



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

Ανάπτυξη μοντέλων μηχανικής μάθησης για την ανίχνευση της νόσου COVID-19 μέσω ηχητικών καταγραφών αναπνοής

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κατερίνα Α. Πετρούλια

Επιβλέπων : Κωνσταντίνα Σ. Νικήτα
Καθηγήτρια, ΕΜΠ

Αθήνα, Ιούλιος 2024



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

Ανάπτυξη μοντέλων μηχανικής μάθησης για την ανίχνευση της νόσου COVID-19 μέσω ηχητικών καταγραφών αναπνοής

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κατερίνα Α. Πετρούλια

Επιβλέπων : Κωνσταντίνα Σ.Νικήτα
Καθηγήτρια ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 9^η Ιουλίου 2024.

.....

Κωσταντίνα Νικήτα
Καθηγήτρια ΕΜΠ

.....

Αθανάσιος Βουλόδημος
Καθηγητής ΕΜΠ

.....

Γιώργος Στάμου
Καθηγητής ΕΜΠ

Αθήνα, Ιούλιος 2024

.....

Κατερίνα Α. Πετρούλια

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

Copyright © Κατερίνα Πετρούλια 2024

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Στο πλαίσιο της παρούσας διπλωματικής εργασίας αναπτύσσονται μοντέλα μηχανικής και βαθιάς μάθησης για την ανίχνευση της νόσου COVID-19, χρησιμοποιώντας ηχητικές καταγραφές αναπνοής. Η πανδημία της νόσου Covid-19 έχει έναν τεράστιο αντίκτυπο στην παγκόσμια υγεία και οικονομία και η ανάγκη για έναν εναλλακτικό τρόπο διάγνωσης είναι επιτακτική. Για το σκοπό αυτό εφαρμόζονται τεχνικές μη επιβλεπόμενης μηχανικής μάθησης (k-means, DBSCAN, Spectral Clustering, Hierarchical Clustering και Meanshift) για τον εντοπισμό και την κατηγοριοποίηση των κορυφών σε εισπνοές, εκπνοές ή θόρυβο. Η διαδικασία προ-επεξεργασίας περιλαμβάνει το φιλτράρισμα για την αφαίρεση θορύβου, τη λογαριθμική συμπίεση, τον υπολογισμό έντασης από το φασματόγραμμα και την αναγνώριση κορυφών στην κυματομορφή της έντασης μετά την χρήση φίλτρου Gauss. Για την ανάπτυξη του μοντέλου ανίχνευσης της νόσου COVID-19 υιοθετείται μια υβριδική προσέγγιση βασισμένη στη χρήση συνελκτικού νευρωνικού δικτύου που τροφοδοτείται με φασματογράμματα και μοντέλου XGBoost που δέχεται ως είσοδο μέσες διάρκειες εισπνοών, εκπνοών και συνολικών αναπνοών. Για την ανάπτυξη και την αξιολόγηση του μοντέλου χρησιμοποιείται το πληθοποριστικό σύνολο δεδομένων SMARTY4COVID. Το σύνολο δεδομένων κρίθηκε περιορισμένο και μη ισορροπημένο ενώ σε πολλές περιπτώσεις ο θόρυβος και ελλιπής καταγραφή των αναπνοών δυσκόλεψαν τη διαδικασία.

Λέξεις Κλειδιά: COVID-19 διάγνωση, Επεξεργασία σήματος, Μηχανική μάθηση, Λόγος διάρκειας εισπνοής/εκπνοής, Συσταδοποίηση, Υβριδικό μοντέλο διάγνωσης, Convolutional Neural Network (CNN), XGBoost

Abstract

In the context of this thesis, machine learning and deep learning models are developed for the detection of COVID-19 using respiratory sound recordings. The COVID-19 pandemic has had a tremendous impact on global health and the economy, making the need for an alternative diagnostic method imperative. For this purpose, unsupervised machine learning techniques (k-means, DBSCAN, Spectral Clustering, Hierarchical Clustering, and MeanShift) are applied to identify and classify peaks into inhalations, exhalations, or noise. The preprocessing procedure includes noise filtering, logarithmic compression, power envelope calculation from the spectrogram, and peak recognition in the intensity waveform after applying a Gaussian filter. A hybrid approach is adopted for developing the COVID-19 detection model, based on the use of a Convolutional Neural Network (CNN) fed with spectrograms and an XGBoost model that takes as input the average durations of inhalations, exhalations, and total breaths. The crowdsourced SMARTY4COVID dataset is used for the development and evaluation of the model. The dataset was found to be limited and unbalanced while in many cases, noise and incomplete breath recordings complicated the process.

Keywords: COVID-19 diagnosis, Signal processing, Machine learning, Inhalation/exhalation duration ratio, Clustering, Hybrid diagnostic model, Convolutional Neural Network (CNN), XGBoost.

Acknowledgements

Θα ήθελα να εκφράσω τις βαθύτατες ευχαριστίες μου στη σύμβουλο της διπλωματικής μου εργασίας καθ. Κωνσταντίνα Νικήτα. Επίσης θα ήθελα να ευχαριστήσω θερμά τον Δρ. Φάνη Γανιτίδη για την πολύτιμη βοήθειά, την καθοδήγηση και τις γνώσεις που μου προσέφερε καθόλη τη διάρκεια εκπόνησης της διπλωματικής μου εργασίας. Θα ήθελα ακόμη να ευχαριστήσω τους γονείς μου, την αδερφή μου και τους φίλους μου που πάντα με στηρίζουν.

