχνικές τόσο κατηγοριοποίησης όσο και παλινδρόμησης που ανήκουν στην οικογένεια των επιβλεπόμενων αλγορίθμων μάθησης, καθώς εφαρμόζουν μία σαφή, ξεκάθαρη λογική που μπορεί να προσαρμοστεί σε ένα μεγάλο εύρος προβλημάτων.

Η λειτουργία των δένδρων αποφάσεων καθορίζεται από μια σειρά ερωτήσεων για τις ιδιότητες των προτύπων εισόδου του συνόλου εκπαίδευσης. Για κάθε δείγμα που εισάγεται στο μοντέλο, αυτό του θέτει μια σειρά ερωτήσεων που αναπαριστούν ένα σύνολο διαφορετικών χαρακτηριστικών του συνόλου δεδομένων. Κάθε φορά που παίρνει απάντηση θέτει μία νέα ερώτηση, μέχρι να φτάσει στα φύλλα του δέντρου όπου συμπεραίνουν την κατηγορία του δείγματος.

Πιο συγκεκριμένα, το σύνολο των ερωτήσεων και των απαντήσεων μπορούν να οργανωθούν σε μια δομή δένδρου αποφάσεων, όπου είναι μια ιεραρχική δομή από εσωτερικούς κόμβους, κατευθυνόμενους συνδέσμους και κόμβους-φύλλα. [19]. Κάθε μη τερματικός κόμβος περιέχει ερωτήσεις - συνθήκες για τις ιδιότητες των προτύπων εισόδου, με σκοπό να κατευθύνει τα πρότυπα με διαφορετικές ιδιότητες προς διαφορετικούς κόμβους αποφάσεων. Η επιλογή του κόμβου-ρίζα καθορίζεται από το ποιο χαρακτηριστικό κατηγοριοποιεί καλύτερα τα δεδομένα, ακολουθώντας την ίδια λογική για τους επόμενους κόμβους. Κάθε κόμβος-φύλλο αντιστοιχεί σε μία κατηγορία ή μια συνεχή τιμή αν πρόκειται για πρόβλημα παλινδρόμησης, όπου αναπαρίσταται με μια περιοχή του χώρου των δεδομένων. Όσο πιο βαθύ είναι το δένδρο, τόσο πιο περίπλοκοι είναι οι κανόνες αποφάσεων και τόσο πιο ακριβές είναι το μοντέλο. Η χρήση των δένδρων αποφάσεων έχει πολλά πλεονεκτήματα. Είναι ένας αλγόριθμος που γίνεται εύκολα κατανοητός καθώς ακολουθεί την ίδια προσέγγιση με την οποία ο άνθρωπος παίρνει αποφάσεις, καθώς και οπτικοποιείται πολύ εύκολα με τη χρήση διαγράμματος. Δεν απαιτεί μεγάλο όγκο δεδομένων για να εκπαιδευτεί και μπορεί να δώσει καλά αποτελέσματα στις περιπτώσεις κατηγοριοποίησης με πολλά δυνατά αποτελέσματα. Τα μειονεκτήματα του δένδρου αποφάσεων συμπεριλαμβάνουν την αστάθεια που μπορεί να προκληθεί στα δένδρα από μικρές αλλαγές στα δεδομένα εισόδου, καθώς και την περίπτωση να φτιαχτούν υπερβολικά σύνθετα δένδρα που δεν γενικεύουν καλά τα δεδομένα (overfitting).



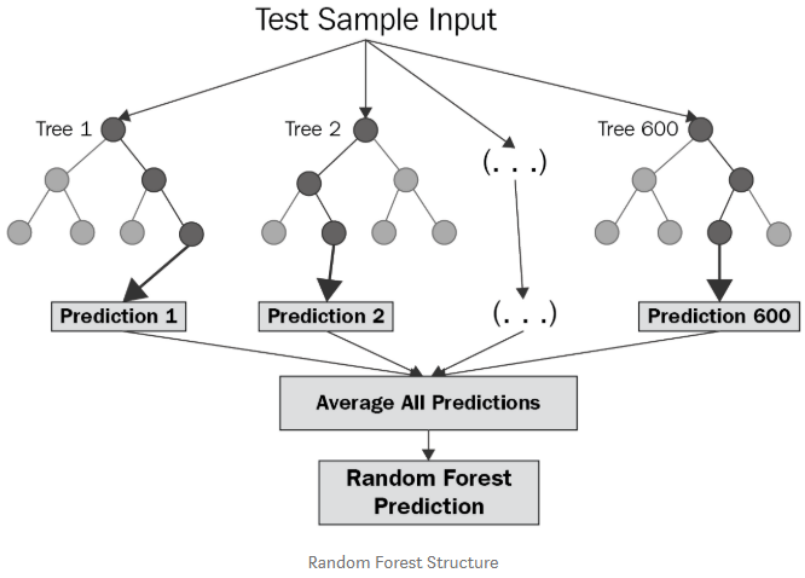
Σχήμα 2.4: Αναπαράσταση δέντρου απόφασης

2.3.2 Random Forest Algorithm

Ο αλγόριθμος Random Forest, όπως πιθανώς θα μπορούσε να καταλάβει κανείς από το όνομα του, αποτελεί μία επέκταση των δένδρων αποφάσεων. Είναι και αυτός ένας αλγόριθμος επιβλεπόμενης μάθησης, όπου δημιουργεί ένα «δάσος», ένα σύνολο δηλαδή δένδρων αποφάσεων. Όπως και ένα δάσος είναι πιο πυκνό αν περιέχει πολλά δένδρα, έτσι και ο αλγόριθμος μας δίνει μεγαλύτερη ακρίβεια όσο περισσότερα δένδρα αποφάσεων δημιουργούνται.

Ο αλγόριθμος ανήκει στην κατηγορία του 'Ensemble Learning', όπου συνδυάζει πολλά μοντέλα για να λύσει ένα πρόβλημα. Δημιουργεί πολλά δένδρα ταξινόμησης ή παλινδρόμησης αντίστοιχα, που εκπαιδεύονται και κάνουν προβλέψεις ανεξάρτητα το ένα από το άλλο πάνω σε ένα ανεξάρτητο τυχαίο δείγμα δεδομένων εισόδου το καθένα. Αυτές οι προβλέψεις συνδυάζονται, ώστε να πραγματοποιήσουν μια μεγάλη πρόβλεψη, όπου θα είναι καλύτερη ή τουλάχιστον τόσο καλή όσο η πρόβλεψη του κάθε ταξινομητή.

Ο αλγόριθμος Random Forest δημιουργεί τυχαία δένδρα αποφάσεων, κάποια από τα οποία είναι χρήσιμα για την πρόβλεψη που θέλουμε να πραγματοποιηθεί και κάποια όχι. Με τον τρόπο αυτό, όταν δίνεται στον αλγόριθμο ένα αντικείμενο όλα τα δένδρα που έχουν δημιουργηθεί θα πραγματοποιήσουν μια πρόβλεψη. Οι προβλέψεις θα συγκεντρωθούν και συμψηφιστούν, προκειμένου να υπολογιστεί η τελική πρόβλεψη. Στην περίπτωση προβλήματος classification, το κάθε δέντρο ψηφίζει και η κλάση με τις περισσότερες ψήφους είναι και η κλάση νικητής, ενώ στο regression σαν τελική τιμή παίρνουμε τον μέσο όρων των αποτελεσμάτων κάθε δένδρου. [20]



Σχήμα 2.5: Αναπαράσταση αλγορίθμου Random Forest

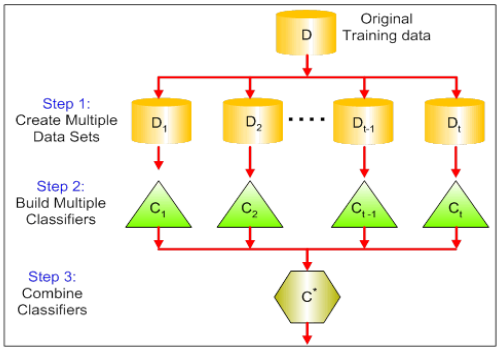
Ο αλγόριθμος αυτός είναι πιο απλός και αποδοτικός σε σχέση με άλλους αλγορίθμους ταξινόμησης μη γραμμικών προβλημάτων καθώς τα περισσότερα δένδρα που θα δώσουν άκυρες προβλέψεις θα ακυρώνονται μεταξύ τους και άρα θα είναι τα χρήσιμα δένδρα αυτά που θα καθορίσουν, εν τέλει, την τελική πρόβλεψη. Επιπροσθέτως το γεγονός πως κάθε δέντρο εκπαιδεύεται σε ένα διαφορετικό τυχαίο σύνολο δειγμάτων προσθέτει ένα επιπλέον στοιχείο τυχαιότητας στο συνολικό μοντέλο, και σε συνδυασμό με το γεγονός πως το τελικό αποτέλεσμα προκύπτει από τον συνδυασμό των επιμέρους δένδρων που απαλείφει τυχόν μεγάλες αποκλίσεις, καταπολεμά το πρόβλημα της υπερεκπαίδευσης (overfitting). Τέλος, όπως αναφέραμε στα δένδρα απόφασης, τα πιο

σημαντικά χαρακτηριστικά που συνεισφέρουν σε μεγαλύτερο βαθμό στην έκβαση του αποτελέσματος, αναδεικνύονται στις ρίζες των δένδρων παίζοντας σημαντικότερο ρόλο στην ταξινόμηση ή παλινδρόμηση των δειγμάτων. Ωστόσο παρουσιάζει και κάποια μειονεκτήματα όπως κάθε αλγόριθμος. Η όλη διαδικασία καταναλώνει πολυ χρόνο καθώς για την έκβαση του αποτελέσματος πρέπει να δώσουν μια τελική τιμή όλα τα δένδρα για να προκύψει η προβλεπόμενη τιμή, ενώ είναι δύσκολη η ερμηνεία τους καθώς έχουμε να ερμηνεύσουμε ένα ολόκληρο ‘δάσος’ από δένδρα έναντι ενός δένδρου.

2.4 Μάθηση Ensemble

Η δυνατότητα ένα μοντέλο να βελτιώσει τις δυνατότητες ταξινόμησης (classification) του ή παλινδρόμησης (regression) αντίστοιχα θα μπορούσε να γίνει αν συνδυαστούν δύο ή περισσότερα μοντέλα ταυτόχρονα. Αυτό εκφράζει το Ensemble Learning, τον συνδυασμό ευφυών τεχνικών ώστε να αυξηθεί η ικανότητα γενίκευσης του τελικού ταξινομητή/εκτιμητή και επομένως η απόδοση του μοντέλου. Η απόδοση του μοντέλου αυξάνεται με διαφορετικό τρόπο ανάλογα ποιον ensemble αλγόριθμο χρησιμοποιούμε, είτε δηλαδή μειώνοντας την διακύμανση (variance) αν εφαρμόζουμε αλγόριθμο bagging, είτε την απόκλιση (bias) αν εφαρμόζουμε αλγόριθμο boosting, είτε βελτιώνοντας το προβλεπόμενο αποτέλεσμα με αλγόριθμο voting. Διαθέτουν ιδιαίτερα θετικά χαρακτηριστικά που τους κάνουν δημοφιλείς. Ένα από αυτά είναι η δυνατότητα να εκτελούνται παράλληλα οι αλγόριθμοι μηχανικής μάθησης που εφαρμόζουμε και να εξετάζονται στα δοκιμαστικά δεδομένα ταυτόχρονα. Επίσης, σπάνια αποδίδουν χαμηλότερα από τους μεμονωμένους αλγορίθμους που τους αποτελούν, χωρίς ωστόσο αυτό να αποτελεί γενικό κανόνα. Παρόλο που υπάρχουν αρκετά είδη μεθόδων ensemble μάθησης, θα παρουσιάσουμε τα 3 πιο συχνά χρησιμοποιούμενα [23]:

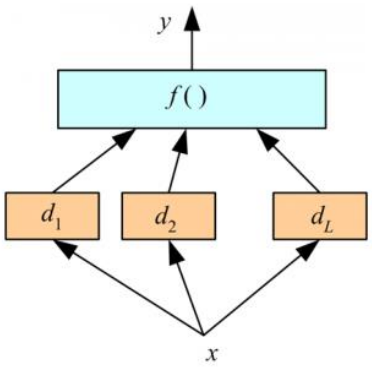
1. Bagging: Με την μέθοδο bagging ή αλλιώς bootstrap aggregation, συναθροίζουμε ένα σύνολο από μοντέλα, το καθένα εκπαιδευμένο σε ένα υποσύνολο δεδομένων bootstrap, παίρνοντας τελικά τον μέσο όρο τους, επιτυγχάνοντας έτσι να καταπολεμήσουμε το πρόβλημα της αυξημένης διακύμανσης (variance). Πιο συγκεκριμένα, στην μέθοδο αυτή εκπαιδεύουμε ένα πλήθος μοντέλων σε διαφορετικα μικρά υποσύνολα του συνόλου προτύπων εισόδου, τα οποία έχουν επιλεγεί τυχαία και με αντικατάσταση με όλα τα δείγματα να ειναι ισοπίθανα να επιλεγούν (Bootstrap Sampling). Τέλος για να συναθροίσουμε τις εξόδους των επιμέρους μοντέλων, αν έχουμε πρόβλημα ταξινόμησης εφαρμόζουμε μέθοδο ψηφοφορίας (voting) κατά την οποία επιλέγεται η κλάση που επιλέχθηκε από τους περισσότερους ταξινομητές, ενώ αν έχουμε πρόβλημα παλινδρόμησης παίρνουμε σαν τελικό αποτέλεσμα τον μέσο όρο των επιμέρους αποτελεσμάτων.



Σχήμα 2.6: Μέθοδος Bagging

2.Boosting: Στην μέθοδο αυτοί ανήκουν οι αλγόριθμοι που χρησιμοποιούν τον σταθμισμένο μέσο όρο για να μετατρέψουν αδύναμους εκτιμητές, σε ισχυρούς. Είναι μια μέθοδος ακολουθιακής λογικής, όπου το κάθε μοντέλο εκπαιδεύεται σε ολόκληρο το δείγμα εκπαίδευσης, ενώ τα επόμενα μοντέλα προσαρμόζουν στην εκπαίδευσή τους το τετραγωνικό σφάλμα από το προηγούμενο μοντέλο. Με αυτόν τον τρόπο, η μέθοδος αυτή δίνει υψηλότερο βάρος/αξία στις παρατηρήσεις που υποτιμήθηκαν σε προηγούμενα μοντέλα. Μόλις λοιπόν ολοκληρωθεί η εκπαίδευση των ακολουθιακών μοντέλων, οι τελικές προβλέψεις προκύπτουν από τον σταθμισμένο μέσο όρων όλων των επιμέρους μοντέλων, με βάρη που καθορίζουν οι εκτιμήσεις ακριβείας (accuracy score) των μοντέλων αυτών.

3. Voting: Η μέθοδος αυτή είναι μια από τις πιο ευθείς της Ensemble μάθησης, στις οποίες οι προβλέψεις από κάθε εκτιμητή συνδυάζονται. Πιο συγκεκριμένα, εκπαιδεύονται δύο ή παραπάνω μοντέλα πάνω στο ίδιο dataset παράλληλα και για να πάρουμε την τελική πρόβλεψη πάνω στα νέα δεδομένα συνδυάζουμε τις επιμέρους προβλέψεις μέσω μιας συνάρτησης της επιλογής μας βέλτιστης ως προς την συμπεριφορά των μοντέλων, όπως ο μέσος όρος, σταθμισμένος ή μη. Στην μέθοδο αυτή συνηθίζεται τα μοντέλα να εκπαιδεύονται πάνω στα ίδια δείγματα αλλα εκμεταλλευόμενα κάθε φορα διαφορετικά ασυσχέτιστα χαρακτηριστικά τους.



Σχήμα 2.7: Μέθοδος Voting

2.5 Βαθιά Νευρωνικά Δίκτυα

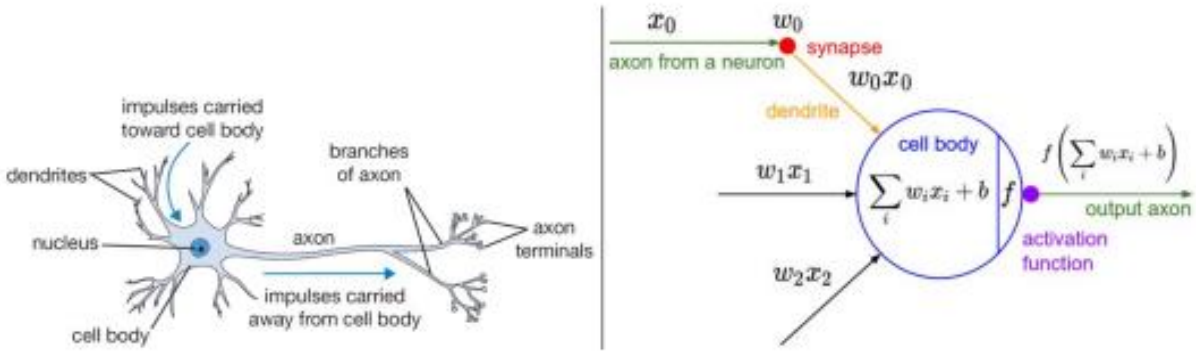
2.5.1 Εισαγωγή

Η βαθιά μάθηση (deep learning) είναι ένα σύνολο μεθόδων μάθησης που προσπαθούν να μοντελοποιήσουν δεδομένα με περίπλοκες αρχιτεκτονικές συνδυάζοντας διαφορετικούς μη-γραμμικούς μετασχηματισμούς. Τα θεμέλια της βαθιάς μάθησης είναι τα νευρωνικά δίκτυα, τα οποία συνδυάζονται και δημιουργούν τα βαθιά νευρωνικά δίκτυα. Οι τεχνικές αυτές έχουν επιτρέψει σημαντική πρόοδο στα πεδία της αυτόματης επεξεργασίας ήχου και εικόνας, που περιλαμβάνουν την αναγνώριση χαρακτηριστικών προσώπου, την αναγνώριση φωνής, την όραση υπολογιστών, την αυτόματη επεξεργασία φυσικής γλώσσας, την ταξινόμηση κειμένου. Υπάρχει πληθώρα πρακτικών εφαρμογών. Ένα εντυπωσιακό παράδειγμα είναι το πρόγραμμα AlphaGo, το οποίο έμαθε να παίζει το παιχνίδι “go” με μεθόδους βαθιάς μάθησης, κερδίζοντας τον παγκόσμιο πρωταθλητή το 2016.

2.5.2 Τεχνητά νευρωνικά Δίκτυα

Ένα Τεχνητό Νευρωνικό Δίκτυο (Artificial Neural Network - ANN) είναι ένα υπολογιστικό μοντέλο εμπνευσμένο από τη βιολογία, το οποίο δημιουργεί μοτίβα βασισμένο στη δομή και τη λειτουργία των νευρώνων που υπάρχουν στον ανθρώπινο εγκέφαλο. Η πρώτη απόπειρα για δημιουργία ενός τεχνητού νευρώνα έγινε από τους McCulloch και Pits το 1943 [24] και αναπτύχθηκαν την δεκαετία του 1960 με κορύφωση το βιβλίο των Minsky και Papert το 1969 [25]. Ωστόσο η λειτουργία του εγκεφάλου δεν είναι ούτε δυαδική ούτε σταθερή, με λογικό ακόλουθο να μην μπορεί να προσεγγιστεί επαρκώς από ANNs και με αποτέλεσμα να στραφεί το ενδιαφέρον των ANNs σαν αντικείμενο έρευνας από μηχανικούς στην επίλυση προβλημάτων που δεν μπορούν να επιλυθούν από την παραδοσιακή υπολογιστική. Πολλές φορές είναι χρήσιμο να αντιμετωπίζουμε τα ANN ως κατευθυνόμενους γράφους με συνάψεις που διαθέτουν συναρτήσεις ενεργοποίησης.

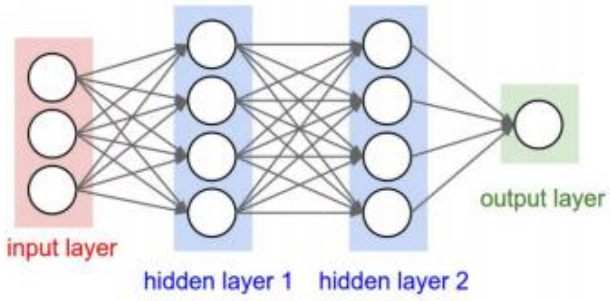
Θεμελιώδες στοιχείο των ANNs αποτελεί ο νευρώνας Perceptron, ο οποίος περιεγράφηκε αρχικά από τον Rosenblatt το 1958 [26]. Το Perceptron αποτελεί εμπρόσθια τροφοδοτούμενο (Feedforward) ANN. Στην γενική τους μορφή τα δίκτυα αυτά προσπαθούν να προσεγγίσουν μέσω της εξόδου τους μια συνάρτηση ή τιμή αναφοράς και ανήκουν στην κατηγορία των γραμμικών ταξινομητών (linear classifiers). Η έξοδος ενός Feedforward δικτύου δίνεται από μια σχέση $y = f(x; \theta)$ με θ ένα σύνολο παραμέτρων της διάταξης. Το μοντέλο Perceptron δέχεται σαν είσοδο ένα διάνυσμα $x = [x_0, x_1, \dots, x_n] \in R_n$ και παράγει μια έξοδο $y \in R_n$. Όπως μπορούμε να παρατηρήσουμε και από το σχήμα 2.10, το διάνυσμα εισόδου πολλαπλασιάζεται με ένα διάνυσμα βαρών $W \in R_n$ και το αποτέλεσμα αφού προστεθεί μια πόλωση (bias) b, εισάγεται σε μια μη γραμμική συνάρτηση ενεργοποίησης (Activation Function) από την οποία παράγεται η έξοδος y του δικτύου.



Σχήμα 2.8: Αριστερά ένας βιολογικός νευρώνας και Δεξιά ο μαθηματικός του συμβολισμός

Ο τεχνητός νευρώνας Perceptron αποτελεί έναν καθαρά γραμμικό ταξινομητή που δεν μπορεί να προσεγγίσει μη γραμμικά προβλήματα. Έτσι λοιπόν για την επίλυση μη γραμμικών προβλημάτων μπορούν να συνδυαστούν πολλοί τεχνητοί νευρώνες και να δημιουργηθεί ένα πολυεπίπεδο δίκτυο νευρώνων Perceptron: Multi-Layer Perceptron (MLP). Η δημιουργία πολυεπίπεδων νευρώνων σε συνδυασμό με την χρήση μη γραμμικών συναρτήσεων ενεργοποίησης μας φέρνει ένα βήμα πιο κοντά στην λειτουργία του ανθρωπίνου εγκεφάλου που λειτουργεί με πολλά στρώματα διασυνδεδεμένων νευρώνων. Στη συγκεκριμένη αρχιτεκτονική κάθε επίπεδο νευρώνων συνδέεται με τους νευρώνες του επόμενου επιπέδου χωρίς να υπάρχουν διασυνδέσεις νευρώνων στο ίδιο επίπεδο. Σε άλλες αρχιτεκτονικές συναντάμε και αλληλεπιδράσεις μεταξύ νευρώνων ίδιου επιπέδου. Τα ενδιάμεσα στρώματα μετά το αρχικό *εισόδου στρώμα* (Input Layer) του MLP ονομάζονται

κρυφά στρώματα (Hidden Layers), ενώ το τελικό ονομάζεται *στρώμα εξόδου* (Output Layer). Το *επίπεδο εισόδου* λαμβάνει τα δεδομένα εισόδου, περνώντας απλά την πληροφορία των χαρακτηριστικών των δεδομένων στο πρώτο κρυφό επίπεδο. Έπειτα, μέσω των *κρυφών στρωμάτων* εξάγονται τα χαρακτηριστικά και όσο κινούμαστε προς ανώτερα κρυμμένα επίπεδα εξάγονται χαρακτηριστικά ανωτέρου σημασιολογικού περιεχομένου. Μια σημαντική επισήμανση είναι πως για να προκύψουν αυτά τα χαρακτηριστικά με την πάροδο των επιπέδων, πρέπει οι συναρτήσεις ενεργοποίησης των κρυφών επιπέδων να είναι μη γραμμικές ώστε το μοντέλο να μαθαίνει περίπλοκες σχέσεις μεταξύ των χαρακτηριστικών που διαφορετικά με γραμμικές σχέσεις δεν θα αναδύονταν ποτέ. Ούτως ή άλλως αν οι συναρτήσεις ενεργοποίησης ήταν όλες γραμμικές, δεν θα υπήρχε λόγος για πολυεπίπεδα συστήματα αφού στο τέλος η γραμμική τους σχέση θα μπορούσε να αναπαρασταθεί από μία συνάρτηση μόνο (ένα επίπεδο νευρώνων). Τέλος, το επίπεδο εξόδου αφού λάβει τα χαρακτηριστικά που εξήγαγε ο τελευταίος κρυφός νευρώνας, λαμβάνει την τελική απόφαση του δικτύου μέσω της κατάλληλης συνάρτησης, ανάλογα αν πρόκειται για σύστημα ταξινόμησης (μέσω μη γραμμικής σχέσης) ή παλινδρόμησης (συνήθως μέσω γραμμικής σχέσης) .



Σχήμα 2.9: Νευρωνικό Δίκτυο 3 επιπέδων

2.5.3 Εκπαίδευση Νευρωνικών Δικτύων

Είναι χρήσιμο να καθορίσουμε μερικά από τα βασικά χαρακτηριστικά που πρέπει να διέπουν ένα λειτουργικό και αποτελεσματικό Νευρωνικό δίκτυο έτσι ώστε να γίνει αντιληπτή η σημασία της επιλογής κατάλληλων εργαλείων και τιμών που θα αξιοποιήσουμε αργότερα για την εκπαίδευση των μοντέλων μας. Ένα νευρωνικό δίκτυο πρέπει να έχει την δυνατότητα να περατώνει την εκπαίδευση σε πεπερασμένο χρόνο, όσο επιτεύξιμο είναι αυτό, καταναλώνοντας όσο το δυνατόν λιγότερη υπολογιστική ισχύ. Συνεπώς τόσο η επιλογή κατάλληλων εργαλείων, όσο και κατάλληλου πλήθους δεδομένων εκπαίδευσης είναι ιδιαίτερα σημαντική. Ταυτόχρονα πρέπει να διατηρεί την ιδιότητα του να γενικεύει δηλαδή να διατηρεί μικρό σφάλμα πρόβλεψης τόσο στα δεόμενα εκπαίδευσης, όσο και στα δεδομένα ελέγχου. Η εκπαίδευση ενός νευρωνικού δικτύου αποτελεί μια επαναληπτική διαδικασία κατά την οποία οι παράμετροι του δικτύου προσαρμόζονται ώστε να έχουμε την επιθυμητή έξοδο/πρόβλεψη. Η κάθε επανάληψη της διαδικασίας ονομάζεται εποχή (epoch) και το πλήθος των εποχών επηρεάζει σημαντικά τόσο την δυνατότητα κατηγοριοποίησης (ή παλινδρόμησης) του μοντέλου όσο και την ικανότητα του να γενικεύει.

Υπάρχουν δύο είδη εκπαίδευσης που διαχωρίζονται με βάση την ανανέωση των παραμέτρων του μοντέλου. Η πρώτη ονομάζεται On-line Learning και τα βάρη του δικτύου ανανεώνονται παράδειγμα με παράδειγμα από τα δεδομένα εκπαίδευσης και η δεύτερη αντίθετα, ονομάζεται εκπαίδευση με πακέτα (**Batch Learning**) και τα βάρη του

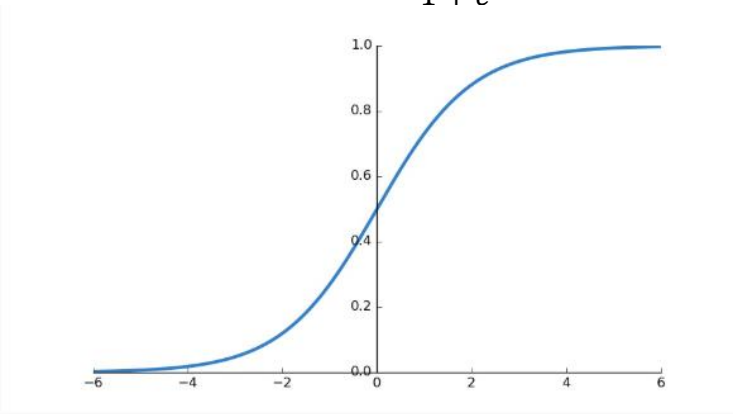
δικτύου ανανεώνονται μετά την είσοδο του συνόλου των δεδομένων εκπαίδευσης που καθορίζει το batch size. Η διαδικασία της εκπαίδευσης λειτουργεί κυρίως με τον υπολογισμό διαφορών και παραγώγων, ενώ εξαρτάται από διάφορες παραμέτρους όπως ο αριθμός των επαναλήψεων, η συνάρτηση ενεργοποίησης των νευρώνων, η συνάρτηση κόστους και η συνάρτηση βελτιστοποίησης. Παρακάτω θα αναπτύξουμε τις συναρτήσεις αυτές που θα αξιοποιήσουμε στα μοντέλα μας.

2.5.3.1 Συνάρτηση Ενεργοποίησης (Activation Function)

Η συνάρτηση ενεργοποίησης είναι ουσιαστικά η πύλη μετάβασης των χαρακτηριστικών των αντικειμένων από το ένα επίπεδο στο επόμενο ενός νευρωνικού δικτύου, δημιουργώντας μέσω της μη γραμμικότητάς της, ανώτερες περίπλοκες σχέσεις μεταξύ των χαρακτηριστικών αυτών. Στην ουσία αποτελεί έναν κόμβο 'απόφασης' που ενεργοποιείται και μεταφέρει το κατάλληλο αποτέλεσμα είτε στον επόμενο κόμβο είτε στην έξοδο του δικτύου ανάλογα με την τιμή που εισάγεται σε αυτόν. Μερικές από τις πιο βασικές συναρτήσεις ενεργοποίησης είναι οι εξής:

- **Σιγμοειδής Συνάρτηση (Sigmoid Function)**: Αποτελεί μία από τις πρώτες συναρτήσεις που χρησιμοποιήθηκαν χρονικά. Η λειτουργία της εγγυείται στην αντιστοίχιση της τιμής εισόδου σε μια τιμή στο διάστημα (0,1) με την ιδιαιτερότητα όμως να μεταφέρει τις μικρότερες τιμές κοντά στο 0 και τις μεγαλύτερες κοντά στο 1 ,πάντα όμως ασυμπτωτικά. Το γεγονός αυτό όμως οδηγεί σε πολύ μικρές τιμές κλίσης, ακόμα και μηδενικές, κάνοντας τη διαδικασία της μάθησης αρκετά αργή ή ενδεχομένως και να σταματήσει πρόωρα. Το φαινόμενο αυτό ονομάζεται Εξασθένιση Κλίσης (Vanishing Gradient) και μας δημιουργεί αρκετά προβλήματα στην διαδικασία μάθησης. Η σιγμοειδής συνάρτηση δίνεται από τον τύπο:

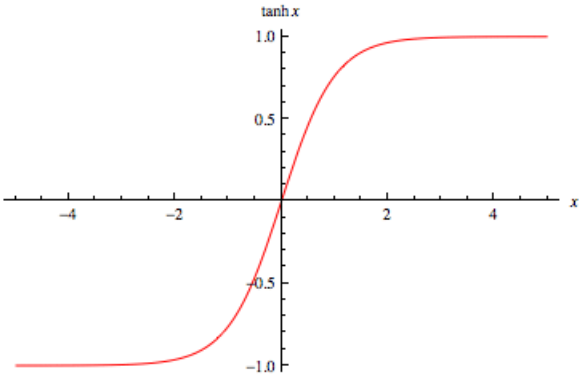
$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$$



Σχήμα 2.10: Σιγμοειδής Συνάρτηση Ενεργοποίησης

- **Υπερβολική Εφαπτομένη (Hyperbolic Tangent)**: Όμοια με την σιγμοειδή, η υπερβολική εφαπτομένη αντιστοιχεί την είσοδο με βάση ένα κατώφλι στο διάστημα (-1,1). Το πλεονέκτημα της έναντι της σιγμοειδούς είναι πως δεν μεταβάλει αισθητά τις τιμές που βρίσκονται κοντά στο 0 και επομένως βοηθούν τον επόμενο νευρώνα κατά τη διαδικασία διάδοσης. Αυτό αποτελεί και τον βασικό λόγο που χρησιμοποιείται σε επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent Neural Networks). Η υπερβολική εφαπτομένη δίνεται από τον τύπο:

$$f(x) = tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

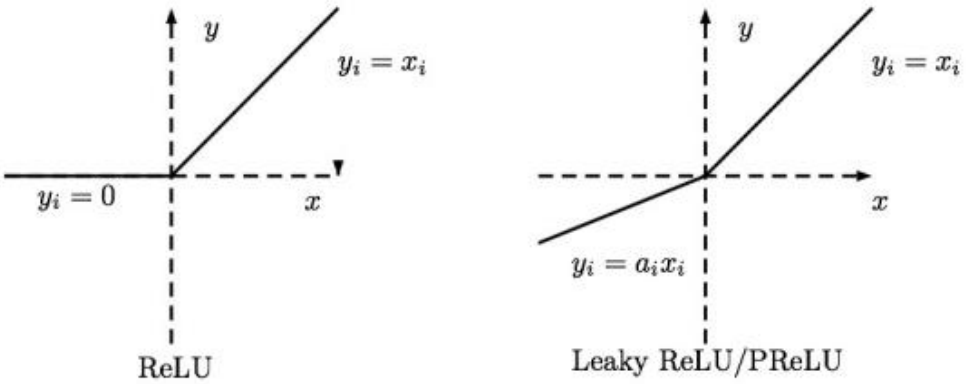


Σχήμα 2.11: Υπερβολική Εφαπτομένη

Όπως μπορούμε να δούμε στο σχήμα 2.11, οι μικρές τιμές μετατοπίζονται κοντά στο -1, ενώ οι μεγάλες στο 1. Όμοια με την σιγμοειδή συνάρτηση, αντιμετωπίζουμε το πρόβλημα του Vanishing Gradient.

- **Rectified Linear Unit (ReLU):** Αποτελεί την πλέον ευρέως χρησιμοποιούμενη συνάρτηση ενεργοποίησης και σύμφωνα με αυτή, οι τιμές εισόδου που είναι μικρότερες του μηδενός παύουν να λαμβάνουν μέρος στην διαδικασία της μάθησης. Ένα βασικό μειονέκτημα της ReLU είναι πως για θετικές εισόδους δεν είναι οριοθετημένη με αποτέλεσμα για ιδιαίτερα μεγάλες εισόδους να λαμβάνουμε τεράστιες εξόδους. Το γεγονός αυτό όμως σε συνδυασμό με μεθόδους κανονικοποίησης που εφαρμόζονται στην προεπεξεργασία των δεδομένων, υστερεί έναντι του βασικού πλεονεκτήματος της και την καθιστά βασική συνάρτηση ενεργοποίησης τόσο για MLP, όσο και για DNN. Η ReLU δίνεται από τον τύπο:

$$f(x) = \max(0, x)$$



Σχήμα 2.12: ReLU

Η ReLU έχει ιδιαίτερα χαμηλή πυκνότητα (low sparsity) αφού νεκρώνει το σύνολο των νευρώνων οι οποίοι έχουν αρνητικές τιμές. Το γεγονός αυτό είναι ιδιαίτερα χρήσιμο υπολογιστικά αφού μειώνει το χρόνο εκμάθησης του δικτύου κάνοντάς το ιδιαίτερα αποδοτικό. Ταυτόχρονα όμως δημιουργούνται προβλήματα στους νευρώνες οι οποίοι έχουν ΄νεκρωθεί', αφού για αρνητικές τιμές θα παίρνουμε σταθερή και μηδενική παράγωγο καθ' όλη τη διαδικασία εκμάθησης, μη δίνοντας στον νευρώνα τη δυνατότητα να αλλάξει τιμή. Το πρόβλημα αυτό αναφέρεται στη βιβλιογραφία ως 'dying ReLU' δηλώνοντας την αδυναμία της ReLU να επαναχρησιμοποιήσει κάποιο νευρώνα που θα λάβει αρνητική τιμή. Για να αντιμετωπιστεί το παραπάνω πρόβλημα χρησιμοποιούμε μια παραλλαγή της ReLU που αντικαθιστά την οριζόντια γραμμή των

αρνητικών τιμών με μια γραμμική συνάρτηση με πολύ μικρή κλίση ώστε να δίνει την δυνατότητα σε έναν νευρώνα που έχει λάβει αρνητική τιμή σε κάποιο διάστημα της μάθησης να επανέλθει (recover). Η παραλλαγή αυτή καλείται Leaky ReLU.

- **Softmax**: Στην πραγματικότητα η συνάρτηση Softmax δεν αποτελεί κατ'ουσίαν συνάρτηση ενεργοποίησης. Εφαρμόζεται στο στρώμα εξόδου των περισσοτέρων Νευρωνικών δικτύων όταν πρόκειται για πρόβλημα κατηγοριοποίησης, ανεξάρτητα από τη συνάρτηση ενεργοποίησης που εφαρμόζεται στους νευρώνες του δικτύου. Οι νευρώνες στο στρώμα εξόδου λαμβάνουν οποιαδήποτε τιμή, αλλά το γεγονός πως πρέπει να δώσουν ως έξοδο μια πιθανότητα οπού ανήκει το κάθε δείγμα σε μια από τις N κλάσεις όταν αντιμετωπίζουμε πρόβλημα κατηγοριοποίησης, δημιούργησε την ανάγκη να κανονικοποιήσουμε τις τιμές εξόδου έτσι ώστε να κατανεμηθούν στο διάστημα [0,1] και το άθροισμα τους να ισούται με 1. Στην ουσία θέλουμε η έξοδος κάθε νευρώνα του στρώματος εξόδου να ισούται με $\hat{y}_i = P(y = i|x)$. Η Softmax δίνεται από τον τύπο:

$$Softmax(z_j) = \frac{e^{z_j}}{\sum_{j=0}^{N} e^{z_j}}$$

2.5.3.2 Συνάρτηση Κόστους (Cost Function)

Η συνάρτηση κόστους (Cost/Loss Function) αποσκοπεί στον έλεγχο της επαναληπτικής διαδικασίας εκπαίδευσης. Συνήθως την συμβολίζουμε με J(θ) και υπολογίζει πόσο κοντά βρίσκεται η έξοδος του δικτύου με την επιθυμητή τιμή για δεδομένες παραμέτρους. Η πιο γανωτή συνάρτηση κόστους χρησιμοποιεί την εντροπία και ονομάζεται Απώλεια Διατροπικής Εντροπίας (**Crossentropy Loss**). Το κόστος υπολογίζεται πάνω στα βάρη-παραμέτρους του δικτύου και εκφράζεται από την εξίσωση:

$$J(\theta) = -H(y, p) = -\sum_{i=0}^{N} \hat{y}_i \log (p_i j)$$

με N το πλήθος των κατηγοριών-κλάσεων των δεδομένων, $\hat{y}_i$ η εκτιμώμενη τιμή για την παρατήρηση i και $p_i j = p(\hat{y}_i = j|x)$ η posteriori πιθανότητα το i δείγμα να ανήκει στην j κλάση [40]. Η παραπάνω εξίσωση υπολογίζει το κόστος ενός δείγματος εισόδου επομένως για τον υπολογισμό του συνολικού κόστους των δεδομένων εισόδου αρκεί να υπολογίσουμε τον αριθμητικό μέσο όρο των επιμέρους σφαλμάτων. Ουσιαστικά η λειτουργία αυτής της συνάρτησης κόστους είναι η σύγκριση των δύο πιθανοτηκών κατανομών, της πρόβλεψης και της αναμενόμενης εξόδου.

Μια επίσης δημοφιλής συνάρτηση κόστους , η οποία χρησιμοποιείται κατα κόρον σε προβλήματα παλινδρόμησης, όπως και στην περίπτωση της δικιάς μας μελέτης, είναι η **Ρίζα Μέσου Τετραγωνικού Σφάλματος (Root Mean Squared Error-RMSE)** η οποία υπολογίζει την ρίζα τετραγωνικής απόστασης μεταξύ της επιθυμητής τιμής και της πρόβλεψης:

$$J(\theta) = \sqrt{\frac{1}{n}\sum_{i=0}^{N}(Y_i - \hat{y}_i)^2}$$

Όπως μπορεί να γίνει εύκολα αντιληπτό αυτή η συνάρτηση κόστους δεν υπολογίζει πιθανοτικές διαφορές αλλά καθαρά αριθμητικές. Ωστόσο πολλές φορές η συνάρτηση αυτή

συγχέεται με την παρόμοια της, την **Mean Squared Error (MAE),** η οποία εκφράζεται με τον τύπο:

$$J(\theta) = \frac{1}{n}\sum_{i=0}^{N} |Y_i - \hat{y}_i|$$

Η διαφορά των δύο συναρτήσεων που μας προέτρεψε να προτιμήσουμε την RMSE έναντι της MAE, είναι περνώντας το συνολικό κόστος σε ριζικό, δίνει μεγαλύτερη βαρύτητα στα μεγαλύτερα κόστη απ' ότι στα μικρότερα, επιτρέποντας τη σωστή διαχείριση των πολύ απομακρυσμένων προβλεπόμενων τιμών από την μέση τιμή (outliers). [41] Μια πιο μαθηματική επεξήγηση είναι η εξής: Όταν η κατανομή των παρατηρήσεων μας είναι ασύμμετρη, δεν ακολουθεί δηλαδή την κανονική κατανομή, όπως στην περίπτωση μας, τότε το κόστος RMSE θα προσπαθεί να ελαχιστοποιηθεί από τον μέσο όρο των τιμών των παρατηρήσεων μας, ενώ το κόστος MAE από την διάμεσο (median) των παρατηρήσεων μας η οποία δεν συμπίπτει απαραίτητα με τον μέσο όρο τους (μπορεί να τείνει σε χαμηλές τιμές ή υψηλές του εύρους τιμών) και έτσι τείνει να εκπαιδεύσει το μοντέλο να προβλέπει τιμές κοντά στην διάμεσο, προκαλώντας μεγαλύτερη συνολική απόκλιση τιμών (biased fit).

2.5.3.3 Αλγόριθμος Βελτιστοποίησης (Optimization Algorithm)

Η διαδικασία προσαρμογής των βαρών για την εξαγωγή ακριβέστερων προβλέψεων είναι γνωστή ως βελτιστοποίηση παραμέτρων. Αφαιρετικά, η διαδικασία αυτή είναι μία μέθοδος κατά την οποία δημιουργείται μία υπόθεση, η υπόθεση συγκρίνεται με την πραγματικότητα και με βάση τη σύγκριση αυτή η υπόθεση βελτιώνεται ή αντικαθίσταται με σκοπό να προσεγγίσει περισσότερο την πραγματικότητα. Κάθε σύνολο βαρών του νευρωνικού δικτύου αντιπροσωπεύει μία συγκεκριμένη υπόθεση για το τι σημαίνουν τα δεδομένα εισόδου. Τα βάρη αντιπροσωπεύουν εικασίες σχετικά με τις συσχετίσεις της εισόδου και της εξόδου που επιδιώκουν να υποθέσουν. Όλα τα πιθανά βάρη και οι συνδυασμοί τους μπορούν να περιγραφούν ως ο υποθετικός χώρος αυτού του προβλήματος. Η προσπάθειά να διαμορφωθεί η καλύτερη υπόθεση είναι θέμα αναζήτησης μέσα από αυτόν τον χώρο και γίνεται με χρήση συναρτήσεων κόστους και αλγορίθμων βελτιστοποίησης. Όσο περισσότερες είναι οι παράμετροι εισόδου, τόσο μεγαλύτερος είναι ο χώρος αναζήτησης του προβλήματος. Μεγάλο μέρος της διαδικασίας μάθησης αποτελεί η απόφαση για το ποιες παράμετροι είναι σημαντικές και ποιες όχι.

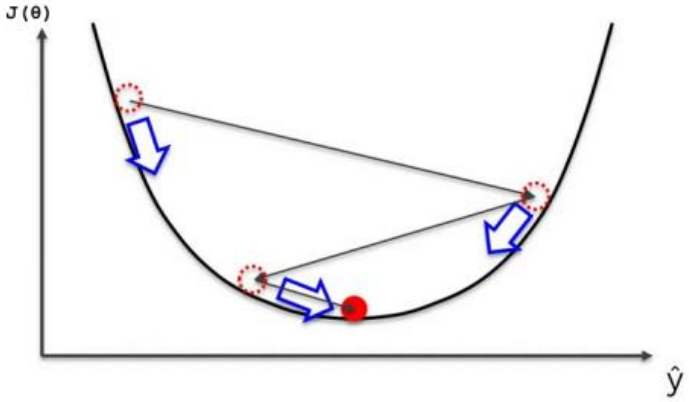
Υπάρχουν δύο κύριοι αλγόριθμοι βελτιστοποίησης οι οποίοι χρησιμοποιούνται ευρέως. Ο πρώτος χρονικά είναι ο αλγόριθμος Οπίσθιας Ανατροφοδότησης Σφάλματος (**Back Propagation**), ο οποίος επαναυπολογίζει και προσαρμόζει τις τιμές βαρών του δικτύου με τον υπολογισμό της κλίσης του κόστους ως προς την κάθε παράμετρο του δικτύου. Περισσότερες πληροφορίες στο [42].

Ένας άλλος, εξίσου διαδεδομένος είναι ο Στοχαστικός Αλγόριθμος Τάχιστης Κατάβασης (**Stohastic gradient Descent**). Ανήκει στην μεγάλη κατηγορία αλγορίθμων Gradient Descent που βασίζουν την λειτουργία τους στην ελαχιστοποίηση ή μεγιστοποίηση μιας συνάρτησης κόστους με την χρήση της κλίσης (gradient) των παραμέτρων του προβλήματος. Ο αλγόριθμος αυτός εκτελείται επαναληπτικά μέχρι να επέλθει σύγκλιση ή να ολοκληρωθεί ένας ορισμένος αριθμός επαναλήψεων (Termination Criteria) και οι παράμετροι του δικτύου ανανεώνονται με βάση την εξίσωση:

$$\theta_{t+1} = \theta_t - \lambda \nabla_\theta J(\theta)$$

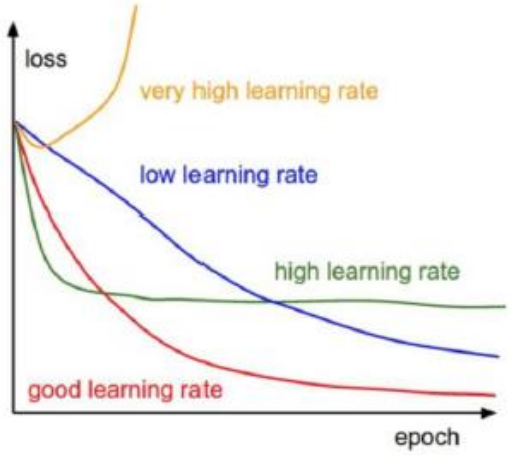
με θ το σύνολο των παραμέτρων του προβλήματος, λ τον ρυθμό μάθησης και J(θ) την συνάρτηση κόστους. Στην πράξη χρησιμοποιείται με τη μορφή μικρών πακέτων δεδομένων (mini-batches) ώστε να έχει ταχύτερη σύγκληση και να μπορεί να εκτελεστεί παράλληλα σε όλους τους πυρήνες των υπολογιστικών συστημάτων. Έτσι κάθε επανάληψη του αλγορίθμου εκτελείται σε ένα τυχαίο δείγμα του συνόλου των δεδομένων εκπαίδευσης, προκαθορισμένου μεγέθους. Ο αλγόριθμος αυτός έχει ορισμένα μειονεκτήματα που μας οδηγούν στην χρήση ορισμένων παραλλαγών του. Το γεγονός πως διατηρεί σταθερό τον ρυθμό μάθησης λ καθώς και η πολύ αργή του σύγκληση σε προβλήματα με μεγάλο πλήθος δεδομένων λόγω της μεγάλης διασποράς της κλίσης (Σχ.2.13) οδήγησαν στη δημιουργία αλγορίθμων προσαρμοσμένων για προβλήματα με μεγάλο πλήθος δεδομένων.

Ένας από αυτούς είναι ο **Adam (adaptive moment estimation)** [43] τον οποίο και θα αξιοποιήσουμε στη μελέτη μας. Ο Adam βασίζεται στη λειτουργία του κλασσικού SGD χρησιμοποιώντας όμως μεταβλητό ρυθμό μάθησης για κάθε παράμετρο-βάρος του δικτύου χρησιμοποιώντας τόσο την πρώτη όσο και την δεύτερη βαθμίδα κλίσης για την λειτουργία του. Ταυτόχρονα για να αποφύγει φαινόμενα ταλάντωσης όπως στο Σχ. 2.13, χρησιμοποιεί δυο μεταβλητές παράληψης (Forget Variables) που επιταχύνουν τη διαδικασία σύγκλησης.



Σχήμα 2.13: Πρόβλημα Σύγκλισης SGD

Δύο σημαντικές παράμετροι του Adam που θα μας απασχολήσουν στη μελέτη μας είναι ο ρυθμός μάθησης (learning rate) και ο ρυθμός απόσβεσης του ρυθμού μάθησης (decay), ίσως και οι πιο σημαντικές υπερπαράμετροι ενός νευρωνικού δικτύου. Ο ρυθμός μάθησης επηρεάζει την ταχύτητα σύγκλισης του δικτύου στις βέλτιστες τιμές βαρών των νευρώνων του, ελέγχει δηλαδή το πόσο μεγάλη ή μικρή θα είναι η αλλαγή στα βάρη του μοντέλου ανάλογα με το σφάλμα που προκύπτει κατά την εκπαίδευση του δικτύου κάθε φορά που ανανεώνονται τα βάρη. Μεγάλος ρυθμός μάθησης μπορεί να οδηγήσει σε γρηγορότερη σύγκλιση και σε ταλάντωση γύρω από τις βέλτιστες τιμές βαρών. Μικρός ρυθμός μάθησης έχει ως αποτέλεσμα πιο αργή σύγκλιση, ενώ μπορεί να οδηγήσει με μεγαλύτερη πιθανότητα σε παγίδευση σε τοπικά ακρότατα. Εδώ λοιπόν παίρνει θέση ο ρυθμός απόσβεσης (decay), ο οποίος καθορίζει το βήμα αλλαγής του ρυθμού μάθησης κατά τη διάρκεια της εκπαίδευσης είτε ανά εποχή είτε ακολουθώντας τη συμπεριφορά κάποιας συνάρτηση [70], ώστε ο ρυθμός μάθησης να μαθαίνει και να συγκλίνει σωστά βάση των απαιτήσεων των παραμέτρων του μοντέλου κατά την εκπαίδευση.



Σχήμα 2.14: Γραφική παράσταση του αλγορίθμου μείωσης κλίσης.

2.5.3.4 Κανονικοποίηση Δικτύου

Είναι σημαντικό σε αυτό το σημείο να καταγραφούν ορισμένες τεχνικές που χρησιμοποιούνται με σκοπό την μείωση του υπολογιστικού κόστους της εκπαίδευσης αλλά και της ικανότητας της γενίκευσης του μοντέλου.

Μια σημαντική τεχνική λοιπόν για την ιδιότητα των μοντέλων να γενικεύουν και να μην υπερπροσαρμόζονται (Overfit) στα δεδομένα εκπαίδευσης είναι ο τυχαίος μηδενισμός βαρών (**Dropout)** [44]. Η τεχνική αυτή μηδενίζει στοχαστικά ορισμένα από τα βάρη του δικτύου με σκοπό να 'νεκρώνονται' ορισμένοι νευρώνες κατά την εκπαίδευση. Αυτό αποσκοπεί στην εκπαίδευση νευρώνων σε ορισμένα χαρακτηριστικά ώστε η απόδοση του μοντέλου να είναι υψηλή για οποιοδήποτε δεδομένο εισόδου.

2.6 Συνελικτικά Νευρωνικά Δίκτυα

2.6.1 Γενική Προσέγγιση

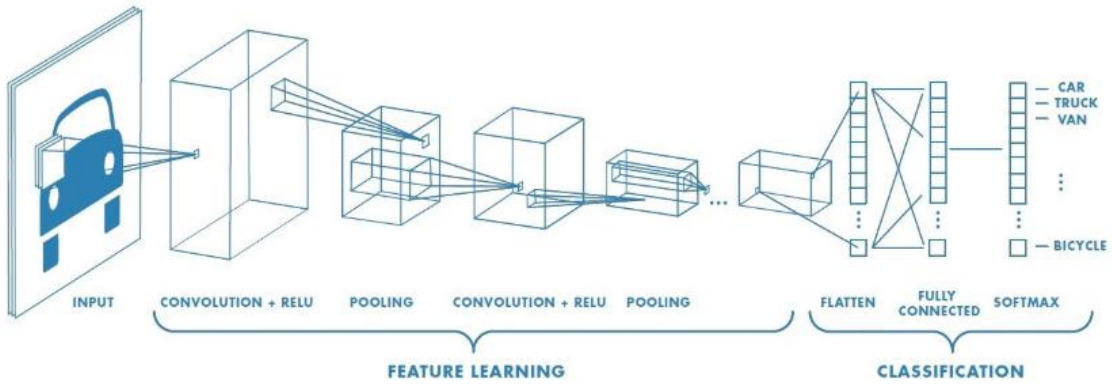
Στην περιοχή της αναγνώρισης σημάτων, εικόνων και ήχων κατά κόρον, έχει καθιερωθεί η χρήση των Συνελικτικών Νευρωνικών Δικτύων (Convolutional Neural Networks-CNN) τα οποία μοιάζουν αρκετά με τα συνηθισμένα Νευρωνικά Δίκτυα. Υπάρχουν λεπτές διαφορές ανάμεσα σε αυτά τα δύο είδη δικτύων, που καθιστούν τα CNN πιο ελκυστικά για χρήση. Πιο συγκεκριμένα, σε ένα παράδειγμα αναγνώρισης εικόνων, τα κανονικά Νευρωνικά Δίκτυα δεν κλιμακώνουν αποδοτικά σε μεγάλες εικόνες. Έστω πώς έχουμε πρότυπα εισόδου εικόνες μεγέθους, μόνο 32 x 32 x 3 (πλάτος x ύψος x χρωματικά κανάλια), και επομένως ένας πλήρως συνδεδεμένος νευρώνας στο πρώτο κρυφό επίπεδο ενός νευρωνικού δικτύου θα είχε 32*32*3 = 3072 βάρη. Παρά το γεγονός ότι αυτός ο αριθμός δείχνει διαχειρίσιμος, για είσοδο εικόνας μεγαλύτερων διαστάσεων, π.χ. 200 x 200 x 3 θα είχαμε νευρώνες με 200*200*3 = 120.000 βάρη ο καθένας. Έτσι ο συνολικός αριθμός των παραμέτρων θα μεγάλωνε ραγδαία. Συνεπώς η πλήρης συνδεσιμότητα είναι σπάταλη, και μπορεί λόγω των πολλών παραμέτρων να οδηγήσει εύκολα σε υπερπροσαρμογή του δικτύου.

Εδώ λοιπον παρεμβαίνουν τα CNN. Γενικότερα τα CNN, διαθέτουν νευρώνες οι οποίοι έχουν την ιδιαιτερότητα να συνδέονται σε μία μικρή περιοχή του προηγούμενου

επιπέδου, αντί με όλους τους νευρώνες όπως θα γινόταν σε μία πλήρη σύνδεση, εξάγοντας έτσι τοπικά χαρακτηριστικά τα οποία θα μπορούσε να αναγνωρίσει σε οποιοδήποτε άλλο σημείου ενός δείγματος εισόδου. Είναι εύκολα κατανοητή λοιπόν η μη ευαισθησία που παρουσιάζουν στη μετατόπιση, κλιμάκωση, στρέβλωση και άλλες μορφές παραμόρφωσης σημάτων.

2.6.2 Συστατικά Μέρη

Ένα Συνελικτικό δίκτυο είναι μία νευρωνική αρχιτεκτονική πολλών επιπέδων ειδικά σχεδιασμένη ώστε να αναγνωρίζει είτε δισδιάστατα είτε μονοδιάστατα σήματα, ακολουθώντας μια καθιερωμένη δομή η οποία ωστόσο ανάλογα τις συνθήκες του προβλήματος μπορεί να μεταβάλλεται. Έτσι λοιπόν τα δομικά βασικά επίπεδα ενός CNN, τα οποία παρουσιάζουμε συνοπτικά σε ένα παράδειγμα στο Σχήμα 2.15 είναι τα εξής:



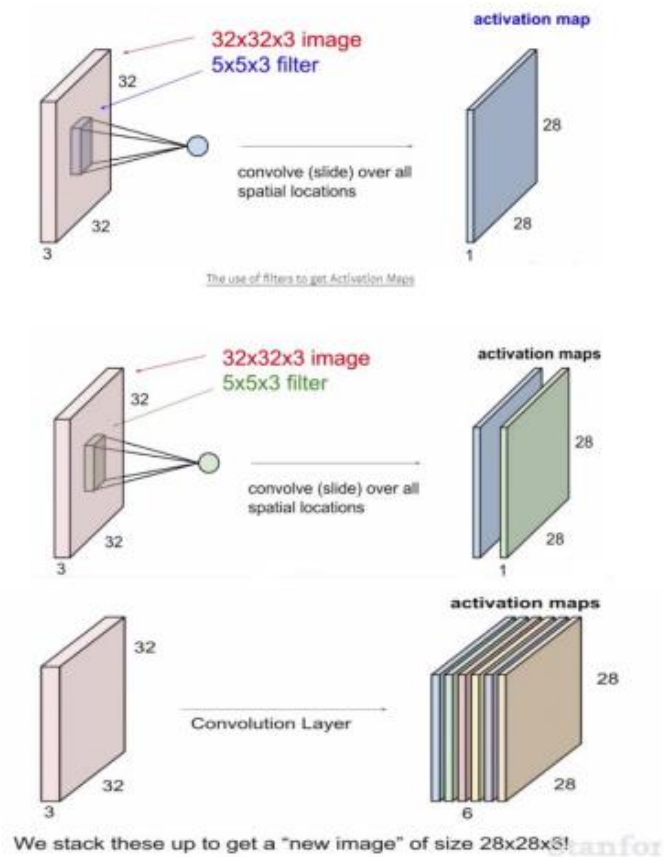
Σχήμα 2.15: Παράδειγμα αρχιτεκτονικής CNN σε πρόβλημα αναγνώρισης οχημάτων

1. Συνελικτικό Επιπεδο (Convolution Layer): Το επίπεδο Συνέλιξης χρησιμοποιεί ένα σύνολο από φίλτρα τα οποία εντοπίζουν την παρουσία συγκεκριμένων χαρακτηριστικών ή μοτίβων που παρουσιάζονται στο ακατέργαστο σήμα που δίνεται στην είσοδο (input). Συνήθως έχουν μικρότερες διαστάσεις από αυτές της εισόδου, αλλά διατηρούν τη διάσταση του βάθους ίδια με αυτές. Το κάθε φίλτρο λοιπόν "γλιστρά" κατά πλάτος και κατά ύψος της εισόδου αν είναι δισδιάστατη, και ένα εσωτερικό γινόμενο υπολογίζεται για να δώσει ένα χάρτη χαρακτηριστικών (feature map). Διαφορετικά φίλτρα τα οποία εντοπίζουν διαφορετικά χαρακτηριστικά περιστρέφονται στην είσοδο είτε είναι τα αρχικα δεδομένα είτε ένας χάρτης χαρακτηριστικών που προέκυψε από προηγούμενο επίπεδο και ένα σύνολο από χάρτες ενεργοποίησης προκύπτει ως έξοδος, το οποίο περνά στο επόμενο επίπεδο του CNN, όπως φαίνεται στο σχήμα 2.15. Τα φίλτρα εφαρμόζονται στα δεδομένα εισόδου με τη λογική ενός κινούμενου παραθύρου. Η έξοδος του φίλτρου υπολογίζεται παράγοντας το άθροισμα του ένα-προς-ένα πολλαπλασιασμού (element-wise product) των στοιχείων του φίλτρου και της περιοχής του πίνακα εισόδου. Οι παράμετροι του συνελικτικού επιπέδου του δικτύου καθορίζουν το σύνολο των φίλτρων που θα χρησιμοποιηθούν. Η αρχιτεκτονική του συνελικτικού δικτύου ορίζεται με τέτοιο τρόπο ούτως ώστε τα παραγόμενα φίλτρα να εξάγουν την ισχυρότερη ενεργοποίηση σε χωρικά τοπικά πρότυπα εισόδου. Αυτό σημαίνει ότι τα φίλτρα έχουν μάθει ότι θα ενεργοποιηθούν σε μοτίβα (ή χαρακτηριστικά) μόνο όταν τα μοτίβα εμφανίζονται στα δεδομένα εκπαίδευσης στο αντίστοιχο πεδίο. Σε πιο βαθιά επίπεδου του δικτύου τα φίλτρα μπορούν να αναγνωρίσουν μη γραμμικούς συνδυασμούς χαρακτηριστικών και είναι ολοένα και πιο

αφηρημένα για το πώς μπορούν να ανιχνεύσουν πρότυπα σε οποιαδήποτε θέση των δεδομένων εισόδου. Για το λόγο αυτό παρατηρείται ότι τα συνελικτικά δίκτυα με πολύ υψηλή απόδοση έχουν μεγάλο βάθος.

Υπερπαράμετροι Συνελικτικού Επιπέδου (hyperparameters):
Οι υπερπαράμετροι που καθορίζουν τη χωρική διάταξη και το μέγεθος των δεδομένων εξόδου από ένα συνελικτικό επίπεδο είναι:

- *Μέγεθος φίλτρου*: Κάθε φίλτρο είναι μικρό χωροταξικά σε σχέση με το πλάτος και το ύψος του μεγέθους της εισόδου. Για παράδειγμα, το μέγεθος του φίλτρου μπορεί να είναι 5x5x3, ενώ οι διαστάσεις της εισόδου είναι 32x32x3.
- *Βάθος εξόδου*: Αντιστοιχεί στο πλήθος των φίλτρων που θα χρησιμοποιηθούν, για τον εντοπισμό διαφορετικών χαρακτηριστικών της εισόδου.
- *Βήμα (stride):* Ρυθμίζει πόσο μακριά θα μετακινηθεί το παράθυρο συρόμενου φίλτρου ανά εφαρμογή της λειτουργίας φίλτρου. Όταν το βήμα έχει την τιμή 1, τότε το φίλτρο μετακινείται κατά ένα pixel της φορά (εικόνα). Αύξηση του βήματος επιφέρει χωρική μείωση του όγκου εξόδου.
- *Γέμισμα ακραίων τιμών:* Το μέγεθος του γεμίσματος γύρω από τα σύνορα εισόδου. Συνήθης πρακτική είναι η χρήση μηδενικών (zero-padding).

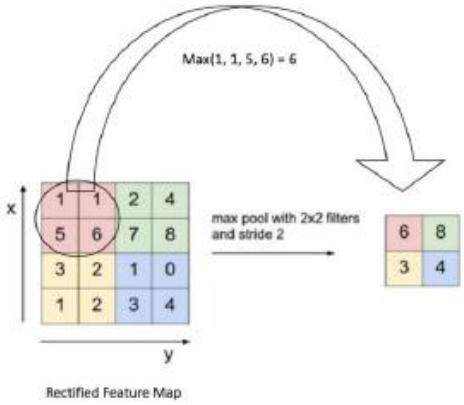


Σχήμα 2.16: Εφαρμογή φίλτρων σε μια εικόνα και παραγωγή χαρτών χαρακτηριστικών/ενεργοποίησης

2. Συνάρτηση Ενεργοποίησης (Μη γραμμική – ReLU): Πρακτικά αποτελούν έναν κόμβο που τοποθετείται μετά το επίπεδο της συνέλιξης. Δεν μεταβάλει το μέγεθος της

εισόδου του και βοηθά στο να παρθεί η απόφαση για το αν θα υδροδοτηθεί ένας νευρώνας ή όχι. Ουσιαστικά η συνάρτηση ενεργοποίησης είναι ένας επιπλέον μη-γραμμικός μετασχηματισμός που εφαρμόζουμε στο σήμα εισόδου. Η μετασχηματισμένη έξοδος στέλνεται στο επόμενο επίπεδο, για το οποίο και αποτελεί σήμα εισόδου. Όπως προαναφέραμε υπάρχουν πολλές συναρτήσεις ενεργοποίησης, όμως η πιο συχνά χρησιμοποιούμενη στις μέρες μας στα Νευρωνικά Δίκτυα είναι η ReLU (rectified linear unit). Ο λόγος είναι οτι αυτή η συνάρτηση δεν ενεργοποιεί όλους τους νευρώνες ταυτόχρονα, δηλαδή τις αρνητικές τιμες τις μετατρέπει σε μηδέν κι έτσι ο εκάστοτε νευρώνας δεν ενεργοποιείται. Αυτό κάνει το σύστημά μας πιο αποδοτικό μιας και δεν λειτουργούν όλοι οι νευρώνες κάθε φορά και μάλιστα στην πράξη συγκλίνει πολύ πιο γρήγορα από τη σιγμοειδή συνάρτηση και την συνάρτηση υπερβολικής εφαπτομένης.

3. Συγκεντρωτικό Επίπεδο (Pooling Layer): Τα επίπεδα Συγκέντρωσης συναντώνται ανάμεσα από τα διαδοχικά επίπεδα Συνέλιξης των CNN. Στόχος είναι η προοδευτική μείωση του χωρικού μεγέθους της αναπαράστασης δεδομένων και των παραμέτρων του δικτύου, έτσι ώστε να ελεγχθεί η υπερπροσαρμογή (overfitting) του μοντέλου. Επομένως τα συγκεντρωτικά επίπεδα μειώνουν τις διαστάσεις κάθε χάρτη χαρακτηριστικών αλλά διατηρεί τις πιο σημαντικές πληροφορίες. Οι πιο συνηθισμένες λειτουργίες αυτού του επιπέδου είναι η Μέση Συγκέντρωση (Average Pooling) και η Μέγιστη Συγκέντρωση (Max Pooling). Θα εστιάσουμε όμως περισσότερο στη δεύτερη μιας και θα την αξιοποιήσουμε στην έρευνά μας. Στη μέθοδο **Max-pooling**, όπως λέει και το όνομά της, ορίζεται ένας πίνακας-παράθυρο (πχ. διαστάσεων 2 ή 2x2) που μετακινείται πάνω στο χάρτη χαρακτηριστικών και διατηρεί μόνο το μεγαλύτερο στοιχείο (max pooling) από την εκάστοτε περιοχή. Μέσω αυτής της διαδικασίας γίνεται υπό-δειγματοληψία του δικτύου. Ένα παράδειγμα φαίνεται στο παρακάτω σχήμα:

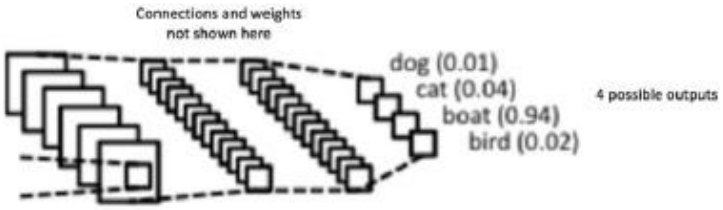


Σχήμα 2.17: Παράδειγμα εφαρμογής της λειτουργίας max pooling

Παράδειγμα εφαρμογής της λειτουργίας max pooling σε χάρτη χαρακτηριστικών που προέκυψε μετά από συνέλιξη και εφαρμογή της συνάρτησης ReLU. Το φίλτρο έχει διαστάσεις 2x2 και το βήμα (stride) έχει τιμή 2. Η είσοδος έχει διαστάσεις 4x4 και μετά από την εφαρμογή της συγκεντρωτικής λειτουργίας η έξοδος έχει διαστάσεις 2x2

4. Πλήρως συνδεδεμένο επίπεδο (Fully Connected Layer): Το πλήρως συνδεδεμένο επίπεδο είναι μία παραδοσιακή αρχιτεκτονική πολλών επιπέδων με νευρώνες, η οποία χρησιμοποιεί μία συνάρτηση ενεργοποίησης (συνήθως τη Softmax) στην έξοδό της. Κάθε επίπεδο που ανήκει σε αυτή την αρχιτεκτονική έχει την ιδιότητα πως κάθε νευρώνας που περιλαμβάνει συνδέεται με όλους τους νευρώνες του προηγούμενου επιπέδου. Η έξοδος των συνελικτικών και συγκεντρωτικών επιπέδων αναπαριστά χαρακτηριστικά υψηλών στρωμάτων. Στη βασική περίπτωση, που το πρόβλημα ανήκει στην κατηγορία της ταξινόμησης (classification), ο σκοπός του πλήρως συνδεδεμένου επιπέδου είναι να

χρησιμοποιήσει αυτά τα χαρακτηριστικά έτσι ώστε να ταξινομήσει την εικόνα εισόδου σε διάφορες κλάσεις, βασιζόμενο στο σύνολο δεδομένων που χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου. Εκτός από την ταξινόμηση, η προσθήκη ενός πλήρως συνδεδεμένου στρώματος είναι επίσης ένας (συνήθως) «φτηνός» τρόπος εκμάθησης μη γραμμικών συνδυασμών αυτών των χαρακτηριστικών. Τα περισσότερα χαρακτηριστικά από τα στρώματα συνένωσης και συγκέντρωσης μπορεί να είναι καλά για την εργασία ταξινόμησης, αλλά οι συνδυασμοί αυτών των χαρακτηριστικών μπορεί να είναι ακόμη καλύτεροι. Το άθροισμα των πιθανών εξόδων από το πλήρως συνδεδεμένο επίπεδο είναι 1, αν πρόκειται για πρόβλημα ταξινόμησης καθώς αποτελούν το άθροισμα όλων των πιθανοτήτων. Αυτό επιτυγχάνεται με τη χρήση της Softmax ως της συνάρτησης ενεργοποίησης στο επίπεδο εξόδου του πλήρως συνδεδεμένου στρώματος. Η συνάρτηση Softmax δέχεται σαν είσοδο ένα διάνυσμα από τυχαίες πραγματικές τιμές και τις αντιστοιχίζει σε ένα διάνυσμα τιμών από 0 έως 1. Το άθροισμα των στοιχείων του διανύσματος εξόδου ισούται με 1. Στην περίπτωση προβλήματος παλινδρόμησης (regression), όπως στα πλαίσια της μελέτης μας, στο τελευταίο επίπεδο πρέπει να υπάρχει ένας μόνο νευρώνας του οποίου η έξοδος να αποτελεί την τελική έξοδο του μοντέλου. Επομένως σαν συνάρτηση ενεργοποίησης χρησιμοποιείται η γραμμική (Linear) καθώς αναζητούμε σαν έξοδο μια συνεχή τιμή που έχει προκύψει από τους προηγούμενους μη γραμμικούς συνδυασμούς των εξαγόμενων χαρακτηριστικών.



Σχήμα 2.18: Παράδειγμα ενός πλήρως συνδεδεμένου δικτύου (FC).

Η έξοδος του δικτύου είναι 4 τιμές που δηλώνουν την πιθανότητα να ανήκει η εικόνα εισόδου σε κάθε μία από τις 4 κλάσεις (σκύλος, γάτα, βάρκα, πουλί). Το μοντέλο εξάγει το συμπέρασμα ότι με μεγάλη πιθανότητα (0.94) στην εικόνα υπάρχει μία βάρκα.

2.7 Επεξεργασία Φυσικής Γλώσσας

2.7.1 Εφαρμογές της Επεξεργασίας Φυσικής Γλώσσας

Η Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP) αποτελεί πεδίο της επιστήμης υπολογιστών, της τεχνητής νοημοσύνης και της γλωσσολογίας με βασικό ερευνητικό ενδιαφέρον την κατανόηση, την παραγωγή αλλά και την επεξεργασία της φυσικής γλώσσας. Είναι η διαδικασία του υπολογιστή κατά τη οποία εξάγει σημαντικές πληροφορίες από την φυσική γλώσσα που δέχεται ως είσοδο και/ή παράγει φυσική γλώσσα ως έξοδο. Είναι η ανάλυση της ανθρώπινης γλώσσας βάσει της σημασιολογίας και διάφορων τεχνικών ανάλυσης [45]. Το NLP αποσκοπεί στο να προσδιορίσει τον υπολογιστικό μηχανισμό που απαιτείται ώστε να παρουσιάζει διάφορες μορφές γλωσσικής συμπεριφοράς. Αποσκοπεί επίσης στο σχεδιασμό, την υλοποίηση και την αξιολόγηση συστημάτων που επεξεργάζονται τις φυσικές γλώσσες για πρακτικές εφαρμογές. Το 1950 ο Turing πρωτοασχολήθηκε με την δημιουργία υπολογιστών κατάλληλων να επεξεργαστούν δεδομένα κειμένου [46], ενώ το 1954 ξεκίνησε η

προσπάθεια δημιουργίας μεταφραστικών μηχανών, χωρίς όμως να έχουν ιδιαίτερη επιτυχία. Το 1964 έμελλε να είναι μια ημερομηνία σταθμός για την επεξεργασία φυσικής γλώσσας αφού δημιουργήθηκε μια μηχανή, η ELIZA, που μπορούσε να αλληλοεπιδρά με τον άνθρωπο, κάνοντας λογικές ερωτήσεις ανάλογα με τις απαντήσεις του ανθρώπου [47]. Η ELIZA δημιουργήθηκε σε συνεργασία ψυχολόγων και επιστημών του εργαστηρίου τεχνητής νοημοσύνης του MIT, που θεώρησαν τη δημιουργία μιας μηχανής ικανής να επικοινωνήσει με τον άνθρωπο ιδιαίτερα σημαντική για ανθρώπους που έπασχαν από κατάθλιψη. Οι πιο δημοφιλείς εφαρμογές του NLP στη σύγχρονη κοινωνία είναι οι εξής:

- Παραγωγή Κειμένου (Text Generation): Μία μέθοδος για τη δημιουργία φυσικών προτάσεων από "λέξεις-κλειδιά" σε διαλογικό σύστημα.
- Περίληψη Κειμένου (Text Summarization): Για την δημιουργία σύντομων περιλήψεων δεδομένων κειμένου, όπως τα βιβλία.
- Μηχανική Μετάφραση (Machine Translation): Για τη δημιουργία αυτόματων μεταφραστών από μια γλώσσα σε μια άλλη.
- Εξόρυξη Κειμένου και Ανάκτηση Πληροφορίας (Text Mining and Information Retrieval): Για την δημιουργία μηχανών που επεξεργάζονται συλλογές κειμένων με σκοπό την εξαγωγή χρήσιμων πληροφοριών από αυτές.
- Γλωσσολογική και Συντακτική Ανάλυση (Language and Syntactic Analysis): Για την γραμματική αλλά και συντακτική ανάλυση μιας πρότασης ή την δημιουργία συντακτικών δέντρων για την απόδοση της πληροφορίας μιας πρότασης.
- Σημασιολογική και Πραγματολογική Ανάλυση (Semantics and Pragmatics Analysis): Για την ανάλυση των λέξεων της πρότασης που περιέχουν την μεγαλύτερη πληροφορία και για την ανάλυση του περιεχομένου και του νοήματος της πρότασης αντίστοιχα.

Όσον αφορά την **Ανάλυση Συναισθήματος (Sentiment Analysis),** μηχανισμοί όπως η Εξόρυξη Κειμένου και Ανάκτηση Πληροφορίας, η Γλωσσολογική και Συντακτική Ανάλυση και η Σημασιολογική και Πραγματολογική Ανάλυση συνδυάζονται για την εξόρυξη συναισθημάτων του ανθρώπου πάνω σε ένα μεγάλο εύρος εφαρμογών, όπως προτιμήσεις σε προϊόντα, υπηρεσίες, οργανισμούς, στην θεραπεία ψυχολογικών διαταραχών και σε πολλές άλλες. Με την εκρηκτική αύξηση χρηστών των μέσων μαζικής ενημέρωσης (social media), που περιλαμβάνουν αξιολογήσεις προϊόντων, συζητήσεις σε φόρουμ, μπλογκ, σχόλια και δημοσιεύσεις σε Twitter, Facebook κ.λπ. στο ίντερνετ, τόσο οι άνθρωποι ατομικά όσο και οι οργανισμοί τα χρησιμοποιούν όλο και περισσότερο για τη λήψη αποφάσεων. Για ένα οργανισμό, δεν είναι πια απαραίτητο να διεξάγει έρευνες για να συλλέξει απόψεις χρηστών, διότι αυτές οι πληροφορίες βρίσκονται πια στο ίντερνετ και είναι δημόσιες. Ωστόσο, η εύρεση και παρακολούθηση απόψεων στο ίντερνετ, καθώς και η αξιολόγηση των πληροφοριών που περιέχουν συνεχίζει να είναι ένα δύσκολο πρόβλημα, λόγω της πληθώρας ιστοσελίδων και χρηστών.

Ωστόσο ο κλάδος της Ανάλυσης Συναισθήματος, συμβάλει κυρίως στην μελέτη της πολικότητας μιας πρότασης/ενός κειμένου, αν δηλαδή εκφράζει θετική ή αρνητική άποψη. Έτσι λοιπόν, εδώ παίρνει θέση ο κλάδος της **Αναγνώρισης Συναισθήματος (Emotion Recognition),** ο οποίος συμβάλει στην εξόρυξη πιο στοιχειωδών και λεπτομερών συναισθημάτων του χρήστη βάση της γραπτής έκφρασης συναισθημάτων του, όπως λύπη, χαρά, φόβος, αηδία, θυμό κλπ. Το ίντερνετ περιέχει τεράστια συλλογή εγγράφων που με τη σειρά τους περιέχουν μεγάλες ποσότητες κειμένου. Οι πηγές κειμένου που είναι χρήσιμες για την αναγνώριση συναισθημάτων είναι οι κριτικές προϊόντων, άρθρα εφημερίδων, ανάλυση χρηματιστηρίου, προσωπικά blog, περιοδικά, ιστοσελίδες κοινωνικής δικτύωσης, φόρουμ, κριτικές, πολιτικά debate; στην πραγματικότητα,

οπουδήποτε οι άνθρωποι εκφράζουν και μοιράζονται τις απόψεις τους ελεύθερα. Ανάμεσα στα πολλά προβλήματα με τα οποία ασχολείται η αυτόματη αναγνώριση κειμένου, η αυτόματη αναγνώριση συναισθημάτων είναι ένα από τα πιο ταχέως αναπτυσσόμενα και ενδιαφέροντα. Μιας και το συναίσθημα είναι σημαντικό για να κατανοήσουμε την ανθρώπινη εμπειρία και επικοινωνία, τα συναισθήματα έχουν μελετηθεί από τις επιστήμες της ψυχολογίας και της συμπεριφορικής ανάλυσης. Μέσω της αναγνώρισης κειμένου, μπορούμε πλέον αυτόματα να αναγνωρίζουμε και ταξινομούμε τα συναισθήματα στο γραπτό λόγο; ωστόσο, οι μέθοδοι δεν είναι πάντα συνεπείς και υπάρχουν πολλές προκλήσεις ακόμα στη σύγκριση διάφορων προσεγγίσεων.

2.7.2 Μηχανική Μάθηση σε Κείμενο

Οι αλγόριθμοι μηχανικής μάθησης είναι δομημένοι ώστε να κατατάσσουν και να επεξεργάζονται αριθμητικά δεδομένα. Για να εφαρμόσουμε λοιπόν τους αλγορίθμους ML σε δεδομένα που έχουν την μορφή κειμένου πρέπει αυτά να μετασχηματιστούν σε αριθμητικά δεδομένα, ως διάνυσμα χαρακτηριστικών (Feature Vector), ώστε να τεθούν ως είσοδοι στους αλγορίθμους αυτούς. Η πιο διαδεδομένη τεχνική στην αναπαράσταση του λεξιλογίου ενός κειμένου είναι η ενσωμάτωση λέξεων (Word Embeddings), σύμφωνα με την οποία η αναπαράσταση κάθε λέξης στο λεξιλόγιο ενός κειμένου γίνεται με ένα διάνυσμα. Για τον υπολογισμό αυτών των διανυσμάτων υπάρχουν διάφορες μεθοδολογίες, τις οποίες παρουσιάζουμε στη συνέχεια.

2.7.2.1 Σύνολο από λέξεις (Bag of Words)

Η μέθοδος Bag of Words (BoW) [48] είναι η πιο απλή τεχνική για τη δημιουργία διανύσματος χαρακτηριστικών. Το BoW βασίζεται σε ένα αραιό διάνυσμα (sparse vector) αναπαράστασης της κάθε λέξης, με το μήκος του διανύσματος να ορίζεται όσο και το μέγεθος του λεξικού. Με αυτή την προσέγγιση κάθε λέξη του λεξικού λαμβάνει ένα προσδιοριστικό αύξοντα αριθμό id, οπότε το sparse vector της λέξης αυτής λαμβάνει την τιμή 1 στην θέση id και 0 στις υπόλοιπες. Η κωδικοποίηση αυτή ονομάζεται One-hot encoding και η ονομασία προέρχεται από την τιμή 1 που λαμβάνει το διάνυσμα της κάθε λέξης. Στη μεθοδολογία αυτή το κείμενο αντιμετωπίζεται ως ένας σάκος (bag) από λέξεις, χωρίς να λαμβάνεται υπόψιν η γραμματική, το συντακτικό του κειμένου, η σειρά των λέξεων ή η εξάρτησή τους από γειτονικές λέξεις. Λαμβάνεται υπόψιν μόνο η παρουσία ή η συχνότητα εμφάνισης των λέξεων. Το γεγονός ότι δεν επηρεάζεται από συμφραζόμενα κάνει την δημιουργία χαρακτηριστικών ιδιαίτερα απλή με ελάχιστο υπολογιστικό κόστος, γεγονός που κάνει το μοντέλο αυτό ιδιαίτερα δημοφιλές. Σε αυτό το σημείο είναι χρήσιμο να ορίσουμε το Uni-gram μοντέλο κατά το οποίο μια λέξη μέσα σε ένα κείμενο αποτελεί αυθαίρετο στοιχείο με μηδενική εξάρτηση από την προηγούμενη και την επόμενη λέξη και αυτή είναι η λογική της μεθόδου BoW. Το BoW είναι ευρέως διαδεδομένο για μεθόδους κατηγοριοποίησης κειμένου, όπου η συχνότητα εμφάνισης μιας λέξης χρησιμοποιείται για την εκπαίδευση ενός ταξινομητή. Παρόλα αυτά το μοντέλο αυτό εμφανίζει και διάφορες αδυναμίες όταν για την ταξινόμηση απαιτείται η κατανόηση του κειμένου [49].

Ένας τρόπος επιμέρους αντιμετώπισης του προβλήματος αυτού είναι το σπάσιμο της πρότασης σε n-grams. Η χρήση των n-grams (συνήθως bigram και trigram), δηλαδή ακολουθίες από n συνεχόμενους όρους, αντί για μεμονωμένες λέξεις βοηθάει στη διατήρηση σημασιολογικών σχέσεων μεταξύ των λέξεων. Ωστόσο η προσέγγιση αυτή προσδίδει μεγάλη κατανάλωση χώρου καθώς αυξάνεται το μέγεθος του λεξιλογίου,

παρόλα αυτα παραμένει επικρατέστερη μέθοδος της απλής BoW. Τα μοντέλα αυτά χρησιμοποιούνται στα Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models) [50], αλλά παρουσιάζουν προβλήματα όταν εμφανίζονται λέξεις που δεν έχουν ξαναπαρουσιαστεί στο λεξικό εκπαίδευσης (Out of Vocabulary-OoV).

2.7.2.2 Term Frequency-Inverse Document Frequency (TF-IDF)

Το μοντέλο TF-IDF στην ουσία αποτελεί επέκταση του BoW. Όπως και το μοντέλο BoW έτσι και το TF-IDF βασίζεται στην δημιουργία διανύσματος χαρακτηριστικών σε αραιούς πίνακες που όμως λαμβάνουν μια πραγματική τιμή στις σημαντικές θέσεις της πρότασης σε αντίθεση με τα BoW που οι τιμές αντιπροσωπεύουν απλές συχνότητες. Ένα πρόβλημα το οποίο παρατηρήθηκε στο BoW μοντέλο και ώθησε τους ερευνητές στην δημιουργία του TF-IDF είναι πως μερικές φορές λέξεις οι οποίες χρησιμοποιούνται ιδιαίτερα συχνά σε πολλά κείμενα δεν φέρουν την πληροφορία των προτάσεων σε αντίθεση με λέξεις που εμφανίζονται συχνά σε κάποιο μεμονωμένο κείμενο. Το γεγονός αυτό διατύπωσε η Sparks [51] δημιουργώντας την έννοια του Inverse Document Frequency σε μια προσπάθεια να ενισχυθούν οι λέξεις που εμφανίζονται συχνά σε ένα κείμενο και να επιβληθεί μια ποινή στις πιο συχνές λέξεις σε ένα σύνολο δεδομένων κατά την διαδικασία προεπεξεργασίας του. Το IDF υπολογίζει το πόσο σπάνια εμφανίζεται μια λέξη μέσα σε ένα σύνολο κειμένων, οπότε αν συνδυαστεί με την Term Frequency- TF δημιουργείται ένα μοντέλο που μπορεί να ανιχνεύσει, σε έναν βαθμό, λέξεις που περιέχουν την μεγαλύτερη πληροφορία σε ένα κείμενο. Η τιμή TF-IDF για μια λέξη w σε ένα κείμενο d από ένα σύνολο κειμένων D, υπολογίζεται από τον τύπο:

$$TF - IDF(w, d, D) = TF(w, d) * IDF(w, d, D)$$

Όπου,

$$TF(w, d) = frew(w, d)$$

$$IDF(w, d, D) = \log\left(\frac{N}{count(d \in D : w \in d)}\right)$$

Παρόλα αυτά και αυτό το μοντέλο παρουσιάζει αρκετές αδυναμίες. Όπως και στο BoW, το μοντέλο δεν μπορεί να αντλήσει από τα συμφραζόμενα την πληροφορία για τη σειρά των λέξεων, ενώ φράσεις που είναι παρόμοιες σημασιολογικά αναγνωρίζονται ως τελείως ξένες. Ταυτόχρονα η πολυπλοκότητα του μοντέλου είναι σαφώς μεγαλύτερη σε σύγκριση με την απλοϊκή προσέγγιση του BoW και σε πολύ μεγάλα δεδομένα κειμένων η αραιή μορφή πινάκων να καταλαμβάνει μεγάλο χώρο μνήμης και να μην είναι λειτουργική.

2.7.2.3 Διανύσματα Λέξεων (Word Embeddings)

Η δημιουργία των Word Embeddings ήταν η πρώτη που επήλθε χρονικά στη δημιουργία αναπαραστάσεων για δεδομένα κειμένου από τον Bengio το 2003 [52]. Αποτελούν Τεχνητά Νευρωνικά Δίκτυα τα οποία εκπαιδεύονται με αλγορίθμους Stohastic Gradient Descent-SGD σε ένα μεγάλο σε έκταση κείμενο (corpus), της τάξης των δισεκατομμυρίων λέξεων, χρησιμοποιώντας τεχνικές μη επιβλεπόμενης μάθησης. Το δίκτυο του Bengio αποτελείται από ένα επίπεδο και χρησιμοποιεί την συνάρτηση Softmax στην έξοδό του ώστε να υπολογίζει τις n-gram πιθανότητες. Η ιδέα πίσω από τα διανύσματα λέξεων που έχουν δημιουργηθεί με μη επιβλεπόμενη μάθηση είναι ότι θα θέλαμε τα διανύσματα παρόμοιων λέξεων να έχουν κοντινές μεταξύ τους τιμές. Ενώ η ομοιότητα λέξεων είναι δύσκολο να προσδιοριστεί και εξαρτάται από το εκάστοτε πρόβλημα, οι σύγχρονες προσεγγίσεις αντλούν έμπνευση από την κατανεμημένη υπόθεση

(distributional hypothesis) [53], υποστηρίζοντας ότι οι λέξεις έχουν παρόμοιο νόημα όταν εμφανίζονται σε παρόμοιο περιεχόμενο. Δηλαδή, οι τεχνικές δημιουργίας word embeddings επιχειρούν να παράγουν κατανεμημένες αριθμητικές αναπαραστάσεις λέξεων, οι οποίες κωδικοποιούν την ομοιότητα των λέξεων. Πολλές διαφορετικές μέθοδοι δημιουργούν επιβλεπόμενα παραδείγματα εκπαίδευσης, στόχος των οποίων είναι είτε να προβλέψουν τη λέξη βάσει του περιεχομένου, ή να προβλέψουν το περιεχόμενο βάσει της λέξης. Παρακάτω παρουσιάζουμε δύο από αυτές που θα μας απασχολήσουν στην έρευνά μας.

2.7.2.3.1 Word2Vec

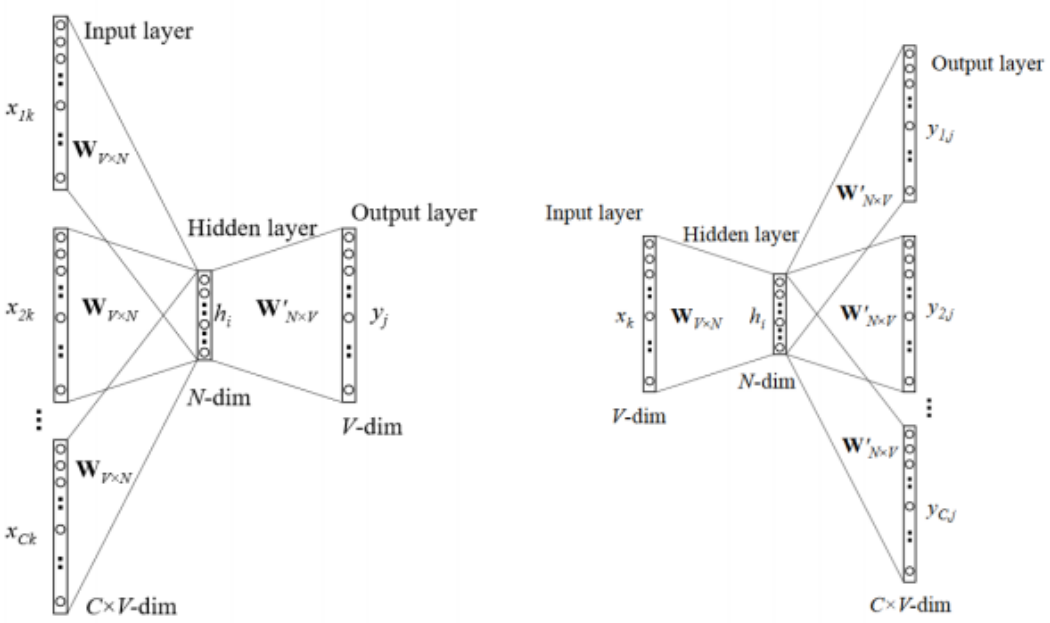
Το 2013 ο Mikolov [54] εισήγαγε την έννοια του αλγορίθμου Word2Vec που συνέβαλε στη δημιουργία μεγάλης ερευνητικής περιοχής γύρω από την ομοιότητα κειμένων (Text Similarity). Το Word2Vec αποτελεί τεχνητό νευρωνικό δίκτυο δύο στρωμάτων νευρώνων και επεξεργάζεται ένα μεγάλο κείμενο (corpus) για να εξάγει διανύσματα χαρακτηριστικών για τις λέξεις του κειμένου. Κάθε μοναδική λέξη που ανήκει στο corpus αποκτά μια διανυσματική αναπαράσταση με τέτοιο τρόπο ώστε λέξεις με παρόμοια συμφραζόμενα να βρίσκονται κοντά στον διανυσματικό χώρο του κειμένου R^N. Τα διανύσματα των λέξεων στην τεχνική Word2Vec είναι κωδικοποιημένα στη μορφή One-hot, που σημαίνει ότι κάθε διάνυσμα έχει μήκος V , όσο και το μέγεθος του λεξιλογίου, αποτελείται από μηδενικά σε όλα τα στοιχεία του εκτός από το στοιχείο που αντιπροσωπεύει την συγκεκριμένη λέξη στο λεξιλόγια, στη θέση του οποίου έχει 1. Στο κρυφό επίπεδο, αθροίζονται τα γινόμενα των εισόδων με τους πίνακες παραμέτρων και στο επίπεδο της εξόδου, εφαρμόζεται η συνάρτηση Softmax για την ομοιότητα και τη συσχέτιση μεταξύ των λέξεων που προσπαθεί να προβλέψει τις σωστές θέσεις των άσσων, δηλαδή τις σωστές λέξεις, στα διανύσματα One-hot της εξόδου. Οι μετρικές που χρησιμοποιούνται κυρίως είναι η Ομοιότητα Συνημιτόνου (Cosine Similarity), αλλά και η Ευκλείδεια απόσταση. Ο αλγόριθμος χρησιμοποιεί δύο βασικές τεχνικές του NLP για τη δημιουργία κατανεμημένων αναπαραστάσεων των λέξεων κειμένου (Distributed Representation of Words), έναν για την εύρεση της λέξης των συμφραζομένων (Continuous Bag Of Words-CBoW) και έναν για την εύρεση των συμφραζομένων με βάση τη λέξη (Skip-Gram), όπως απεικονίζεται στο Σχήμα 2.19. Αναλυτικότερα, οι δύο τεχνικές που προαναφέρθηκαν μπορούν να μεταφραστούν μέσω της συνάρτησης κόστους στα δύο εξής προβλήματα ελαχιστοποίησης:

- Συνάρτηση Κόστους CBoW:

$$J(\theta) = \frac{1}{T}\sum_{t=1}^{T} p(w_t | w_{t-n}, w_{t-n-1}, \dots, w_{t-1}, w_{t+1}, \dots, w_{t+n-1}, w_{t+n})$$

- Συνάρτηση Κόστους Skip-Gram:

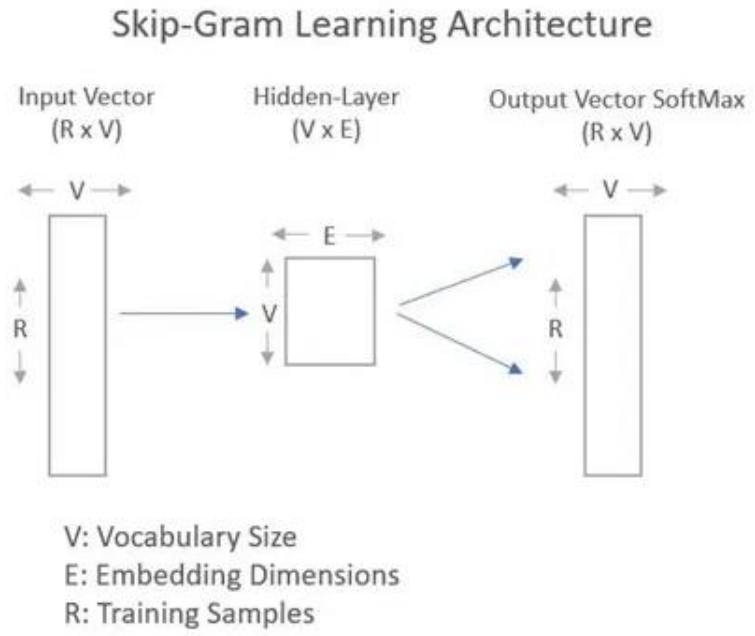
$$J(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=-n}^{n} p(w_{t+j} | w_t)$$



Σχήμα 2.19: CBoW vs Skip-Gram

Όπως γίνεται αντιληπτό, η συνάρτηση κόστους για το μοντέλο CBoW αποσκοπεί στην πρόβλεψη της λέξης w_t χρησιμοποιώντας την γνώση των n προηγούμενων και των n επόμενων λέξεων ενώ αντίθετα το μοντέλο Skip-gram αποσκοπεί στην πρόβλεψη των n γειτονικών λέξεων που συνοδεύουν την λέξη w_t.

Τελικά, οι κατανεμημένες αναπαραστάσεις των λέξεων προκύπτουν από τον πίνακα του κρυφού επιπέδου VxE, όπου V το πλήθος των διακριτών λέξεων του κειμένου και E η διάσταση της κάθε αναπαράστασης της λέξης που συνήθως είναι μικρότερη του V. Παρακάτω παρατίθεται μια σχηματική απεικόνιση των διαστάσεων των πινάκων των επιπέδων που παίρνουν μέρος στην εκπαίδευση ενός Word2Vec Skip-Gram μοντέλου. Τις ίδιες διαστάσεις συναντάμε και σε ένα μοντέλο CBoW:



Σχήμα 2.20: Αρχιτεκτονική Πινάκων ενός Skip-Gram μοντέλου

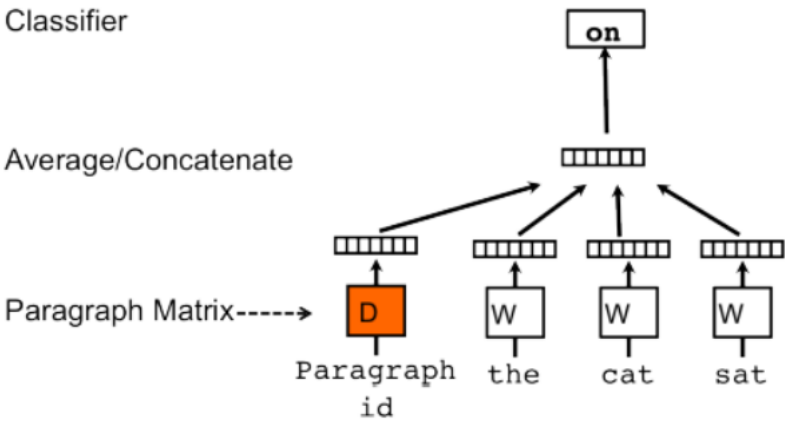
Και τα δύο μοντέλα έχουν προτερήματα και μειονεκτήματα. Το μοντέλο Skip-gram δουλεύει καλύτερα για μικρό όγκο δεδομένων και καταφέρνει να αναπαραστήσει σπάνιες λέξεις καλύτερα. Από την άλλη το μοντέλο CBoW αναπαριστά καλύτερα πιο συχνές λέξεις και είναι πιο γρήγορο.

2.7.2.3.2 Doc2Vec

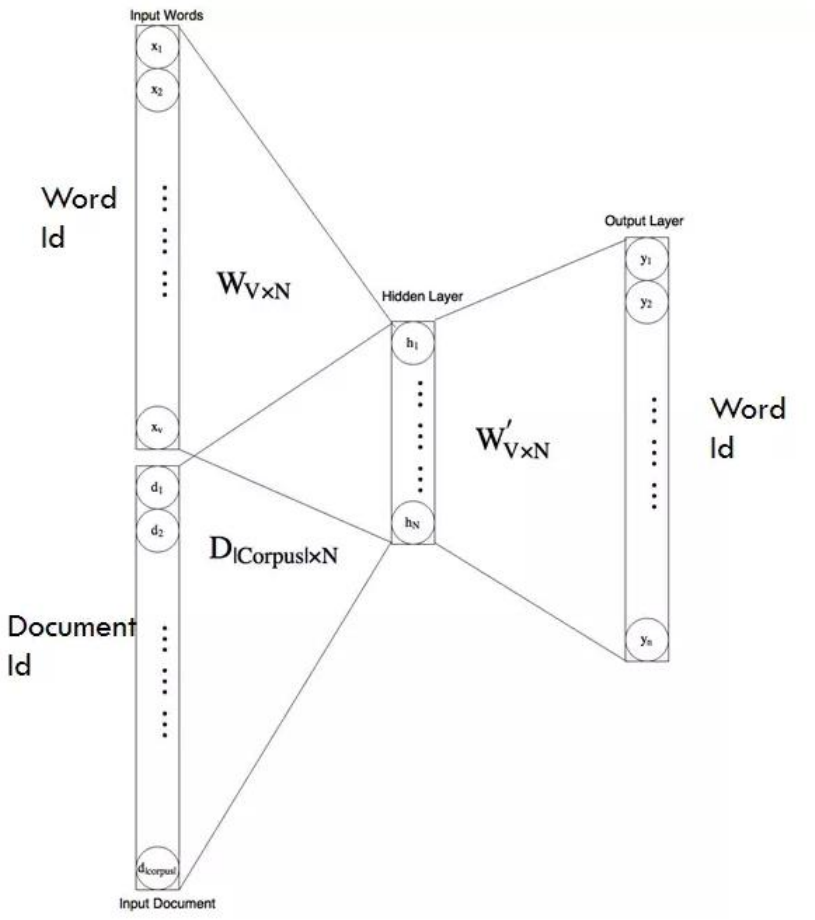
Οι Le και Mikolov στο [55] πρότειναν ένα νέο νευρωνικό μοντέλο για την εκπαίδευση διανυσματικών αναπαραστάσεων κειμένου που αποτελεί άμεση γενίκευση του μοντέλου Word2Vec. Είναι το μοντέλο Paragraph Vector (PV), ωστόσο συχνά το ίδιο μοντέλο αναφέρεται με το όνομα Doc2Vec, ονομασία που προδίδει τη στενή σχέση του με το μοντέλο Word2Vec αλλά και τη δυνατότητα του μοντέλου να παράγει αναπαραστάσεις για οποιασδήποτε μορφής έγγραφα, από φράσεις και προτάσεις μέχρι παραγράφους και ολόκληρα κείμενα.

Η βασική διαφορά του μοντέλου με το Word2Vec είναι ότι δεν υπολογίζει διανυσματικές αναπαραστάσεις μόνο για τις λέξεις αλλά και για ολόκληρα κομμάτια κειμένου, επιτρέποντας την σύγκριση προτάσεων, παραγράφων ακόμη και ολόκληρων κειμένων, ανεξαρτήτως μήκους τους, καθώς καταφέρνει να αναπαραστήσει την σημασιολογική ερμηνεία τους. Χρησιμοποιεί τις διανυσματικές αναπαραστάσεις κειμένου (document vectors) μαζί με τα word vectors για να κάνει προβλέψεις είτε λέξεων είτε συμφραζομένων τους και εκπαιδεύει και τα δύο με κριτήριο την ελαχιστοποίηση του κόστους πρόβλεψης. Όπως και στο Word2Vec, υπάρχουν δύο τεχνικές εκπαίδευσης διανυσματικών αναπαραστάσεων:

1. Ας υποθέσουμε ένα σώμα κειμένου που οργανώνεται σε M έγγραφα και περιέχει N διαφορετικές λέξεις. Το κάθε έγγραφο αναπαρίσταται από ένα μοναδικό διάνυσμα στήλη σε έναν πίνακα D και αντίστοιχα κάθε λέξη από ένα μοναδικό διάνυσμα στήλη σε έναν πίνακα W. Το μοντέλο αρχικοποιεί τυχαία διανυσματικές αναπαραστάσεις διάστασης V για κάθε ένα από τα M έγγραφα και κάθε μία από τις N λέξεις. Έπειτα διατρέχει το σώμα κειμένου με παράθυρα, όπως και το μοντέλο Word2Vec και σε κάθε παράθυρο χρησιμοποιεί τα διανύσματα των λέξεων του παραθύρου αλλά και το διάνυσμα του εγγράφου στο οποίο αναφέρεται το παράθυρο για να προβλέψει την λέξη που ακολουθεί μετά το παράθυρο. Έτσι τα διανύσματα των λέξεων μοιράζονται σε όλο το κείμενο και τα διανύσματα των εγγράφων μοιράζονται μεταξύ παραθύρων του ίδιου εγγράφου. Η προσέγγιση αυτή καλείται Distributed Memory (PV-DM) και δίνεται σχηματικά στο σχήμα 2.21, καθώς και στο 2.22 με αναγραφόμενες τις διαστάσεις των μεγεθών.

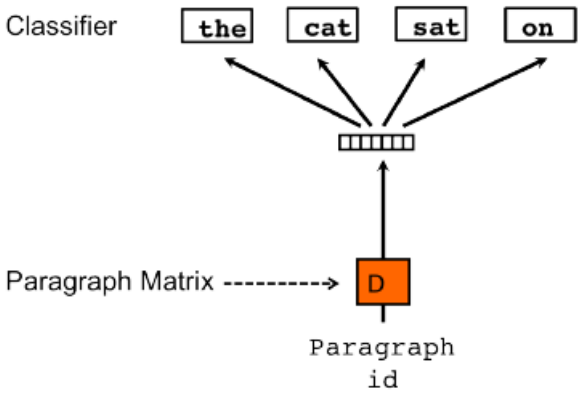


Σχήμα 2.21: Αναπαράσταση λειτουργίας PV-DM



Σχήμα 2.22: Αναπαράσταση διαστάσεων των διανυσματικών μεγεθών λειτουργίας PV-DM [56]

2. Μία εναλλακτική προσέγγιση καλείται Distributed Bag-of-Words (PV-DBOW) και σχηματικά αναπαρίσταται στο σχήμα 2.23. Σε κάθε παράθυρο το δίκτυο τροφοδοτείται από το διάνυσμα του εγγράφου, στο οποίο ανήκει το παράθυρο, και οι παράμετροι προσαρμόζονται ώστε το δίκτυο να μπορεί να προβλέπει επιτυχώς τυχαία επιλεγμένες λέξεις του παραθύρου.



Σχήμα 2.23: PV-DBOW.

Εδώ το διάνυσμα παραγράφου εκπαιδεύεται για να προβλέπει λέξεις σε ένα μικρό παράθυρο εγγράφου

Οι δύο διαφορετικές προσεγγίσεις εμπνέονται από τις παραλλαγές Skip-Gram και Continuous Bag-of-Words του μοντέλου Word2Vec. Σε κάθε περίπτωση το μοντέλο εκπαιδεύεται με τρόπο εντελώς αντίστοιχο του μοντέλου Word2Vec και παράγει word vectors και document vectors σταθερής διάστασης *V*. Με αυτό τον τρόπο τα έγγραφα απεικονίζονται στον ίδιο διανυσματικό χώρο με τις λέξεις και ιδανικά, ενσωματώνουν όλη τη σημασιολογική πληροφορία που προέρχεται από τις λέξεις που τα συνθέτουν αλλά και τη σειρά των λέξεων στα επιλεγμένα παράθυρα, τουλάχιστον στην τεχνική Distributed Memory. Ουσιαστικά ακολουθεί την λογική των n-gram μοντέλων με n αρκετά μεγάλο, ωστόσο υπερτερεί αφού τα n-gram μοντέλα δημιουργούν αναπαραστάσεις πολύ μεγάλων διαστάσεων οπότε καταναλώνουν μεγάλη υπολογιστική χωρητικότητα.

Έπειτα από πειράματα οι Le και Mikolov αναφέρουν η τεχνική PV-DM δίνει ικανοποιητικές εκτιμήσεις στα περισσότερα προβλήματα, αλλά ο συνδυασμός της με την τεχνική PV-DBOW είναι συνήθως πιο αποδοτικός.

Μετά την εκτίμηση των document vectors, αυτά μπορούν να χρησιμοποιηθούν σαν χαρακτηριστικά σε προβλήματα ταξινόμησης, καθώς και παλινδρόμησης. Οι Le και Mikolov αναφέρουν καλές επιδόσεις τόσο σε προβλήματα sentiment analysis, όσο και σε προβλήματα εξόρυξης πληροφορίας.

2.8 Αξιολόγηση Αλγορίθμων Επιβλεπόμενης Μάθησης

Η προετοιμασία των δεδομένων και εκπαίδευση είναι αντικείμενα υψίστης σημασίας για την ανάπτυξη ενός μοντέλου μηχανικής μάθησης, εξίσου σημαντική ομως είναι και η παρακολούθηση της επίδοσης του μοντέλου, δηλαδή το πόσο καλά μπορεί να γενικεύει το μοντέλο για άγνωστα δεδομένα. Οι μετρικές αξιολόγησης εξυπηρετούν αυτό το σκοπό, συγκεκριμένα μετρούν κάποιο μέγεθος του εκπαιδευμένου μοντέλου ως προς κάποιο χαρακτηριστικό. Χωρίς την χρήση μετρικών αξιολόγησης η βελτίωση της προβλεπτικής ικανότητας του μοντέλου ή η σύγκριση του με άλλα μοντέλα δεν θα ήταν εφικτή. Αξίζει να σημειωθεί ότι η χρήση των μετρικών αξιολόγησης δεν είναι καθολική αλλά η φύση του κάθε προβλήματος ενθαρρύνει την χρήση διαφορετικών μετρικών αξιολόγησης. Στη συνέχεια θα παρουσιάσουμε τις πιο συχνά χρησιμοποιούμενες μετρικές τόσο στα προβλήματα ταξινόμησης όσο και παλινδρόμησης.

2.8.1 Μετρικές Αξιολόγησης

2.8.1.1 Μετρικές μοντέλων κατηγοριοποίησης

Οι μετρικές των μοντέλων κατηγοριοποίησης παρουσιάζουν μεγάλη ποικιλία. Αρχικά όμως θα παρουσιάσουμε τις κλάσεις στις οποίες μπορούν να ανήκουν οι προβλέψεις ενός μοντέλου βάση των οποίων προκύπτουν και οι μετρικές.

- *True Positive (TP):* Το σύνολο της εξόδου για το οποίο η πρόβλεψη είναι σωστή και η προβλεπόμενη κλάση θετική.
- *True Negative (TN):* Το σύνολο της εξόδου για το οποίο η πρόβλεψη είναι σωστή και η προβλεπόμενη κλάση αρνητική.
- *False Positive (FP):* Το σύνολο της εξόδου για το οποίο η πρόβλεψη είναι λανθασμένη και η προβλεπόμενη κλάση θετική.
- *False Negative (FN):* Το σύνολο της εξόδου για το οποίο η πρόβλεψη είναι λανθασμένη και η προβλεπόμενη κλάση αρνητική.

		Προβλεπόμενη κλάση	
		Θετική (P)	Αρνητική (N)
Πραγματική κλάση	Θετική (P)	TP	FN
	Αρνητική (N)	FP	TN

Στη συνέχεια παρουσιάζουμε τις πιο σημαντικές μετρικές αξιολόγησης τις οποίες θα αξιοποιήσουμε και στην μελέτη μας.

1. *Accuracy*: Η μετρική Accuracy ή Ακρίβεια περιγράφει τον λόγο των σωστά ταξινομημένων δειγμάτων προς το σύνολο όλων των δειγμάτων. Ειναι η πιο σημαντική μετρική και δινει μια άμεση και απλή αξιολόγηση του μοντέλου. Η χρήση της συνίσταται για δεδομένα που είναι καλά ισορροπημένα:

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

2. *Precision*: Η μετρική αυτή εκφράζει τον λόγο των σωστά ταξινομημένων θετικών προβλέψεων προς το σύνολο των προβλέψεων που έχουν ταξινομηθεί ως θετικές. Χρησιμοποιείται σε προβλήματα που η εγκυρότητα της πρόβλεψης είναι μεγάλης σημασίας.

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

3. *Recall*: Εκφράζει το λόγο των σωστά ταξινομημένων θετικών προβλέψεων προς το σύνολο των θετικών προβλέψεων. Η μετρική Recall χρησιμοποιείται σε προβλήματα που έχουν ως σκοπό την μεγιστοποίηση των θετικών προβλέψεων.

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

4. *F1 Score*: Η μετρική αξιολόγησης F1 Score εκφράζει τον αρμονικό μέσο όρο των μετρικών Precision και Recall. Η χρήση της μετρικής F1 Score γίνεται όταν το πρόβλημα απαιτεί καλό Precision και Recall.

$$F1\ Score = 2 \cdot \frac{Precision\ \cdot Recall}{Precision + Recall}$$

5. *Crossentropy*: Η μετρική Crossentropy ή Log Loss λαμβάνει υπόψιν την αβεβαιότητα της πρόβλεψης βασιζόμενη στο πόσο διαφέρει από την πραγματική τιμή. Χρησιμοποιείται σε προβλήματα δυαδικής ταξινόμησης.

$$Crossentropy = -(ylog(p) + (1 - y)\log(1 - p))$$

όπου p είναι η πιθανότητα η πρόβλεψη να είναι 1 και y είναι η πρόβλεψη του μοντέλου. Η μετρική αυτή χρησιμοποιείται όταν η έξοδος του μοντέλου είναι πιθανοτικές προβλέψεις.

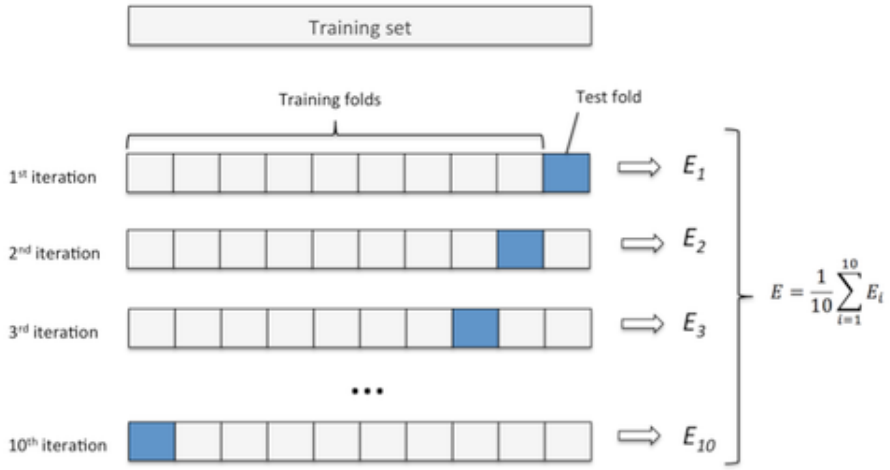
2.8.1.2 Μετρικές μοντέλων παλινδρόμησης

Για τα μοντέλα παλινδρόμησης δύο είναι οι κύριες μετρικές, τις οποίες και μελετήσαμε στην ενότητα 2.5.3.2, το μέσο τετραγωνικό σφάλμα (RMSE) και το μέσο απόλυτο σφάλμα (MAE).

2.8.2 Μέθοδοι Αξιολόγησης

Για την πραγματοποίηση πειραμάτων με ένα μοντέλο επιβλεπόμενης μάθησης υπάρχουν διαφορετικές μέθοδοι σχετικά με τον τρόπο που γίνεται η εκπαίδευση και η αξιολόγηση του συστήματος. Μία από τις πιο διαδεδομένες μεθόδους είναι ο k-fold cross-validation τον οποίο και θα αξιοποιήσουμε σε κάποια μοντέλα. Κατά τη διαδικασία του k-fold cross-validation τα δεδομένα χωρίζονται σε k υποσύνολα. Σε κάθε έναν από τους k γύρους της μεθόδου αυτής, τα k-1 υποσύνολα χρησιμοποιούνται ως training set για τον αλγόριθμο και το υποσύνολο που απομένει χρησιμοποιείται ως test set με βάση το οποίο εξάγονται οι επιθυμητές μετρικές. Συνηθισμένες τιμές είναι το 5 και το 10 (5-fold και 10-fold CV). Στο τέλος όλης της διαδικασίας υπολογίζεται ο μέσος όρος για κάθε μετρική και οι προκύπτοντες αριθμοί αποτελούν τον δείκτη απόδοσης του εκάστοτε αλγορίθμου μηχανικής μάθησης.

Τα πλεονεκτήματα της μεθόδου k-fold cross-validation είναι ότι υπάρχει μικρή μεροληψία (bias) στη διαδικασία μάθησης του αλγορίθμου καθώς ένα μεγάλο ποσοστό των δεδομένων χρησιμοποιούνται για την εκπαίδευση του αλγορίθμου και έτσι εκμεταλλεύεται πλήρως όλη η διαθέσιμη πληροφορία που σε άλλες περιπτώσεις θα παρέμενε ανεκμετάλλευτη καθώς θα αξιοποιούταν ένα μεγάλο κομμάτι της ως δεδομένα επαλήθευσης. Επιπλέον μειώνεται η διακύμανση (variance) καθώς όλα τα δεδομένα περνούν από το test set. Με άλλα λόγια, το k-fold cross-validation προσφέρει αντικειμενικά αποτελέσματα για την επίδοση ενός μοντέλου αποφεύγοντας την υπερεκπαίδευση (overfitting) του αλγορίθμου στα δεδομένα μέσω της χρήσης των διαφορετικών folds, και προσφέρει έναν τρόπο εξαγωγής συμπερασμάτων για το πόσο καλά γενικεύει αυτό το μοντέλο και πόσο μπορεί δυνητικά να αποδώσει καλά σε καινούρια δεδομένα.



Σχήμα 2.24: k-fold cross validation

2.8.3 Αναζήτηση Πλέγματος

Κατά την διεξαγωγή πειραμάτων με έναν αλγόριθμο επιβλεπόμενης μάθησης, χρειάζεται πολύ συχνά η εύρεση των υπερπαραμέτρων του αλγορίθμου που οδηγούν στην βέλτιστη απόδοση του. Οι υπερπαράμετροι ορίζονται πριν τη διαδικασία μάθησης και καθορίζουν ιδιότητες του μοντέλου όπως η πολυπλοκότητα του, η μορφή της επιφάνειας απόφασης κ.α. Μια δημοφιλής τεχνική που χρησιμοποιείται για αυτό το σκοπό είναι η

αναζήτηση πλέγματος (grid search) [57]. Κατά την τεχνική αυτή πραγματοποιείται εξαντλητικός έλεγχος σε συνδυασμούς υπερπαραμέτρων με προκαθορισμένο εύρος τιμών προκειμένου να βρεθεί ο συνδυασμός που έχει την καλύτερη απόδοση.

Κεφάλαιο 3

3. Επεξεργασία Δεδομένων

Όπως προαναφέραμε, τα πειράματά μας εκτελούνται πάνω στο αποθετήριο δεδομένων Distress Analysis Interview Corpus-Wizard of Oz (DAIC-WOZ) [1]. Το αποθετήριο αυτό αποτελείται από κλινικές συνεντεύξεις ανθρώπων σχεδιασμένες να υποστηρίζουν την διάγνωση ψυχικών διαταραχών, όπως η κατάθλιψη στην περίπτωση μας, που έχουν μαγνητοσκοπηθεί κάτω από όσο το δυνατόν πιο όμοιες συνθήκες περιβάλλοντος, προσπαθώντας να εξαλείψουν τον θόρυβο που θα κληθούμε να αντιμετωπίσουμε στην προεπεξεργασία των δεδομένων μας.

Εξαρχής τα δεδομένα του dataset μας ήταν κατανεμημένα σε δεδομένα εκπαίδευσης (train set), επαλήθευσης (dev set) και εξέτασης (test set). Έτσι λοιπόν το κάθε σύνολο δεδομένων αποτελείται από τα δείγματα που του έχουν ανατεθεί, το φύλο τους και τις ακόλουθες τιμές που παίζουν τον ρόλο των ταμπελών (labels) για την επίβλεψη του μοντέλου που είναι οι τιμές των PHQ8 scores, τόσο των επιμέρους, όσο και του τελικού, καθώς και η δυαδική τιμή του τελικού PHQ8 score που καθορίζει την ταξινόμηση σε κατάθλιψη ή μη. Στην εικόνα 3.1 παραθέτουμε μια ένδειξη του πως διατάσσονται τα datasets, ενώ στον πίνακα 3.1 παραθέτουμε την κατανομή των δειγμάτων στα διάφορα datasets.

Participan	PHQ8_Bin	PHQ8_Scc	Gender	PHQ8_No	PHQ8_De	PHQ8_Sle	PHQ8_Tire	PHQ8_Ap	PHQ8_Fai	PHQ8_Cor	PHQ8_Moving
303	0	0	0	0	0	0	0	0	0	0	0
304	0	6	0	0	1	1	2	2	0	0	0
305	0	7	1	0	1	1	2	2	1	0	0
310	0	4	1	1	1	0	0	0	1	1	0
312	0	2	1	0	0	1	1	0	0	0	0
313	0	7	1	1	1	1	1	1	1	1	0
315	0	2	1	0	0	0	1	1	0	0	0

Σχήμα 3.1: Δείγμα διοργάνωσης των datasets

Gender	Train Set	Dev Set	Test Set
Male	63	35	47
Female	44	19	24
All	107	35	47

Πίνακας 3.1: Κατανομή ανδρών/γυναικών στα datasets

Το κάθε δείγμα συνέντευξης λοιπόν αποτελείται από τις καταγραφές ήχου και βιντεο, αλλά και από τις απομαγνητοφωνημένες ερωτοαπαντήσεις τα οποία προέρχονται απο τις συνεδρίες ψυχανάλυσης του κάθε ασθενή με έναν εικονικό πράκτορα, την Έλλη. Ωστόσο βάση της νομοθεσίας της ΕΕ για τη συλλογή και χρήση δεδομένων προσωπικού χαρακτήρα από οργανισμούς GDPR, ήταν αδύνατον να διαμοιραστούν τα πρωτότυπα

βιντεοσκοπημένα αρχεία. Έτσι λοιπόν για την δημιουργία της παραπάνω βάσης δεδομένων, συλλέχθηκαν κάποια περιγραφικά δεδομένα ήχου και εικόνας καθώς και τα απομαγνητοφωνημένα κείμενα συνεντεύξεων, των οποίων την δομή θα αναλύσουμε στην συνέχεια. Τα χαρακτηριστικά αυτά είναι οι λεγόμενοι περιγραφητές, που κάνουν εφικτή τη μοντελοποίηση κάθε είδους δεδομένων παραγωγής πληροφορίας.

Επιπλέον για να δώσουν στους χρήστες τη δυνατότητα να εξάγουν οι ίδιοι οποιοδήποτε μη λεκτικό χαρακτηριστικό ομιλίας ή εικόνας που δεν δίνεται στη συγκεκριμένη βάση δεδομένων, παραχωρούν και την πρωτότυπη ηχογράφηση από τις συνεντεύξεις, καθώς και τις εικόνες σε συντεταγμένες 2D και 3D pixels από την δειγματοληψία των βίντεο σαν πρωτότυπα αρχεία.

3.1 Προεπεξεργασία Συνολικού Dataset

3.1.1 Μέθοδοι Προεπεξεργασίας

Μια διαδικασία ML αποτελείται από μια σειρά μετασχηματιστών πάνω στα χαρακτηριστικά των δεδομένων εισόδου που τελειώνει σε έναν εκτιμητή. Οι μετασχηματιστές χρησιμοποιούνται για να κανουν την προεπεξεργασία (μέσω μετασχηματισμού) των δεδομένων. Τα βήματα προεπεξεργασίας στοχεύουν αρχικά στην αφαίρεση η αντικατάσταση απουσιάζουσων τιμών από το dataset, έπειτα στη μετατροπή των κατηγορικών μεταβλητών κατάλληλα ώστε να μπορούν να τους διαχειριστούν αλγόριθμοι μηχανικής μάθησης και τέλος στην επιλογή ή εξαγωγή των κατάλληλων χαρακτηριστικών για το μοντέλο μας. Σκοπός είναι η κανονικοποίηση της εισόδου για τη μείωση της αναπαράστασης με ταυτόχρονη απόρριψη της πλεονάζουσας-άχρηστης πληροφορίας που καλείται θόρυβος. Έτσι, μειώνεται η διάσταση του προβλήματος και ο κίνδυνος του overfitting και ταυτόχρονα μειώνεται και ο απαιτούμενος χρόνος εκπαίδευσης του μοντέλου. Οι πιο γνωστοί μετασχηματιστές τους οποίους και θα μελετήσουμε τόσο ατομικά όσο και συνδυαστικά μεταξύ τους είναι οι εξής:

- *Επιλογή χαρακτηριστικών Variance Threshold*: Μια πολύ σημαντική παράμετρος για την απόδοση των εκτιμητών είναι η διάσταστικότητα των δεδομένων, ιδιαίτερα σε σχέση με τον διαθέσιμο αριθμό δειγμάτων. Γενικά και ανεξάρτητα από το μοντέλο του εκτιμητή, η απόδοση αυξάνεται όσο αυξάνεται το πλήθος και η ποιότητα των δεδομένων και όσο μειώνεται η διαστατικότητα. Αντίστροφα, τα προβλήματα δυσκολεύουν όσο η διαστατικότητα αυξάνεται και τα δείγματα δεν επαρκούν για να καλύψουν όλες τις κατηγορίες του προβλήματος. Αναφερόμαστε στο πρόβλημα αυτό ως την κατάρα της διαστατικότητας (the curse of dimensionality): όσο αυξάνει η διαστατικότητα, τόσο τα διαθέσιμα δεδομένα γίνονται αραιά (sparse). Για να μειωθεί η διαστατικότητα των δεδομένων χρησιμοποιούμε τεχνικές μείωσης διαστατικότητας (dimensionality reduction). Μια απλή τεχνική επιλογής χαρακτηριστικών είναι το ελάχιστο κατώφλι της διακύμανσης (Variance threshold). Σε γενικές γραμμές αν η διακύμανση ενός χαρακτηριστικού εισόδου είναι πολύ χαμηλή, δεν μπορεί να προσφέρει σημαντικά στη διαχωριστική ικανότητα του εκτιμητή. Ειδικά στην περίπτωση που η διακύμανση είναι 0, δηλαδή το χαρακτηριστικό εχει σταθερή τιμή για όλα τα δείγματα εκπαίδευσης, δεν χρησιμεύει καθόλου στον εκτιμητή για να αποφασίσει αν ένα δείγμα ανήκει σε μία κλάση ή σε μια άλλη ή να προβλέψει κάποια τιμή και επιπλέον μπορεί να δυσκολέψει άλλες διαδικασίες της προεπεξεργασίας όπως η κανονικοποίηση των χαρακτηριστικών.

- *Οι δύο μετασχηματιστές κανονικοποίησης (ο scaler και ο min_max_scaler):* Χαρακτηριστικά με πολύ μεγάλες διαφορές στις απόλυτες τιμές τους μπορούν να προκαλέσουν προβλήματα στην εκπαίδευση και να δώσουν εκτιμητές με μη βέλτιστη απόδοση. Η κανονικοποίηση μετασχηματίζει τις τιμές των χαρακτηριστικών ώστε να αμβλυνθούν αυτές οι διαφορές. Η κανονικοποίηση των χαρακτηριστικών μπορεί να γίνει με 2 βασικούς τρόπους, γνωστούς και από τη στατιστική. Με την διαίρεση με τη διαφορά μεγίστου-ελαχίστου (feature scaling) οπότε οι τιμές όλων των χαρακτηριστικών κλιμακώνονται γραμμικά στο διάστημα [0,1] ή με το z-score (ή standard score) του κάθε χαρακτηριστικού (standardization), που κάνει το χαρακτηριστικό να έχει μέση τιμή μηδέν και διακύμανση μονάδα, σαν την κανονική κατανομή. Η μετατροπή μεγίστου ελαχίστου γίνεται με τον τύπο:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Η μετατροπή σε standard score γίνεται με τον τύπο:

$$z = \frac{X - \mu}{\sigma}$$

Η μετατροπή σε Standard score είναι πιο ανθεκτική από την min max σε τιμές outliers δηλαδή σποραδικές τιμές που είναι πολύ μακριά απο τη μέση τιμή και τις υπόλοιπες τιμές του χαρακτηριστικού (η min max θα συμπιέσει τις περισσότερες τιμές σε ένα μικρό διάστημα). Από την άλλη η κλιμάκωση σε [0,1] είναι λιγότερο ευαίσθητη σε πολυ μικρές αποκλίσεις και επίσης σε αραιά (sparse) διανύσματα χαρακτηριστικών (δηλαδή με πολλές μηδενικές τιμές) η εφαρμογή της διατηρεί τα μηδέν, κάτι που μπορεί να είναι καθοριστικό για την ταχύτητα εκπαίδευσης.

- *Ο εξισορροπητής με τυχαία υπερδειγματοληψία RandomOverSampler :* Με τον ορό μη ισορροπημένο dataset εννοούμε ένα dataset στο οποίο τα πλήθη των δειγμάτων της κάθε κλάσης διαφέρουν σημαντικά μεταξύ τους. Χωρίς να υπάρχει κάποια συνολική απάντηση, όταν ο λόγος μεταξύ του αριθμού των δειγμάτων δύο κλάσεων αρχίζει να είναι μεγαλύτερος από 2:3, μπορούμε να αρχίζουμε να θεωρούμε το dataset μη ισορροπημένο (imbalanced). Στα πραγματικά datasets αυτό είναι κάτι πολύ κοινό. Οι περισσότεροι ταξινομητές ωστόσο εκπαιδεύονται καλύτερα όταν τα δείγματα όλων των κλάσεων είναι σχετικά ισάριθμα. Έχουμε δύο βασικούς τρόπους βασικούς τρόπους για να εξισορροπούμε ένα dataset, την υποδειγματοληψία (undersampling) και την υπερδειγματοληψία (oversampling). Εν ολίγοις, στο undersampling απλά αφαιρούμε τυχαία δείγματα από όλες τις κατηγορίες που έχουν μεγαλύτερο πλήθος από τη μικρότερη, ενώ στο oversampling επιλέγουμε τυχαία ορισμένα παραδείγματα από τις λιγότερο συχνές κατηγορίες και τα επαναλαμβάνουμε. Στην πρώτη δηλαδή αφαιρούμε δεδομένα ενώ στην άλλη προσθέτουμε. Γενικά το oversampling ενδείκνυται περισσότερο, αφού δεν χάνουμε δεδομένα εκπαίδευσης.

- *Η εξαγωγή χαρακτηριστικών PCA*: Όπως είπαμε και προηγουμένως για να μειώσουμε τις διαστάσεις των μεταβλητών μας μπορούμε να κάνουμε δύο πράγματα: να αφαιρέσουμε κατηγορίες που δεν προσφέρουν σημαντική πληροφορία, δηλαδή να κάνουμε επιλογή μεταβλητών (feature selection)**.** Εναλλακτικά, μπορούμε να κάνουμε εξαγωγή νέων χαρακτηριστικών σε ένα χώρο μικρότερων διαστάσεων (feature extraction). Η βασικότερη τεχνική feature extraction είναι η ανάλυση σε κύριες συνιστώσες (principal components analysis - PCA) όπου αναλύουμε τα δεδομένα σε κύριες συνιστώσες και δουλεύουμε με τελείως νέες, γραμμικά ασυσχέτιστες μεταβλητές μικρότερης διαστατικότητας.

Οι μετασχηματιστές έχουν δύο βασικές μεθόδους, την fit και την transform. Με την fit μαθαίνουν κάποιες παραμέτρους (πχ τη μέση τιμή) με βάση τα δεδομένα train και με την transform μπορούν να μετασχηματίσουν τα δεδομένα (train ή test) με βάσει τις παραμέτρους που έχουν μάθει.

3.1.2 Ανασκόπηση του Dataset

Όσον αφορά το αποθετήριο δεδομένων DAIC-WOZ της έρευνας αυτής που περιλαμβάνει τα αρχεία csv[2] με την κατανομή των δειγμάτων και τις ετικέτες τους, μελετήσαμε τα δεδομένα αυτά για τυχόν αδυναμίες. Εφόσον τα κατευθυντήρια datasets περιλαμβάνουν μόνο τις ετικέτες των δειγμάτων, ελέγξαμε το κάθε dataset για τυχόν απουσιάζουσες τιμές, γεγονός σύνηθες καθώς τα δεδομένα αυτά προκύπτουν από μετρήσεις ή αντικείμενα του πραγματικού κόσμου. Όπως προέκυψε το train set είχε μια απουσιάζουσα τιμή σε ένα δείγμα, που αντιπροσώπευε την τιμή του PHQ8 Sleep Score και συγκρίνοντας το άθροισμα των υπολοίπων εφτά PHQ8 score με το τελικό, παρατηρήσαμε πως η απουσιάζουσα τιμή αναφερόταν στο 0, την οποία και αντικαταστήσαμε. Στα dev set και test set δεν εντοπίστηκε κάποια απουσιάζουσα τιμή.

```
Number of NaN values in train labels:  1
Number of NaN values in new train labels :  0
Number of NaN values in Dev labels :  0
```

Στη συνέχεια ελέγχουμε τα πλήθη των δειγμάτων της κάθε κλάσης (Depressed/ Not Depressed) και παρατηρούμε πως διαφέρουν σημαντικά μεταξύ τους με την κλάση των πασχόντων από κατάθλιψη να αποτελείται από 30 δείγματα, ενώ των μη πασχόντων από 77 δείγματα. Επομένως εφαρμόζουμε την μέθοδο της υπερδειγματοληψίας (oversampling) για να αποτρέψουμε τα μοντέλα ML που θα δημιουργήσουμε να συγκλίνουν σε τιμές των δειγμάτων που ανήκουν στην κυρίαρχη κλάση.

```
Resample train set...
class frequencies:  [77 30]
total samples:  107
class percentage:  [71.96261682 28.03738318]
Train_set is imbalanced!!!!!!
New train set shape:  (154, 12)
class frequencies:  [77 77]
total samples:  154
class percentage:  [50. 50.]
Train_set is balanced!!!!!!
(154, 12)
```

Τελικά, το πλήθος των δειγμάτων μας αυξάνεται φτάνοντας στο σύνολο των 154 δειγμάτων, γεγονός που προσδίδει μεγαλύτερες πιθανότητες γενίκευσης του μοντέλου μας, καθώς εξαρχής το πλήθος των διαφορετικών δειγμάτων είναι μικρό στα όρια του ανεπαρκούς.

[2] https://en.wikipedia.org/wiki/Comma-separated_values

3.2 Επεξεργασία Οπτικοακουστικών Δεδομένων

3.2.1 Χαρακτηριστικά Φωνής

3.2.1.1 Μελέτη Δεδομένων

Στην ενότητα αυτή θα παρουσιάσουμε τις διάφορες κατηγορίες ακουστικών χαρακτηριστικών που έχουν εξαχθεί από τα ηχογραφημένα στιγμιότυπα των συνεντεύξεων, με σκοπό την αναγνώριση των συναισθημάτων τους. Το μέσο εξαγωγής των ακουστικών χαρακτηριστικών είναι το COVAREP toolbox, ένας συνδυασμός περιβαλλόντων Matlab και Octave ανοιχτού κώδικα σχεδιασμένο για την ανάλυση του ήχου [27]. Επιλέχθηκε λοιπόν αυτό το εργαλείο καθώς παρουσιάζει ακρίβεια και ποιότητα στην επεξεργασία της φωνής καθώς και στα προσωδιακά χαρακτηριστικά του ομιλητή. Σκοπός είναι η ανίχνευση όλων εκείνων των στοιχείων που επηρεάζονται από την συναισθηματική έκφραση. Σύμφωνα με τα [28],[29] έχει αποδειχθεί πως οι μέθοδοι εξαγωγής των ακουστικών χαρακτηριστικών που εφαρμόζει το εργαλείο αυτό είναι στενά συσχετισμένοι με την αναγνώριση της κατάθλιψης και των ψυχικών διαταραχών ευρύτερα.

Θεωρώντας ότι κάθε δεδομένο αντιστοιχεί σε μια ηχητική εκφώνηση ενός ομιλητή, η εξαγωγή χαρακτηριστικών μπορεί να γίνει σε επίπεδο πλαισίου, ομάδας πλαισίου ή και εκφώνησης. Έτσι λοιπόν τα διανύσματα χαρακτηριστικών διακρίνονται σε δύο κατηγορίες: short-time και long-time. Η πρώτη κατηγορία αντιστοιχεί στα εξαγόμενα χαρακτηριστικά ανά πλαίσιο, όπου η εκφώνηση έχει χωριστεί σε πλαίσια ίσης διάρκειας (συνήθως 10-50 msec) με χρήση τεχνικών παραθύρωσης. Στη δεύτερη κατηγορία, αντίθετα, ανήκουν τα χαρακτηριστικά εκείνα που εξάγονται από σήμα μεγαλύτερης διαρκείας, ακόμα και από ολόκληρη την εκφώνηση.

Μια άλλη πιθανή κατηγοριοποίηση είναι ο διαχωρισμός χαμηλού επιπέδου περιγραφητών (low-level descriptors ή LLDs) και συναρτησιακών (functionals). Οι πρώτοι περιγραφητές περιλαμβάνουν χαρακτηριστικά φασματικά, προσωδιακά και ποιότητας φωνής, καθώς και τις παραγώγους τους. Με βάση αυτούς τους χαμηλού-επιπέδου περιγραφητές, είναι δυνατός ο υπολογισμός στατιστικών, με σκοπό την παραγωγή χαρακτηριστικών ανά ομάδα πλαισίων ή ανά εκφώνηση. Τα στατιστικά αυτά αντιστοιχούν στα λεγόμενα συναρτησιακά (functionals).

	LLDs
Προσωδιακά	Fundamental Frequency F0, Voicing VUV
Φασματικά	Mel cepstral coefficients (MCEP0-24), harmonic model and phase distortion mean (HMPDM0-24) and deviations (HMPDD0-12), Formant Frequencies (0-4).
Ποιότητα Φωνής	Normalized amplitude quotient (NAQ), quasi open quotient (QOQ), the difference in amplitude of the first two harmonics of the differentiated glottal source spectrum (H1H2), parabolic spectral parameter (PSP), maxima dispersion quotient (MDQ), spectral tilt/slope of wavelet responses (peak-slope), and shape parameter of the Liljencrants-Fant model of the glottal pulse dynamics (Rd)

Πίνακας 3.2: Ακουστικά χαρακτηριστικά χαμηλού επιπέδου (LLDs)

Παρακάτω λοιπόν δίνουμε μια σύντομη περιγραφή των ακουστικών χαρακτηριστικών που έχουμε στη διάθεση μας για την Αξιολόγηση της Κατάθλιψης, αφού καταγράψαμε ονομαστικά τους χαμηλού-επιπέδου περιγραφητές στον Πίνακα 3.2. Όλα τα ακουστικά χαρακτηριστικά έχουν δειγματοληφθεί στη συχνότητα των 100Hz, δηλαδή σε παράθυρα των 10ms.

Προσωδιακά Χαρακτηριστικά

Με τον όρο 'χαρακτηριστικά προσωδίας' εννοούμε τα χαρακτηριστικά εκείνα που σχετίζονται με την τονικότητα και την ένταση της φωνής. Τα κυριότερα είναι η θεμελιώδης συχνότητα (pitch ή F0) και η ενέργεια της φωνής. Τα χαρακτηριστικά προσωδίας έχουν χρησιμοποιηθεί ευρέως στην έρευνα της αναγνώρισης συναισθήματος μέσω φωνής με επιτυχία. [30] [31] [32]

1. Θεμελιώδης Συχνότητα(F0/Pitch):

Με τον όρο F0 συμβολίζουμε μια φυσική ιδιότητα του ήχου, την θεμελιώδη του συχνότητα που στην περίπτωση της ομιλίας είναι ο αριθμός των παλμών της γλώσσας ανά δευτερόλεπτο (συχνότητα των κυμάτων που παράγουν τον ήχο) και μετριέται σε Hz. Αντίστοιχα με τον όρο Pitch αναφερόμαστε στην θεμελιώδη συχνότητα του ήχου που γίνεται αντιληπτή από το ανθρώπινο αυτί και μετριέται σε Mel. Η κλίμακα Mel (από τη λέξη melody) δίνει την υποκειμενική αντίληψη των συχνοτήτων από τον άνθρωπο, έτσι ώστε να είναι ισοκατανεμημένες σύμφωνα με την ακοή του. Επειδή όμως η διακριτική ικανότητα του ανθρώπου είναι μεγαλύτερη στις χαμηλές συχνότητες από τις υψηλές, η αντιστοιχία της κλίμακας Mel με τις συχνότητες σε Hz προκύπτει λογαριθμική.[36] Το Pitch συνήθως συνδέεται με την απόσταση των συχνοτήτων των αρμονικών η οποία είναι ίση με την ελάχιστη-βασική συχνότητα F0 του σήματος. Έχει καθιερωθεί λοιπόν να μετράμε σαν pitch την θεμελιώδη συχνότητα ενός σήματος, στην περίπτωση που αυτό έχει περιοδική ή περίπου περιοδική μορφή. Όσον αφορά στη συμπεριφορά τη θεμελιώδους συχνότητας στα διάφορα συναισθήματα, οι περισσότερες έρευνες έχουν καταλήξει στο ότι το F0 έχει μεγαλύτερη μέση τιμή και εύρος τιμών στη χαρά και στο θυμό, ενώ μικρότερες τιμές στη λύπη και την αποστροφή. Στον θυμό έχουν βρεθεί απότομες διακυμάνσεις του F0, ενώ στη χαρά έχει πιο ομαλή πορεία. Φθίνουσα πορεία των τιμών του παρατηρείται στα συναισθήματα της λύπης και της αποστροφής. [33]

2. VUV:

Ο όρος αυτός είναι ένα δυαδικό μέγεθος που παίρνει είτε την τιμή 1 είτε την τιμή 0, αν στο παράθυρο ήχου στο οποίο αναφέρεται υπάρχει ομιλία από άνθρωπο ή όχι αντίστοιχα. Αν δηλαδή VUV=0, τότε οι φωνητικές χορδές του ομιλητή δεν πάλλονται και επομένως τα μεγέθη που αναφέρονται στην ποιότητα της φωνής και πιθανώς έχουν λάβει κάποιες μετρήσεις από εξωτερικούς παράγοντας πρέπει να μηδενιστούν χειροκίνητα.

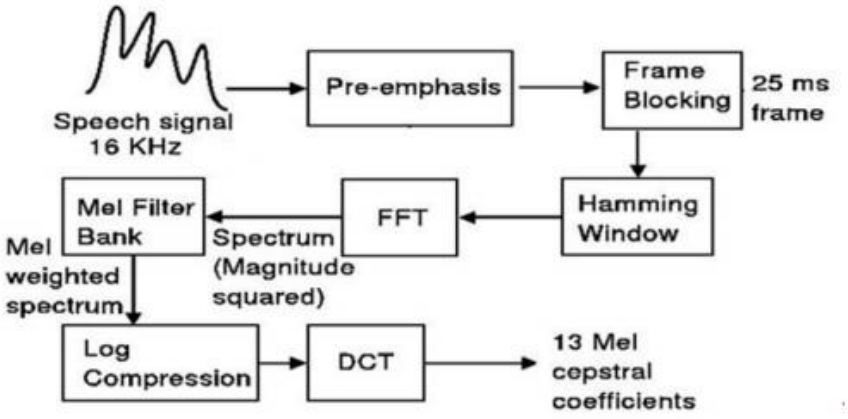
Φασματικά Χαρακτηριστικά

Τα χαρακτηριστικά φάσματος προκύπτουν έπειτα από επεξεργασία του σήματος στο πεδίο των συχνοτήτων και αποτελούν την πιο συνηθισμένη κατηγορία χαρακτηριστικών. Στην επεξεργασία φωνής, συχνά εξάγονται χαρακτηριστικά φάσματος τα οποία σχετίζονται με τον σχηματισμό της φωνητικής οδού κατά την διάρκεια της ομιλίας. Στην έρευνα αναγνώρισης συναισθήματος έχουν χρησιμοποιηθεί ευρέως σαν χαρακτηριστικά οι

συχνότητες Formants [34] [35], οι συντελεστές LPC οι οποίοι περιέχουν πληροφορία για την περιβάλλουσα του φάσματος της φωνής και τέλος, οι συντελεστές Mel frequency cepstral coefficients (MFCC). Στα δικά μας πειράματα θα ασχοληθούμε με την πρώτη και την τελευταία κατηγορία, οι οποίες και έχουν ευρύτερα τις πιο επιτυχημένες εφαρμογές, καθώς επίσης και με μια άλλη κατηγορία φασματικών και ενεργειακών χαρακτηριστικών, τους μέσους παραμόρφωσης φάσης (HMPDM0-24) και τις αποκλίσεις τους (HMPDD0-12), των οποίων ο ρόλος δεν είναι ιδιαίτερα καθοριστικός.

1. Mel cepstral coefficients (MCEP):

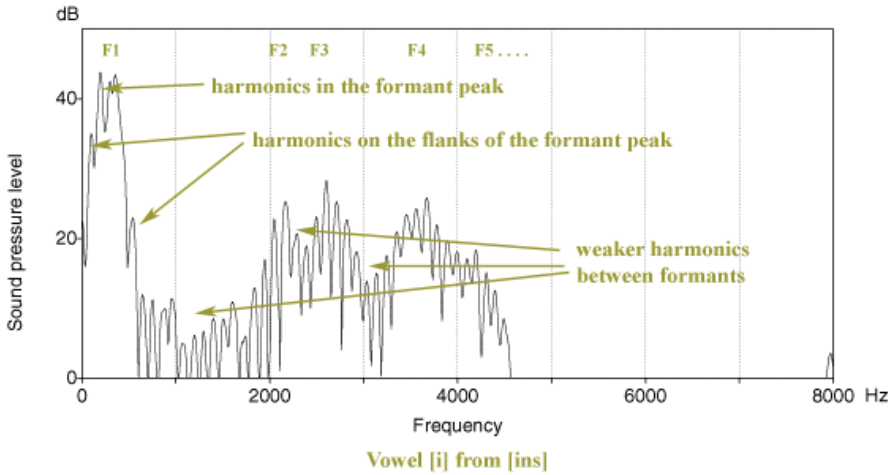
Οι συντελεστές MCEP αποτελούν το περισσότερο διαδεδομένο σύνολο χαρακτηριστικών σε εφαρμογές αναγνώρισης φωνής. Αποτελούν μια συμπαγή αναπαράσταση του φάσματος του σήματος φωνής και περιέχουν μεικτή πληροφορία καθώς σχετίζονται τόσο με τον ομιλητή και την κατάσταση στην οποία βρίσκεται, όσο και με τα λεγόμενά του. Βασικό στοιχείο τους είναι ότι παρέχουν μια καλύτερη αναπαράσταση του σήματος, αξιοποιώντας ιδιότητες της ακοής του ανθρώπου. Στο σχήμα 3.2 δίνουμε την διαδικασία εξαγωγής τους από το σήμα ήχου. Οι συντελεστές MCEP είναι στην ουσία μια ομάδα συντελεστών cepstrum που εξάγονται μετά από ανάλυση του σήματος με μια ειδικά σχεδιασμένη συστοιχία φίλτρων (Mel Filter Bank) [36].



Σχήμα 3.2: Διαδικασία εξαγωγής χαρακτηριστικών MCEP

2. Formant Frequency:

Στο χώρο των συχνοτήτων, οι συχνότητες φωνοσυντονισμού (formants) αντιστοιχούν στους συντονισμούς της φωνητικής οδού και σχετίζονται με τη μορφή και τις διαστάσεις της [37]. Κάθε μορφή της φωνητικής οδού χαρακτηρίζεται από ένα σύνολο τέτοιων συχνοτήτων. Η μορφή αυτή μεταβάλλεται με το χρόνο, με σκοπό τη διαμόρφωση του επιθυμητού ήχου κάθε φορά. Κάθε συχνότητα formant χαρακτηρίζεται από μια κεντρική συχνότητα και ένα εύρος ζώνης. Εφόσον το σχήμα της φωνητικής οδού επηρεάζεται και από την συναισθηματική κατάσταση του ομιλητή, οι συγκεκριμένες συχνότητες είναι ιδιαίτερα χρήσιμα χαρακτηριστικά για την αναγνώριση συναισθήματος. Για παράδειγμα, έχει βρεθεί ότι τα άτομα σε κατάσταση άγχους ή κατάθλιψης δεν αρθρώνουν έμφωνους ήχους με τον ίδιο τρόπο όπως στην ουδέτερη κατάσταση [37].



Σχήμα 3.3: Απεικόνιση των 5 πρώτων formant συχνοτήτων για το φωνήεν "i" [38]

Ποιότητα Φωνής

Ενδιαφέρον παρουσιάζουν, επίσης και επιπλέον παράμετροι που εκφράζουν την ποιότητα φωνής, μέσω της γλωττιδικής ροής αέρα. Η γλωττιδική ροή όπως είναι αυτονόητο εκφράζει την ροή του αέρα που περνάει από τις φωνητικές χορδές. Ωστόσο η χρήση τους στο παρελθόν αποφευγόταν λόγω της δυσκολίας να προσεγγισθούν ,αφού απαιτούν περίπλοκες μαθηματικές πράξεις. Κάποιες από αυτές χρησιμοποιούνται για να εκτιμήσουν το εύρος των παραμέτρων γλωττίδας όπως η NAQ (normalized amplitude quotient) που εκφράζει τον συντελεστή κανονικοποιημένου πλάτους και παραμετροποιεί την φάση κλεισίματος της γλωττίδας χρησιμοποιώντας μετρήσεις πλάτους των κυματομορφών από αντίστροφο φιλτράρισμα. Πειράματα με διαφορετικά φωνήεντα (αναπνευστικά, φυσιολογικά, πατημένα) έχουν δείξει ότι τα χαρακτηριστικά αυτά μπορούν να διαχωρίσουν τον τύπο της 'φωνοποίησης' του φωνήεντος αποδοτικά. Οι υπόλοιπες παράμετροι που αξιοποιούνται παρατίθενται στον πίνακα 3.1.

Στατιστικά Χαρακτηριστικά (Functionals)

Σε όλα τα παραπάνω χαρακτηριστικά που προαναφέραμε, είναι δυνατή και μάλιστα συνίσταται η επιβολή στατιστικών σχέσεων. Ακραίες τιμές (Μέγιστη και Ελάχιστη), Μέσες τιμές (Αριθμητικός ή Γεωμετρικός μέσος όρος), Διάμεσος, Ροπές (Τυπική Απόκλιση, Διακύμανση, Κύρτωση, Λοξότητα), Εκατοστημόρια, Τεταρτημόρια, Κεντροειδή, Κλίση, Μέση τιμή Τετραγωνικού Σφάλματος και Διάρκεια/Χρόνος είναι στατιστικές συναρτήσεις που δοκιμάσθηκαν στην εργασία μας αλλά δεν απέδωσαν όλες το ίδιο. Τελικά αξιοποιήθηκαν μόνο οι τρεις βασικές στατιστικές συναρτήσεις Μέγιστης (Max), Ελάχιστης (Min) και Μέσης (Mean) Τιμής.

Συνολικά, συγκεντρώνοντας τα παραπάνω χαρακτηριστικά και εφαρμόζοντας τις στατιστικές συναρτήσεις, καταλήγουμε σε πλήθος 237 χαρακτηριστικών φωνής υψηλού επιπέδου που θα καθορίζουν την έκβαση του ακουστικού μοντέλου της κατάθλιψης έπειτα από την διαδικασία της προεπεξεργασίας και της βαθιάς μάθησης. Ο αριθμός 237 προκύπτει από το άθροισμα των 5 συχνοτήτων Formant και των 74 χαρακτηριστικών που περιλαμβάνουν τα προσωδιακά, του φάσματος και της ποιότητας φωνής, και έπειτα από την εφαρμογή των τριών στατιστικών συναρτήσεων σε κάθε χαρακτηριστικό (75 * 3 = 237).

Στον ακόλουθο πίνακα 3.3 παρουσιάζουμε την επιρροή των δύο βασικών συναισθημάτων που μας απασχολούν (χαρά, λύπη) σε κάποια βασικά ακουστικά χαρακτηριστικά βάση του [39].

	Χαρά	***Λύπη***
Θεμελιώδης Συχνότητα	Αύξηση της μέσης τιμής	Κάτω από την κανονική μέση τιμή
Ένταση	Αύξηση	Μείωση
Ενέργεια Υψηλών Συχνοτήτων	Αύξηση	Μείωση
Ρυθμός Ομιλίας	Αύξηση	Ελαφρά αργός
Ποιότητα Φωνής	Τεταμένη, Ξεψυχισμένη, Πολύ δυνατή	Χαλαρή, Βαθιά

Πίνακας 3.3: Ακουστικά χαρακτηριστικά χαμηλού επιπέδου (LLDs)

3.2.1.2 Προεπεξεργασία Δεδομένων

3.2.1.2.1 Προτεινόμενοι Μέθοδοι Προεπεξεργασίας

Στην ενότητα αυτή θα εφαρμόσουμε μια σειρά από μετασχηματιστές για την προετοιμασία των ακουστικών περιγραφητών πριν εισέλθουν στον τελικό εκτιμητή, ο οποίος θα είναι ένας βασικός (Baseline) αλγόριθμος παλινδρόμησης, ο Random Forest Regressor. Η επιλογή του baseline αλγορίθμου αυτού έγινε λόγω του μεγάλου όγκου χαρακτηριστικών που έχουμε τα οποία μπορεί να χειρίζεται αξιοπρεπώς εφαρμόζοντας την Ensemble μάθηση. Στη συνέχεια λοιπόν θα παρουσιάσουμε την διαδικασία εφαρμογής μετασχηματιστών για την εύρεση του αποδοτικότερου pipeline μετασχηματιστών, καθώς και την διαδικασία βελτιστοποίησης των υπερπαραμέτρων κάποιων μετασχηματιστών με την μέθοδο αναζήτησης πλέγματος (grid search).

1. Αρχικά θα καθαρίσουμε τα δεδομένα μας από τιμές που αντιστοιχούν είτε σε χρονικά παράθυρα που μιλάει ο εικονικός πράκτορας Ellie, είτε σε παράθυρα που δεν υπάρχει ομιλία από καμία από τις δύο πλευρές. Αυτό θα γίνει με την βοήθεια των απομαγνητοσκοπήσεων των συνεντεύξεων. Κάθε πρόταση οποιουδήποτε ομιλητή συνοδεύεται από την χρονική στιγμή που ξεκινάει καθώς και από την χρονική στιγμή που τερματίζει. Γνωρίζοντας επίσης πως τα χαρακτηριστικά ήχου χαμηλής τάξης που έχουν εξαχθεί για κάθε δείγμα, έχουν δειγματοληφθεί ανά παράθυρα των 30ms, κάνουμε τις χρονικές αντιστοιχίες και πετάμε τα χαρακτηριστικά ήχου που ανήκουν στα παράθυρα ομιλίας του εικονικού πράκτορα.

2. Στη συνέχεια, όπως αναφέρεται στο [1], υπάρχει μια παράμετρος που προαναφέραμε στην ενότητα 3.2.1.1, η VUV η οποία μας πληροφορεί πότε εντοπίζεται ήχος (VUV=1) και πότε όχι (VUV=0). Έτσι λοιπόν όταν δεν εντοπίζονται ηχητικά κύματα, δεν έχει νόημα και η ανάθεση τιμών στους περιγραφητές ήχου F0, NAQ, QOQ, H1H2, PSP, MDQ, peakSlope, Rd, γι αυτό και τους μηδενίζουμε χειροκίνητα.

3. Όπως λοιπόν πολύ συχνά εντοπίζουμε απουσιάζουσες τιμών χαρακτηριστικών, έτσι και πολλές φορές ο υπολογιστής αδυνατεί να καταγράψει τιμές που υπερβαίνουν κάποιο ανώτατο όριο, οπότε και τις αντικαθιστά με τον συμβολισμό του απείρου. Τα συστήματα μηχανικής μάθησης ωστόσο αναγνωρίζουν μόνο μαθηματικές αναπαραστάσεις, με αποτέλεσμα η εκπαίδευση να παγώνει όταν συναντάει το σύστημα μη αριθμητικές τιμές όπως η αναπαράσταση του απείρου. Μια εύκολη και αποδοτική λύση είναι η αντικατάσταση του με τον μεγαλύτερο πεπερασμένο δεκαδικό αριθμό που μπορεί να υποστηρίξει.

```
TrainData have infinite values
Process......
No more infinite values
DevData have infinite values
Process......
No more infinite values
```

Στο σημείο λοιπόν αυτό αξίζει να επισημάνουμε πως η εφαρμογή των στατιστικών συναρτήσεων που αναφέραμε στην ενότητα 3.2.1.1 γίνεται αφού εκτελεστούν τα βήματα 1,2,3.

4. Μια επόμενη μέθοδος μετασχηματιστή που εφαρμόζουμε είναι η Feature Selection με την τεχνική του Variance Threshold. Μελετώντας τις διακυμάνσεις των χαρακτηριστικών ήχου υψηλού επιπέδου, παρατηρούμε πως 24 από τα 237 έχουν μηδενική διακύμανση δηλαδη έχουν σταθερή τιμή για όλα τα δείγματα εκπαίδευσης. Έτσι λοιπόν τα πετάμε από την διαδικασία της εκπαίδευσης και εξετάζουμε την συμπεριφορά του εκτιμητή για κατώφλια λίγο μεγαλύτερα του μηδενός, αλλά ακόμη πολύ μικρής κλίμακας. Ωστόσο έπειτα από δοκιμές παρατηρούμε πως η επίδοση του μοντέλου φθίνει, κρατώντας σαν τελικό κατώφλι την τιμή 0.

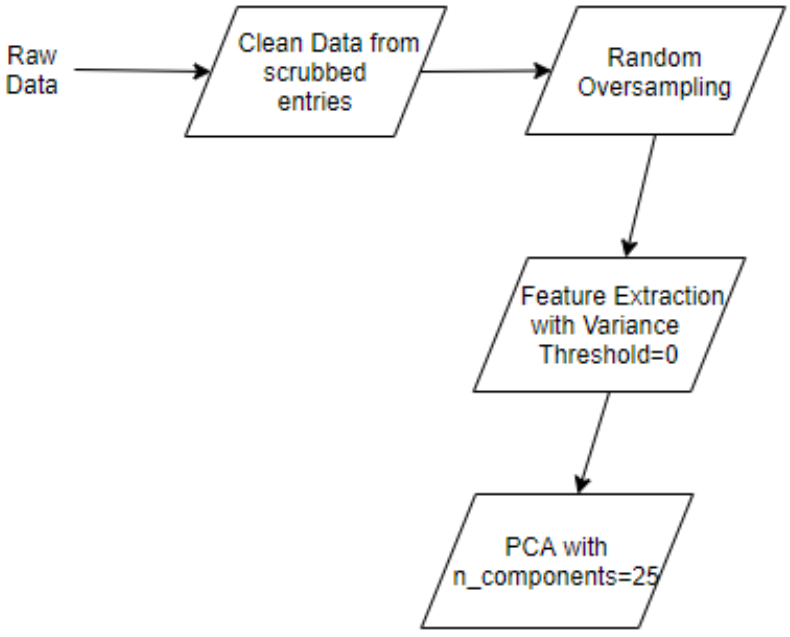
5. Μια εναλλακτική μέθοδος feature extraction που ωστόσο μπορεί να εφαρμοστεί μετά την προαναφερθείσα, είναι η τεχνική PCA κατα την οποία κάνουμε εξαγωγή νέων χαρακτηριστικών σε ένα χώρο μικρότερων διαστάσεων που καθορίζουμε εμείς. Κάνοντας λοιπόν δοκιμές σε ένα εύρος τιμών [10,60] με βήμα 5, καταλήξαμε στην βέλτιστη προσέγγιση με n_components=25.

3.2.1.2.2 Τελικός Συνδυασμός Μεθόδων Προεπεξεργασίας

Για να αποφανθούμε τον τελικό ιδανικό συνδυασμό μετασχηματιστών που θα τροφοδοτήσουν τα τελικώς επεξεργασμένα δεδομένα στον εκτιμητή μας κάναμε δοκιμές συνδυασμών τους οποίους συγκρίναμε βάση των μετρικών αξιολόγησης RMSE και MAE. Επιπλέον για την βελτιστοποίηση των υπερπαραμέτρων τόσο των μετασχηματιστών όσο και του εκτιμητή αξιοποιήσαμε την αναζήτηση πλέγματος. Τελικώς προέκυψε ο πίνακας σύγκρισης 3.4 που παρουσιάζει τα μοντέλα με αύξουσα σειρά απόδοσης και δίνει σαν βέλτιστο pipeline προεπεξεργασίας αυτό που παρουσιάζεται σαν διάγραμμα ροής στο σχήμα 3.4.

	MAE	RMSE
ROS-Scaler-VarThres-PCA(n=80)-RF(n=5)	6.258235	7.341370
Voting Regressor-RF(n=10)	5.826471	6.968479
VarianceThreshold-RF(n=10)	5.945294	6.896430
ROS-RF(n=10)	5.840882	6.762307
Voting Regressor-PCA-RF(n=10)	5.314706	6.487114
Scaler-VarThres-PCA(n=35)-RF(n=100)	5.162647	6.191514
ROS-VarianceThreshold-RF(n=10)	5.034043	6.036414
PCA(n=25)-RF(n=10)	4.801765	5.832229
ROS-VarThres-PCA(n=25)-RF(n=10)	4.802647	5.562506

Πίνακας 3.4:Σύγκριση μεθόδων Pipeline Προεπεξεργασίας δεδομένων



Σχήμα 3.4: Διάγραμμα ροής βέλτιστου συνδυασμού μετασχηματιστών προεπεξεργασίας περιγραφητών ήχου

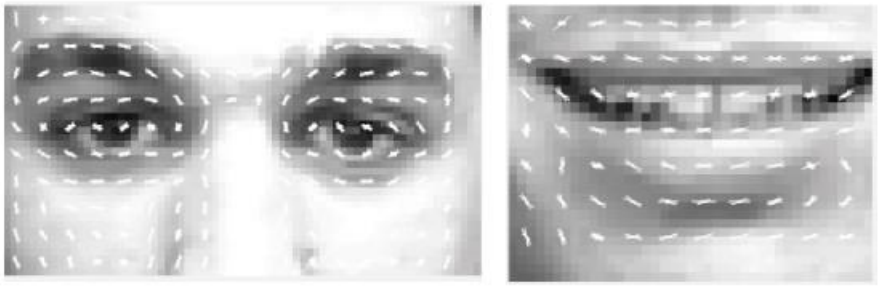
3.2.2 Οπτικά Χαρακτηριστικά

3.2.2.1 Μελέτη Δεδομένων

Όπως προαναφέραμε στην ενότητα 1.2.2, οι άνθρωποι για να εκφράσουν την ψυχολογική τους κατάσταση πέραν της φωνής, χρησιμοποιούν ως μέσο και τις εκφράσεις των προσώπων τους. Ένα ισχυρό σύστημα κωδικοποίησης δράσης του προσώπου είναι αυτό των περιγραφητών AUs [7], λεπτομέρειες των οποίων δίνουμε στην ενότητα 1.2.2 και οι οποίοι έχουν εξαχθεί για κάθε δείγμα του dataset μας ανά παράθυρο των 33ms καθ' όλη τη διάρκεια της συνέντευξης. Επιπλέον όμως με την χρήση του εργαλείου OpenFace[3], έχει εξαχθεί μια πληθώρα περιγραφικών χαρακτηριστικών για την επεξεργασία των βιντεοσκοπημένων εκφράσεων των ασθενών και την δημιουργία των κατάλληλων προτύπων τα οποία καλύπτουν όλες τις περισσότερες πιθανές μεθόδους εξαγωγής χαρακτηριστικών που θα μπορούσαν να εφαρμοστούν στις εκφράσεις του προσώπου. Έτσι λοιπόν, συνολικά προκύπτουν τα παρακάτω χαρακτηριστικά [1]:

1. HOG χαρακτηριστικά (Histogram of oriented gradients):

Η μέθοδος αυτή αποτελεί μια πυκνή κατανομή τοπικών περιγραφητών, ένα ιστόγραμμα κατευθύνσεων για κάθε υποπεριοχή του προσώπου, η οποία τεχνική χρησιμοποιείται ευρύτατα στην περιοχή της όρασης υπολογιστών. Εφαρμόζοντας λοιπόν την τεχνική αυτή στην περιοχή της κεφαλής με διαστάσεις 112x112 pixels, εξάγεται για κάθε παράθυρο των 33 ms, ένα διάνυσμα μεγέθους 4464 θέσεων.



Σχήμα 3.5: Οπτικοποίηση των HOG χαρακτηριστικών στην περιοχή των ματιών και του στόματος

2. Χαρακτηριστικά κατεύθυνσης βλέμματος (Eye Gaze) & θέσης κεφαλιού (Head Pose)

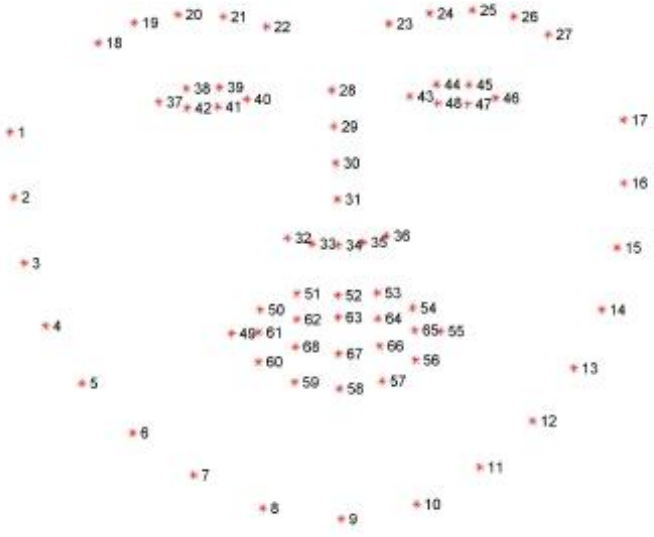
Ομοίως, ανά παράθυρο των 33ms, για τα χαρακτηριστικά Eye Gaze παίρνουμε 4 διαφορετικα διανύσματα, τα δύο πρώτα εκφράζουν την κατεύθυνση των ματιών στο πραγματικό σύστημα συντεταγμένων, ενώ τα δύο επόμενα την κατεύθυνση των ματιών σε σχέση με την θέση του κεφαλιού (π.χ. αν τα μάτια κάνουν κίνηση προς τα πάνω, τότε τα δύο τελευταία διανύσματα θα υποδείξουν κίνηση προς τα πάνω ακόμη και αν το κεφάλι είναι στραμμένο προς τα κάτω). Αντίστοιχα, τα χαρακτηριστικά Head Pose δίνουν ένα διάνυσμα 6 θέσεων, του οποίου οι τρεις πρώτες αντιπροσωπεύουν την θέση του κεφαλιού

[3] http://github.com/TadasBaltrusaitis/OpenFace

στο τρισδιάστατο σύστημα συντεταγμένων, ενώ οι τρεις τελευταίες τις συντεταγμένες περιστροφής του κεφαλιού στο ίδιο σύστημα συντεταγμένων. Η θέση του κεφαλιού δίνεται σε μονάδες millimeters ενώ η περιστροφή του σε radians κατα σύμβαση γωνίας Euler.

3. Χαρακτηριστικά συντεταγμένων σημείων του κεφαλιού και AUs

Τέλος, εκτός από τα χαρακτηριστικά AUs,δηλαδή περιγραφητές προσώπου που προκύπτουν από το σύστημα κωδικοποίησης δράσης του προσώπου (FACS) που αναλύσαμε στην ενότητα 1.2.2, δίνονται και οι συντεταγμένες 68 σημείων του προσώπου στο κανονικό δισδιάστατο σύστημα συντεταγμένων, καθώς και στο τρισδιάστατο με παράθυρο δειγματοληψίας 33ms. Ωστόσο για να υπάρχει μέτρο σύγκρισης ανάμεσα στα χαρακτηριστικά που έχουν εξαχθεί από το κάθε δείγμα και να μην επηρεαστεί η απόδοση του συστήματος αξιολόγησης της κατάθλιψης από εξωτερικούς τεχνικούς παράγοντες, η κάμερα βιντεοσκόπησης έχει τοποθετηθεί στο κέντρο του συστήματος συντεταγμένων του προσώπου κάθε δείγματος. Μονάδα μέτρησης είναι το pixel.



Σχήμα 3.6: Τα 68 σημεία του προσώπου στο 2D σύστημα συντεταγμένων

Συνολικά, έχοντας μια τεράστια ποικιλία από διαφορετικούς περιγραφητές χαρακτηριστικών των εκφράσεων του προσώπου, είμαστε σε θέση να πειραματιστούμε σε ποικίλες μεθοδολογίες, τις οποίες θα παρουσιάσουμε στην επόμενη υποενότητα 3.2.2.2.

3.2.2.2 Προεπεξεργασία Δεδομένων

3.2.2.2.1 Προτεινόμενοι Μέθοδοι Προεπεξεργασίας

Όπως προ είπαμε, πριν την κατασκευή των feature vectors που θα τροφοδοτηθούν σε κάποιον εκτιμητή, συνήθως εφαρμόζουμε κάποιες τεχνικές προεπεξεργασίας των δεδομένων, κάποιες από τις οποίες αναπτύξαμε στην ενότητα 3.1.1. Έτσι λοιπόν παρουσιάζουμε στη συνέχεια τις μεθόδους εξαγωγής χαρακτηριστικών βίντεο που μελετήσαμε.

1. Ομοίως με τα χαρακτηριστικά ήχου, παρατηρούμε πως σε κάποια παράθυρα αποτυπώσεων των συντεταγμένων σημείων του προσώπου, δεν καταγράφονται πραγματικές τιμές, αλλά ένα σύμβολο που υποδεικνύει την αδυναμία του υπολογιστή να καταγράψει τα επιθυμητά δεδομένα. Ωστόσο μια πιθανή εξήγηση αυτού του φαινομένου θα μπορούσε να είναι κάποια έντονη μετακίνηση του ασθενή μπροστά στην κάμερα ή ακόμη και πλήρη μετακίνηση του εκτός πεδίου εγγραφής. Έτσι λοιπόν διορθώνουμε τα δεδομένα μας πετώντας τα παράθυρα που παρουσιάζουν αυτό το χαρακτηριστικό. Σε κάθε δείγμα για κάθε οπτικό χαρακτηριστικό του έχουν εξαχθεί κατά μέσο όρο 30.000 παράθυρα κατα τη διάρκεια της συνεδρίας, γεγονός που προδίδει την ασημαντότητα της τεχνικής να πετάξουμε τα ελαττωματικά παράθυρα, εφόσον υπάρχουν.

2. Εκμεταλλευόμενοι το γεγονός πως τα χαρακτηριστικά εισόδου αποτελούν διαδοχικά παράθυρα μιας αλληλουχίας από εικόνες (frames), δηλαδή ενός βίντεο, γίνεται εύκολα αντιληπτό πως μεμονωμένα frames δεν φέρουν σημαντική πληροφορία για τα δείγματα μας αν πάρουμε υπόψιν μας και το γεγονός πως κάθε δείγμα αποτελείται από χιλιάδες frames. Βάση αυτής της παρατήρησης και εμπνευσμένοι από το [58], υπολογίζουμε τις μεταβολές των χαρακτηριστικών κατεύθυνσης βλέμματος (gaze) και θέσης κεφαλιού (pose) ανά frame, τόσο δηλαδή την πρώτη παράγωγο, όσο και την δεύτερη για εξόρυξη περαιτέρω πληροφορίας. Τέλος τόσο τα πρωτότυπα χαρακτηριστικά gaze και pose, όσο και η ‘ταχύτητα’ και ‘επιτάχυνσή’ τους δίνουν σαν τελική έξοδο χαρακτηριστικών τον μέσο όρο των frames τους πριν συναθροιστούν και δημιουργήσουν έξι (6) νέα feature vectors, τα οποία θα ονομάσουμε FacialMarker2 για ευκολία στην αναφορά τους. Στον πίνακα 3.5 παρουσιάζουμε τις προκύπτουσες διαστάσεις της μεθόδου αυτής.

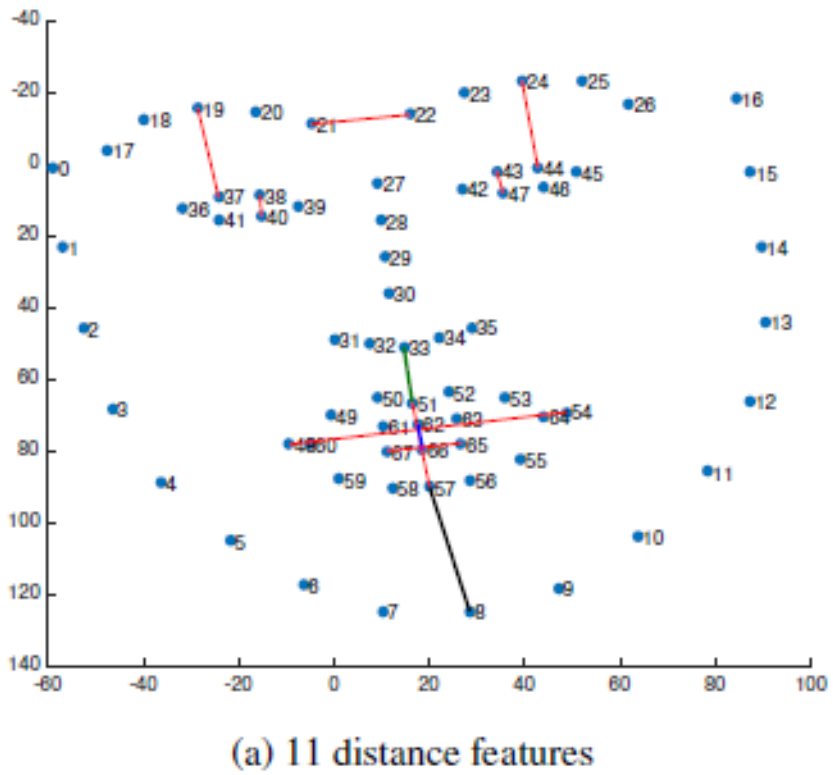
Feature Set Name	*Description*	*Dimension*
FacialMarker1	Eye Gaze	12
	Head Pose	6
	Total	18
FacialMarker2	FacialMarker1 + Δ + Δ-Δ	54

Πίνακας 3.5: Δημιουργία Χαρακτηριστικών video από Eye Gaze και Head Pose Features

3. Βασιζόμενοι στην προσέγγιση του [21], θα εκμεταλλευτούμε τις δοσμένες συντεταγμένες του προσώπου στο δισδιάστατο σύστημα που αντιστοιχούν σε σημεία του προσώπου καθοριστικών θέσεων για την ερμηνεία εκφράσεων βασισμένων σε συναισθηματικές αντιδράσεις. Έτσι εξάγουμε τα λεγόμενα Γεωμετρικά χαρακτηριστικά (Geometrical features) τα οποία ουσιαστικά προκύπτουν από συγκεκριμένες αποστάσεις μεταξύ υψηλής σημασίας σημείων του προσώπου, καθώς επίσης και από τις μεταβολές των αποστάσεων αυτών σε διάρκεια δέκα (10) παραθύρων. Η μεταβλητή της διάρκειας, δηλαδη του πλήθους των παραθύρων επιλέχθηκε ως η βέλτιστη μετά από δοκιμές. Πιο συγκεκριμένα, οι αποστάσεις που επιλέχθηκαν φέρουν πληροφορία σχετικά με το ανοιγοκλείσιμο του στόματος τόσο στον κάθετο όσο και στον οριζόντιο άξονα, αντίστοιχα για τα μάτια και τα φρύδια. Λεπτομερής απεικόνισή τους φαίνεται στο σχήμα 3.5. οι αποστάσεις αυτές κανονικοποιούνται βάση του πλάτους του προσώπου καθώς αυτό παραμένει πάντα σταθερό, ώστε οι αποστάσεις που υπολογίζονται να μην παρουσιάζουν αναληθείς τιμές. Αφου λοιπόν εξάγουμε και τις πρώτες παραγώγους των αποστάσεων αυτών, για κάθε θέση του τελικού γεωμετρικού διανύσματος παίρνουμε τον μέσο όρο κατα μήκος των τελικών πλαισίων που προέκυψαν για κάθε δείγμα. Στον πίνακα 3.6 παρουσιάζουμε τις διαστάσεις των χαρακτηριστικών αυτών.

Feature Set Name	*Description*	*Dimension*
Geometrical	Distance	11
	Δ Distance	11
	Total	22

Πίνακας 3.6: Δημιουργία Γεωμετρικών Χαρακτηριστικών από τα Facial Landmarks 2D

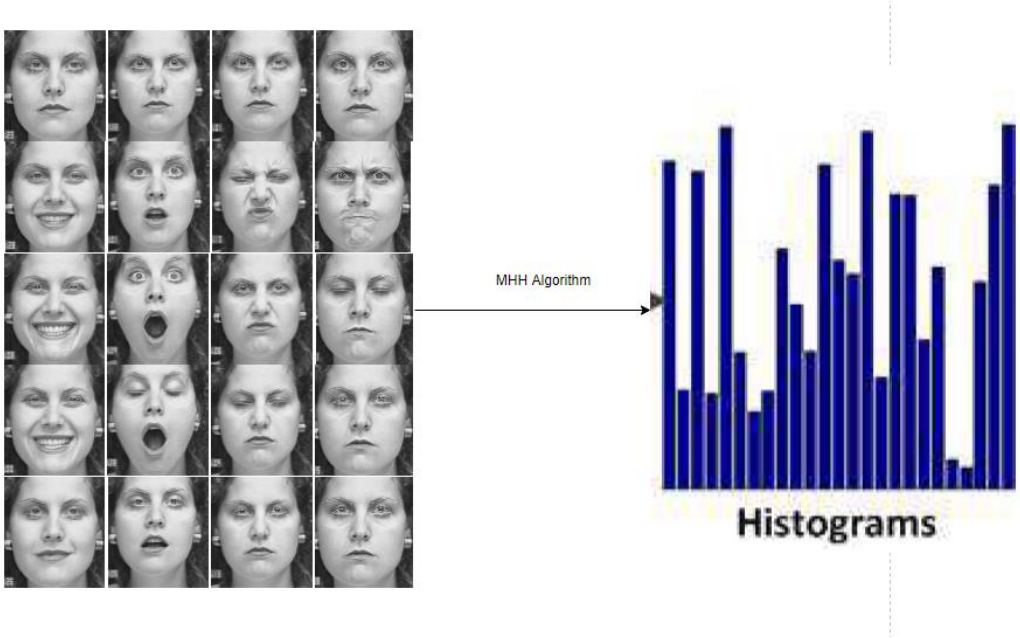


(a) 11 distance features

Σχήμα 3.7: Απεικόνιση των Γεωμετρικών χαρακτηριστικών πάνω στα facial landmarks

4. Ωστόσο την απόδοση της τεχνικής αυτής έρχεται να ξεπεράσει ο Yang κ.α. [60] εκμεταλλευόμενος την προσέγγιση του Meng κ.α. [59] πάνω στην δημιουργία Motion History Histogram (MHH) για την εξαγωγή δυναμικών χαρακτηριστικών από περιγραφητές video. Σε συνδυασμό λοιπόν με την επίδοση των περιγραφητών προσώπου AUs που προκύπτουν από το σύστημα κωδικοποίησης δράσης του προσώπου (FACS) όπως αναλύσαμε στην ενότητα 1.2.2 και που η χρήση τους έχει αποδειχθεί πως κατέχει κυρίαρχο ρόλο στην αναγνώριση της κατάθλιψης [62], ο Yang προτείνει την εξαγωγή χαρακτηριστικών μέσω MHH από τα περιγραφικά χαρακτηριστικά προσώπου AUs για την εκμάθηση της δυναμικής της συμπεριφοράς τους βάση συναισθήματος. Η τεχνική MHH είναι μια περιγραφική αναπαράσταση της οπτικής κίνησης που αρχικά εφαρμόστηκε για την αναγνώριση της ανθρώπινης δράσης [61]. Αυτή η προσέγγιση λοιπόν, κάνει μια εκτίμηση των αλλαγών κάθε περιγραφητή AU κατά μήκος όλων των τιμών των παραθύρων που έχει διαχωριστεί η ακολουθία εικόνων (βίντεο), με αποτέλεσμα να δημιουργείται ένα ιστόγραμμα από B εξίσου διαχωρισμένα διαστήματα, $\{R_b, b = 1, \dots, B\}$, ανάμεσα σε ένα

συνολικό διάστημα εύρους [-8,8], καθώς τα δυναμικά των χαρακτηριστικών κυμαίνονται σε αυτό το εύρος. Για την εκτίμηση των δυναμικών των AUs , υπολογίζουμε τις αριθμητικές διαφορές των τιμών των AUs σε διαφορετικά χρονικά πλαίσια $M_k, k = 1, \dots, K$, επομένως το κάθε AU διάνυσμα αντιπροσωπεύεται από kxB τιμές, όπου η κάθε τιμή προκύπτει από το άθροισμα των τιμών των δυναμικών που ανήκουν στο αντίστοιχο εύρος b και ακολουθούν το αντίστοιχο βήμα παραθύρων k. Δηλαδή, για χρονικό πλαίσιο M_k, υπολογίζουμε την δυναμική των AU ως D(i,j) = $AU^j_{i+M_k} - AU^j_i$, για την τιμή του AU^j μεταξύ των frames i και $i + M_k$ και στο τέλος κάθε κλάδος (bin) του ιστογράμματος αντιπροσωπεύεται από το πλήθος των D(i,j) που ανήκουν στο διάστημα του κλάδου αυτού, με $i = 1, \dots, N, j = 1, \dots, N_{AU}$, όπου N είναι ο αριθμός των frames που προκύπτουν από τα χρονικά πλαίσια M_k και N_{AU}=20 το πλήθος των AUs που μας δίνεται. Στα πειράματα μας λοιπόν θα χρησιμοποιήσουμε 5 χρονικά πλαίσια M_k=10,20,30,40,50 frames, ενώ 4 ίσα διαστήματα R_i μέσα στο εύρος [-8,8]. Τελικώς προκύπτει διάνυσμα χαρακτηριστικών μήκους 400 για κάθε δείγμα.



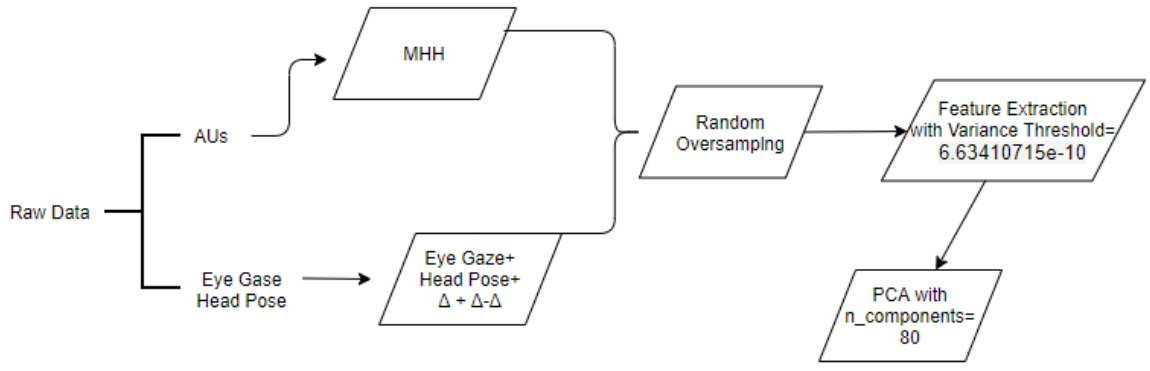
Σχήμα 3.8: Διάγραμμα μετατροπής των AUs σε HMM

3.2.2.2.2 Τελικός Συνδυασμός Μεθόδων Προεπεξεργασίας

Για να αποφανθούμε τον τελικό ιδανικό συνδυασμό μετασχηματιστών που θα τροφοδοτήσουν τα τελικώς επεξεργασμένα δεδομένα στον εκτιμητή μας κάναμε δοκιμές συνδυασμών τους οποίους συγκρίναμε βάση των μετρικών αξιολόγησης RMSE και MAE. Επιπλέον για την βελτιστοποίηση των υπερπαραμέτρων τόσο των μετασχηματιστών όσο και του εκτιμητή αξιοποιήσαμε την αναζήτηση πλέγματος. Τελικώς προέκυψε ο πίνακας σύγκρισης 3.7 που παρουσιάζει τα μοντέλα με αύξουσα σειρά απόδοσης και δίνει σαν βέλτιστο pipeline προεπεξεργασίας αυτό που παρουσιάζεται σαν διάγραμμα ροής στο σχήμα 3.9.

Feature Extraction Combinations	MAE	RMSE
AUs Dynamics-Oversample	6.079412	7.347849
[AUs Dynamics, facialMarker2]-Oversample	5.750000	7.165296
facialMarker1	5.570588	6.789395
[facialMarker1, var_skew_bool]	5.605882	6.752647
facialMarker2- Oversample	5.358824	6.716704
Geometrical Features	5.452941	6.690599
AUs Dynamics -PCA(n=25)	5.294118	6.611977
facialMarker1- Oversample	5.214706	6.591907
AUs Dynamics	5.323529	6.582999
AUs Dynamics-Variance Threshold=0	5.323529	6.582999
AUs Dynamics -PCA(n=50)	5.358824	6.550079
[AUs Dynamics, facialMarker2]	4.882353	6.355498
facialMarker2	4.955882	6.274903
AUs Dynamics -Variance Threshold =0.00072865	5.085294	6.242902
[AUs Dynamics, facialMarker2]-Variance Threshold=6.63410715e-10-PCA(n=80)	4.924706	5.953014

Πίνακας 3.7:Σύγκριση μεθόδων Προεπεξεργασίας δεδομένων στον baseline RF Regressor(n=10)



Σχήμα 3.9: Διάγραμμα ροής βέλτιστου συνδυασμού μετασχηματιστών προεπεξεργασίας περιγραφητών βίντεο

3.3 Επεξεργασία Κειμένου

3.3.1 Προτεινόμενοι Μέθοδοι Προεπεξεργασίας

Πέραν των μη λεκτικών περιγραφητών που εξήγαμε από τα οπτικοακουστικά δεδομένα, οι απομαγνητοφωνημένες συνεντεύξεις μπορούν να μας δώσουν σημαντική πληροφορία για την ψυχική κατάσταση του ασθενή. Κάθε συνέντευξη καθοδηγείται από τον εικονικό πράκτορα που σημαίνει πως μόνο αυτός μπορεί να εισάγει καινούριο θέμα συζήτησης και όχι ο ίδιος ο ασθενής ο οποίος απαντάει μόνο στις ερωτήσεις που του θέτει ο υπολογιστής. Οι ερωτήσεις αυτές έχουν επιλεχθεί έτσι ώστε οι απαντήσεις τους να προδίδουν συμπτώματα σχετιζόμενα με την ψυχολογική πλευρά της κατάθλιψης, όπως διαταραχή ύπνου, έλλειψη όρεξης κ.α. Εμπνευσμένοι λοιπόν από το [17] θα εφαρμόσουμε μια μέθοδο εξαγωγής χρήσιμης πληροφορίας από την ανάλυση περιεχομένου των απαντήσεων των ασθενών πάνω σε συγκεκριμένες καθοδηγούμενες ερωτήσεις που αφορούν πέντε (5) διαφορετικά συμπτώματα της κατάθλιψης:

1. Πρώην διάγνωση της Κατάθλιψης (Ναι/Οχι)
2. Πρώην μετα-τραυματικό στρες (PTSD) (Ναι/Οχι)
3. Διαταραχή Ύπνου (Ναι/Οχι)
4. Αισθήματα (Ευχάριστα/Δυσάρεστα)
5. Προσωπικότητα (Εσωστρεφής/Εξωστρεφής)

Όπως γνωρίζουμε όμως, οποιοσδήποτε αλγόριθμος ML μπορεί να εκπαιδευτεί μόνο σε δεδομένα που αναπαρίστανται αριθμητικά, επομένως πρέπει να μετατρέψουμε την πληροφορία του κειμένου σε μαθηματικές αναπαραστάσεις. Έτσι λοιπόν στόχος μας είναι να εντοπίσουμε την παρουσία ή μη των παραπάνω συμπτωμάτων για μετέπειτα χρήση στην ταξινόμηση και εκτίμηση της κατάθλιψης. Το τελικό διάνυσμα χαρακτηριστικών κειμένου θα είναι ένα διάνυσμα μήκους 5 με δυαδικές τιμές στην κάθε του θέση, π.χ. [0,0,1,0,0].

Για την δημιουργία των διανυσμάτων αυτών θα μελετήσουμε δύο (2) προσεγγίσεις, πρώτα την Semantic Context Analysis κατά την οποία θα δημιουργήσουμε ένα λεξικό με όσο το δυνατόν περισσότερες πιθανές απαντήσεις στις πέντε αυτές ερωτήσεις που προαναφέραμε και χειροκίνητα θα δίνουμε την αντίστοιχη ετικέτα στην απάντηση βάση της αντιστοιχίας με το λεξικό (0/1) και δεύτερον την προσέγγιση με τα Paragraph Vectors κατά την οποία θα αναπαραστήσουμε τις απαντήσεις των ασθενών με PV περιγραφητές τους οποίους έπειτα θα περάσουμε από SVM ταξινομητές για την κατηγοριοποίηση σε 0/1 τιμές.

Ωστόσο η διαδικασία αυτή ακολουθεί τη λογική της μη επιβλεπόμενης μάθησης καθώς δεν μας δίνεται κάποια ετικέτα για τα συμπτώματα αυτα, επομένως αφού μάλιστα βασιζόμαστε πλήρως στον υποκειμενικό παράγοντα στην μελέτη αυτη είναι προφανές πως θα υπάρχουν σημαντικές αποκλίσεις.

3.3.1.1 Προσέγγιση Semantic Context

Στην προσέγγιση αυτή, όπως προ είπαμε θα κατηγοριοποιήσουμε σε κάθε ασθενή τα παραπάνω πέντε συμπτώματα βάση λέξεων κλειδιών των απαντήσεων τους που θα αντιστοιχίζουμε σε ένα λεξικό που δημιουργήσαμε χειροκίνητα. Ουσιαστικά θα υλοποιήσουμε ένα σύστημα ανάλυσης περιεχομένου (Content Analysis) και θα δίνουμε την κατάλληλη ετικέτα στο κάθε σύμπτωμα ανάλογα της παρουσίας του ή μη.

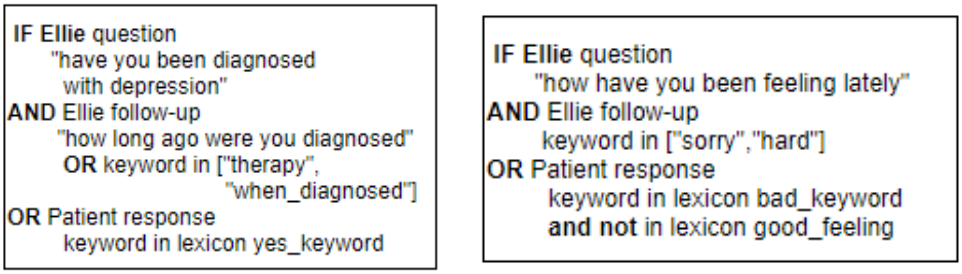
3.3.1.1.1 Δημιουργία Λεξικού

Πιο συγκεκριμένα, όσον αφορά τις ερωτήσεις που η απάντηση τους είναι άμεσα κατηγοριοποιήσιμη, όπως η πρώην διάγνωση της κατάθλιψης και του PTSD, η απάντηση δηλαδή είναι ως επι τω πλείστων Ναι ή Οχι, οι ετικέτες τους αυτόματα συμπληρώνονται ως Ναι (1) ή Οχι (0). Ωστόσο, οι τρεις επόμενες ερωτήσεις των οποίων οι απαντήσεις είναι πιο έμμεσες, χρειάζονται περαιτέρω ανάλυση. Παραδείγματος χάριν, στην ερώτηση περί συμπτώματος διαταραχής ύπνου, ο ασθενής μπορεί να δώσει μια απάντηση που να περιλαμβάνει τις ακόλουθες πιθανές λέξεις ή φράσεις: 'not had a good sleep', 'really hard', 'kind a difficult', 'never easy'. Στην περίπτωση αυτή κατηγοριοποιούμε την απάντηση ως αρνητική (1). Αν η απάντηση περιλαμβάνει όμως φράσεις όπως: 'no problem','pretty good','get a good night's sleep','it varies' τότε δίνουμε την ετικέτα θετική (0). Κάνοντας επομένως μελέτη των κειμένων μας πάνω στο σετ εκπαίδευσης (Train Set), καθώς επίσης και εισάγοντας δικές μας πιθανές απαντήσεις ώστε να καλύψουμε όσο το δυνατόν μεγαλύτερο εύρος απαντήσεων και από τα δεδομένα επαλήθευσης (dev set και test set), δημιουργούμε τα ακόλουθα λεξικά (Σχήμα 3.10) καθώς και διαγράμματα ροής, κάποια από τα οποία παρατίθενται στο Σχήμα 3.11. Τα διαγράμματα αυτά κάνουν έλεγχο της ροής της συζήτησης μεταξύ ασθενή και εικονικού πράκτορα καθώς μεταξύ της απάντησης του ασθενή και της ερώτησης μπορεί να παρεμβάλλονται τόσο αναστεναγμοί, γέλια κ.α. του ίδιου ασθενή όσο και σχολιασμοί του εικονικού πράκτορα οι οποίοι ωστόσο αν υπάρχουν είναι συγκεκριμένοι και γνωστοί.

```
no_keyword=["no","not","nope","never"]
yes_keyword=["yes","yeah"]
fairsleep_keyword=["depends","it's a little difficult","it varies","challenging"]
easysleep_keyword=["sleep good","it's reasonably okay","somewhat okay","pretty good","heavy sleeper","pretty easy",
                   "don't have problems","pretty well","really good","good night's rest","it's easy","good sleeper",
                   "really easy","it's great","very easy","extremely easy","fairly easy","pretty much","dead asleep",
                   "it's easy", "no problem","extremely easy","sleep well","sometimes good","it's not that hard"]
hardsleep_keyword=["pretty hard","it's been hard","really bad","not easy","haven't had a good night's sleep","difficult",
             "impossible","a little while","it isn't","don't sleep that good",
             "help me sleep","the hard part","that is terrible","awful","not really super easy","i don't sleep",
             "crying","not very easy","not too easy","not so easy","not that easy","i'm never","it's kinda hard",
             "don't get much sleep","i can't remember","insomniac","not getting any sleep","sometimes it's not",
             "never feel rested","hardly ever","very hard","not so good","it's hard","i can't","the only problem",
             "kinda hard","not as good","part of the problem"]
good_feeling=["pretty good","feeling good","positive","alright","feeling fine","happy","pretty good","optimistic",
              "fine","feeling okay","feeling well","feel good","great","happy","stoked","been okay","feel okay","excited",
              "confident","i'm okay","content"]
bad_feeling=["uneasy","feeling depressed","it's kinda hard"," down ","downs","not okay","tired","worried","anxious",
             "depressed","difficulty","frustrated","not good","didn't feel so good","very stressed","trapped","helpless",
             "burdened","worried","upset","irritable","anger","sad"]
ellie_response=["sorry","hard"]

extrovert_personality=["no ","i'm not","outgoing","extrovert","outspoken"]
introvert_personality=["yes","yeah","sure","at times","introvert","i guess so","quiet","shy","introverted"]
both=["both","the middle","somewhat","depends","yeah uh a little bit","varies","in the median","in between"]
```

Σχήμα 3.10: Λεξικά πιθανώς απαντήσεων πάνω στα 5 συμπτώματα κατάθλιψης



Σχήμα 3.11: Διαγράμματα ροής σημασιολογικού περιεχομένου για την ένδειξη συμπτωμάτων πρώην κατάθλιψης και συναισθημάτων

Τα δεδομένα των κειμένων μας μας έχουν δοθεί για χάρη ευκολίας καθαρισμένα, χωρίς κεφαλαίους χαρακτήρες ή σημεία στίξης και καθώς μας ενδιαφέρει οι προτάσεις να είναι πλήρως διατηρημένες όπως δίνονται για ανίχνευση συγκεκριμένων φράσεων, παραλείπουμε το βήμα της προεπεξεργασίας του κειμένου.

Επιπλέον αξίζει να επισημάνουμε πως σε κάθε ασθενή γίνονται μερικές από τις παραπάνω 5 ερωτήσεις ανάλογα τη ροή της συζήτησης, σε κάποιες περιπτώσεις και όλες, επομένως έχουμε εφαρμόσει σαν γενικό κανόνα, σε όσα συμπτώματα δεν γίνεται αναφορά να δίνουμε την ετικέτα 0 που δηλώνει την μη ύπαρξη του συμπτώματος υπέρ της κατάθλιψης.

3.3.1.1.3 Συμπεράσματα

Αφού λοιπόν εξάγουμε τα διανύσματα ένδειξης συμπτωμάτων των ασθενών, παρουσιάζουμε κάποια στατιστικά δεδομένα πάνω στα δείγματα εκπαίδευσης που στο σύνολο ειναι 107. Στο Σχήμα 3.12 παραθέτουμε το πλήθος των συνεντεύξεων που μελετάτε το κάθε σύμπτωμα, ενώ στο Σχήμα 3.13 τα ποσοστά των ανθρώπων που έχουν κάποιο σύμπτωμα το οποίο συνέβαλε στην εμφάνιση της κατάθλιψης καθώς και αυτών που δεν πάσχουν από κατάθλιψη ούτε εμφανίζουν το αντίστοιχο σύμπτωμα.

```
Interviews with depression question: 78
Interviews with ptsd: 86
Interviews with sleep question: 82
Interviews with feeling question: 72
Interviews with personality question: 80
```

Σχήμα 3.12: Πλήθος εμφανίσεων ερωτήσεων κάθε συμπτώματος στα δείγματα εκπαίδευσης

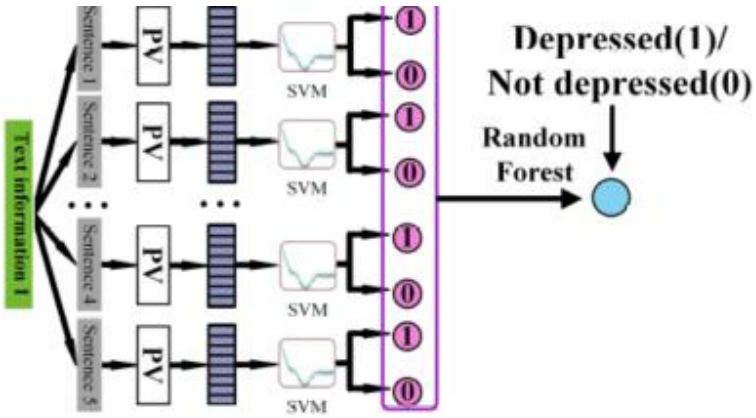
```
30  people suffer now from depression
77  people do not suffer from depression
23.33%  people with depression suffer from ptsd
97.40%  people without depression do not suffer from ptsd
23.33%  people with depression suffered from depression in the past
80.52%  people without depression did not suffer from depression in the past
43.33%  people with depression suffer from sleep disorder
87.01%  people without depression do not suffer from sleep disorder
36.67%  people with depression do not feel well
87.01%  people without depression feel well
16.67%  people with depression are introvert
89.61%  people without depression are extrovert
```

Σχήμα 3.13: Στατιστικά συμπτωμάτων στα δείγματα εκπαίδευσης

3.3.1.2 Προσέγγιση Paragraph Vector

Όπως προείπαμε, η πληροφορία κειμένου για να αναγνωριστεί από κάποιον αλγόριθμο μηχανικής μάθησης πρέπει να μετατραπεί σε διανύσματα χαρακτηριστικών. Στην προσέγγιση αυτή λοιπόν θα εξάγουμε Paragraph Vector (PV) περιγραφητές (2.7.2.3.2) για κάθε απάντηση που δίνει ο κάθε ασθενής στα πέντε συμπτώματα που αναλύσαμε προηγουμένως. Έπειτα για κάθε σύμπτωμα θα εκπαιδεύσουμε ένα μοντέλο ταξινόμησης SVM για την παρουσία ή μη του συμπτώματος. Ο ταξινομητή αυτός όμως ανήκει στους αλγορίθμους επιβλεπόμενης μάθησης, επομένως σαν ετικέτες των δειγμάτων μας πάνω στα οποία θα εκπαιδευτούν οι ταξινομητές θα αξιοποιήσουμε τα διανύσματα που εξήγαμε στην προσέγγιση σημασιολογικής ανάλυσης περιεχομένου 3.3.1.1. Ο λόγος που θα χρησιμοποιήσουμε την τεχνική των PV είναι πως μπορούν να αναπαραστήσουν προτάσεις

ανεξαρτήτου μεγέθους και παράλληλα εκμεταλλεύονται τα πλεονεκτήματα της προσέγγισης Word2Vec σχετικά με την σημασιολογία των λέξεων, ενώ επιπλέον χρησιμοποιεί την πληροφορία που δίνει η σειρά των λέξεων σε μια πρόταση. Παρακάτω παρουσιάζουμε στο Σχήμα 3.14 την αρχιτεκτονική της προσέγγισης PV για την εξαγωγή αναπαραστάσεων των προκαθορισμένων συμπτωμάτων.



Σχήμα 3.14: Αρχιτεκτονική προσέγγισης PV-SVM [17]

Για την δημιουργία των Paragraph Vectors θα εκπαιδεύσουμε ένα μοντέλο πάνω σε πραγματικά δεδομένα που εκφράζουν συναισθηματικές καταστάσεις που έχουν εξαχθεί από αναρτήσεις ανθρώπων στα κοινωνικά μέσα δικτύωσης.

3.3.1.2.1 Δημιουργία μοντέλων Doc2Vec ή PV

- ***Βήμα 1ο: Συλλογή Δεδομένων***

Για την δημιουργία του μοντέλου Doc2Vec θα αξιοποιήσουμε τέσσερα (4) διαφορετικά αποθετήρια δεδομένων, καθώς δυστυχώς δεν υπάρχουν διαθέσιμα ελεύθερα για χρήση αποθετήρια δεδομένων πάνω στην ταξινόμηση της κατάθλιψης. Έτσι λοιπόν επιλέγουμε τα παρακάτω datasets, σε πλήθος 4 για λόγους ποικιλίας αλλά και αφθονίας δεδομένων εκπαίδευσης.

1. Το πρώτο dataset depressive_tweets_processed.csv [63] αποτελείται από tweets που έχουν επιλεχθεί τυχαία από το twitter_sentiment.csv [64], στο σύνολο 2313 από τα οποία τα μισά δεν παρουσιάζουν ενδείξεις κατάθλιψης του χρήστη, ενώ τα άλλα μισά δείχνουν υπόνοιες κατάθλιψης του χρήστη.

	Text
0	The lack of this understanding is a small but ...
1	i just told my parents about my depression and...
2	depression is something i don't speak about ev...
3	Made myself a tortilla filled with pb&j. My de...
4	@WorldofOutlaws I am gonna need depression med...

2. Το δεύτερο dataset amazon_alexa.csv [65] αποτελείται από κριτικές χρηστών, στο σύνολο 3150, πάνω σε προϊόντα του Amazon Alexa εκφράζοντας μια πληθώρα από συναισθήματα.

	Text
0	Love my Echo!
1	Loved it!
2	Sometimes while playing a game, you can answer...
3	I have had a lot of fun with this thing. My 4 ...
4	Music

3. Το τρίτο dataset twitter_sentiment.csv [66] αποτελείται από τυχαία tweets στο σύνολο 1578612 τα οποία καλύπτουν ένα μεγάλο εύρος συναισθημάτων και όχι επικεντρωμένων γύρω από την κατάθλιψη, όπως το πρώτο dataset.

	Text
0	is so sad for my APL frie...
1	I missed the New Moon trail...
2	omg its already 7:30 :O
3	.. Omgaga. Im sooo im gunna CRy. I'...
4	i think mi bf is cheating on me!!! ...

4. Το τέταρτο dataset imdb_sentiments.csv [67] αποτελείται από κριτικές χρηστών του imdb, στο σύνολο 25000, πάνω σε ταινίες, οι οποίες έχουν εξαχθεί από το Stanford AI Repository.

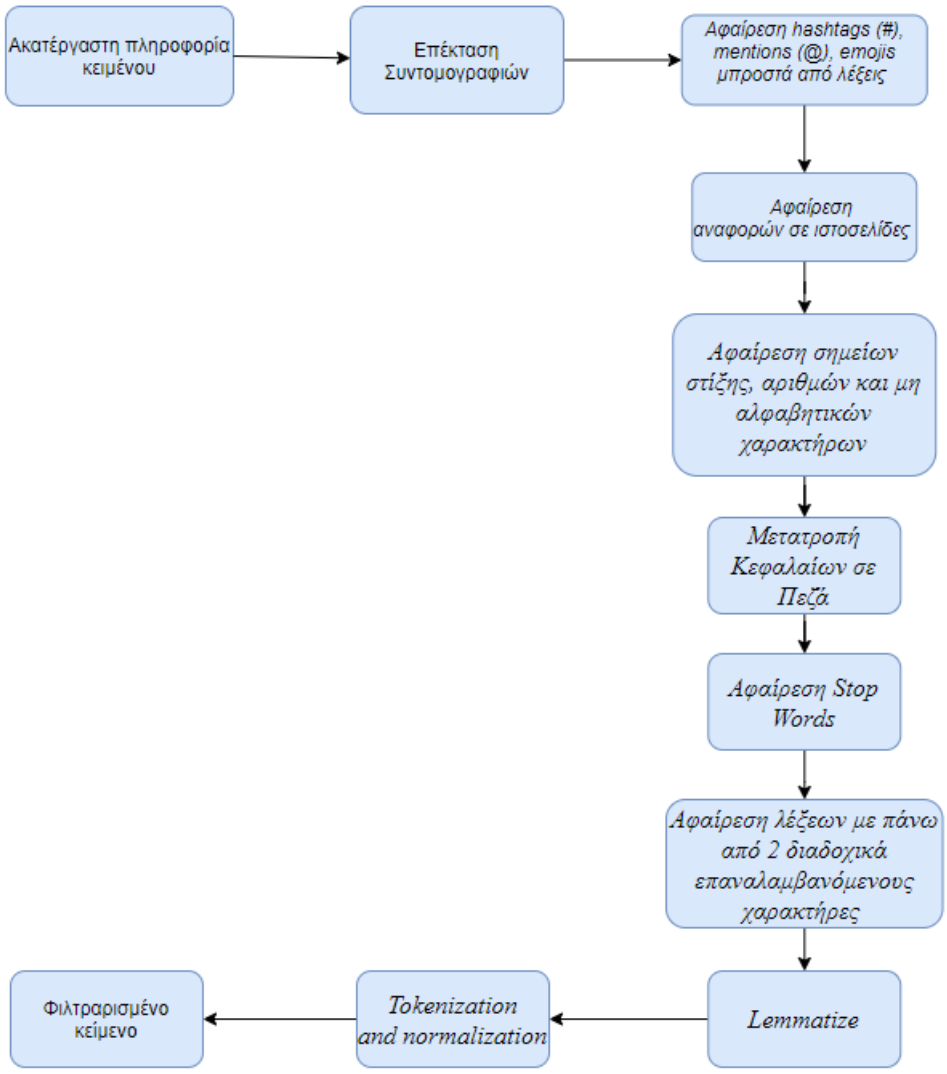
	Text
0	For a movie that gets no respect there sure ar...
1	Bizarre horror movie filled with famous faces ...
2	A solid, if unremarkable film. Matthau, as Ein...
3	It's a strange feeling to sit alone in a theat...
4	You probably all already know this by now, but...

Τέλος συγκεντρώνουμε όλα τα δεδομένα και επιλέγουμε τυχαία 1/3 αυτών να αποτελεί το τελικό μας dataset. Αφού λοιπόν γίνει η επιλογή των δειγμάτων καταλήγουμε στο τελικό dataset μεγέθους 482722 δειγμάτων.

	Text
201574	@inkscar I could cook you something. You might...
923953	if god wanted me to be awake this early.... he...
1211561	sorry for last night. whatever it is, i didn't...
517543	800 followers and counting (@ecostin eÅŸti pre...
1370403	Am slacking at work. Need to get motivated to...
23470	@11394607 I'm not yet in Manila. I'll be movin...
1072542	Ok got my money straight yesssss!! Now if I co...
368313	@paper_chaserz Get 100 followers a day using w...
817759	hello everyone i just looked at some vid @smo...
1342042	Will go to the media markt now, wish you a won...
266092	@koshzor yessshhh, I already want this game SO...
1164870	Poor cubby...he got called in today to work at...
409558	@samarowais samarrrrr thank you so much for th...
793215	@filmutopia wow sounds great
953776	just got back from jeff's. showering and then ...
962842	is back from holiday
1139450	On the way to my sweet sweet bed
901841	I still don't have a voice but am feeling much...
1119732	Rise and shine my twitters! Going to feed the ...
1341621	Why the hell r we having Down-Time? We've had ...

- ***Βήμα 2ο: Προεπεξεργασία Δεδομένων***

Πριν την κατασκευή των Paragraph Vectors και γενικότερα οποιονδήποτε feature vectors, συνήθως εφαρμόζουμε κάποιες τεχνικές προεπεξεργασίας των δεδομένων. Ο σκοπός είναι η κανονικοποίηση της εισόδου για τη μείωση της αναπαράστασης με ταυτόχρονη απόρριψη της πλεονάζουσας-άχρηστης πληροφορίας που καλείται θόρυβος. Έτσι, μειώνεται η διάσταση του προβλήματος και ο κίνδυνος του overfitting και ταυτόχρονα μειώνεται και ο απαιτούμενος χρόνος ταξινόμησης/παλινδρόμησης. Ο αλγόριθμος προεπεξεργασίας των δεδομένων που εφαρμόσαμε φαίνεται στο ακόλουθο διάγραμμα ροής.



Σχήμα 3.15: Διάγραμμα ροής αλγορίθμου προεπεξεργασίας κειμένου

Τα επιμέρους βήματα, τα οποία υλοποιήθηκαν μέσω Regular Expressions [68], περιγράφονται ακολούθως:

1. Επέκταση Συντομογραφιών

Οι Συντομογραφίες στις οποίες αναφερόμαστε είναι αυτές κατά τις οποίες κόβουμε κομμάτια των λέξεων και τις ενώνουμε με τις γειτονικές προφέροντάς τες σαν μια ενιαία

λέξη. Π.χ. αντί για *'Do not'* προτιμούμε το *'Don't'*. Έτσι λοιπόν δημιουργούμε ένα λεξικό που αντιστοιχίζει τέτοιες συντομογραφίες, στην πρωτότυπή τους μορφή και μέσω μιας συνάρτησης αντιστοίχισης αντικαθιστούμε σε όλο το dataset τις συντομογραφίες αυτές. Σκοπός της διαδικασίας αυτής είναι να αποτρέψουμε το μοντέλο μας να θεωρήσει σαν καινούριες λέξεις τις συντομογραφίες, επομένως να μειώσουμε το μέγεθος του τελικού λεξιλογίου μας αλλά και να διατηρήσουμε πληροφορία, όπως αυτή που δίνει η άρνηση *not*, που διαφορετικά θα χανόταν αφού π.χ. η συντομογραφία *'Don't'* δεν θα αντιπροσώπευε άρνηση αλλά μια καινούρια λέξη *'Don't'*. Στο Σχήμα 3.16 παρουσιάζουμε τις συντομογραφίες που εντοπίστηκαν στο dataset μας.

```
dict_keys(["ma'am", "he'd've", "I'd", "you're", "we're", "it'd", "y'all're", "m
ightn't", "wouldn't've", "couldn't", "it'll", "to've", "that'd", "shan't've",
"it'll've", "it'd've", "haven't", "will've", "what'll've", "I'll", "shan't", "c
an't've", "they'll", "'cause", "I've", "there'd", "where'd", "she'll", "he'll",
"he's", 'wouldve', "y'all'd've", "who've", "why've", "would've", "we'll", "ther
e's", "mayn't", "so've", "couldn't've", "must've", "should've", "shouldn't've",
"what's", "can't", "who'll've", "needn't've", "aren't", "didn't", "when's", "wh
ere's", "mightn't've", "isn't", "there'd've", "we'd", "could've", "sha'n't", "s
he's", "y'all've", "he'll've", "so's", "it's", "they've", "let's", "you'd",
"I'd've", "y'alls", "she'll've", "they'll've", "why's", "you'll've", "needn't",
"oughtn't've", "I'm", "wouldn't", "they'd've", "wasn't", "what've", "how'll",
"hasn't", "how'd'y", "he'd", "doesn't", "won't've", "she'd've", "they're", "ho
w's", "shouldn't", "weren't", "won't", "where've", "she'd", "who'll", "you've",
"y'all'd", "what're", "mustn't", "mustn't've", "don't", "that's", "who's", "ho
w'd", "ain't", "hadn't've", "they'd", "you'd've", "we've", "oughtn't", "tha
t'd've", "I'll've", "we'd've", "when've", "o'clock", "what'll", "y'all", "you'l
l", "we'll've", "hadn't", "might've"])
```

Σχήμα 3.16: Συντομογραφίες του τελικού dataset από τα tweets

2. Αφαίρεση hashtags (#), mentions (@), emojis μπροστά από λέξεις

Όταν αντιμετωπίζουμε τόσο κριτικές, όσο και εκφράσεις απόψεων στα μέσα κοινωνικής δικτύωσης πολλές φορές αυτά συνοδεύονται από αναφορές σε άλλους χρήστες με το πρόθεμα '@', αναφορές σε σχετικά θέματα με το πρόθεμα '#' αλλα και από συνδυασμούς σημείων στίξης που εκφράζουν τη διάθεση του χρήστη εκείνη τη στιγμη (γνωστά ως 'emojis'). Αυτή η πληροφορία λοιπόν είναι περιττή στα πλαίσια της μελέτης μας, γι αυτό και την αφαιρούμε από τα κείμενα μας.

3. Αφαίρεση αναφορών σε ιστοσελίδες

Ομοίως με παραπάνω παρατηρούμε συχνές αναφορές την σχολίων σε ιστοσελίδες, πληροφορία που μας είναι περιττή στη συγκεκριμένη περίπτωση. Έτσι όσες συμβολοσειρές ξεκινούν με τον συμβολισμό 'http' τις αφαιρούμε.

4. Αφαίρεση σημείων στίξης, αριθμών και μη αλφαβητικών χαρακτήρων

Όπως τα κεφαλαία γράμματα, έτσι και τα σημεία στίξης και οι αριθμοί δεν επιδρούν άμεσα την πολικότητα του συναισθήματος . Μερικές φορές απλώς, η χρήση τους (πχ. του «!») μπορεί να ενισχύσει το συναίσθημα. Έτσι, μπορούν να αγνοηθούν για τα πλαίσια του προβλήματος μας.

5. Μετατροπή Κεφαλαίων σε Πεζά

Μετατρέπουμε όλα τα κεφαλαία γράμματα σε πεζά διότι η διαφοροποίηση πχ. των λέξεων Great και great δεν παίζει κάποιο ρόλο στο πρόβλημα του sentiment analysis, μιας

και δεν αλλάζει η πολικότητα των λέξεων παρά μόνο ίσως η έντασή της (με τη χρήση κεφαλαίων).

6. Αφαίρεση Stop Words

Τα Stop Words είναι λέξεις ασήμαντες για την κατηγοριοποίηση κειμένων που δεν προσθέτουν καμία πληροφορία στην πρόταση που ανήκουν, αλλά και λέξεις που εμφανίζονται σε μεγάλη συχνότητα. Μπορεί να είναι λέξεις συνδετικές, άρθρα, αντωνυμίες, καθώς επίσης και λέξεις που θεωρούμε εμείς ασήμαντες και σπάνιες για τα δεδομένα μας. Στη συγκεκριμένη περίπτωση εμείς προσθέσαμε λέξεις μηδενικού περιεχομένου αλλά και λέξεις που προέκυψαν από τον διαχωρισμό κάποιων σημείων στίξης που δεν στέκουν μόνες τους στο λόγο.

7. Αφαίρεση λέξεων με πάνω από 2 διαδοχικά επαναλαμβανόμενους χαρακτήρες

Εκτός από τα emojis που προαναφέραμε, οι χρήστες πολλές φορές εκφράζονται και μέσω επιφωνημάτων της μορφής ‘awww’,’xxxx’ κ.λπ. Ωστόσο γνωρίζοντας πως υπάρχουν πολύ συχνά λέξεις που η σύνταξή τους περιλαμβάνει δύο ίδιους διαδοχικούς χαρακτήρες, αποφασίζουμε να αφαιρέσουμε τις λέξεις που εμφανίζουν πάνω από 2 διαδοχικούς όμοιους χαρακτήρες, καθώς στην πλειοψηφία τους αυτές οι λέξεις θα αντιπροσωπεύουν επιφωνήματα.

8. Lemmatize

Στη συνέχεια πρέπει να κανονικοποιήσουμε τη μορφή των λέξεων εφαρμόζοντας stemming (πχ με τον αλγόριθμο Porter) ή λημματοποίηση (lemmatization). Και οι δύο τεχνικές αποσκοπούν στην αποκοπή των μορφολογικών καταλήξεων των λέξεων. Η διαφορά τους είναι ότι η τεχνική του stemming εφαρμόζει μία πιο ακατέργαστη (crude) αποκοπή, συχνά δεν δίνει ως αποτέλεσμα έγκυρες λέξεις και εξαρτάται μόνο από τη λέξη αγνοώντας το context, ενώ η τεχνική του lemmatization από την άλλη επιστρέφει το λήμμα μίας λέξης, δηλαδή μία έγκυρη λέξη και εξαρτάται από το μέρος του λόγου (POS tag) της λέξης. Εμείς επιλέξαμε την λημματοποίηση κυρίως για να αποφύγουμε τις μη έγκυρες λέξεις. Για την λημματοποίηση χρησιμοποιήσαμε τον WordNet Lemmatizer από το NLTK ο οποίος λαμβάνει υπόψη την POS ετικέτα των λέξεων.

9. Tokenization and normalization

Τέλος, εφαρμόζουμε την διαδικασία του tokenization, δηλαδη τον διαχωρισμό των κειμένων σε tokens-λέξεις πάνω στα οποία θα εκπαιδευτεί ο αλγόριθμος δημιουργίας διανυσμάτων λέξεων.

Στο Σχήμα 3.17 παρουσιάζουμε ένα παράδειγμα του αρχικού κειμένου και του τελικού έπειτα από την παραπάνω προεπεξεργασία.

```
[["@inkscar I could cook you something. You might not recover from that. I've b
een told that my grilled fish is &quot;to die for&quot; and several do. "
  list(['could', 'cook', 'something', 'might', 'not', 'recover', 'told', 'grill
ed', 'fish', 'die', 'several'])]
 ['if god wanted me to be awake this early.... he wouldve made it bright outsid
e '
  list(['god', 'wanted', 'awake', 'early', 'would', 'made', 'bright', 'outsid
e'])]
 ["sorry for last night. whatever it is, i didn't mean to. don't remember a thi
ng so yea. love love, have a great weekend "
  list(['sorry', 'last', 'night', 'whatever', 'not', 'mean', 'not', 'remember',
'thing', 'yea', 'love', 'love', 'great', 'weekend'])]
 ['800 followers and counting (@ecostin eÅŸti pregÄƒtit sÄƒ rÄƒmÃ¢i Ã®n spate?)
'
  list(['follower', 'counting', 'preg', 'tit', 'spate'])]
 ['Am slacking at work.  Need to get motivated to get something constructive do
ne around the office.'
  list(['slacking', 'work', 'need', 'get', 'motivated', 'get', 'something', 'co
nstructive', 'done', 'around', 'office'])]
 ["@11394607 I'm not yet in Manila. I'll be moving in my new apartment by next
week, probably on Monday. "
  list(['not', 'yet', 'manila', 'moving', 'new', 'apartment', 'next', 'week',
'probably', 'monday'])]
 ['Ok got my money straight yesssss!! Now if I could just find my check book..
Got money just to give it away '
  list(['got', 'money', 'straight', 'could', 'find', 'check', 'book', 'got', 'm
oney', 'give', 'away'])]
 ['@paper_chaserz Get 100 followers a day using www.tweeteradder.com Once you a
dd everyone you are on the train or pay vip '
  list(['get', 'follower', 'day', 'using', 'tweeteradder', 'com', 'add', 'every
one', 'train', 'pay', 'vip'])]
 ['hello everyone  i just looked at some vid @smoshian looked at and it looks l
ike fun  http://bit.ly/UNcLY  xD LOL'
  list(['hello', 'everyone', 'looked', 'vid', 'looked', 'look', 'like', 'fun',
'lol'])]
 ["Will go to the media markt now, wish you a wonderful stunning day. Here it b
egins to thunder and I'm afraid of thunderstorms "
  list(['medium', 'markt', 'wish', 'wonderful', 'stunning', 'day', 'begin', 'th
under', 'afraid', 'thunderstorm'])]]
```

Σχήμα 3.17: Κείμενο πριν και μετά την προεπεξεργασία

Αφού λοιπόν μετατρέψαμε το κείμενο κάθε δείγματος σε μια λίστα από τις λέξεις με την μεγαλύτερη πληροφορία που το αποτελούν, στη συνέχεια θα οπτικοποιήσουμε στο Σχήμα 3.18 μέσω της συνάρτησης WordCloud ένα ποσοστό των 121586 μοναδικών λέξεων που αποτελούν το λεξικό μας δίνοντας βαρύτητα στις κυρίαρχες λέξεις, αυτές δηλαδή που εμφανίζονται πιο συχνά.

Σχήμα 3.18: Οπτικοποίηση με βαρύτητες του λεξικού που δημιουργήσαμε

- ***Βήμα 3ο: Εφαρμογή αλγορίθμου PV-DM***

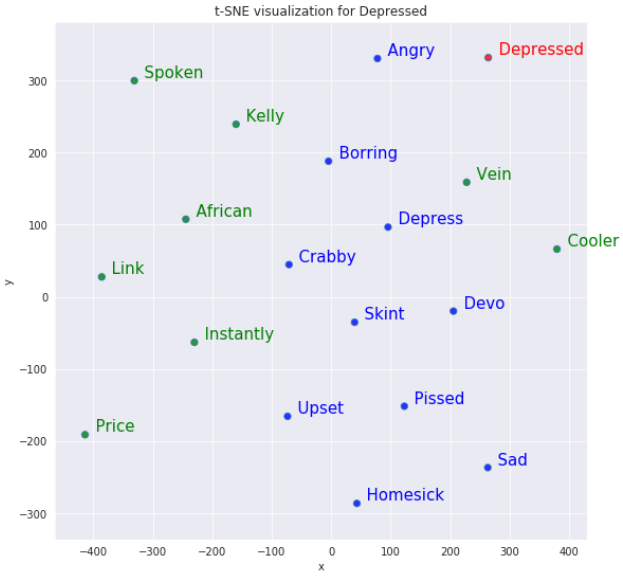
Αφού λοιπόν προηγήθηκε η διαδικασία της προεπεξεργασίας του κειμένου, τροφοδοτούμε τα δεδομένα μας στον αλγόριθμο μάθησης Paragraph Vector-Distributed Memory,απ' όπου προκύπτει και το μοντέλο "doc2vec.model" , από το οποίο πλέον μπορούμε να πάρουμε διανυσματικές αναπαραστάσεις μήκους=150 (αριθμός που επιλέχθηκε έπειτα από δοκιμές), τόσο μεμονωμένων λέξεων, όσο και προτάσεων. Θέλοντας να ελέγξουμε την αποδοτικότητα του μοντέλου μας, του τροφοδοτούμε την λέξη "*depressed*" ζητώντας να μας επιστρέψει τις πιο όμοιες λέξεις που διαθέτει στο λεξικό του με την λέξη αυτή. Το αποτελεσμα που παρουσιάζουμε στη συνέχεια μαρτυρά μια αξιοπρεπή λειτουργία του μοντέλου μας.

```
[('anxious', 0.35393786430358887),
 ('sluggish', 0.3510551452636719),
 ('sad', 0.33833229541778564),
 ('upset', 0.32091450691223145),
 ('nauseous', 0.31342846155166626),
 ('peeved', 0.3071664571762085),
 ('loopy', 0.3055908679962158),
 ('pissed', 0.30393052101135254),
 ('stressed', 0.3036532998085022),
 ('homesick', 0.29522132873535156)]
```

3.3.1.2.2 Οπτικοποίηση απόδοσης του μοντέλου Doc2Vec

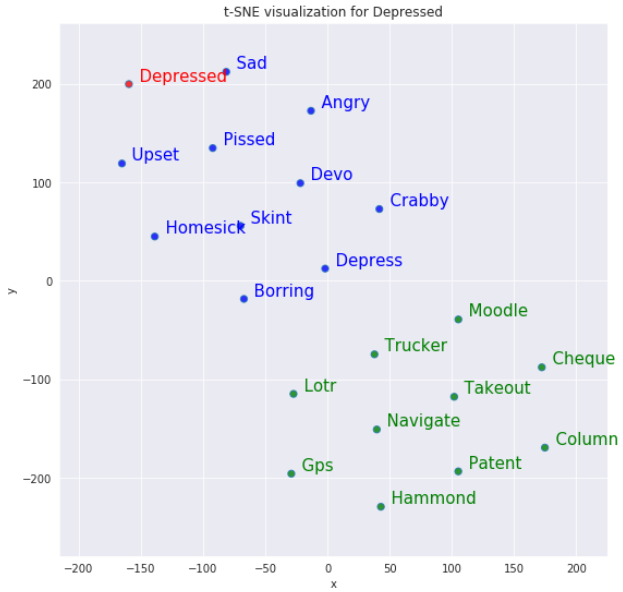
Για την οπτικοποίηση της λειτουργικότητας του μοντέλου μας, μέσω του αλγορίθμου t-SNE [69] ο οποίος κάνει μη γραμμική μείωση των διαστάσεων των δεδομένων εισόδου του, θα προβάλουμε διανύσματα λέξεων υψηλής διάστασης σε χώρο δυο (2) διαστάσεων, με στόχο την, κατανοητή στο ανθρώπινο μάτι, παρατήρηση ενδιαφερόντων και με σημαντική πληροφορία μοτίβων:

Αρχικά θα παρουσιάσουμε στο ίδιο σύστημα συντεταγμένων τη σχέση της λέξης “depressed” με τις 8 πιο όμοιες της, αλλά και με 8 τυχαίες λέξεις που επιλέξαμε εμείς.



Σχήμα 3.19: Οπτικοποίηση σχέσεων όμοιων λέξεων και τυχαίων του μοντέλου Doc2Vec

Έπειτα θα παρουσιάσουμε στο ίδιο σύστημα συντεταγμένων τη σχέση της λέξης “depressed” με τις 8 πιο όμοιες της, αλλά και με τις 8 πιο αντίθετες ορολογικά λέξεις που ανήκουν στο λεξικό του μοντέλου.



Σχήμα 3.20: Οπτικοποίηση σχέσεων όμοιων λέξεων και πλήρως αντίθετων ορολογικά λέξεων του μοντέλου Doc2Vec

Η απόδοση του μοντέλου αυτού είναι αρκετά αξιοπρεπής δεδομένου των όχι τόσο αντιπροσωπευτικών δειγμάτων πάνω στα οποία εκπαιδεύτηκε, ωστόσο έχει αρκετές αδυναμίες οι οποίες θα φανούν στην συνέχεια που θα εφαρμόσουμε τις διανυσματικές αναπαραστάσεις του μοντέλου αυτού στα δείγματα προς αξιολόγηση της κατάθλιψης.

3.3.1.2.3 Εφαρμογή PV-SVM στις προτεινόμενες προτάσεις

Αφού δημιουργήσαμε το Doc2Vec μοντέλο από το οποίο θα εξάγουμε τις διανυσματικές αναπαραστάσεις των προτάσεων των ασθενών πάνω στα προκαθορισμένα συμπτώματα, θα περάσουμε τα δεδομένα από ολα τα δείγματα για κάθε σύμπτωμα από κατάλληλα διαμορφωμένα δίκτυα SVM για την παρουσία ή μη αυτών. Καθώς πραγματευόμαστε με αλγόριθμο επιβλεπόμενης μάθησης, θα χρησιμοποιήσουμε σαν ετικέτες την συμπτωμάτων τις τιμές που εξήγαμε από την Context Analysis στην ενότητα 3.3.1.1.

1.Πρώην διάγνωση της Κατάθλιψης

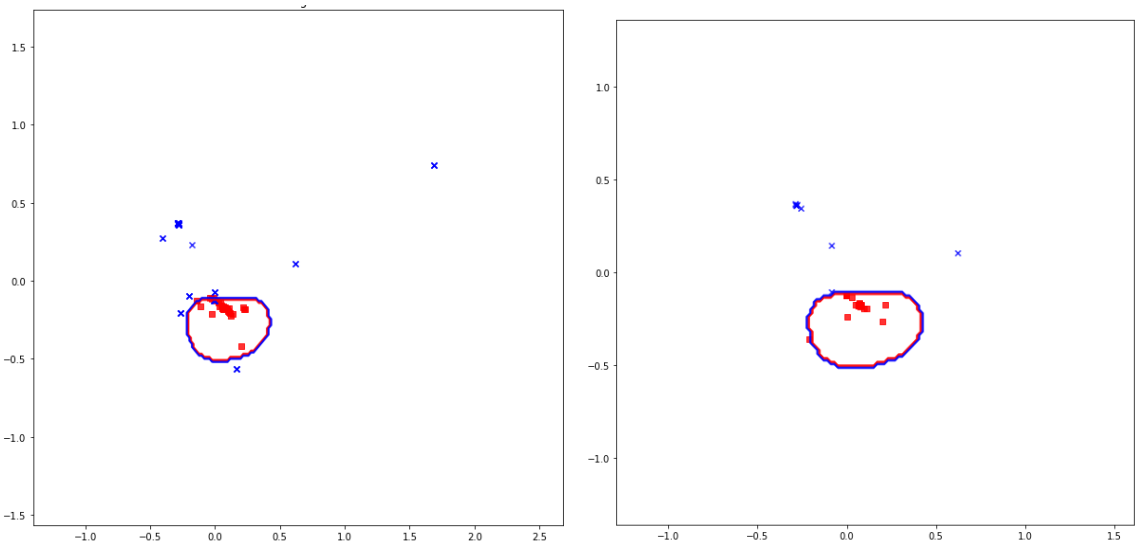
Θα εφαρμόσουμε τον SVM ταξινομητή με πυρήνα rbf και παραμέτρους γ=10, C=10 οι οποίες προέκυψαν ως οι βέλτιστες από την μέθοδο grid search.

```
The best parameters are {'gamma': 10.0, 'C': 10.0} with a score of 0.93
Accuracy:96.97%
[[23  1]
 [ 0  9]]
              precision    recall  f1-score   support

           0       1.00      0.96      0.98        24
           1       0.90      1.00      0.95         9

    accuracy                           0.97        33
   macro avg       0.95      0.98      0.96        33
weighted avg       0.97      0.97      0.97        33

[0 0 0 0 0 0 1 1 0 0 1 1 0 0 0 0 0 0 1 1 0 0 0 1 0 0 0 0 1 1 0 0 0]
[0 1 0 0 0 0 1 1 0 0 1 1 0 0 0 0 0 0 1 1 0 0 0 1 0 0 0 0 1 1 0 0 0]
```



Σχήμα 3.21: Διάγραμμα ταξινόμησης δεδομένων από PV-SVM μοντέλο (αριστερά στα δεδομένα εκπαίδευσης και δεξιά στα δεδομένα επαλήθευσης)

2. Πρώην μετα-τραυματικό στρες (PTSD)

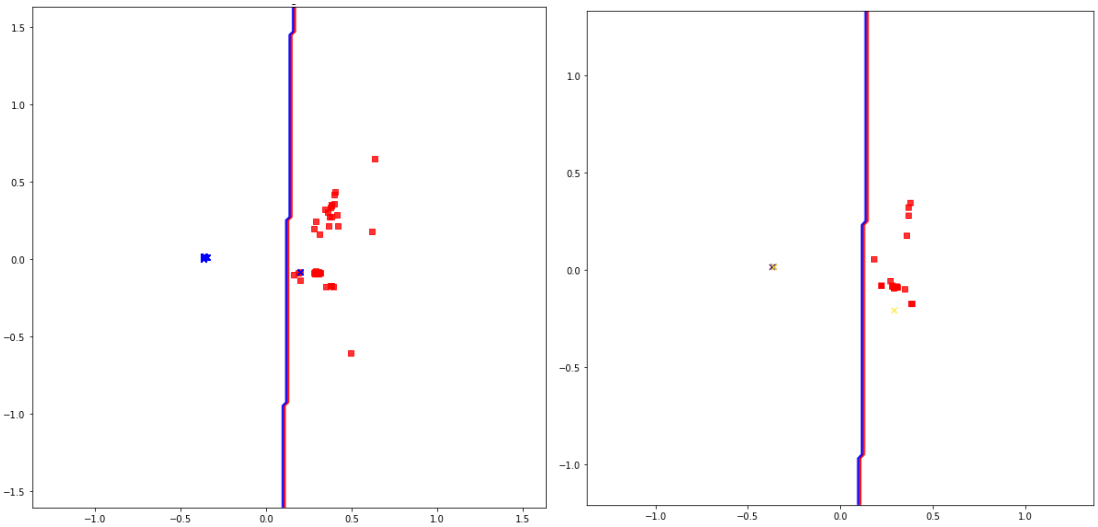
Θα εφαρμόσουμε τον SVM ταξινομητή με πυρήνα rbf και παραμέτρους γ=0.001, C=0.001 οι οποίες προέκυψαν ως οι βέλτιστες από την μέθοδο grid search.

```
The best parameters are {'gamma': 0.001, 'C': 0.001} with a score of 0.97
Accuracy:96.97%
[[29  0]
 [ 1  3]]
              precision    recall  f1-score   support

           0       0.97      1.00      0.98        29
           1       1.00      0.75      0.86         4

    accuracy                           0.97        33
   macro avg       0.98      0.88      0.92        33
weighted avg       0.97      0.97      0.97        33

[0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0]
[0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0]
```



Σχήμα 3.22: Διάγραμμα ταξινόμησης δεδομένων από PV-SVM μοντέλο (αριστερά στα δεδομένα εκπαίδευσης και δεξιά στα δεδομένα επαλήθευσης)

3. Διαταραχή Ύπνου

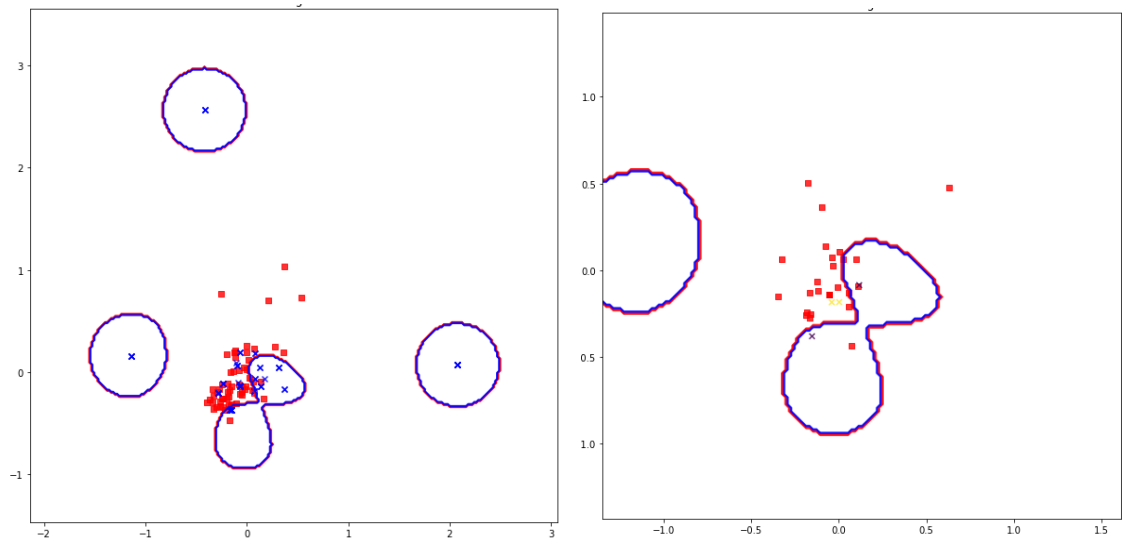
Θα εφαρμόσουμε τον SVM ταξινομητή με πυρήνα rbf και παραμέτρους γ=10, C=10 οι οποίες προέκυψαν ως οι βέλτιστες από την μέθοδο grid search.

```
The best parameters are {'gamma': 10.0, 'C': 10.0} with a score of 0.66
Accuracy:84.85%
[[26  3]
 [ 2  2]]
              precision    recall  f1-score   support

           0       0.93      0.90      0.91        29
           1       0.40      0.50      0.44         4

    accuracy                           0.85        33
   macro avg       0.66      0.70      0.68        33
weighted avg       0.86      0.85      0.86        33

[0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0]
```



Σχήμα 3.23: Διάγραμμα ταξινόμησης δεδομένων από PV-SVM μοντέλο (αριστερά στα δεδομένα εκπαίδευσης και δεξιά στα δεδομένα επαλήθευσης)

4. Αισθήματα

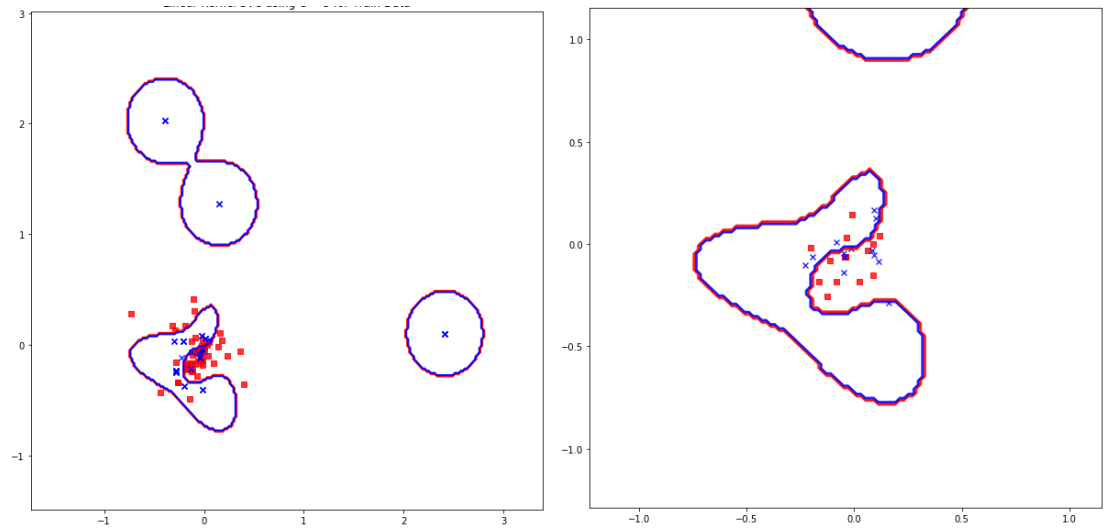
Θα εφαρμόσουμε τον SVM ταξινομητή με πυρήνα rbf και παραμέτρους γ=10, C=10 οι οποίες προέκυψαν ως οι βέλτιστες από την μέθοδο grid search.

```
The best parameters are {'gamma': 10.0, 'C': 10.0} with a score of 0.72
Accuracy:69.70%
[[17  3]
 [ 7  6]]
              precision    recall  f1-score   support

           0       0.71      0.85      0.77        20
           1       0.67      0.46      0.55        13

    accuracy                           0.70        33
   macro avg       0.69      0.66      0.66        33
weighted avg       0.69      0.70      0.68        33

[0 0 0 0 0 1 1 1 0 1 1 1 1 1 0 0 0 0 1 0 1 0 1 1 0 0 1 0 0 0 0 0 0]
[0 0 0 0 0 1 1 1 0 0 0 0 1 0 0 0 1 0 0 1 1 1 0 0 0 0 1 0 0 0 0 0 0]
```



Σχήμα 3.24: Διάγραμμα ταξινόμησης δεδομένων από PV-SVM μοντέλο (αριστερά στα δεδομένα εκπαίδευσης και δεξιά στα δεδομένα επαλήθευσης)

5. Προσωπικότητα

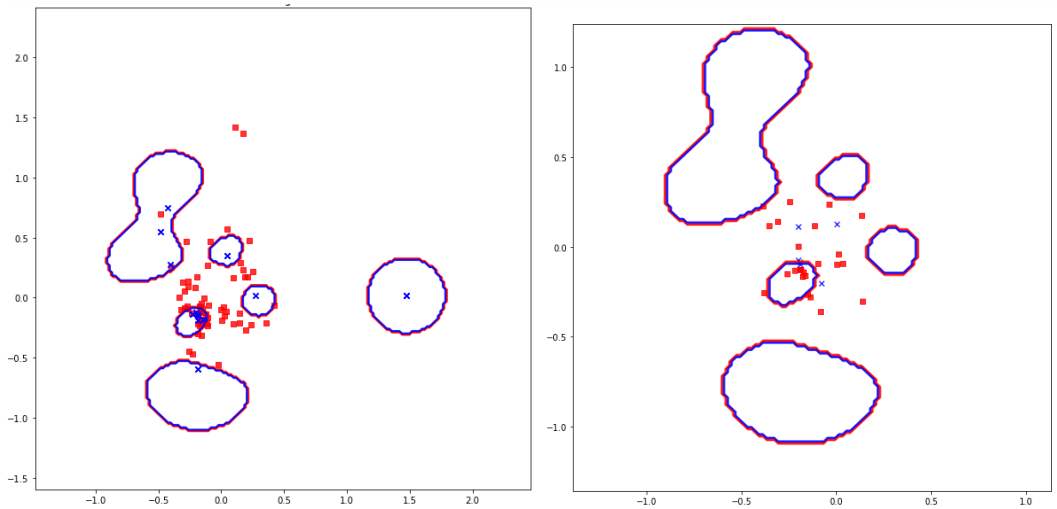
Θα εφαρμόσουμε τον SVM ταξινομητή με πυρήνα rbf και παραμέτρους γ=10, C=10 οι οποίες προέκυψαν ως οι βέλτιστες από την μέθοδο grid search.

```
The best parameters are {'gamma': 10.0, 'C': 10.0} with a score of 0.78
Accuracy:54.55%
[[17 10]
 [ 5  1]]
              precision    recall  f1-score   support

           0       0.77      0.63      0.69        27
           1       0.09      0.17      0.12         6

    accuracy                           0.55        33
   macro avg       0.43      0.40      0.41        33
weighted avg       0.65      0.55      0.59        33

[0 0 0 0 0 0 1 1 0 0 1 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0]
[0 1 0 1 1 0 0 0 0 0 1 1 1 0 0 1 0 0 1 1 0 0 0 0 0 1 0 0 0 0 0 1 0]
```



Σχήμα 3.25: Διάγραμμα ταξινόμησης δεδομένων από PV-SVM μοντέλο (αριστερά στα δεδομένα εκπαίδευσης και δεξιά στα δεδομένα επαλήθευσης)

Τέλος παρουσιάζουμε τα ποσοστά ακριβείας των επιμέρους μοντέλων PV-SVM για κάθε σύμπτωμα. Όπως ήταν αναμενόμενο στα συμπτώματα που αναφέρονται στα συναισθήματα και την προσωπικότητα αδυνατούν τα αντίστοιχα μοντέλα να μάθουν τα σωστά μοτίβα, λόγω της εμμεσότητας των απαντήσεων που δίνουν οι ασθενείς, της όχι τόσο αξιόπιστης καταγραφής των ετικετών από την σημασιολογική ανάλυση αλλά και του ελλιπούς πλήθους δειγμάτων εκπαίδευσης.

	Accuracy
PTSD	96.969697
Depression Diagnosis	96.969697
Sleep	84.848485
Feelings	69.696970
Personality	54.545455

Πίνακας 3.8: Ακρίβεια των επιμέρους μοντέλων PV-SVM

4. Εκπαίδευση Συστημάτων

4.1 Διαδικασία Εκπαίδευσης Μοντέλων

Μια ολοκληρωμένη διαδικασία pipeline ML αποτελείται τόσο απο την αρχιτεκτονική του, δηλαδή τον συνδυασμό μετασχηματιστών και την επιλογή του τελικού εκτιμητή (το pipeline), όσο και από τις (βέλτιστες) τιμές των υπερπαραμέτρων όλων των προηγουμένων που προκύπτουν είτε με χειροκίνητη μελέτη είτε με την μέθοδο αναζήτησης πλέγματος (grid search). Έτσι λοιπόν η γενική διαδικασία εκπαίδευσης ήταν η ίδια για όλα τα μοντέλα, αλλάζοντας απλά τις διαδικασίες προεπεξεργασίας ανάλογα με το είδος της πληροφορίας που είχαμε και οι υπερπαράμετροι ανάλογα με τον τύπο του προβλήματος (κατηγοριοποίηση ή παλινδρόμηση).

Προτού όμως πειραματιστούμε σε οποιαδήποτε αρχιτεκτονική, θα τροφοδοτήσουμε τα διανύσματα χαρακτηριστικών που εξήγαμε στην ενότητα 3.2 σε έναν Baseline εκτιμητή και συγκεκριμένα τον Random Forest Regressor για να έχουμε ένα μέτρο σύγκρισης της απόδοσης των μοντέλων που θα μελετήσουμε με έναν γενικευμένο συμβατικό εκτιμητή απλής αρχιτεκτονικής.

Για κάθε ένα από τα συστήματα ανάλυσης Ήχου, Εκφράσεων του προσώπου και Κειμένου θα μελετήσουμε Συνελικτικές Βαθιές αρχιτεκτονικές των οποίων τα πλεονεκτήματα αναλύσαμε στην ενότητα 2.5. Η διαδικασία εκπαίδευσης λοιπόν κάθε συστήματος περιγράφεται από τα εξής βήματα:

- ✓ **Αρχικοποίηση Μοντέλου**

Στο στάδιο αυτό αρχικά ορίζεται η αρχιτεκτονική του νευρωνικού δικτύου, δηλαδή στρώματα, αρχικοποιήσεις βαρών. Όλες οι αρχιτεκτονικές που εξετάστηκαν περιεγράφηκαν στο Θεωρητικό Υπόβαθρο. Στα μοντέλα παλινδρόμησης χρησιμοποιήθηκε ως συνάρτηση σφάλματος (loss function) αυτή της ρίζας του μέσου τετραγωνικού σφάλματος (Root Mean Squared Error) ενώ σε κάθε βήμα εκπαίδευσης υπολογίζονταν επιπροσθέτως το μέσο απόλυτο σφάλμα (Mean Absolute Error) και το μέσο τετραγωνικό σφάλμα (Mean Squared Error) για να παρακολουθούμε την πορεία του συστήματος ανά εποχή. Στα μοντέλα κατηγοριοποίησης χρησιμοποιήθηκε η κατηγορική εντροπία (categorical cross entropy) ως συνάρτηση σφάλματος ενώ επίσης υπολογίζονταν η κατηγορική ακρίβεια (categorical accuracy), ένας άλλος τύπος ακρίβειας (precision), η ανάκληση (recall) και το f1 score. Ως συνάρτηση βελτιστοποίησης (optimizer) της εκπαίδευσης όλων των μοντέλων χρησιμοποιήθηκε ο ‘Adam’.

- ✓ **Πέρασμα Δεδομένων στο Μοντέλο**

Έχοντας εξάγει στην ενότητα 3 τις διανυσματικές αναπαραστάσεις των τριών ειδών πληροφορίας που διαθέτουμε, είμαστε έτοιμοι να την τροφοδοτήσουμε στα συστήματα μας χωρίς περαιτέρω επεξεργασία. Ωστόσο βασισμένοι στη μελέτη του [17], για να αυξήσουμε την απόδοση των μοντέλων μας κάναμε διαχωρισμό των ανδρών δειγμάτων και των

γυναικών, εκπαιδεύοντας για το κάθε φύλο διαφορετικά μοντέλα με αμυδρά διαφορετικές υπερπαραμέτρους προς όφελος του καθενός χωριστά. Μελετώντας τα χαρακτηριστικά μας παρατηρήσαμε πως οι άνδρες έκαναν λιγότερες κινήσεις τόσο του κεφαλιού, όσο και των ματιών τους από τις γυναίκες, ενώ στο κομμάτι του ήχου, πολλά ακουστικά χαρακτηριστικά που προαναφέραμε παρουσίαζαν τιμές εντός διαφορετικού εύρους μεταξύ ανδρών και γυναικών. Η διαφορά στα ακουστικά χαρακτηριστικά γίνεται εύκολα αντιληπτή από τον άνθρωπο από το γεγονός πως η φωνή της γυναίκας πιάνει υψηλότερες συχνότητες από του άνδρα σε ένα απλό άκουσμα της, ενώ του άνδρα συνηθίζει να είναι πιο βαριά. Η τροφοδότηση των χαρακτηριστικών αυτών άνδρα και γυναίκας μαζί στο ίδιο μοντέλο ML προκαλούσε πολλά σφάλματα στο σύστημα καθώς είναι δύσκολη η κανονικοποίηση τους στην συγκεκριμένη περίπτωση αφού αναγκάζει τις τιμές τους να συσσωρευτούν μεταξύ μικρών διαστημάτων, εμποδίζοντας την εμφάνιση μοτίβων που οφείλονται σε έντονες διακυμάνσεις των τιμών. Επομένως όλα τα μοντέλα εκτός αυτό της ανάλυσης του κειμένου εκπαιδεύτηκαν χωριστά τόσο για τους άνδρες όσο και για τις γυναίκες.

- ✓ **Διαδικασία Αποθήκευσης των Βέλτιστων Βαρών**

Κατά την εκτέλεση της εκπαίδευσης των μοντέλων, στο τέλος κάθε εποχής, γινόταν έλεγχος της επίδοσης στα δεδομένα επικύρωσης. Αυτό το σύνολο δεδομένων δεν χρησιμοποιούνταν κατά τη διάρκεια της εκπαίδευσης. Μετά από κάθε βήμα εκπαίδευσης γινόταν λοιπόν έλεγχος αν βελτιώνεται η απόδοση του μοντέλου και σε περίπτωση που αυτό γινόταν, αποθηκεύαμε τα βάρη του μοντέλου. Έτσι, στο τέλος της διαδικασίας είχαν αποθηκευτεί τα βάρη που έδιναν το βέλτιστο αποτέλεσμα στο σύνολο επικύρωσης (validation set ή dev set). Αυτή η προσέγγιση ακολουθήθηκε ώστε να αντιμετωπιστεί το πιθανό πρόβλημα της υπερπροσαρμογής του μοντέλου (overfitting).

- ✓ **Διαδικασία Παραγωγής Γραφικών Παραστάσεων**

Για την αποδοτικότερη εκπαίδευση των μοντέλων ήταν σημαντική η δυναμική απεικόνιση διαφόρων μετρικών. Η συνεχής εποπτεία του συστήματος βοήθησε στην ορθή επιλογή υπερπαραμέτρων της εκπαίδευσης. Χρησιμοποιήθηκε το εύχρηστο εργαλείο pyplot της βιβλιοθήκης matplotlib, με το οποίο αναπαραστήσαμε τις μετρικές MAE και RMSE από κάθε εποχή τα οποία αποθηκεύονταν στο αντικείμενο history κατα τη διάρκεια της εκπαίδευσης.

- ✓ **Διαδικασία Μεταβολής Υπερπαραμέτρων της Αρχιτεκτονικής Δικτύου**

Έπειτα από κάθε αρχικοποίηση μοντέλου, και έχοντας σαν οδηγό αξιολόγησης την απόδοση των δυναμικών απεικονίσεων των μετρικών των μοντέλων, αλλά επιπροσθέτως και την διατήρηση της ισορροπίας μεταξύ απόκλισης και διακύμανσης (bias-variance trade off) [72], το οποίο το εξετάζουμε από τις γραφικές απεικονίσεις των προβλεπόμενων και πραγματικών τιμών PHQ8-scores στο ίδιο διάγραμμα, δοκιμάζουμε αρκετές αρχιτεκτονικές δικτύων μέχρι να συγκλίνουμε στην καλύτερη δυνατή. Όπως αναλύσαμε στην παράγραφο 2.6.2, τα συνελικτικά δίκτυα έχουν μια πληθώρα από υπερπαραμέτρους (πλήθος στρωμάτων, φίλτρων, μέγεθος πυρήνων, συναρτήσεις ενεργοποίησης κ.λπ.), επομένως τα περνάμε από αρκετούς συνδυασμούς υπερπαραμέτρων μέχρι να καταλήξουμε στη βέλτιστη δυνατή αρχιτεκτονική.

✓ **Διαδικασία Μεταβολής του Ρυθμού Μάθησης**

Είναι συνηθισμένη πρακτική να αυξομειώνεται ο ρυθμός μάθησης του μοντέλου σε περίπτωση που δε βελτιώνεται κάποια μετρική απόδοσης. Στην παρούσα εργασία χρησιμοποιήθηκε αυτή η προσέγγιση στην εκπαίδευση όλων των μοντέλων. Αν μετά από 10 εποχές δεν είχε βελτιωθεί το μέσο τετραγωνικό σφάλμα στο σύνολο επικύρωσης για τα μοντέλα παλινδρόμησης και η κατηγορική εντροπία για τα μοντέλα κατηγοριοποίησης τότε είτε μειωνόταν ο ρυθμός μάθησης πολλαπλασιαζόμενος με το παράγοντα 0.1 είτε αυξανόταν πολλαπλασιαζόμενος με το παράγοντα 10. Είναι σημαντικό να σημειωθεί ότι η αρχική τιμή (default) του ρυθμού μάθησης ήταν 0.001, καθώς αυτή ήταν η προτεινόμενη τιμή από τη δημοσίευση [71] που παρουσίασε τον αλγόριθμο 'Adam'.

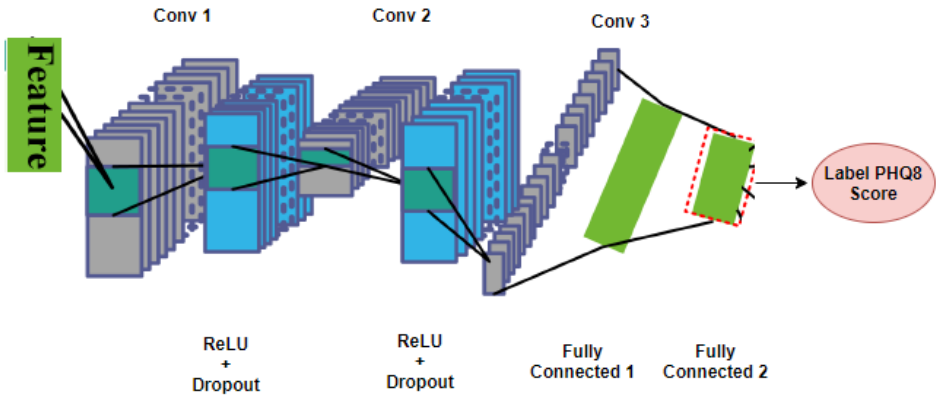
✓ **Προσαρμογή του Μοντέλου**

Στο τελευταίο στάδιο της διαδικασίας συνδυάζονταν όλα τα προηγούμενα στάδια που περιεγράφηκαν για να γίνει η τελική προσαρμογή του μοντέλου. Το πλήθος των εποχών εκπαίδευσης διέφερε σε κάθε τύπο προβλήματος. Η επιλογή του αριθμού των βημάτων εκπαίδευσης έγινε με βάση τις δυναμικές απεικονίσεις των μετρικών των μοντέλων. Η εκπαίδευση σταμάταγε όταν η γενική απόδοση σταματούσε να βελτιώνεται.

4.2 Προτεινόμενες Αρχιτεκτονικές και Παράμετροι Μοντέλων

Εμπνευσμένοι λοιπόν από την ικανότητα των βαθιών μοντέλων να μαθαίνουν σχέσεις μεταξύ δεδομένων που κανένα άλλο μοντέλο δεν μπορούσε προηγουμένως, θα αξιοποιήσουμε τα βαθιά συνελικτικά νευρωνικά μοντέλα (DCNN) για κάθε ένα από τα τρία είδη πληροφορίας που μας δίνεται (ήχος, εικόνα, κείμενο) χωριστά, ώστε να εξάγουμε χαμηλού επιπέδου διανυσματικά χαρακτηριστικά εκμεταλλευόμενοι τον δυναμικό χαρακτήρα των δεδομένων αυτών, αλλα και να βελτιώσουμε την ακρίβεια πρόβλεψης των PHQ8 Scores.. Μελέτες που έχουν διεισδύσει σε παρόμοιες αρχιτεκτονικές έχουν αποδείξει πως η βαθιά μάθηση έχει συνεισφέρει σημαντικά στον τομέα της αναγνώρισης. Τα βαθιά συνελικτικά δίκτυα (DCNN) τα τελευταία χρόνια χρησιμοποιούνται κατά κόρον σε έρευνες όπως η αναγνώριση και ανίχνευση αντικειμένων [73], αναγνώριση προσώπων [74], αναγνώριση συναισθήματος [75] και κατηγοριοποίησης της κατάθλιψης [76], δίνοντας εντυπωσιακές επιδόσεις στα συστήματα αυτά. Παρά το γεγονός πως μπορούν να εντοπίσουν μεγάλες διακυμάνσεις των δεδομένων, μαθαίνουν επίσης να διακριτοποιήσουν συμπαγείς αναπαραστάσεις χαρακτηριστικών, ειδικά όταν το μέγεθος των δεδομένων είναι τεράστιο.

Όπως είπαμε τα δίκτυα DCNN του ήχου, της εικόνας και του κειμένου θα εκπαιδευτούν χωριστά επιδιώκοντας το καθένα να προσεγγίσει όσο το δυνατόν καλύτερα την ετικέτα του κάθε δείγματος PHQ8 score. Για τον συνδυασμό των επιμέρους μοντέλων θα μιλήσουμε στην επόμενη ενότητα. Κάθε ένα από αυτά τα δίκτυα αποτελείται από *n* συνελικτικά επίπεδα (convolutional layers), ακολουθούμενα το καθένα από τη συνάρτηση ενεργοποίησης ReLU καθώς και από ένα επίπεδο Dropout, ενώ το τελευταίο συνελικτικό επίπεδο ακολουθείται από δύο (2) πλήρως συνδεδεμένα επίπεδα (fully connected layers-FC) που το τελευταίο καταλήγει σε ένα ακόμη FC που υπολογίζει την προβλεπόμενη τιμή του κάθε δείγματος. Σε όλες τις αρχιτεκτονικές των οπτικοακουστικών μοντέλων σαν συνάρτηση κόστους χρησιμοποιούμε τη ρίζα του μέσου τετραγωνικού σφάλματος (RMSE). Στο Σχήμα 4.1 παραθέτουμε μια γενική εικόνα των DCNN δικτύων που θα μελετήσουμε.



Σχήμα 4.1: DCNN αρχιτεκτονική για κάθε είδος δεδομένων χωριστά για αναγνώριση της κατάθλιψης

Για την υλοποίηση των DCNN δικτύων για τα δεδομένα ήχου και εικόνας, με διαχωρισμό των δύο φύλων, εξετάσαμε διάφορες αρχιτεκτονικές και επιλέξαμε τις καλύτερες δυνατές βάση των μετρικών αξιολόγησης RMSE και MAE πάνω στα δεδομένα επαλήθευσης (dev set), αλλα και έχοντας υπόψιν την εξισορρόπηση της απόκλισης και διακύμανσης (bias and variance trade-off). Για το πλήθος των συνελικτικών επιπέδων (CV), τα οποία να τονίσουμε ήταν μονοδιάστατα λόγω της μορφής της πληροφορίας που τους τροφοδοτούσαμε, δοκιμάσαμε τιμές από 2 έως 5 επίπεδα και για το μέγεθος του πυρήνα των φίλτρων τους πειραματιστήκαμε με τις τιμές 3 και 5, ενώ το stride ήταν σταθερά ίσο με 1. Το πλήθος των φίλτρων κάθε συνελικτικού επιπέδου κυμαινόταν μεταξύ των τιμών {12,24,48,64,128} και τέλος έπειτα από κάθε συνελικτικό επίπεδο ακολουθούσε ένα επίπεδο εφαρμογής της συνάρτησης ενεργοποίησης ReLU και του Dropout με τιμές εντός εύρους [0.2,0.5]. Σύνηθες στα DCNN είναι και η εφαρμογή επιπέδων Pooling, όπως αναλύσαμε στο θεωρητικό υπόβαθρο. Ωστόσο στις δικές μας αρχιτεκτονικές, το επίπεδο αυτό οδηγούσε στην απώλεια σημαντικής πληροφορίας δεδομένης της δυσαναλογίας δειγμάτων-πλήθους χαρακτηριστικών.

Όσον αφορά την διαδικασία εκπαίδευσης των δεδομένων μας, αξίζει να σημειώσουμε πως τα δεδομένα επαλήθευσης (dev set) τροφοδοτήθηκαν σε μοντέλα εκπαιδευμένα πάνω στα δεδομένα εκπαίδευσης μόνο, ενώ τα δεδομένα εξέτασης (test set) τροφοδοτήθηκαν πάνω στις τελικές αρχιτεκτονικές μοντέλων που επιλέχθηκαν βάση απόδοσης των δεδομένων επαλήθευσης, με την διαφορά ότι πλέον τα μοντέλα αυτά είχαν εκπαιδευτεί τόσο πάνω στα δεδομένα εκπαίδευσης όσο και επαλήθευσης.

4.3 Συστήματα Ανάλυσης Ήχου

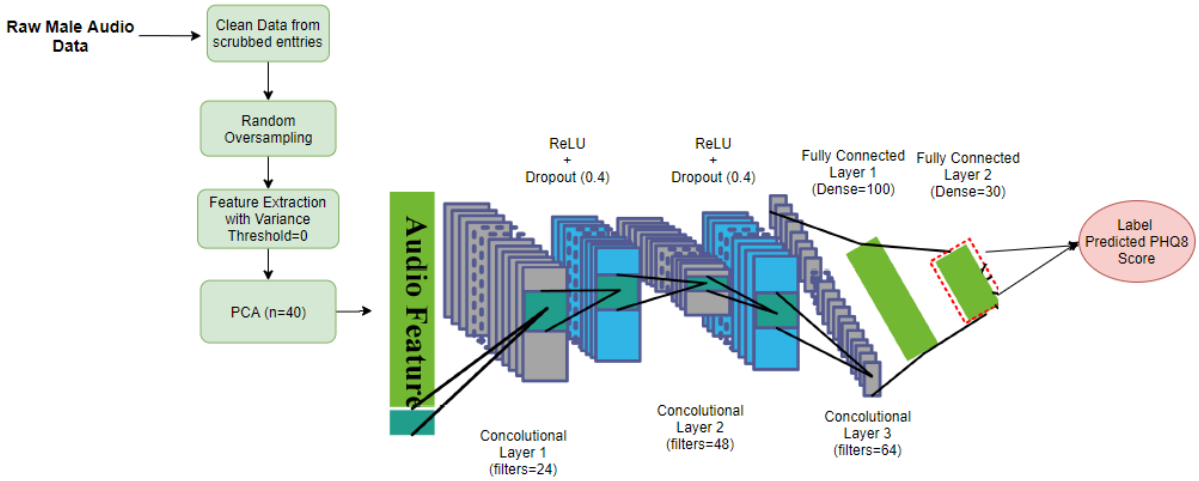
4.3.1 Μοντέλα για το Ανδρικό φύλο

Αφού εξάγουμε τα δεδομένα που επεξεργαστήκαμε στην ενότητα 3.2.1 και τα φιλτράρουμε αποκομίζοντας τα δείγματα μόνο του ανδρικού φύλου εξετάζουμε μια πληθώρα από αρχιτεκτονικές με την διαδικασία εκπαίδευσης από την ενότητα 4.1, επιλέγοντας την αποδοτικότερη. Στον ακόλουθο πίνακα 4.1 θα παρουσιάσουμε τις επιδόσεις κάποιων βαθιών αρχιτεκτονικών που μελετήσαμε πάνω στα δεδομένα επαλήθευσης, καθώς και της Baseline αρχιτεκτονικής, ενώ στη συνέχεια θα δώσουμε λεπτομερή αποτελέσματα και γραφικές απεικονίσεις της τελικής μας επιλογής πάνω στα δεδομένα επαλήθευσης (dev set).

	Random Forest Regressor				MAE	RMSE
Baseline Model	n_trees=10				6.22937	7.34263
	DCNN models				**MAE**	**RMSE**
	Conv	**Dropout**	**Fully-Conv**	**'Adam' parameters**		
Model 1	{12,24,48,64}	{0.3,0.3,0.3}	{1000,500,50,1}	Default	4.64316	5.95889
Model 2	{24,48,128}	{0.4,0.4}	{100,40,1}	Default	4.46050	5.75685
Model 3	{24,48,64,128}	{0.3,0.3,0.3}	{1000,500,30,1}	Default	4.44884	5.72035
Model 4	{24,48,64}	{0.3,0.3}	{100,30,1}	lr=default (0.001), decay=0.001	4.20801	5.54527
Model 5	{24,48,64}	{0.4,0.4}	{50,20,1}	lr=default (0.001), decay=0.001	4.18543	5.49272
Model 6	{12,24,48}	{0.4,0.4}	{50,20,1}	Default	4.13264	5.45204

Πίνακας 4.1: Σύγκριση Baseline μοντέλου με DCNN μοντέλα στα δεδομένα ήχου για τους άνδρες

Η τελική αρχιτεκτονική DCNN που καταλήξαμε για τα δεδομένα ήχου στα δείγματα του ανδρικού φύλου παρουσιάζεται στο διάγραμμα 4.2:



Σχήμα 4.2: Αρχιτεκτονική μοντέλου DCNN για τα δεδομένα ήχου στα δείγματα ανδρικού φύλου

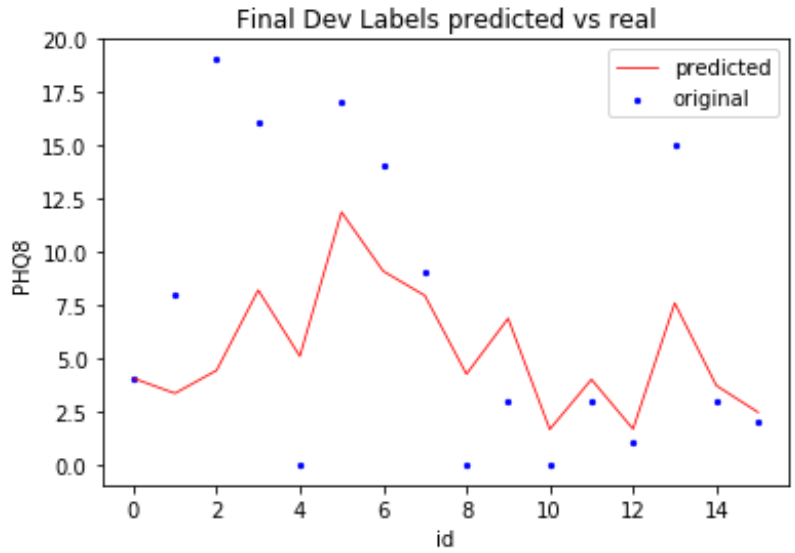
Στην τελική αρχιτεκτονική του μοντέλου αυτού παρατηρούμε πως η παράμετρος του PCA για την μείωση της διαστατικότητας των δεμένων εισόδου έχει αλλάξει από την βέλτιστη που βρήκαμε στην παράγραφο 3.2.1.2.2., από 25 που ήταν πριν, τώρα χρησιμοποιούμε διάσταση n=40. Αυτο συμβαίνει διότι μελετώντας τα βαθιά νευρωνικά μοντέλα πειραματιστήκαμε και με τις υπερπαραμέτρους των δεδομένων προεπεξεργασίας καθώς ανάλογα τον εκτιμητή του μοντέλου τα δεδομένα συμπεριφέρονται διαφορετικά, επομένως καταλήξαμε πως το βέλτιστο n_components για τον PCA ήταν το n=40.

Το τελικό μοντέλο λοιπόν αποτελείται από τρία (3) συνελικτικά επίπεδα με αριθμό φίλτρων {24,48,64} αντίστοιχα, ενδιάμεσες στρώσεις ReLU και Dropout με ποσοστά dropout {0.4,0.4} ενδιάμεσα των συνελικτικών επιπέδων και τρεις (3) FC στρώσεις με αριθμώ νευρώνων {100,30,1} αντίστοιχα. Στη συνέχεια παρουσιάζουμε μια λεπτομερή

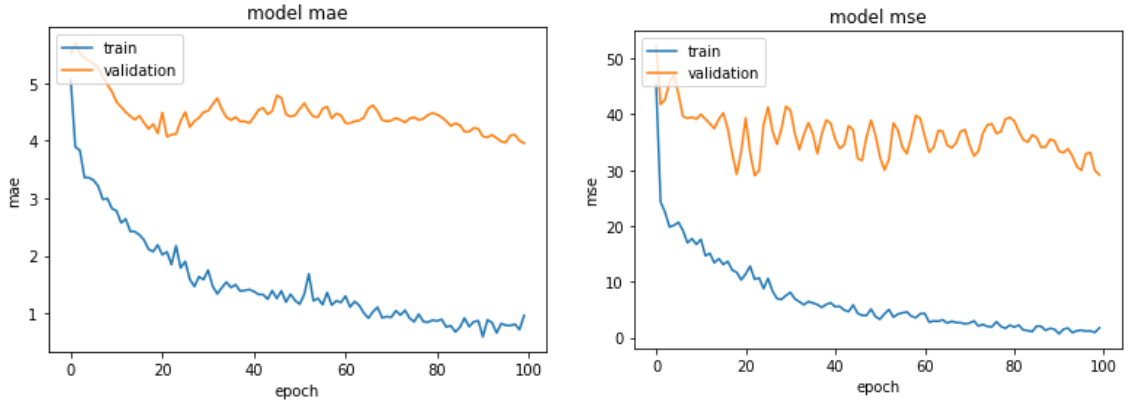
επισκόπηση της εκτέλεση της εκπαίδευσης του μοντέλου μας, μαζί με τις προβλεπόμενες τελικές τιμές PHQ8 score, ενώ στο σχήμα 4.3 βλέπουμε την γραφική αναπαράσταση των προβλεπόμενων ετικετών των δειγμάτων μας μαζί με τις πραγματικές.

Τέλος στο Σχήμα 4.4 μπορούμε να δούμε την πορεία των μετρικών αξιολόγησης του μοντέλου ανά εποχή, κατά τη διάρκεια εκπαίδευσης του δικτύου για να έχουμε μια καλύτερη και γενικότερη επισκόπηση της απόδοσης του συστήματός μας.

```
Train Data: (63, 237)
Dev Data: (16, 237)
Train Data resampled scaled: (63, 213)
Train PCA: (63, 40)
(63, 40, 1)
Load model...
Model: "sequential_10"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
1 (Conv1D)                   (None, 36, 24)            144
_________________________________________________________________
dropout_3 (Dropout)          (None, 36, 24)            0
_________________________________________________________________
2 (Conv1D)                   (None, 32, 48)            5808
_________________________________________________________________
dropout_4 (Dropout)          (None, 32, 48)            0
_________________________________________________________________
3 (Conv1D)                   (None, 28, 64)            15424
_________________________________________________________________
flatten_2 (Flatten)          (None, 1792)              0
_________________________________________________________________
4 (Dense)                    (None, 100)               179300
_________________________________________________________________
5 (Dense)                    (None, 30)                3030
_________________________________________________________________
6 (Dense)                    (None, 1)                 31
=================================================================
Total params: 203,737
Trainable params: 203,737
Non-trainable params: 0
_________________________________________________________________
['loss', 'mae', 'mse']
63/63 [==============================] - 0s 993us/step
Train evaluate: [0.5269154168310619, 0.526915431022644, 0.5054328441619873]
MAE =  3.9581471905112267
RMSE =  5.399076031496191
Final Labels: [ 4  8 19 16  0 17 14  9  0  3  0  3  1 15  3  2]
Predicted Final Labels: [ 4.0421634  3.3387384  4.413491   8.1776085  5.08028
11.839052
  9.072409   7.9246907  4.239942   6.851593   1.6482472  3.9887466
  1.6636935  7.570982   3.7023368  2.4503238]
```



Σχήμα 4.3: Προβλεπόμενες τιμές PHQ8 έναντι των πραγματικών για το ανδρικό φύλο



Σχήμα 4.4: Γραφικές αναπαραστάσεις των μετρικών αξιολόγησης του δικτύου ανά εποχή για το ανδρικό φύλο

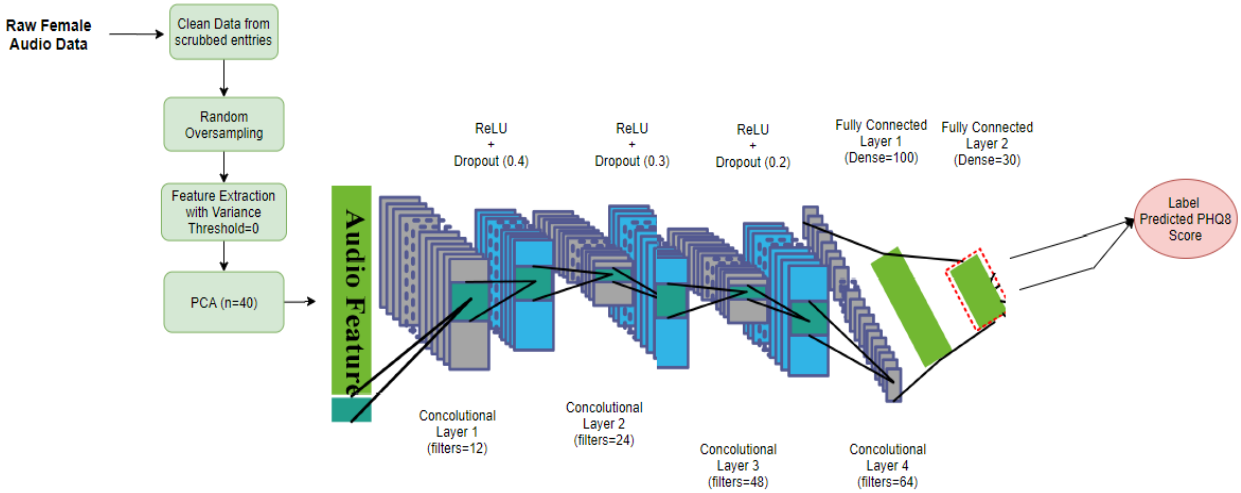
4.3.2 Μοντέλα για το Γυναικείο φύλο

Αφού εξάγουμε τα δεδομένα που επεξεργαστήκαμε στην ενότητα 3.2.1 και τα φιλτράρουμε αποκομίζοντας τα δείγματα μόνο του γυναικείου φύλου εξετάζουμε μια πληθώρα από αρχιτεκτονικές με την διαδικασία εκπαίδευσης από την ενότητα 4.1, επιλέγοντας την αποδοτικότερη. Στον ακόλουθο πίνακα 4.2 θα παρουσιάσουμε τις επιδόσεις κάποιων βαθιών αρχιτεκτονικών που μελετήσαμε πάνω στα δεδομένα επαλήθευσης, καθώς και της Baseline αρχιτεκτονικής, ενώ στη συνέχεια θα δώσουμε λεπτομερή αποτελέσματα και γραφικές απεικονίσεις της τελικής μας επιλογής πάνω στα δεδομένα επαλήθευσης (dev set).

	Random Forest Regressor				MAE	RMSE
Baseline Model	n_trees=10				4.94052	6.01621
	DCNN models				**MAE**	**RMSE**
	Conv	**Dropout**	**Fully-Conv**	**'Adam' parameters**		
Model 1	{24,48,64}, kernel_size=5	{0.4,0.4}	{100,30,1}	lr=0.01, decay=0.001	4.92382	5.89389
Model 2	{24,48}, kernel_size=5	{0.3,0.2}	{100,30,1}	Default	4.40221	5.84196
Model 3	{24,48}, kernel_size=5	{0.1}	{100,30,1}	Default	4.32672	5.72420
Model 4	{12,24}, kernel_size=5	-	{100,30,1}	lr=0.01, decay=0.001	4.57261	5.72075
Model 5	{12,24,48,64}, kernel_size=5	{0.4,0.3,0.2}	{100,30,1}	Default	4.51356	5.55587
Model 6	{12,24,48,64}, kernel_size=3	{0.2,0.2,0.2}	{100,30,1}	Default	4.32190	5.43105

Πίνακας 4.2: Σύγκριση Baseline μοντέλου με DCNN μοντέλα στα δεδομένα ήχου για τις γυναίκες

Η τελική αρχιτεκτονική DCNN που καταλήξαμε για τα δεδομένα ήχου στα δείγματα του ανδρικού φύλου παρουσιάζεται στο διάγραμμα 4.5:



Σχήμα 4.5: Αρχιτεκτονική μοντέλου DCNN για τα δεδομένα ήχου στα δείγματα γυναικείου φύλου

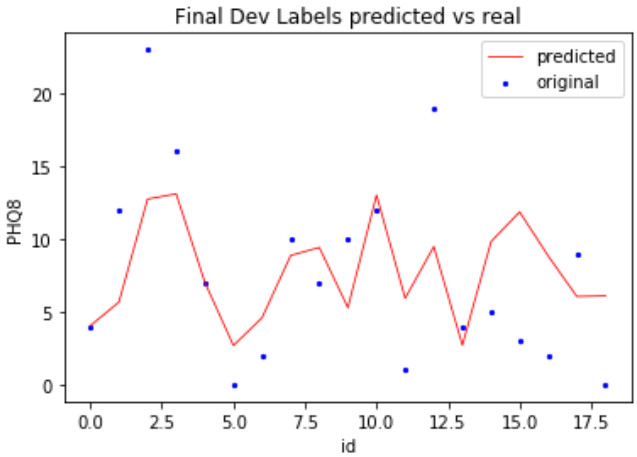
Στην τελική αρχιτεκτονική του μοντέλου αυτού παρατηρούμε πως η παράμετρος του PCA για την μείωση της διαστατικότητας των δεμένων εισόδου έχει αλλάξει από την βέλτιστη που βρήκαμε στην παράγραφο 3.2.1.2.2., από 25 που ήταν πριν, τώρα χρησιμοποιούμε διάσταση n=40. Αυτο συμβαίνει διότι μελετώντας τα βαθιά νευρωνικά μοντέλα πειραματιστήκαμε και με τις υπερπαραμέτρους των δεδομένων προεπεξεργασίας καθώς ανάλογα τον εκτιμητή του μοντέλου τα δεδομένα συμπεριφέρονται διαφορετικά, επομένως καταλήξαμε πως το βέλτιστο n_components για τον PCA ήταν το n=40.

Το τελικό μοντέλο λοιπόν αποτελείται από τέσσερα (4) συνελικτικά επίπεδα με αριθμό φίλτρων {12,24,48,64} αντίστοιχα, ενδιάμεσες στρώσεις ReLU και Dropout με ποσοστά dropout {0.4,0.3,0.2} ενδιάμεσα των συνελικτικών επιπέδων και τρεις (3) FC στρώσεις με αριθμώ νευρώνων {100,30,1} αντίστοιχα. Στη συνέχεια παρουσιάζουμε μια λεπτομερή επισκόπηση της εκτέλεση της εκπαίδευσης του μοντέλου μας, μαζί με τις προβλεπόμενες

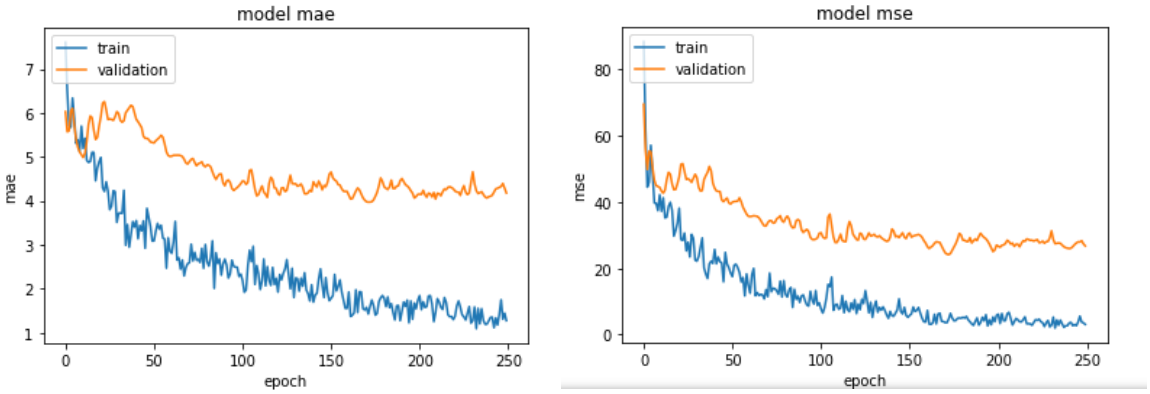
τελικές τιμές PHQ8 score, ενώ στο σχήμα 4.6 βλέπουμε την γραφική αναπαράσταση των προβλεπόμενων ετικετών των δειγμάτων μας μαζί με τις πραγματικές.

Τέλος στο Σχήμα 4.7 μπορούμε να δούμε την πορεία των μετρικών αξιολόγησης του μοντέλου ανά εποχή, κατά τη διάρκεια εκπαίδευσης του δικτύου για να έχουμε μια καλύτερη και γενικότερη επισκόπηση της απόδοσης του συστήματός μας.

```
Train Data: (54, 237)
Dev Data: (19, 237)
Train Data resampled scaled: (54, 209)
Train PCA: (54, 40)
(54, 40, 1)
Load model...
Model: "sequential_15"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_57 (Conv1D)           (None, 38, 12)            48
_________________________________________________________________
dropout_43 (Dropout)         (None, 38, 12)            0
_________________________________________________________________
conv1d_58 (Conv1D)           (None, 36, 24)            888
_________________________________________________________________
dropout_44 (Dropout)         (None, 36, 24)            0
_________________________________________________________________
conv1d_59 (Conv1D)           (None, 34, 48)            3504
_________________________________________________________________
dropout_45 (Dropout)         (None, 34, 48)            0
_________________________________________________________________
conv1d_60 (Conv1D)           (None, 32, 64)            9280
_________________________________________________________________
flatten_15 (Flatten)         (None, 2048)              0
_________________________________________________________________
dense_43 (Dense)             (None, 100)               204900
_________________________________________________________________
dense_44 (Dense)             (None, 30)                3030
_________________________________________________________________
dense_45 (Dense)             (None, 1)                 31
=================================================================
Total params: 221,681
Trainable params: 221,681
Non-trainable params: 0
_________________________________________________________________
['loss', 'mae', 'mse']
54/54 [==============================] - 0s 904us/step
Train evaluate: [0.5754115802270395, 0.5754116177558899, 0.6216983199119568]
MAE =  4.177970434490003
RMSE =  5.173906878452305
Final Labels: [ 4 12 23 16  7  0  2 10  7 10 12  1 19  4  5  3  2  9  0]
Predicted Final Labels: [ 4.034919   5.6664023 12.741362  13.098103   7.0231295
2.6786675
  4.583794   8.86582    9.415555   5.280215  13.016495   5.9221144
  9.478491   2.6985822  9.80824   11.858718   8.824667   6.044425
  6.0885386]
```



Σχήμα 4.6: Προβλεπόμενες τιμές PHQ8 έναντι των πραγματικών για το γυναικείο φύλο



Σχήμα 4.7: Γραφικές αναπαραστάσεις των μετρικών αξιολόγησης του δικτύου ανά εποχή για το γυναικείο φύλο

4.4 Συστήματα Ανάλυσης Εκφράσεων του Προσώπου

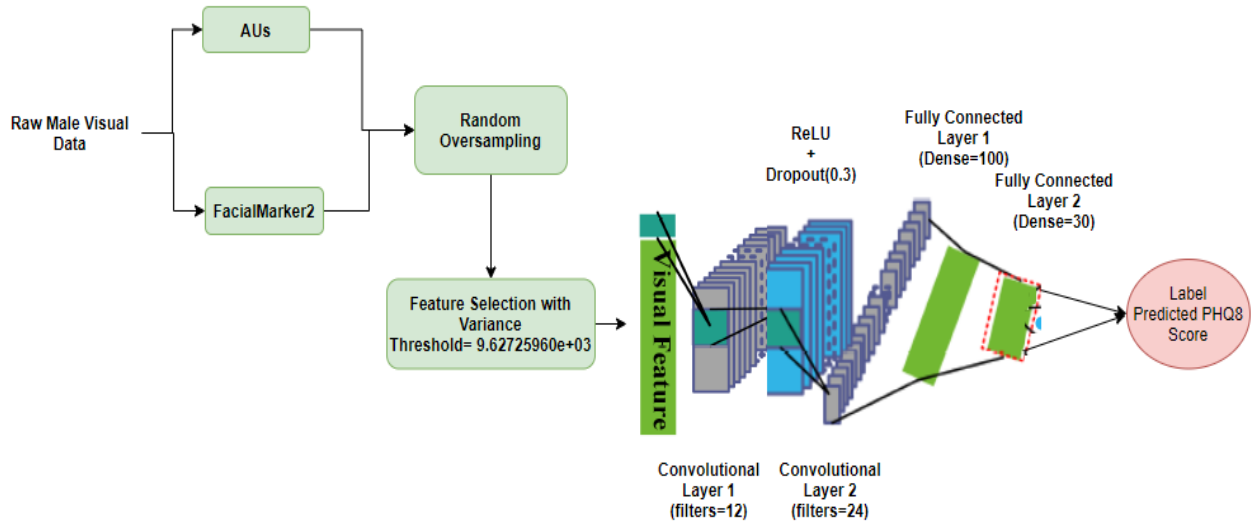
4.4.1 Μοντέλα για το Ανδρικό φύλο

Ομοίως για τα οπτικά χαρακτηριστικά που εξήγαμε στην ενότητα 3.2.2, ακολουθούμε την ίδια μεθοδολογία όπως για τα ακουστικά. Στον ακόλουθο πίνακα 4.3 θα παρουσιάσουμε τις επιδόσεις κάποιων βαθιών αρχιτεκτονικών που μελετήσαμε πάνω στα δεδομένα επαλήθευσης, καθώς και της Baseline αρχιτεκτονικής, ενώ στη συνέχεια θα δώσουμε λεπτομερή αποτελέσματα και γραφικές απεικονίσεις της τελικής μας επιλογής πάνω στα δεδομένα επαλήθευσης (dev set).

	Random Forest Regressor				**MAE**	**RMSE**
Baseline Model	n_trees=10				5.52562	6.70502
	DCNN models				**MAE**	**RMSE**
	Conv	**Dropout**	**Fully-Conv**	**'Adam' parameters**		
Model 1	{24,48}, kernel_size=5	{0.4,0.1}	{500,30,1}	Default	4.52005	5.95337
Model 2	{12,24,48}, kernel_size=5	{0.3,0.3}	{500,30,1}	Default	4.34491	5.76608
Model 3	{24,48,64}, kernel_size=5	{0.3,0.1}	{500,50,1}	Default	4.26667	5.52986

Πίνακας 4.3: Σύγκριση Baseline μοντέλου με DCNN μοντέλα στα οπτικά χαρακτηριστικά για τους άνδρες

Η τελική αρχιτεκτονική DCNN που καταλήξαμε για τα οπτικά δεδομένα στα δείγματα του ανδρικού φύλου παρουσιάζεται στο διάγραμμα 4.8:



Σχήμα 4.8: Αρχιτεκτονική μοντέλου DCNN για τα οπτικά δεδομένα στα δείγματα ανδρικού φύλου

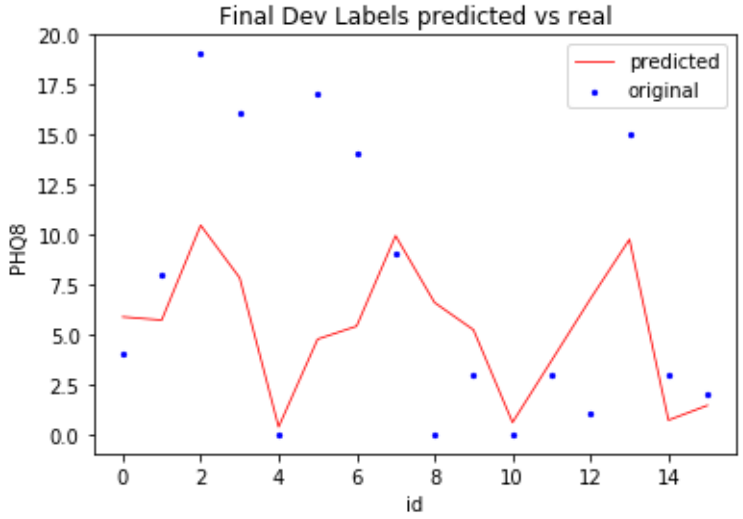
Στην τελική αρχιτεκτονική του μοντέλου αυτού παρατηρούμε πως το βήμα προεπεξεργασίας των τελικών εξαγόμενων χαρακτηριστικών (AUs και FacialMarker2) PCA για την μείωση της διαστατικότητας των δεμένων εισόδου παραλείπεται σε αντίθεση με τις μελέτες που κάναμε ,στην παράγραφο 3.2.1.2.2. Αυτο συμβαίνει διότι μελετώντας τα βαθιά νευρωνικά μοντέλα πειραματιστήκαμε και με τις υπερπαραμέτρους των δεδομένων προεπεξεργασίας καθώς ανάλογα τον εκτιμητή του μοντέλου τα δεδομένα συμπεριφέρονται διαφορετικά, επομένως καταλήξαμε πως το βήμα αυτό προεπεξεργασίας για τα μοντέλα του ανδρικού φύλου είναι περιττό. Επιπρόσθετα με την ίδια λογική στην μέθοδο Feature Reduction με το κατώφλι διακύμανσης (Variance Threshold), έπειτα από πειράματα επιλέξαμε ως βέλτιστο κατώφλι την τιμή 9.62725960e+03.

Το τελικό μοντέλο λοιπόν αποτελείται από δυο (2) συνελικτικά επίπεδα με αριθμό φίλτρων {12,24} αντίστοιχα, ενδιάμεσες στρώσεις ReLU και Dropout με ποσοστά dropout {0.3} ενδιάμεσα των συνελικτικών επιπέδων και τρεις (3) FC στρώσεις με αριθμό νευρώνων {100,30,1} αντίστοιχα. Τέλος χρησιμοποιήσαμε τις παραμέτρους *lr=0.001* και *decay=0.001* στη συνάρτηση βελτιστοποίησης 'Adam'. Στη συνέχεια παρουσιάζουμε μια λεπτομερή επισκόπηση της εκτέλεση της εκπαίδευσης του μοντέλου μας, μαζί με τις προβλεπόμενες τελικές τιμές PHQ8 score, ενώ στο σχήμα 4.9 βλέπουμε την γραφική αναπαράσταση των προβλεπόμενων ετικετών των δειγμάτων μας μαζί με τις πραγματικές.

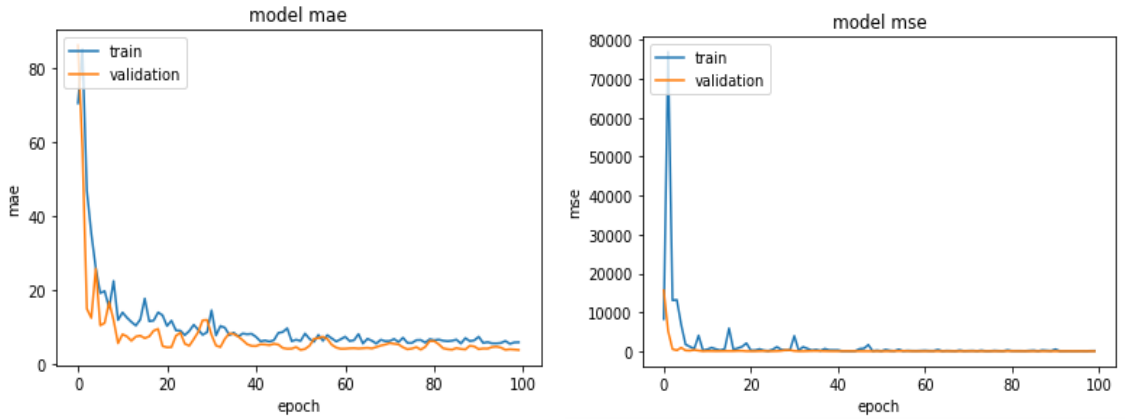
Τέλος στο Σχήμα 4.10 μπορούμε να δούμε την πορεία των μετρικών αξιολόγησης του μοντέλου ανά εποχή, κατά τη διάρκεια εκπαίδευσης του δικτύου για να έχουμε μια καλύτερη και γενικότερη επισκόπηση της απόδοσης του συστήματός μας.

```
(100, 140, 1)
Load model...
Model: "sequential_10"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_19 (Conv1D)           (None, 136, 12)           72
_________________________________________________________________
dropout_10 (Dropout)         (None, 136, 12)           0
_________________________________________________________________
conv1d_20 (Conv1D)           (None, 132, 24)           1464
_________________________________________________________________
flatten_10 (Flatten)         (None, 3168)              0
_________________________________________________________________
dense_28 (Dense)             (None, 100)               316900
_________________________________________________________________
dense_29 (Dense)             (None, 30)                3030
_________________________________________________________________
dense_30 (Dense)             (None, 1)                 31
=================================================================
Total params: 321,497
Trainable params: 321,497
Non-trainable params: 0
```

```
_________________________________________________________________
['loss', 'mae', 'mse']
100/100 [==============================] - 0s 548us/step
Train evaluate: [5.259468784332276, 5.259469032287598, 41.236175537109375]
MAE =  4.186823973432183
RMSE =  5.531515442440604
Final Labels: [ 4  8 19 16  0 17 14  9  0  3  0  3  1 15  3  2]
Predicted Final Labels: [ 5.8649573   5.703561   10.438924    7.8170056   0.369
25116  4.7518544
  5.3988466   9.914989    6.580229    5.2150884   0.5930464   3.676097
  6.766247    9.731981    0.7067554   1.4417937 ]
```



Σχήμα 4.9: Προβλεπόμενες τιμές PHQ8 έναντι των πραγματικών για το ανδρικό φύλο



Σχήμα 4.10: Γραφικές αναπαραστάσεις των μετρικών αξιολόγησης του δικτύου ανά εποχή για το ανδρικό φύλο

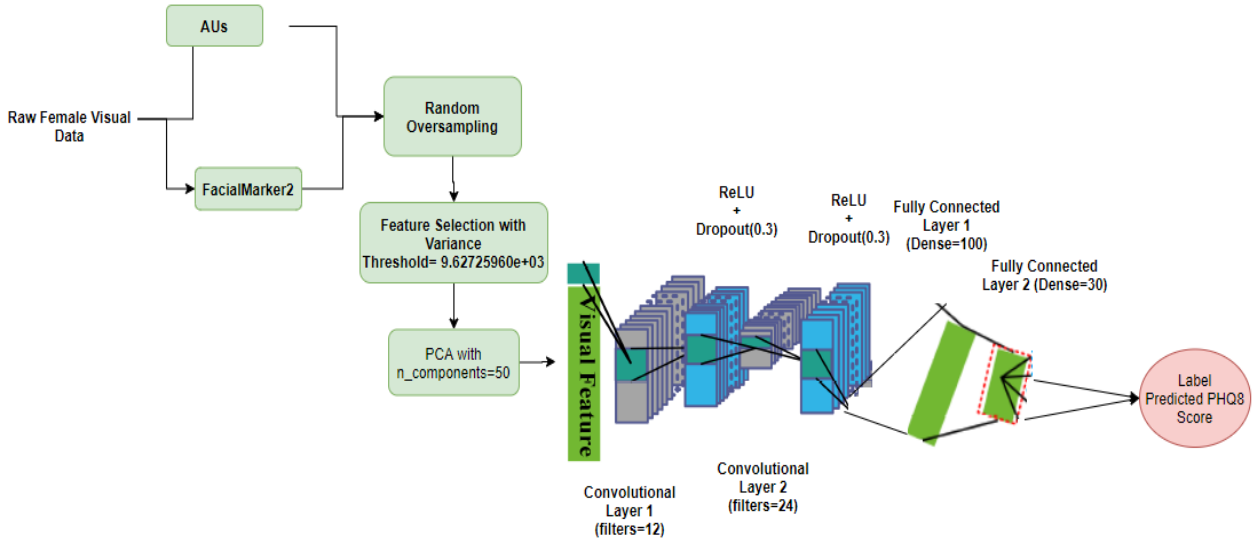
4.4.2 Μοντέλα για το Γυναικείο φύλο

Ομοίως για τα οπτικά χαρακτηριστικά που εξήγαμε στην ενότητα 3.2.2, ακολουθούμε την ίδια μεθοδολογία όπως για τα ακουστικά. Στον ακόλουθο πίνακα 4.4 θα παρουσιάσουμε τις επιδόσεις κάποιων βαθιών αρχιτεκτονικών που μελετήσαμε πάνω στα δεδομένα επαλήθευσης, καθώς και της Baseline αρχιτεκτονικής, ενώ στη συνέχεια θα δώσουμε λεπτομερή αποτελέσματα και γραφικές απεικονίσεις της τελικής μας επιλογής πάνω στα δεδομένα επαλήθευσης (dev set).

	Random Forest Regressor				**MAE**	**RMSE**
Baseline Model	n_trees=10				5.37421	6.98296
	DCNN models				**MAE**	**RMSE**
	Conv	**Dropout**	**Fully-Conv**	**'Adam' parameters**		
Model 1	{24,48,64}, kernel_size=5	{0.3,0.1,0.2}	{100,30,1}	Default	5.34702	6.67336
Model 2	{24,48}, kernel_size=5	{0.3,0.1}	{100,30,1}	Default	5.09300	6.26577
Model 3	{12,24}, kernel_size=5	{0.3,0.3}	{100,50,1}	Default	5.19339	6.16134

Πίνακας 4.4: Σύγκριση Baseline μοντέλου με DCNN μοντέλα στα οπτικά χαρακτηριστικά για τις γυναίκες

Η τελική αρχιτεκτονική DCNN που καταλήξαμε για τα οπτικά δεδομένα στα δείγματα του ανδρικού φύλου παρουσιάζεται στο διάγραμμα 4.11:



Σχήμα 4.11: Αρχιτεκτονική μοντέλου DCNN για τα οπτικά δεδομένα στα δείγματα γυναικείου φύλου

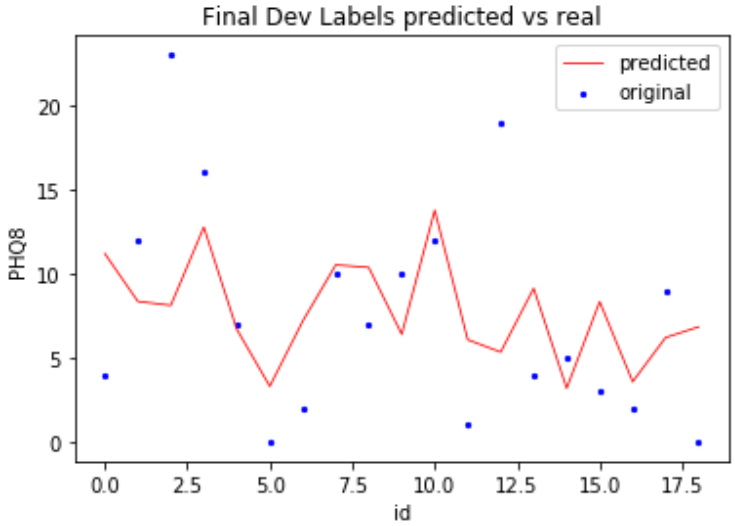
Στην τελική αρχιτεκτονική του μοντέλου αυτού παρατηρούμε πως η παράμετρος του PCA για την μείωση της διαστατικότητας των δεδομένων εισόδου έχει αλλάξει από την βέλτιστη που βρήκαμε στην παράγραφο 3.2.1.2.2., από 25 που ήταν πριν, τώρα χρησιμοποιούμε διάσταση n=50. Αυτο συμβαίνει διότι μελετώντας τα βαθιά νευρωνικά μοντέλα πειραματιστήκαμε και με τις υπερπαραμέτρους των δεδομένων προεπεξεργασίας καθώς ανάλογα τον εκτιμητή του μοντέλου τα δεδομένα συμπεριφέρονται διαφορετικά, επομένως καταλήξαμε πως το βέλτιστο n_components για τον PCA ήταν το n=50. Επιπρόσθετα με την ίδια λογική στην μέθοδο Feature Reduction με το κατώφλι διακύμανσης (Variance Threshold), έπειτα από πειράματα επιλέξαμε ως βέλτιστο κατώφλι την τιμή 0.

Το τελικό μοντέλο λοιπόν αποτελείται από δυο (2) συνελικτικά επίπεδα με αριθμό φίλτρων {12,24} αντίστοιχα, ενδιάμεσες στρώσεις ReLU και Dropout με ποσοστά dropout {0.3,0.3} ενδιάμεσα των συνελικτικών επιπέδων και τρεις (3) FC στρώσεις με αριθμό νευρώνων {100,30,1} αντίστοιχα. Στη συνέχεια παρουσιάζουμε μια λεπτομερή επισκόπηση της εκτέλεση της εκπαίδευσης του μοντέλου μας, μαζί με τις προβλεπόμενες τελικές τιμές PHQ8 score, ενώ στο σχήμα 4.12 βλέπουμε την γραφική αναπαράσταση των προβλεπόμενων ετικετών των δειγμάτων μας μαζί με τις πραγματικές.

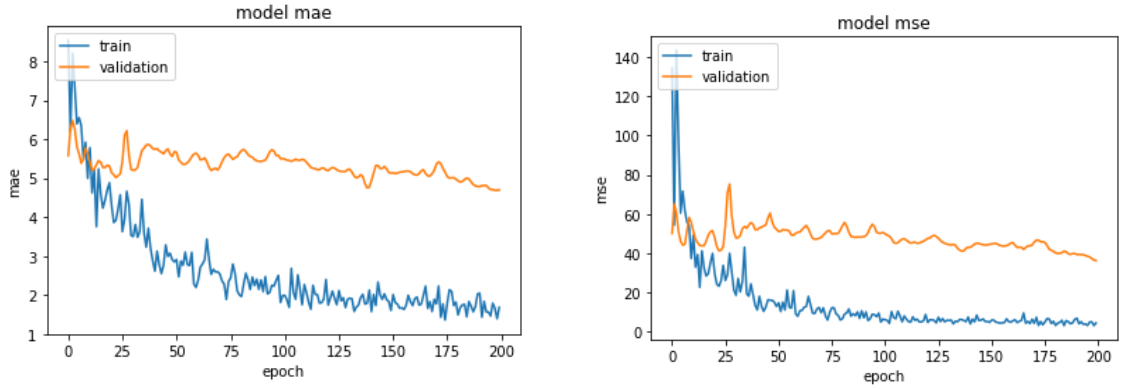
Τέλος στο Σχήμα 4.13 μπορούμε να δούμε την πορεία των μετρικών αξιολόγησης του μοντέλου ανά εποχή, κατά τη διάρκεια εκπαίδευσης του δικτύου για να έχουμε μια καλύτερη και γενικότερη επισκόπηση της απόδοσης του συστήματός μας.

```
(100, 140, 1)
Load model...
Model: "sequential_10"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_19 (Conv1D)           (None, 136, 12)           72
_________________________________________________________________
dropout_10 (Dropout)         (None, 136, 12)           0
_________________________________________________________________
conv1d_20 (Conv1D)           (None, 132, 24)           1464
_________________________________________________________________
flatten_10 (Flatten)         (None, 3168)              0
_________________________________________________________________
dense_28 (Dense)             (None, 100)               316900
_________________________________________________________________
dense_29 (Dense)             (None, 30)                3030
_________________________________________________________________
dense_30 (Dense)             (None, 1)                 31
=================================================================
Total params: 321,497
Trainable params: 321,497
Non-trainable params: 0
```

```
_________________________________________________________________
['loss', 'mae', 'mse']
100/100 [==============================] - 0s 548us/step
Train evaluate: [5.259468784332276, 5.259469032287598, 41.236175537109375]
MAE =  4.186823973432183
RMSE =  5.531515442440604
Final Labels: [ 4  8 19 16  0 17 14  9  0  3  0  3  1 15  3  2]
Predicted Final Labels: [ 5.8649573   5.703561   10.438924    7.8170056   0.369
25116  4.7518544
  5.3988466   9.914989    6.580229    5.2150884   0.5930464   3.676097
  6.766247    9.731981    0.7067554   1.4417937 ]
```



Σχήμα 4.12: Προβλεπόμενες τιμές PHQ8 έναντι των πραγματικών για το γυναικείο φύλο



Σχήμα 4.13: Γραφικές αναπαραστάσεις των μετρικών αξιολόγησης του δικτύου ανά εποχή για το γυναικείο φύλο

4.5 Συστήματα Ανάλυσης Κειμένου

Στο κομμάτι εξαγωγής σημασιολογικής πληροφορίας από τις απομαγνητοφωνημένες συνεντεύξεις των ασθενών δειγμάτων, μελετήσαμε στην ενότητα 3.3.1 δύο διαφορετικές προσεγγίσεις με στόχο να διακρίνουμε την αποδοτικότερη. Στην ενότητα αυτή λοιπόν θα μελετήσουμε την απόδοση των προσεγγίσεων αυτών τροφοδοτώντας τα διανυσματικά χαρακτηριστικά του κειμένου των δειγμάτων από την κάθε προσέγγιση τόσο στον Baseline εκτιμητή Random Forest, όσο και σε βαθιές συνελικτικές αρχιτεκτονικές. Αξίζει να σημειώσουμε πως στην περίπτωση του κειμένου δεν υπάρχει κάποιος παράγοντας που να διαχωρίζει την πληροφορία που δίνει το αντρικό από αυτήν που δίνει το γυναικείο φύλο, επομένως όλα τα μοντέλα μας θα πειραματιστούν πάνω σε όλα τα δείγματα ανεξαρτήτως φύλου.

Στο κομμάτι της σημασιολογικής πληροφορίας θα εκμεταλλευτούμε επίσης, πέρα από το σύστημα παλινδρόμησης, τα πλεονεκτήματα της κατηγοριοποίησης σε ασθενείς με και χωρίς κατάθλιψη.

4.5.1 Πρόβλημα Κατηγοριοποίησης (Classification)

Τα μοντέλα με τα οποία πειραματιστήκαμε που δώσανε τις καλύτερες αποδόσεις και αξίζει να αναφερθούμε ειναι τα εξής:

- ***Model 1***: Αξιοποιήσαμε την διανυσματική πληροφορία που εξήγαμε από την μέθοδο Σημασιολογικής Ανάλυσης Περιεχομένου (Semantic Context Analysis) και την τροφοδοτήσαμε στον ταξινομητή Random Forest με παράμετρο n_estimators=5.

- ***Model 2***: Έπειτα εκπαιδεύσαμε το ίδιο μοντέλο με την παραπάνω αρχιτεκτονική αλλά με δεδομένα εκπαίδευσής που υπέστησαν υπερδειγματοληψία.

- ***Model 3***: Αξιοποιήσαμε την πληροφορία που προέκυψε από την εκπαίδευση των ατομικών SVM ταξινομητών στους PV περιγραφητές των 5 γνωστών συμπτωμάτων και την τροφοδοτήσαμε στον Random Forest ταξινομητή με παράμετρο n_estimators=15.

- ***Model 4:*** Τέλος πειραματιζόμαστε με αρχιτεκτονικές βαθιών νευρωνικών δικτύων, όπου την βέλτιστη εκ αυτών την παρουσιάζουμε στον πίνακα 4.5.

	Preprocessed Text Method	Resampled Data	Classification model	Accuracy
Model 1	Semantic Context Analysis	No	Random Forest(n=5)	84.85%
Model 2	Semantic Context Analysis	Yes	Random Forest(n=5)	81.82%
Model 3	PV-SVM	No	Random Forest(n=15)	78.79%
Model 4	Semantic Context Analysis	No	DNN {layers (30,30,1), activation (relu, relu, sigmoid), Adam, epochs=100}	84.85%

Πίνακας 4.5: Σύγκριση μοντέλων ταξινόμησης για την κατάθλιψη πάνω στα δεδομένα κειμένου για όλα τα δείγματα μαζί

Παρατηρούμε πως την καλύτερη απόδοση την έχουν τα μοντέλα που έχουν εκπαιδευτεί πάνω στα εξαγόμενα διανύσματα από την μέθοδο Semantic Context Analysis. Ωστόσο τόσο ο ταξινομητής Random Forest (Model 1), όσο και ο ταξινομητής με το βαθύ νευρωνικό δίκτυο (Model 4) δίνουν ίδια ποσοστά ακριβείας. Έτσι επιλέγουμε το μοντέλο αυτό με την πιο απλή αρχιτεκτονική για εξοικονόμηση χώρου και χρόνου το οποίο είναι το **Model 1**. Η προσαρμογή του μοντέλου αυτού πάνω στα δεδομένα επαλήθευσης μας δίνουν τις ακόλουθες μετρικές απόδοσης:

```
Accuracy:84.85%
[[20  1]
 [ 4  8]]
              precision    recall  f1-score   support

           0       0.83      0.95      0.89        21
           1       0.89      0.67      0.76        12

    accuracy                           0.85        33
   macro avg       0.86      0.81      0.83        33
weighted avg       0.85      0.85      0.84        33

[0 0 0 1 1 1 1 1 0 1 1 0 0 0 0 0 1 0 1 0 1 0 0 1 0 0 0 0 0 0 1 0 0]
[0 0 1 0 1 0 1 1 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 1 0 0]
```

4.5.2 Πρόβλημα Παλινδρόμησης (Regression)

Τα μοντέλα με τα οποία πειραματιστήκαμε που δώσανε τις καλύτερες αποδόσεις και αξίζει να αναφερθούμε ειναι τα εξής:

- ***Model 1***: Αξιοποιήσαμε την διανυσματική πληροφορία που εξήγαμε από την μέθοδο Σημασιολογικής Ανάλυσης Περιεχομένου (Semantic Context Analysis) και την τροφοδοτήσαμε στον Random Forest Regressor με παράμετρο n_estimators=60.

- ***Model 2***: Έπειτα εκπαιδεύσαμε το ίδιο μοντέλο με την παραπάνω αρχιτεκτονική αλλά με δεδομένα εκπαίδευσής που υπέστησαν υπερδειγματοληψία.

- ***Model 3***: Αξιοποιήσαμε την πληροφορία που προέκυψε από την εκπαίδευση των ατομικών SVM ταξινομητών στους PV περιγραφητές των 5 γνωστών συμπτωμάτων και την τροφοδοτήσαμε στον Random Forest Regressor με παράμετρο n_estimators=60.

- ***Model 4:*** Πειραματιστήκαμε με αρχιτεκτονικές βαθιών νευρωνικών δικτύων, όπου την βέλτιστη εκ αυτών την παρουσιάζουμε στον πίνακα 4.6.

- ***Model 5:*** Τέλος εκπαιδεύσαμε το ίδιο μοντέλο με την παραπάνω αρχιτεκτονική αλλά με δεδομένα εκπαίδευσης που υπέστησαν υπερδειγματοληψία.

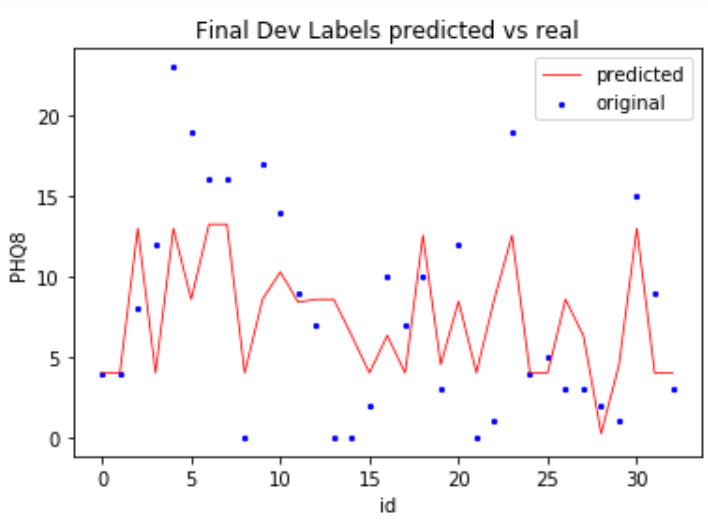
	Preprocessed Text Method	**Resampled Data**	**Regression model**	**MAE**	**RMSE**
Model 1	Semantic Context Analysis	No	Random Forest (n=60)	4.04943	4.97114
Model 2	Semantic Context Analysis	Yes	Random Forest (n=60)	4.19898	5.05807
Model 3	PV-SVM	No	Random Forest (n=60)	4.76353	6.09545
Model 4	Semantic Context Analysis	No	DNN {layers (30,30,30,15,1), activation (relu,.., relu, linear), Adam, epochs=100}	3.93107	4.86799
Model 5	Semantic Context Analysis	Yes	DNN {layers (30,30,30,15,1), activation (relu,.., relu, linear), Adam, epochs=100}	4.16887	5.09619

Πίνακας 4.6: Σύγκριση μοντέλων παλινδρόμησης για την κατάθλιψη πάνω στα δεδομένα κειμένου για όλα τα δείγματα μαζί

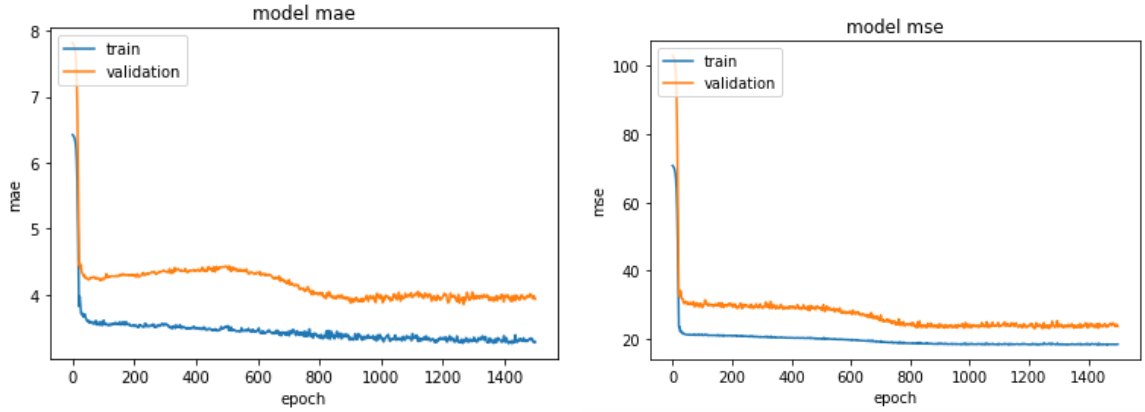
Παρατηρούμε πως την καλύτερη απόδοση την έχουν τα μοντέλα που έχουν εκπαιδευτεί πάνω στα εξαγόμενα διανύσματα από την μέθοδο Semantic Context Analysis. Ωστόσο παρόλο που τα δύο μοντέλα Model 1 και Model 4 δίνουν τις βέλτιστες μετρικές απόδοσης οι οποίες είναι και αρκετά κοντά μεταξύ τους και οι προβλεπόμενες τιμές που δίνουν παρουσιάζουν παρόμοια συμπεριφορά, επιλέξαμε σαν τελικό μοντέλο αξιολόγησης της κατάθλιψης το **Model 4** καθώς υπερτερεί κατ' ελάχιστον από το Model 1. Η προσαρμογή του μοντέλου αυτού πάνω στα δεδομένα επαλήθευσης μας δίνουν τις ακόλουθες μετρικές απόδοσης:

```
MAE =  3.931075599157449
RMSE =  4.86799164209982
[ 4  4  8 12 23 19 16 16  0 17 14  9  7  0  0  2 10  7 10  3 12  0  1 19
  4  5  3  3  2  1 15  9  3] [ 4.0072727   4.0072727  12.976439    4.0072727
12.976439    8.570945
 13.231009   13.231009    4.0072727   8.570945   10.272028    8.395397
  8.570945    8.570945    6.3231387   4.0072727   6.3231387   4.0072727
 12.543129    4.514457    8.449003    4.0072727   8.570945   12.543129
  4.0072727   4.0072727   8.570945    6.3231387   0.23538074  4.514457
 12.976439    4.0072727   4.0072722 ]
```

Στο σχήμα 4.14 συγκρίνουμε τις πραγματικές τιμές PHQ8 των δειγμάτων σε σύγκριση με την συνάρτηση των προβλεπόμενων τιμών στα δεδομένα επαλήθευσης, ενώ στο 4.15 παραθέτουμε την πορεία των μετρικών MAE και RMSE κατά τη διάρκεια εκπαίδευση του δικτύου μας.



Σχήμα 4.14: Γραφική παράσταση των πραγματικών και προβλεπόμενων τιμών PHQ8 από το τελικό μοντέλο παλινδρόμησης στα δεδομένα κειμένου



Σχήμα 4.15: Γραφικές αναπαραστάσεις των μετρικών αξιολόγησης του τελικού μοντέλου παλινδρόμησης στα δεδομένα κειμένου ανά εποχή για όλα τα δείγματα εισόδου

5. Προτεινόμενη Υβριδική Προσέγγιση για την Εκτίμηση της Κατάθλιψης

5.1 Χρησιμότητα Υβριδικού Αλγορίθμου

Ένας Υβριδικός Αλγόριθμος συνδυάζει δύο ή περισσότερους αλγορίθμους οι οποίοι επιλύουν το ίδιο πρόβλημα, είτε διαλέγοντας τον καλύτερο από αυτούς ανάλογα με την περίπτωση ή συνδυάζοντας τα ξεχωριστά αποτελέσματα τους για την εξαγωγή ενός ενιαίου αποτελέσματος. Μια συνήθης περίπτωση είναι οι επι μέρους αλγόριθμοι να διαφέρουν ως προς την απόδοση τους ανάλογα με τις ξεχωριστές περιπτώσεις του προβλήματος που καλούνται να επιλύσουν, όπως για παράδειγμα όταν εκπαιδεύονται σε διαφορετικού είδους πληροφορία. Η όλη διαδικασία σκοπεύει στον συνδυασμό των επιθυμητών χαρακτηριστικών των επι μέρους αλγορίθμων έτσι ώστε ο Υβριδικός Αλγόριθμος να έχει καλύτερη απόδοση από τις ξεχωριστές συνιστώσες του.

Επιπλέον μελέτες πάνω στην έρευνα της αξιολόγησης της κατάθλιψης έχουν αποδείξει πως όταν οι προσεγγίσεις της κατηγοριοποίησης και της παλινδρόμησης του προβλήματος αυτού συναντιόνται, επιτυγχάνεται καλύτερη απόδοση στο τελικό σύστημα [77] [78].

Αφού λοιπόν στην ενότητα 4 εφαρμόσαμε το πρόβλημα της αξιολόγησης της κατάθλιψης σε κάθε είδος πληροφορίας μεμονωμένα, αλλά και για το κάθε φύλο χωριστά, καλούμαστε τώρα να εκμεταλλευτούμε τα πλεονεκτήματα του κάθε μοντέλου από τα παραπάνω και με μια συνδυαστική λογική να τα αξιοποιήσουμε για την καλύτερη δυνατή συνολική πρόβλεψη των τιμών PHQ8. Η λογική πίσω από αυτή την ιδέα είναι ότι συνδυάζοντας τα αποτελέσματα των επιμέρους μεθόδων αξιολόγησης και ταξινόμησης της κατάθλιψης, θα μπορέσουμε να εκμεταλλευτούμε τις περιπτώσεις ασθενών για τις οποίες μια ή παραπάνω από τις μεθόδους μπορεί να μην παράγουν αξιόπιστες προβλέψεις, λαμβάνοντας έτσι υπόψιν τις προβλέψεις των μεθόδων που είναι πιο προσεγγιστικά κοντά στις πραγματικές τιμές-ετικέτες. Με την υβριδική προσέγγιση, οι αποδόσεις μιας μεθόδου συμπληρώνουν τις αδυναμίες της άλλης και με αυτό τον τρόπο η υβριδική μέθοδος γενικεύει πολύ καλύτερα στις διάφορες περιπτώσεις ασθενών με υψηλότερη συνολική ποιότητα προβλέψεων των PHQ8 τιμών για όλους τους ασθενείς.

5.2 Προτεινόμενη Υβριδική Προσέγγιση

Με βάση τις μελέτες που κάναμε παραπάνω, προτείνουμε τελικά μια υβριδική πολυτροπική προσέγγιση για την αξιολόγηση και ταξινόμηση της κατάθλιψης η οποία αποτελείται από πέντε (5) βασικά δομικά μέρη:

1) Ένα βαθύ συνελικτικό δίκτυο (DCNN) για το κάθε φύλο (άνδρες/γυναίκες) το οποίο εξετάζει την ακουστική πληροφορία που δίνουν τα δείγματα εισόδου και στην έξοδο του δίνει την προβλεπόμενη τιμή PHQ8 που αντιστοιχεί στο κάθε νέο δείγμα. Λεπτομερή ανάπτυξη του μοντέλου αυτού κάναμε στην ενότητα 4.3.

2) Ένα βαθύ συνελικτικό δίκτυο (DCNN) για το κάθε φύλο (άνδρες/γυναίκες) το οποίο εξετάζει την οπτική πληροφορία που δίνουν τα δείγματα εισόδου και στην έξοδο του δίνει την προβλεπόμενη τιμή PHQ8 που αντιστοιχεί στο κάθε νέο δείγμα. Λεπτομερή ανάπτυξη του μοντέλου αυτού κάναμε στην ενότητα 4.4.

3) Ένα βαθύ νευρωνικό δίκτυο (DNN), ενιαίο για τα δύο φύλα, το οποίο εξετάζει την σημασιολογική πληροφορία που δίνουν οι απομαγνητοφωνήσεις των δειγμάτων εισόδου και στην έξοδο του δίνει την προβλεπόμενη τιμή PHQ8 που αντιστοιχεί στο κάθε νέο δείγμα. Λεπτομερή ανάπτυξη του μοντέλου αυτού κάναμε στην ενότητα 4.5.2.

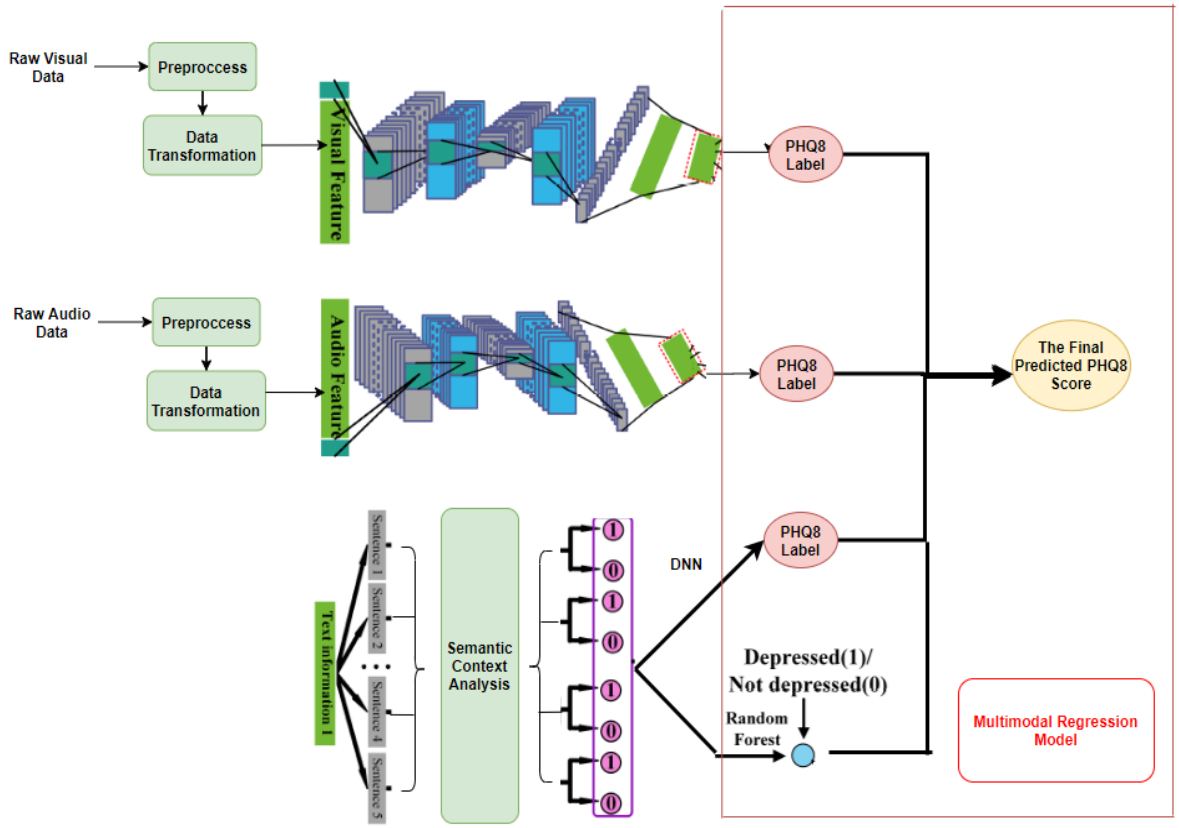
4) Ένα μοντέλο ταξινόμησης της κατάθλιψης με εκτιμητή τον ταξινομητή Random Forest με τροφοδοτούμενα δεδομένα την σημασιολογική πληροφορία που δίνουν οι απομαγνητοφωνήσεις των δειγμάτων εισόδου και με έξοδο την έκβαση του αποτελέσματος αν το κάθε νέο δείγμα ασθενή πάσχει από κατάθλιψη ή οχι. Λεπτομερή ανάπτυξη του μοντέλου αυτού κάναμε στην ενότητα 4.5.1.

5) Για τον τελικό συνδυασμό των παραπάνω μοντέλων, θα εφαρμόσουμε μια συνδυαστική προσέγγιση της κατηγορίας Voting του Ensemble Learning (Ενότητα 2.4). Πιο συγκεκριμένα προτείνεται μια απλή υβριδική πολυτροπική προσέγγιση παλινδρόμησης η οποία δέχεται σαν είσοδο τις εξόδους των προηγούμενων μοντέλων που αναφέραμε και σαν έξοδο δίνει την τελική προβλεπόμενη τιμή PHQ8 του κάθε ασθενή βάση του **εξής υβριδικού αλγορίθμου:**

Για κάθε ασθενή του δείγματος εξέτασης:
Αν το αποτέλεσμα του αλγορίθμου κατηγοριοποίησης (4) είναι θετικό (1), δηλαδή πάσχει από κατάθλιψη:
Για όλα τα μοντέλα (1,2,3) που έδωσαν προβλεπόμενη τιμή>10:
υπολόγισε τον μέσο όρο τους.
Αλλιως πάρε την μεγαλύτερη τιμή που προκύπτει από αυτά.

Αν το αποτέλεσμα του αλγορίθμου κατηγοριοποίησης (4) είναι αρνητικό (0), δηλαδή δεν πάσχει από κατάθλιψη:
Για όλα τα μοντέλα (1,2,3) που έδωσαν προβλεπόμενη τιμή<10:
υπολόγισε τον μέσο όρο τους.
Αλλιως πάρε την μικρότερη τιμή που προκύπτει από αυτά.

Στο ακόλουθο Σχήμα 5.1 δίνουμε μια αναπαράσταση της δομής του συνολικού πολυτροπικού υβριδικού μοντέλου που προτείναμε, ενώ στον πίνακα 5.1 παρουσιάζουμε την δομή και τις παραμέτρους του τελικού ολικού συστήματος:



Σχήμα 5.1: Δομή του προτεινόμενο υβριδικού πολυτροπικού μοντέλου ανίχνευσης και κατηγοριοποίησης της κατάθλιψης.

Οπτικό-Ακουστικά Μοντέλα Βαθιάς Συνελικτικής Αρχιτεκτονικής			
Φύλο	**Είδος Πληροφορίας**	**DCNN**	
		Conv	**Fully-Conv**
Γυναίκα	Ήχος	{12,24,48,64}	{100,30}
Γυναίκα	Βίντεο	{12,24}	{100,30}
Άνδρας	Ήχος	{24,48,64}	{100,30}
Άνδρας	Βίντεο	{12,24}	{100,30}
Μοντέλα Ανάλυσης Κειμένου για Εκτίμηση της Κατάθλιψης			
Φύλο	**Κείμενο**	**DNN**	
Άνδρας + Γυναίκα	5 επιλεγμένες προτάσεις	{30,30,30,15}	
Μοντέλα Ανάλυσης Κειμένου για Κατηγοριοποίηση της Κατάθλιψης			
Φύλο	**Κείμενο**	**Αλγόριθμος Ταξινόμησης**	
Άνδρας + Γυναίκα	5 επιλεγμένες προτάσεις	Random Forest (n=5)	

Πίνακας 5.1: Δομή και Παράμετροι του Ολικού Συστήματος Αξιολόγησης της Κατάθλιψης

5.3 Σύγκριση Baseline προσέγγισης και Προτεινόμενης

Στη συνέχεια θα συγκρίνουμε τα αποτελέσματα της προσέγγισης μας με αυτά του baseline μοντέλου ώστε να έχουμε ένα μέτρο σύγκρισης της αποδοτικότητας του συστήματος μας, τόσο στα δεδομένα επαλήθευσης (dev set) όσο και στα δεδομένα εξέτασης (test set).

- ✓ Υπενθυμίζουμε πως στα Baseline μοντέλα τροφοδοτούμε την αρχική πληροφορία που μας δίνεται για κάθε είδος δεδομένων (ήχος, εικόνα, κείμενο) χωρίς επιπλέον προεπεξεργασία ή παραγωγή νέων διανυσμάτων, σε έναν εκτιμητή Random Forest. Εφόσον λοιπόν εξετάζουμε την συμπεριφορά των αρχικών δοσμένων χαρακτηριστικών, δεν διεξάγουμε κάποια περαιτέρω μελέτη του κειμένου και επομένως δεν δύναται η εκπαίδευση συστήματος πάνω σε σημασιολογικά χαρακτηριστικά. Στον πίνακα 5.2 παρουσιάζουμε τις μετρικές απόδοσης των συστημάτων παλινδρόμησης για τα μοντέλα οπτικό-ακουστικών δεδομένων, ενώ στο τέλος παραθέτουμε και τα αποτελέσματα από την συνένωση των επιμέρους μοντέλων η οποία δίνει τις τελικές προβλεπόμενες τιμές από τον μέσο όρο των επιμέρους τιμών.

Gender	Dataset	MAE	RMSE
Female Audio	Dev	4.699	5.685
Female Audio	test	5.733	6.754
Male Audio	Dev	6.507	7.633
Male Audio	test	5.192	6.251
Female Video	Dev	5.374	6.983
Female Video	test	5.752	6.531
Male Video	Dev	5.526	6.705
Male Video	test	4.666	5.669
All Audio-Video	Dev	5.520	6.620
All Audio-Video	Test	5.660	7.050

Πίνακας 5.2: Μετρικές απόδοσης των Baseline μοντέλων

- ✓ Στους ακόλουθους δύο (2) πίνακες θα παρουσιάσουμε την απόδοση των τελικών επιλεγμένων μοντέλων, τόσο των ατομικών όσο και του τελικού υβριδικού στα δεδομένα επαλήθευσης αλλά και εξέτασης. Στον πίνακα 5.3 παραθέτουμε τις μετρικές απόδοσης των μοντέλων που εκπαιδεύτηκαν στα δεδομένα σημασιολογικού χαρακτήρα, ενώ στον 5.4 την απόδοση όλων των τελικών μοντέλων για τα δύο φύλα χωριστά αλλά και του συνολικού υβριδικού συστήματος που έχουν συνενωθεί όλα τα επιμέρους μοντέλα και των δύο φύλων.

Dataset	State	Accuracy	F1 Score	Precision	Recall
Dev	Depresssed	84.848	0.760	0.890	0.670
Dev	Not Depresssed	84.848	0.890	0.830	0.950
Test	Depresssed	82.609	0.640	0.750	0.640
Test	Not Depresssed	82.609	0.910	0.850	0.910

Πίνακας 5.3: Μετρικές απόδοσης του ταξινομητή πάνω στα δεδομένα κειμένου

Gender	Dataset	MAE	RMSE
Female Audio	Dev	4.178	5.174
Female Audio	Test	4.906	6.071
Male Audio	Dev	3.958	5.399
Male Audio	Test	4.410	5.823
Female Video	Dev	4.696	6.027
Female Video	Test	4.944	6.373
Male Video	Dev	4.187	5.532
Male Video	Test	4.580	5.855
All Text	Dev	3.931	4.868
All Text	Test	3.572	4.583
All Audio-Video-Text	Dev	3.759	5.224
All Audio-Video-Text	Test	3.347	4.543

Πίνακας 5.4: Μετρικές απόδοσης προτεινόμενων επιμέρους και τελικού μοντέλων

Τελος παραθέτουμε τις μετρικές απόδοσης μόνο των τελικών συστημάτων τόσο στα δεδομένα επαλήθευσης (Πίνακας 5.5), όσο και στα δεδομένα εξέτασης (Πίνακας 5.6) για πιο άμεση σύγκριση με το Baseline σύστημα

Model	**Features**	**MAE**	**RMSE**
Baseline	Audio-Video	5.520	6.620
Proposed Hybrid Multimodal Model	Audio-Video-Text	**3.759**	**5.224**

Πίνακας 5.5: Αποτελέσματα τελικού μοντέλου έναντι Baseline στα δεδομένα επαλήθευσης

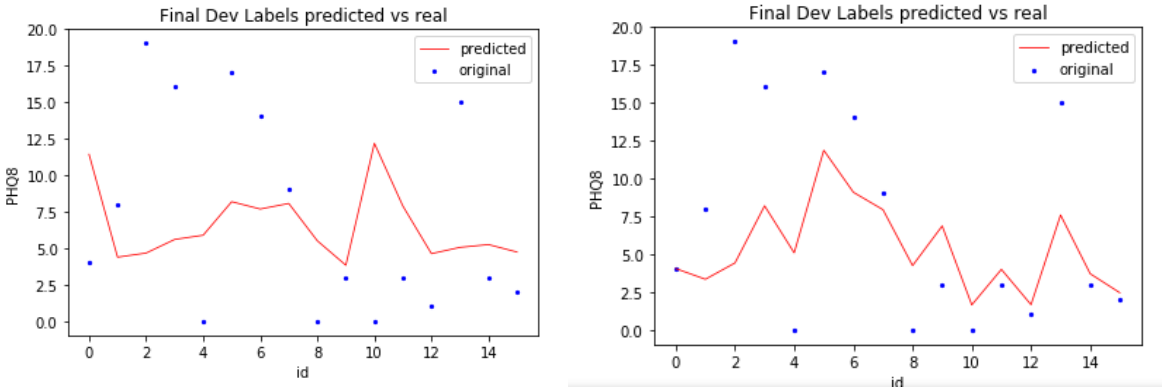
Model	**Features**	**MAE**	**RMSE**
Baseline	Audio-Video	5.660	7.050
Proposed Hybrid Multimodal Model	Audio-Video-Text	**3.347**	**4.543**

Πίνακας 5.6: Αποτελέσματα τελικού μοντέλου έναντι Baseline στα δεδομένα εξέτασης

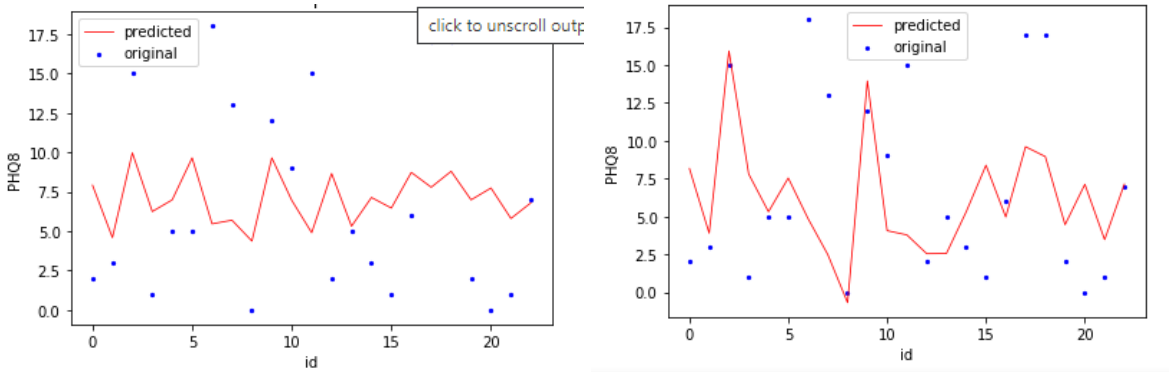
Όπως παρατηρούμε οι μετρικές απόδοσης της προτεινόμενης υβριδικής προσέγγισης μας, τόσο στα δεδομένα επαλήθευσης, όσο και στα δεδομένα εξέτασης παρουσιάζουν κατά πολυ βελτιωμένη απόδοση και μάλιστα την καλύτερη δυνατή ανάμεσα σε προσεγγίσεις του ίδιου προβλήματος [17][21][79][80].

Ωστόσο οι μετρικές απόδοσης δεν αποτελούν μοναδικό κριτήριο αξιολόγησης των μοντέλων μας. Είναι σημαντικό να παρακολουθούμε και γραφικά την σχέση προβλεπόμενων και πραγματικών τιμών για να κάνουμε εκτίμηση της εξισορρόπησης μεταξύ απόκλισης και διακύμανσης των τιμών (bias-variance tradeoff). Ακολούθως θα συγκρίνουμε την απόδοση των μοντέλων μας με τα Baseline παραθέτοντας τις γραφικές παραστάσεις των αποτελεσμάτων των προβλέψεων των μοντέλων αυτών έναντι των πραγματικών τιμών.

1. Μοντέλα πάνω στα χαρακτηριστικά ήχου για τα δείγματα ανδρικού φύλου:

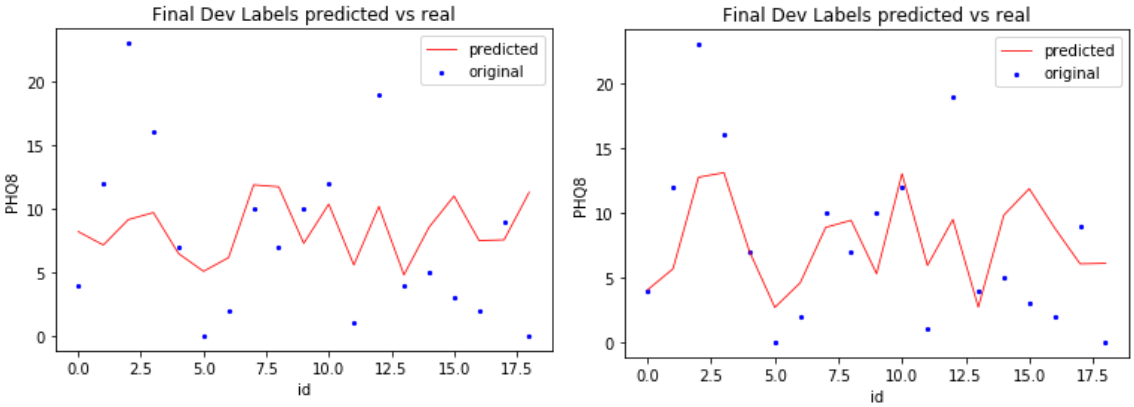


Σχήμα 5.2: Γραφική αναπαράσταση αποτελεσμάτων baseline και προτεινόμενου μοντέλου στα χαρακτηριστικά ήχου για το ανδρικό φύλο στα δεδομένα επαλήθευσης (αριστερά baseline, δεξια proposed)

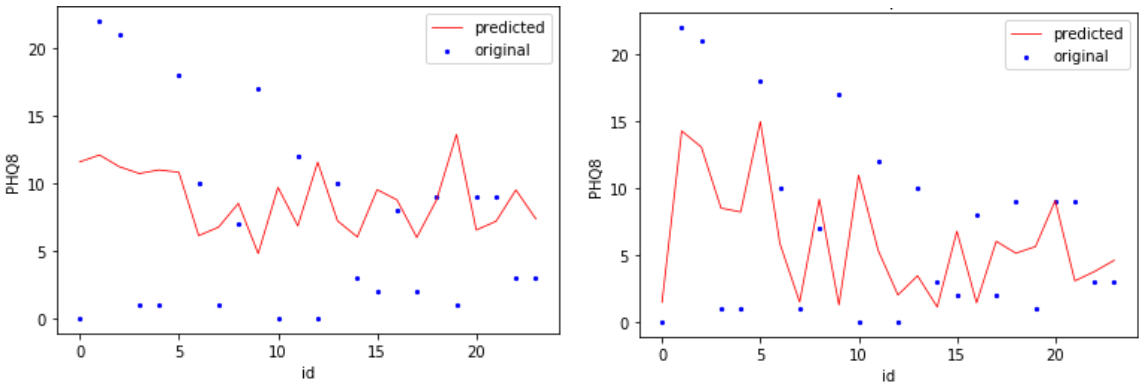


Σχήμα 5.3: Γραφική αναπαράσταση αποτελεσμάτων baseline και προτεινόμενου μοντέλου στα χαρακτηριστικά ήχου για το ανδρικό φύλο στα δεδομένα εξέτασης (αριστερά baseline, δεξια proposed)

2. Μοντέλα πάνω στα χαρακτηριστικά ήχου για τα δείγματα γυναικείου φύλου:

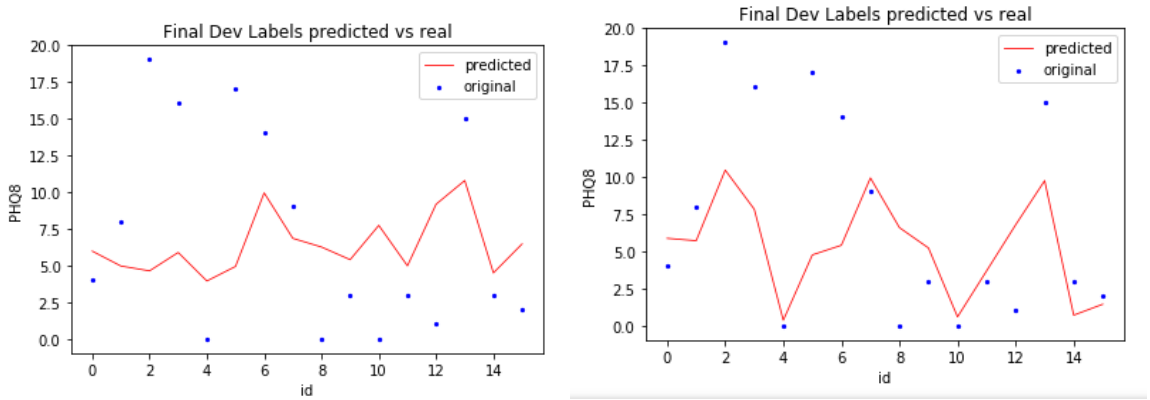


Σχήμα 5.4: Γραφική αναπαράσταση αποτελεσμάτων baseline και προτεινόμενου μοντέλου στα χαρακτηριστικά ήχου για το γυναικείο φύλο στα δεδομένα επαλήθευσης (αριστερά baseline, δεξια proposed)

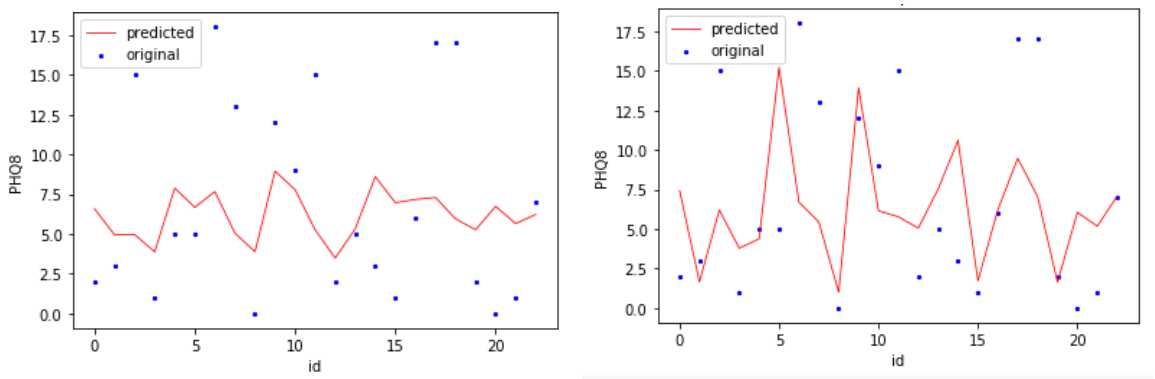


Σχήμα 5.5: Γραφική αναπαράσταση αποτελεσμάτων baseline και προτεινόμενου μοντέλου στα χαρακτηριστικά ήχου για το γυναικείο φύλο στα δεδομένα εξέτασης (αριστερά baseline, δεξια proposed)

3. Μοντέλα πάνω στα οπτικά χαρακτηριστικά για τα δείγματα ανδρικού φύλου:

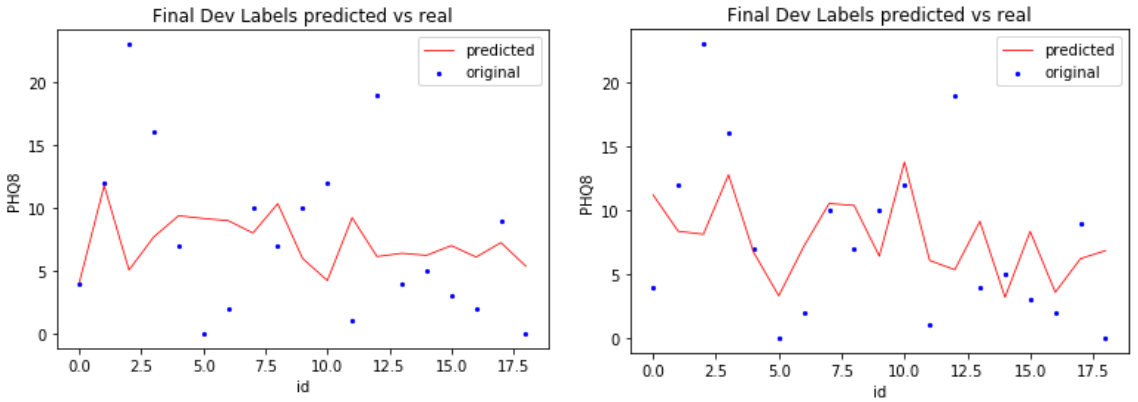


Σχήμα 5.6: Γραφική αναπαράσταση αποτελεσμάτων baseline και προτεινόμενου μοντέλου στα οπτικά χαρακτηριστικά για το ανδρικό φύλο στα δεδομένα επαλήθευσης (αριστερά baseline, δεξια proposed)

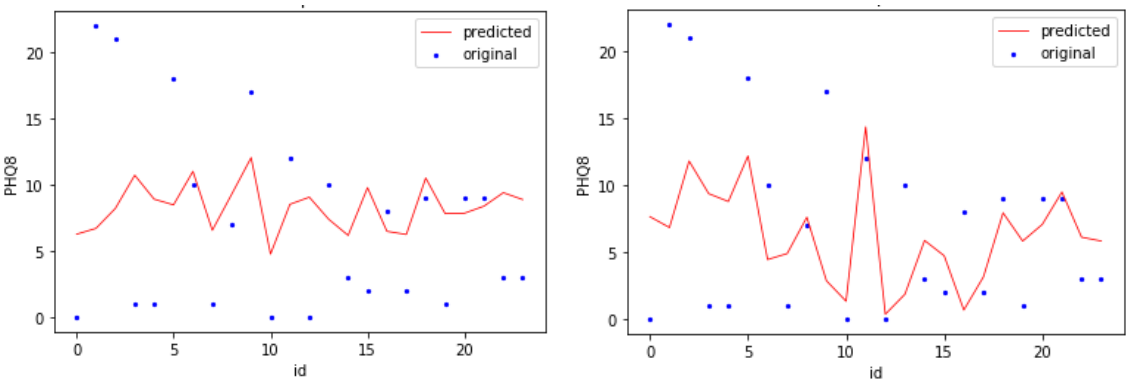


Σχήμα 5.7: Γραφική αναπαράσταση αποτελεσμάτων baseline και προτεινόμενου μοντέλου στα οπτικά χαρακτηριστικά για το ανδρικό φύλο στα δεδομένα εξέτασης (αριστερά baseline, δεξια proposed)

4. Μοντέλα πάνω στα οπτικά χαρακτηριστικά για τα δείγματα γυναικείου φύλου:

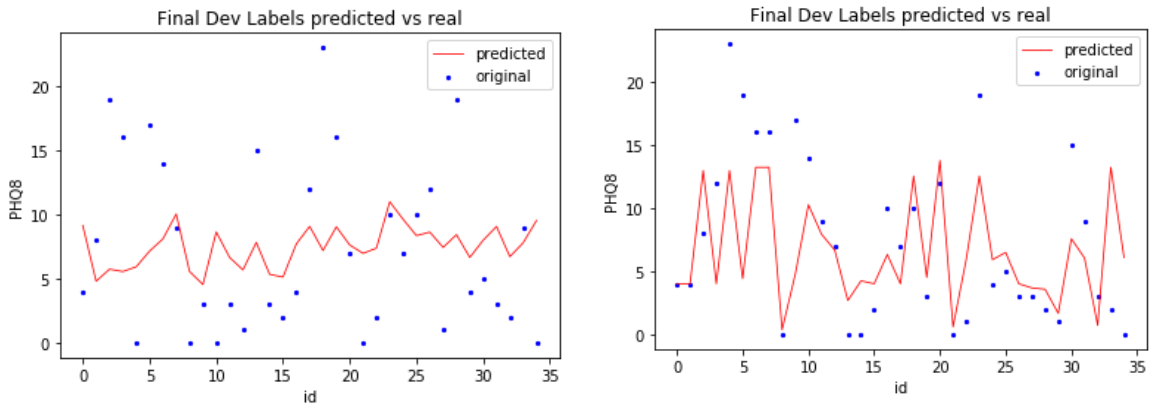


Σχήμα 5.8: Γραφική αναπαράσταση αποτελεσμάτων baseline και προτεινόμενου μοντέλου στα οπτικά χαρακτηριστικά για το γυναικείο φύλο στα δεδομένα επαλήθευσης (αριστερά baseline, δεξια proposed)

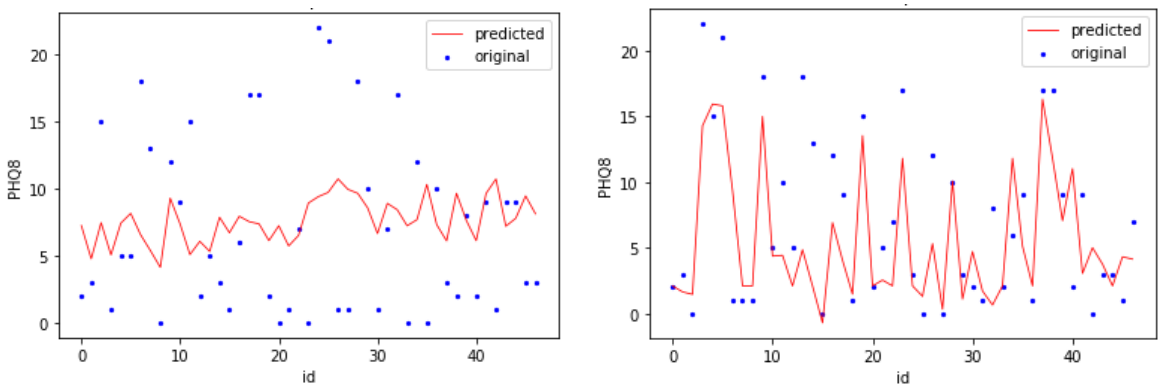


Σχήμα 5.9: Γραφική αναπαράσταση αποτελεσμάτων baseline και προτεινόμενου μοντέλου στα οπτικά χαρακτηριστικά για το γυναικείο φύλο στα δεδομένα εξέτασης (αριστερα baseline, δεξια proposed)

5. Τελικό Baseline και Προτεινόμενο Υβριδικό μοντέλο για όλα τα δείγματα:



Σχήμα 5.10: Γραφική αναπαράσταση αποτελεσμάτων baseline και προτεινόμενου τελικού υβριδικού πολυτροπικού μοντέλου για όλα τα δείγματα και των δύο φύλων στα δεδομένα επαλήθευσης (αριστερά baseline, δεξια proposed)



Σχήμα 5.11: Γραφική αναπαράσταση αποτελεσμάτων baseline και προτεινόμενου τελικού υβριδικού πολυτροπικού μοντέλου για όλα τα δείγματα και των δύο φύλων στα δεδομένα εξέτασης (αριστερα baseline, δεξια proposed)

6. Συμπεράσματα και Προτάσεις

6.1 Συμπεράσματα

Για την ανάπτυξη ενός συστήματος Βαθιάς Μηχανικής Μάθησης που θα πετυχαίνει μεγάλες επιδόσεις στο πρόβλημα της αξιολόγησης της κατάθλιψης μέσω ανάλυσης ηχητικών σημάτων της φωνής, εκφράσεων του προσώπου και σημασιολογικής πληροφορίας του λόγου, πραγματοποιήθηκε σε πρώτο στάδιο εκτενής μελέτη παρεμφερών εργασιών και ερευνών με σκοπό την επιλογή των κατευθύνσεων δράσης. Από τη μελέτη αυτή προέκυψε πως για την ανάπτυξη ενός τέτοιου συστήματος, η εφαρμογή ενός συνδυασμού επιμέρους μοντέλων ανάλυσης ήχου, εικόνας και κειμένου είναι σε θέση να πετύχει καλύτερες επιδόσεις σε σχέση με τα συστήματα που αξιοποιούν ξεχωριστά κάθε μια από αυτές τις αναλύσεις. Με γνώμονα τα παραπάνω πραγματοποιήθηκε ξεχωριστή διαχείριση των παραπάνω τριών ειδών πληροφορίας με σκοπό την βέλτιστη επιλογή αρχιτεκτονικών και αναπαραστάσεων δεδομένων, όπως περιγράφεται στην πειραματική διαδικασία του κεφαλαίου 5.

Η πρώτη διαδικασία μελέτης σχετίζεται με την αξιοποίηση του ηχητικού σήματος από συνελικτικά βαθιά νευρωνικά δίκτυα για την ανάπτυξη ενός συστήματος αξιολόγησης της κατάθλιψης όπως μελετάται εκτενώς στις παραγράφους 3.2.1 και 4.3. Αντικείμενο μελέτης σε αυτό το βήμα ήταν η εξαγωγή των κατάλληλων χαρακτηριστικών ήχου και, αφού τροφοδοτηθούν στο μοντέλο, η επιλογή του συνδυασμού αυτών που θα πετυχαίνουν την καλύτερη επίδοση. Τα χαρακτηριστικά που αποφάσισαν να εξαχθούν, για κάθε ηχητικό κλιπ είναι προσωδιακά και φασματικά, αλλά και χαρακτηριστικά που υποδεικνύουν την ποιότητα της φωνής. Τα χαρακτηριστικά αυτά ενσωματώθηκαν σε μορφή πινάκων αφού εξήχθησαν ανά παράθυρο των 10 ms για κάθε δείγμα και στην συνέχεια συνδυάστηκαν με διάφορους τρόπους για να τροφοδοτήσουν το τελικό μοντέλο ήχου. Στη συγκεκριμένη μελέτη ωστόσο, εμπνευσμένοι από προηγούμενες σχετικές μελέτες προσεγγίσαμε την αναπαράσταση 1D ακουστικών χαρακτηριστικών, αντί της συνήθους 2D, εφαρμόζοντας διάφορες στατιστικές συναρτήσεις στους εξαγόμενους πίνακες των χαμηλού επιπέδου χαρακτηριστικών δημιουργώντας υψηλού επιπέδου αναπαραστάσεις διάστασης 1D. Τα πλεονεκτήματα τις 1D προσέγγισης έναντι της 2D, παρουσιάζονται στο γεγονός πως το νευρωνικό δίκτυο στο οποίο θα τροφοδοτηθούν μαθαίνει να δημιουργεί σχέσεις ανάμεσα στα δεδομένα που του δίνονται απευθείας από το ηχητικό σήμα σε αντίθεση με τις 2D που περνούν απο διαδικασίες εφαρμογής φίλτρων για την εκμάθηση μοτίβων, καθώς επίσης και στο γεγονός πως το υπολογιστικό κόστος μειώνεται άρα και η απαίτηση για πολλά δεδομένα εκπαίδευσης [81]. Έτσι η προσέγγιση αυτή ταιριάζει πλήρως στις συνθήκες της μελέτης μας όπου τα δεδομένα μας ειναι περιορισμένα. Επιπλέον όπως παρουσιάσαμε στον πίνακα 3.2, όταν τα χαρακτηριστικά ήχου περάσουν από την κατάλληλη διαδικασία προεπεξεργασίας, τόσο για τον καθαρισμό τους, όσο και για την αφαίρεση της μη ουσιώδους πληροφορίας μειώνοντας την διαστατικότητα τους, το δίκτυο τροφοδότησης τους μαθαίνει καλύτερα μοτίβα με αποτέλεσμα την αποδοτικότερη κατάταξη των νέων δεδομένων που εισέρχονται.

Η δεύτερη προσπάθεια μελέτης που πραγματοποιήθηκε αφορούσε την δημιουργία ενός συστήματος που θα ‘μετέφραζε’ τις εκφράσεις του προσώπου των ασθενών. Από αυτή την

μελέτη προέκυψαν τέσσερις μεθοδολογίες αναπαράστασης της οπτικής πληροφορίας όπως αναπτύξαμε στην ενότητα 3.2.2. Πιο συγκεκριμένα σκοπός μας ήταν η εξαγωγή της δυναμικής πληροφορίας από τα χαρακτηριστικά του προσώπου, ώστε με την τροφοδότηση τους μετέπειτα σε κάποιο νευρωνικό δίκτυο να μαθευτούν μοτίβα στις εκφράσεις των ανθρώπων ικανά να συνεισφέρουν στην εκτίμηση της κατάθλιψης. Στον πίνακα 3.5 παρουσιάζεται η απόδοση αυτών των τεχνικών αλλά και συνδυασμοί τους, από τους οποίους αποδείχθηκε πως ο συνδυασμός των μεθόδων εξαγωγής MHH των AUs και εξαγωγής της μεταβολής των χαρακτηριστικών Eye Gaze και Head Pose, όσο και των πρώτων και δεύτερων παραγώγων τους, εκτιμούσε αποδοτικότερα τις τιμές PHQ8 των ασθενών.

Και στις δύο προσεγγίσεις επεξεργασίας της οπτικό-ακουστική πληροφορίας, εκμεταλλευτήκαμε την αποδοτικότητα των συνελικτικών βαθιών νευρωνικών δικτύων (DCNN) για να εξάγουμε τοπικά χαρακτηριστικά τα οποία θα μπορούσαν να εμφανιστούν σε κάποιο δείγμα οποιαδήποτε χρονική στιγμή χωρίς ωστόσο αυτό να επηρεάζει τον εντοπισμό τους. Τα φίλτρα που εφαρμόζουν στα δεδομένα εισόδου μπορούν να αναγνωρίσουν μη γραμμικούς συνδυασμούς χαρακτηριστικών και είναι ολοένα και πιο αφηρημένα για το πώς μπορούν να ανιχνεύσουν πρότυπα σε οποιαδήποτε θέση των δεδομένων εισόδου. Η αποδοτικότητα τους γίνεται εύκολα αντιληπτή συγκρίνοντας τις επιδόσεις των συστημάτων αυτών ανάμεσα σε έναν Baseline εκτιμητή και σε βαθιές συνελικτικές αρχιτεκτονικές με διαφορετικές παραμέτρους στους πίνακες 4.1,4.2,4.3,4.4. Γίνεται αντιληπτό λοιπόν πως όταν έχουμε να αντιμετωπίσουμε πληροφορία που ‘απλώνεται’ στο χρόνο, ο συνδυασμός εξαγωγής χαρακτηριστικών από την δυναμική πληροφορία που δίνουν τα δεδομένα εισόδου και η τροφοδότησή τους σε βαθιές συνελικτικές αρχιτεκτονικές βελτιώνει σημαντικά την απόδοση του προβλήματος μας.

Η τρίτη προσέγγιση μας αφορούσε την εξαγωγή σημασιολογικής πληροφορίας από τα κείμενα, τις απομαγνητοφωνημένες συνεντεύξεις δηλαδή των ασθενών. Οι διανυσματικές αναπαραστάσεις λέξεων είναι ένα πολύτιμο εργαλείο για την επεξεργασία κειμένου. Μεταφράζουν σημασιολογικές και συντακτικές σχέσεις μεταξύ λέξεων σε γραμμικές ιδιότητες ενός διανυσματικού χώρου. Είναι ανεξάρτητες εφαρμογής, και ακόμα και απλή άθροισή τους δίνει συγκρίσιμα αποτελέσματα με τις κλασσικές μεθόδους, στο πρόβλημα της ανάλυσης συναισθήματος. Για την εξαγωγή λοιπόν διανυσματικών αναπαραστάσεων πειραματιστήκαμε με δύο βασικές μεθόδους οι οποίες αφορούσαν την δημιουργία διανυσμάτων από τις απαντήσεις που έδιναν οι ασθενείς σε πέντε διαφορετικές ερωτήσεις σχετιζόμενες με πέντε διαφορετικά συμπτώματα που θα μπορούσαν να οδηγήσουν στην έκβαση διάγνωσης για την κατάθλιψη. Η πρώτη μέθοδος την οποία μελετήσαμε στην ενότητα 3.3.1.2 ήταν η εξαγωγή Paragraph Vector περιγραφητών. Για την δημιουργία του μοντέλου όμως από το οποίο εξήγαμε τους περιγραφητές αυτού για τις πέντε προτάσεις, αξιοποιήσαμε έναν συνδυασμό από συλλεγμένα δεδομένα τόσο από το Twitter που αφορούσαν συναισθηματικές καταστάσεις των χρηστών, όσο και από το imdb που αφορούσαν κριτικές ταινιών και έδειχναν την συναισθηματική κατασταση του χρήστη. Αφού λοιπόν αναπαραστήσαμε τις προτάσεις αυτές με διανυσματικές αναπαραστάσεις από το μοντέλο που δημιουργήσαμε, μελετήσαμε την ικανότητα SVM ταξινομητών να κατηγοριοποιήσουν την ύπαρξη ή οχι αυτών των συμπτωμάτων στα δείγματα επαλήθευσης. Ενω λοιπον στα δυο πρώτα συμπτώματα (Prior Depression, PTSD), που οι απαντήσεις των ασθενών ήταν ως επι τω πλείστων απόλυτες Ναι ή Όχι, η εφαρμογη των SVM στους PV περιγραφητές ήταν αποτελεσματική, στα επόμενα τρία συμπτώματα (Sleep, Feeling, Personality) που οι απαντήσεις δεν ήταν άμεσες, αλλά αμφιλεγόμενες, το μοντέλο PV-SVM έδωσε φτωχά αποτελέσματα. Η δεύτερη προσέγγιση η οποία και τελικά καθιερώθηκε στο τελικό μοντέλο, πέτυχε υψηλότερες επιδόσεις, καθώς βασίστηκε στην ανάλυση Σημασιολογικού Περιεχομένου. Ουσιαστικά δημιουργήσαμε ένα λεξικό με πιθανές λέξεις

και εκφράσεις των απαντήσεων των ασθενών και η παρουσία τους καθόριζε την ύπαρξη ή οχι του αντίστοιχου συμπτώματος. Η προσέγγιση αυτή σε έναν μεγαλύτερο όγκο δειγμάτων θα μπορούσε να είναι πολυ δαπανηρή και τελικά να αποδεικνύονταν δυσμενέστερη της PV-SVM, ωστόσο στη δική μας μελέτη που τα δείγματα ασθενών είναι πολυ περιορισμένα σε συνδυασμό και με το εκτενές λεξικό που δημιουργήσαμε, δίνει πολύ καλή και αξιοπρεπή απόδοση στα νέα δείγματα. Ιδιαίτερα συγκρίνοντας την με την όμοια state-of-the-art προσέγγιση στα δεδομένα κειμένου που υλοποίησε ο Sun κ.α. [21], παρουσιάζει καλύτερη απόδοση.

Συνολικά λοιπόν όσον αφορά την προεπεξεργασία των δεδομένων μας μπορούμε να αναφέρουμε κάποια πιο γενικά συμπεράσματα που προέκυψαν από την μελέτη μας. Πιθανότατα το καθοριστικότερο για την αποδοτικότητα των τελικών χαρακτηριστικών που εξήγαμε ήταν η μείωση της διαστατικότητας τους γιατί ο λόγος samples / number of features είναι αρκετά μικρός ,ενώ τυπικά πρέπει έχουμε πολλαπλάσια ή τάξη μεγέθους περισσότερα samples από features για την γενίκευση ενός μοντέλου. Βέβαια τόσο ποιους μετασχηματιστές προεπεξεργασίας θα εφαρμόσουμε, όσο και ποιες τιμές των παραμέτρων τους (variance threshold, αριθμός κυρίων συνιστωσών) θα αποδώσουν καλύτερα δεν το ξέρουμε από την αρχή, και μπορεί να είναι διαφορετικό ανάλογα τον εκτιμητή, ακόμα και στο ίδιο dataset. Έχουμε μόνο κάποιες εμπειρικές γνώσεις όπως ότι η κανονικοποίηση γενικά βοηθάει, ότι τα samples πρέπει να είναι αρκετά περισσότερα από τα features κλπ. Αντίστοιχα στην περίπτωση μας, τόσο για διαφορετικού είδους πληροφορίας, όσο και για διαφορετικά φύλα οι βέλτιστοι μετασχηματιστές προεπεξεργασίας του κάθε μοντέλου διέφεραν μεταξύ τους.

Τέλος λοιπόν αποσκοπώντας στον συνδυασμό των παραπάνω μοντέλων για την ανάδειξη ισχυρών προβλέψεων έναντι ασθενέστερων, προτείναμε στην ενότητα 5.2 μια υβριδική πολυτροπική προσέγγιση του προβλήματος αξιολόγησης της κατάθλιψης. Συνδυάζοντας λοιπόν τις τρεις προαναφερθείσες μεθόδους με ένα σύστημα ταξινόμησης της κατάθλιψης βάση της πληροφορίας του κειμένου, πετύχαμε σημαντικές αποδόσεις στην αξιολόγηση της κατάθλιψης που μεμονωμένες οι παραπάνω μέθοδοι δυσκολευόντουσαν να φτάσουν. Αποδεικνύεται επομένως πως ο συνδυασμός τόσο διαφορετικών μεθόδων της ίδια κατηγορίας (παλινδρόμησης ή ταξινόμησης), όσο και ο συνδυασμός μεθόδων κι από τις δύο κατηγορίες συνεισφέρει σημαντικά στην απόδοση ενός προβλήματος. Με την υβριδική προσέγγιση, οι αποδόσεις μιας μεθόδου συμπληρώνουν τις αδυναμίες της άλλης και με αυτό τον τρόπο η υβριδική μέθοδος γενικεύει πολύ καλύτερα στις διάφορες περιπτώσεις ασθενών με καλύτερη συνολική προσέγγιση των PHQ8 τιμών για όλους τους ασθενείς. Ωστόσο, για την επιλογή των κατάλληλων μοντέλων, λόγω ανεπαρκούς πλήθους δειγμάτων που μας διέθετε ο οργανισμός AVEC2017 [10], έγιναν πολλοί πειραματισμοί τόσο στους μετασχηματιστές των δεδομένων εισόδου όσο και στις υπερπαραμέτρους των εκτιμητών, έχοντας ως στόχο όχι μόνο την βελτίωση των μετρικών απόδοσης αλλά και την αποφυγή της υπερεκπαίδευσης (overfitting) αλλά και της υπεργενίκευσης συγχρόνως. Όταν ένα σύστημα εκπαιδεύεται πάνω σε ανεπαρκή δείγματα, ακόμη και αν κάνουμε υπερδειγματοληψία που ο ρόλος της είναι να εξομαλύνει την κατανομή στις διάφορες κλάσεις (ταξινόμηση) ή τιμές (παλινδρόμηση), τότε δυσκολεύεται να μάθει σχέσεις που να καλύπτουν πιθανές μελλοντικές παρατηρήσεις εισόδου καθώς αρέσκεται σε έναν περιορισμένο αριθμό περιγραφητών βασισμένων στα λιγοστά δείγματα που του τροφοδοτήθηκαν για την εκπαίδευση. Συγκρίνοντας λοιπόν τις γραφικές παραστάσεις στην ενότητα 5.3 παρατηρούμε πως τα Baseline συστήματα ενώ παρουσιάζουν αξιοπρεπείς μετρικές απόδοσης, εμφανίζεται έντονο το πρόβλημα της υπεργενίκευσης καθώς όλες οι προβλεπόμενες τιμές κυμαίνονται κοντά στη μέση τιμή των δειγμάτων εκπαίδευσης. Στις προτεινόμενες μεθόδους ενώ οι μετρικές απόδοσης παρουσιάζουν βελτίωση στις μετρικές απόδοσης, κάποιος θα έλεγε πως η πολυπλοκότητα των βαθιών νευρωνικών δικτύων του

δεν ανταποκρίνεται στο μέγεθος βελτίωσης, η οποία είναι μικρή, των μετρικών αυτών. Εδώ λοιπόν έρχεται η σημασία της μελέτης των γραφικών απεικονίσεων των προβλεπόμενων τιμών με τις πραγματικές. Παρατηρώντας τη σχέση των προβλεπόμενων-πραγματικών τιμών, βλέπουμε πως πλεον η διακύμανση των προβλεπόμενων τιμών αυξάνεται σημαντικά από την σχεδόν μηδενική τιμή που είχε, ενώ η απόκλιση των τιμών παρουσιάζει μείωση προσεγγίζοντας την καλύτερη δυνατή ισορροπία μεταξύ απόκλισης και διακύμανσης (bias-variance trade off).

6.2 Προτάσεις

΄Ενας από τους σημαντικότερους παράγοντες στην αποτελεσματική ανάπτυξη ενός συστήματος Βαθιάς Μηχανικής Μάθησης είναι το μέγεθος του συνόλου δεδομένων που θα χρησιμοποιηθούν. Για το πρόβλημα επιβλεπόμενης μάθησης της εκτίμησης της κατάθλιψης που προσπαθήσαμε να επιλύσουμε, το μοναδικό dataset που μπορούσαμε να αξιοποιήσουμε αποτελούνταν από ένα σύνολο 189 δειγμάτων, υπερβολικά μικρό για τις απαιτήσεις ενός συστήματος βαθιάς μάθησης από τα οποία ένα ποσοστό περίπου 40% εκμεταλλεύθηκε για επαλήθευση και εξέταση. ΄Ετσι, ένας πρώτος στόχος επέκτασης της εργασίας αυτής θα ήταν η αναζήτηση ή ακόμα και η κατασκευή ενός αρκετά μεγαλύτερου συνόλου δεδομένων κατάλληλο για την πολυτροπική εκτίμηση της κατάθλιψης.

Όσον αφορά τα δεδομένα, ιδιαίτερο ενδιαφέρον θα εμφάνιζε η υλοποίηση της προσέγγισης μας αν εφαρμόζαμε συγχρονισμό των οπτικό-ακουστικών δεδομένων μας με την χρονική στιγμή των απαντήσεων των ασθενών στα συμπτώματα που εξετάσαμε.

Επιπλέον βελτίωση της απόδοσης του προβλήματος αυτού θα απέδιδε η καλύτερη αξιοποίηση της πληροφορίας του κειμένου. Πιο συγκεκριμένα η δημιουργία αποδοτικότερων Paragraph Vector περιγραφητών μέσω συλλογής περισσότερο αντιπροσωπευτικών αποθετηρίων δεδομένων αλλά και καλύτερης μελέτης στην προεπεξεργασία τους δίνοντας βαρύτητα στην ερμηνεία της αρνητικής σημασίας προτάσεων που δίνουν λέξεις όπως not, neither κλπ. , θα επέτρεπε εξόρυξη σημασιολογικής πληροφορίας από ολόκληρα τα κείμενα συνεντεύξεων και επομένως καλύτερή εκτίμηση της κατάθλιψης.

Μια διαφορετική προσέγγιση θα ήταν η προσθήκη αρχιτεκτονικής με ανατροφοδοτούμενα νευρωνικά δίκτυα (RNN): Τα αρχικά δεδομένα που χρησιμοποιήθηκαν ήταν σε μορφή frames (πλαισίων) από βίντεο, επομένως τα επί μέρους στιγμιότυπα που εξάχθηκαν ήταν μία ακολουθία (sequence) ανά συνεδρία. Το γεγονός αυτό ενισχύει τις πιθανότητες να πετύχουν ακόμα καλύτερα αποτελέσματα συστήματα υβριδικά που περιλαμβάνουν αρχιτεκτονικές CNN και RNN νευρωνικών δικτύων. Τα ανατροφοδοτούμενα νευρωνικά δίκτυα (RNN) διαθέτουν μνήμη. Η προσθήκη μνήμης, ωστόσο, σε ένα δίκτυο, έχει κάποιο σκοπό. Υπάρχουν πληροφορίες μέσα στις ακολουθίες εισόδου, τις οποίες χρησιμοποιούν τα ανατροφοδοτούμενα νευρωνικά δίκτυα προκειμένου να εκτελούν εργασίες που τα απλά προωθητικά δίκτυα δεν μπορούν. Τα δίκτυα αυτά έχουν την ικανότητα να εξάγουν συσχετίσεις μέσα από ακολουθίες δεδομένων, επομένως η χρήση τους μπορεί να οδηγήσει σε εξαγωγή σημαντικών χαρακτηριστικών που εκφράζουν την αλλαγή τη συναισθηματικής κατάστασης του χρήστη κατά τη διάρκεια της εκάστοτε συνεδρίας.

Τέλος εφικτή θα ήταν η αξιοποίηση και η εφαρμογή του συστήματος που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής εργασίας και σε άλλα προβλήματα. Παραδείγματος χάρη, η χρήση των συστημάτων και τεχνικών που εφαρμόστηκαν σε προβλήματα αναγνώρισής ψυχικών διαταραχών ή συναισθηματικής ανάλυσης γενικότερα. Μια άλλη χρήση του συστήματος θα μπορούσε να είναι η επέκταση της προσέγγισης μας

εκτίμησης της κατάθλιψης σε ένα αυτόματο σύστημα που συνομιλεί με ανθρώπους στα πλαίσια μιας συνεδρίας ψυχανάλυσης και εντοπίζει λεκτικές και μη, ενδείξεις κάποιας ψυχικής διαταραχής χωρίς την παρέμβαση κάποιου ανθρώπου σε πραγματικό χρόνο.

Βιβλιογραφία

[1] Jonathan Gratch, Ron Artstein, Gale M Lucas, Giota Stratou, Stefan Scherer, Angela Nazarian, Rachel Wood, Jill Boberg, David DeVault, Stacy Marsella, et al. 2014. The Distress Analysis Interview Corpus of human and computer interviews. In LREC. 3123–3128.

[2] Rosalind W. Picard. "Affective Computing". MIT press 321 (1995), σσ. 1-16.

[3] Ron Kohavi; Foster Provost (1998). «Glossary of terms». *Machine Learning* 30: 271–274.

[4] Michel Valstar, Jonathan Gratch, Bjorn Schuller, Fabien Ringeval, Dennis Lalanne, Mercedes Torres, Stefan Scherer, Giota Stratou, Roddy Cowie, and Maja Pantic. 2016. Avec 2016: Depression, mood, and emotion recognition workshop and challenge. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge. ACM, 3–10.

[5] Robert Plutchik. Emotion: Theory, research, and experience: Vol. 1. Theories of emotion. New York: Academic, 1980.

[6] Recognizing Action Units for Facial Expression Analysis, Yingli Tian Takeo Kanade Jeffrey F. Cohn

[7] https://imotions.com/blog/facial-action-coding-system/

[8] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jonathan Gratch, Arno Hartholt, Margaux Lhommet, Gal Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Alberto Rizzo, and Louis-Philippe Morency. 2014. SimSensei kiosk: A virtual human interviewer for healthcare decision support. In Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'14. ACM, Paris, France, 1061–1068.

[9] Kroenke K, Strine TW, Spritzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord. 2009; 114(1-3):163-73.

[10] Fabien Ringeval, Bjorn Schuller, Michel Valstar, Jonathan Gratch, Roddy Cowie, Stefan Scherer, Sharon Mozgai, Nicholas Cummins, Maximilian Schmid, and Maja Pantic. 2017. AVEC 2017: Real-life Depression, and Affect Recognition Workshop and Challenge. In Proceedings of the 7th International Workshop on Audio/Visual Emotion Challenge. ACM, 1–8.

[11] Jeffrey F Cohn, Tomas Simon Kruez, Iain Matthews, Ying Yang, Minh Hoai Nguyen, Margara Tejera Padilla, Feng Zhou, and Fernando De la Torre. 2009.Detecting depression from facial actions and vocal prosody. In Affective

Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on. IEEE, 1–7.

[12] Nicholas Cummins, Julien Epps, Michael Breakspear, and Roland Goecke. 2011. An investigation of depressed speech detection: Features and normalization. In Twelfth Annual Conference of the International Speech Communication Association.

[13] Michel Valstar, Jonathan Gratch, Bjorn Schuller, Fabien Ringeval, Dennis Lalanne, Mercedes Torres Torres, Stefan Scherer, Giota Stratou, Roddy Cowie, and Maja Pantic. 2016. Avec 2016: Depression, mood, and emotion recognition workshop and challenge. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge. ACM, 3–10.

[14] Xingchen Ma, Hongyu Yang, Qiang Chen, Di Huang, and Yunhong Wang. 2016. DepAudioNet: An Efficient Deep Model for Audio based Depression Classification. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge. ACM, 35–42.

[15] Anastasia Pampouchidou, Olympia Simantiraki, Amir Fazlollahi, Matthew Pediaditis, Dimitris Manousos, Alexandros Roniotis, Georgios Giannakakis, Fabrice Meriaudeau, Panagiotis Simos, Kostas Marias, et al. 2016. Depression Assessment by Fusing High- and Low-Level Features from Audio, Video, and Text. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, ACM, 27–34.

[16] Varun Jain, James L Crowley, Anind K Dey, and Augustin Lux. 2014. Depression estimation using audiovisual features and fisher vector encoding. In Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge. ACM, 87–91.

[17] L. Yang, H. Sahli, X. Xia, E. Pei, M. C. Oveneke, and D. Jiang, “Hybrid depression classification and estimation from audio video and text information,” in Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge. ACM, 2017, pp. 45–51

[18] URL:https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

[19] URL:https://medium.com/deep-math-machine-learning-ai/chapter-4-decision-trees-algorithms-b93975f7a1f1

[20] URL:https://www.datacamp.com/community/tutorials/random-forests-classifier-python

[21] Bo Sun, Yinghui Zhang, Jun He, Lejun Yu, Qihua Xu, Dongliang Li, and Zhaoying Wang. 2017. A Random Forest Regression Method With Selected-Text Feature For Depression Assessment. In Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge. ACM, 61–68.

[22] Y. Gong and C. Poellabauer, “Topic modeling based multimodal depression detection,” in Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, ser. AVEC ’17. New York, NY, USA: ACM, 2017, pp. 69–76.

[23] URL: https://www.datacamp.com/community/tutorials/ensemble-learning-python

[24] Warren S. McCulloch and Walter Pitts. “A logical calculus of the ideas immanent in nervous activity”. In: The bulletin of mathematical biophysics 5.4 (Dec. 1943), pp. 115–133. issn: 1522-9602

[25] Marvin Minsky and Seymour Papert. Perceptrons: An Introduction to Computational Geometry. Cambridge, MA, USA: MIT Press, 1969.

[26] F. Rosenblatt. “The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain”. In: Psychological Review (1958), pp. 65–386.

[27] Gilles Degottex, John Kane, Thomas Drugman, Tuomo Raitio, and Stefan Scherer. 2014. COVAREP – A collaborative voice analysis repository for speech technologies. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP. IEEE, Florence, Italy, 960–964.

[28] Stefan Scherer, Gale Lucas, Jonathan Gratch, Alberto Rizzo, and Louis-Philippe Morency. 2015. Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews. IEEE Transactions on A.ective Computing 7, 1 (January-March 2015), 59–73.

[29] Stefan Scherer, Giota Stratou, Gale Lucas, Marwa Mahmoud, Jill Boberg, Jonathan Gratch, Albert (Skip) Rizzo, and Louis-Philippe Morency. 2014. Automatic audiovisual behavior descriptors for psychological disorder analysis. Image and Vision Computing 32, 10 (October 2014), 648–658.

[30] M. Bulut, S. Lee, and S. Narayanan, “Recognition for synthesis: Automatic parameter selection for resynthesis of emotional speech from neutral speech,” 2008.

[31] F. Dellaert, T. Polzin, and A. Waibel, “Recognizing emotion in speech,” Proc. ICSLP, Philadelphia, PA, USA, pp. 1970–1973, 1996.

[32] V. Petrushin, “Emotion recognition agents in real world,” in AAAI Fall Symposium on Socially Intelligent Agents: Human in the Loop, 2000

[33] A. Paeschke, “Global trend of fundamental frequency in emotional speech,” ISCA - Speech Prosody, Nara, Japan (March 2004), vol. 18, pp. 671–674, 2004.

[34] D. Cairns and J. Hansen, “Nonlinear analysis and classification of speech under stressed conditions,” J. Acoust. Soc. Am., vol. 96, pp. 3392–3400, 1994.

[35] C. M. Lee, S. Yildirim, M. Bulut, A. Kazemzadeh, C. Busso, Z. Deng, S. Lee, and S. Narayanan, “Emotion recognition based on phoneme classes,” Proceedings of ICSLP, Jeju, Korea, 2004.

[36] Lawrence Rabiner και Ronald Schafer. Theory and Applications of Digital Speech Processing. 1st. Upper Saddle River, NJ, USA: Prentice Hall Press, 2011.

[37] Dimitrios Ververidis και Constantine Kotropoulos. “Emotional speech recognition: Resources, features, and methods”. Speech Communication 48.9 (2006), σσ. 1162-1181.

[38] URL: https://person2.sol.lu.se/SidneyWood/praate/whatform.html

[39] Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz και John G. Taylor. “Emotion recognition in human-computer interaction”. Signal Processing Magazine, IEEE 18.1 (2001), σσ. 32-80

[40] URL:https://machinelearningmastery.com/loss-and-loss-functions-for-training-deep-learning-neural-networks/

[41] URL:https://stats.stackexchange.com/questions/48267/mean-absolute-error-or-root-mean-squared-error

[42] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. “Learning representations by back-propagating errors”. In: Nature 323 (Oct. 1986), p. 533

[43] D. P. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization”. In: ArXiv e-prints (Dec. 2014).

[44] Nitish Srivastava et al. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: J. Mach. Learn. Res. 15.1 (Jan. 2014), pp. 1929–1958.

[45] S. Dhuria, “Natural language processing: An approach to parsing and semantic analysis,” International Journal of New Innovations in Engineering and Technology, 2015.

[46] A. M. TURING. “COMPUTING MACHINERY AND INLIGENCE”. In: Mind LIX.236 (1950), pp. 433–460.

[47] Joseph Weizenbaum. “ELIZA Computer Program for the Study of Natural Language Communication Between Man and Machine”. In: Commun. ACM 9.1 (Jan. 1966), pp. 36–45. issn: 0001-0782

[48] Yin Zhang, Rong Jin, and Zhi-Hua Zhou. “Understanding bag-of-words model: A statistical framework”. In: International Journal of Machine Learning and Cybernetics 1 (Dec. 2010), pp. 43–52.

[49] Jeff Mitchell and Mirella Lapata. “Composition in Distributional Models of Semantics”. In: Cognitive Science 34.8 (2010), pp. 1388–1429.

[50] Leonard E. Baum and Ted Petrie. “Statistical Inference for Probabilistic Functions of Finite State Markov Chains”. In: The Annals of Mathematical Statistics 37.6 (1966), pp. 1554–1563.

[51] KAREN SPARCK JONES. “A STATISTICAL INTERPRETATION OF TERM SPECIFICITY AND ITS APPLICATION IN RETRIEVAL”. In: Journal of Documentation 28.1 (1972), pp. 11–21.

[52] Yoshua Bengio et al. “A Neural Probabilistic Language Model”. In: J. Mach. Learn. Res. 3 (Mar. 2003), pp. 1137–1155.

[53] Z. S. Harris, “Distributional structure,” Word, vol. 10, no. 2-3, pp. 146–162, 1954.

[54] T. Mikolov et al. “Distributed Representations of Words and Phrases and their Compositionality”. In: ArXiv e-prints (Oct. 2013).

[55] Quoc Le and Tomas Mikolov, Distributed Representations of Sentences and Documents, International Conference on Machine Learning, 2014.

[56] URL: http://piyushbhardwaj.github.io/documents/w2v_p2vupdates.pdf

[57] URL:https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

[58] Md Nasir, Arindam Jati, Prashanth Gurunath Shivakumar, Sandeep Nallan Chakravarthula, and Panayiotis Georgiou. 2016. Multimodal and multiresolution depression detection from speech and facial landmark features. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge. ACM, 43–50.

[59] Hongying Meng, Di Huang, Heng Wang, Hongyu Yang, Mohammed AI-Shuraifi, and Yunhong Wang. 2013. Depression recognition based on dynamic facial and vocal expression features using partial least square regression. In Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge. ACM, 21–30.

[60] L. Yang, H. Sahli, X. Xia, E. Pei, M. C. Oveneke, and D. Jiang, “Hybrid depression classification and estimation from audio video and text information,” in Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge. ACM, 2017, pp. 45–51.

[61] H. Meng, N. Pears, and C. Bailey. A human action recognition system for embedded computer vision application. In IEEE Conference on Computer Vision and Pattern Recognition Workshop on Embedded Computer Vision, pages 1–6, 2007.

[62] Giota Stratou, Stefan Scherer, Jonathan Gratch, and Louis-Philippe Morency. 2015. Automatic nonverbal behavior indicators of depression and ptsd: the effect of gender. Journal on Multimodal User Interfaces 9, 1 (2015), 17–29.

[63] URL: https://github.com/eddieir/Depression_detection_using_Twitter_post

[64] URL: https://www.kaggle.com/ywang311/twitter-sentiment/data

[65] URL: https://www.kaggle.com/sid321axn/amazon-alexa-reviews

[66] URL: https://www.kaggle.com/ywang311/twitter-sentiment/data

[67] URL: https://www.kaggle.com/jcblaise/imdb-sentiments#train.csv

[68] URL: https://docs.python.org/3/library/re.html

[69] URL:https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html

[70] URL:https://keras.io/api/optimizers/learning_rate_schedules/

[71] D. P. Kingma, J. Ba: 'Adam a method for stochastic optimization', 2014

[72] URL:https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229

[73] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition. 1–9.

[74] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 815–823.

[75] Veena Mayya, Radhika M Pai, and MM Manohara Pai. 2016. Automatic Facial Expression Recognition Using DCNN. Procedia Computer Science 93 (2016), 453–461.

[76] Xingchen Ma, Hongyu Yang, Qiang Chen, Di Huang, and Yunhong Wang. 2016. DepAudioNet: An Efficient Deep Model for Audio based Depression Classification. In Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge. ACM, 35–42.

[77] James R Williamson, Thomas F Quatieri, Brian S Helfer, Gregory Ciccarelli, and Daryush D Mehta. 2014. Vocal and facial biomarkers of depression based on motor incoordination and timing. In Proceedings of the 4th InternationalWorkshop on Audio/Visual Emotion Challenge. ACM, 65–72.

[78] Mohammed Senoussaoui, Milton Orlando Sarria Paja, Joγo Felipe Santos, and Tiago H. Falk. 2014. Model Fusion for Multimodal Depression Classification and Level Detection. In AVEC@MM.

[79] Y. Gong and C. Poellabauer. Topic modeling based multi-modal depression detection. In Annual Workshop on Audio/Visual Emotion Challenge, 2017.

[80] L. Yang, D. Jiang, X. Xia, E. Pei, M. C. Oveneke, and H. Sahli, "Multimodal measurement of depression using deep learning models," in Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge. ACM, 2017, pp. 53–59

[81] Piczak (2015a) Piczak, K. (2015a). Environmental sound classification with convolutional neural networks. In 25th International Workshop on Machine Learning for Signal Processing (pp. 1-6).