Περιεχόμενα

Εισαγωγή.....	12
A.Θεωρητικό Μέρος	14
1.Covid 19	14
1.1 Παθολογία και διάγνωση.....	15
1.2 Απεικονιστικές εκδηλώσεις της Covid-19.....	18
2.Βήχας και Δείκτης $Ti/Ttot$ (χρόνος εισπνοής/συνολικός χρόνος αναπνοής)	23
2.1 Ανάλυση βήχα.....	23
2.2 Η Κλινική Σημασία του Κλασσικού Δείκτη $Ti/Ttot$	25
3.Επεξεργασία Δεδομένων	30
3.1 Φιλτράρισμα	31
3.2 Λογαριθμική Συμπίεση.....	33
3.3 Μετασχηματισμός Fourier(STFT) και φασματόγραμμα (spectrogram).....	35
3.4 Περίγραμμα ισχύος- Power Envelope	38
3.5 Συσταδοποίηση.....	41
4.Μηχανική Μάθηση.....	45
4.1 Εποπτευόμενη μάθηση	45
4.2 Μη εποπτευόμενη μάθηση	46
4.3 Ενισχυτική μάθηση.....	46
4.4 Βαθιά Μάθηση	47
4.4.2 Extreme Gradient Boosting (XGBoost)	51
5.Μετρικές Αξιολόγησης.....	53
6.Σχετική Έρευνα	55
B. Πρακτικό Μέρος.....	67
7.Τα δεδομένα.....	67
8.Μεθοδολογία	69
8.1 Προεπεξεργασία	69
8.2 Συσταδοποίηση.....	73
8.3 Συνελκτικό Νευρωνικό Δίκτυο.....	77
8.4 Εκπαίδευση του XGBoost Μοντέλου	80
9.Αποτελέσματα	82
9.1 Εντοπισμός κορυφών.....	82
9.2 Συσταδοποίηση.....	83
9.3 Μοντέλο CNN.....	86
9.4 Μοντέλο XGBoost	94
9.5 Συνδυασμός των δύο μοντέλων.....	95

10. Συμπεράσματα.....	96
-----------------------	----

Εικόνα 1: Ένα ενδεικτικό διάγραμμα που δείχνει τυπικές ή κοινές απεικονιστικές εκδηλώσεις τραυματισμού του πνεύμονα COVID-19. GGO[8].....	18
Εικόνα 2 Ένα επεξηγηματικό διάγραμμα που δείχνει τη φυσιολογική κυψελιδική-διάμεση-τριχοειδική διαπαφή με παθολογικά ευρήματα που σχετίζονται με πρόιμη και όψιμη πνευμονική λοίμωξη COVID-19 και σχετικές ακτινολογικές εκδηλώσεις. GGO. [20].....	20
Εικόνα 3 Ανάλυση κορυφών στα ενεργειακά περιβλήματα ηχητικών βήχα: από υγιή άτομο, με άσθμα, με βρογχίτιδα και Covid-19[18].....	24
Εικόνα 4 Φάσμα ενεργειών ήχων βήχα[18]	25
Εικόνα 5 Αρχικό Γράφημα Ήχου	70
Εικόνα 6 Γράφημα ήχου μετά την αποθρομβοποίηση και την λογαριθμική συμπίεση.....	71
Εικόνα 7 Γράφημα με περίγραμμα Έντασης.....	71
Εικόνα 8 Φασματόγραμμα	72
Εικόνα 9 Γράφημα που παρουσιάζει την εύρεση κορυφών και τον υπολογισμό των πλατών τους.....	72
Εικόνα 10 Δομή βασικού μοντέλου CNN	79
Εικόνα 11 Μεθοδολογία.....	81
Εικόνα 12 Συσταδοποίηση με αλγόριθμο K-mean.....	83
Εικόνα 13 Συσταδοποίηση με αλγόριθμο DBSCAN	84
Εικόνα 14 Συσταδοποίηση με αλγόριθμο Spectral	84
Εικόνα 15 Γραφήματα με κατηγοριοποίηση κορυφών με βάση τους αλγόριθμους συσταδοποίησης	85
Εικόνα 16 Γράφημα των μετρικών τιμών Silhouette	86
Εικόνα 17 Αρχική εκπαίδευση	89
Εικόνα 18 Εκπαίδευση αντίστροφης αρχιτεκτονικής.....	94

Εισαγωγή

Η πανδημία της νόσου Covid-19 έχει έναν τεράστιο αντίκτυπο στην παγκόσμια υγεία και οικονομία. Η ανάγκη για έναν εναλλακτικό τρόπο διάγνωσης της Covid-19 που είναι άμεσος, αυτόματος και μπορεί να χρησιμοποιηθεί μέσω κινητού τηλεφώνου χωρίς τη διαμεσολάβηση ειδικού, είναι επιτακτική. Η εξάπλωση της πανδημίας απαιτεί γρήγορη και αξιόπιστη διάγνωση για τον έλεγχο της διασποράς του ιού.

Οι παραδοσιακοί τρόποι διάγνωσης, όπως τα PCR τεστ, μπορεί να χρειάζονται πολλές ώρες ή ακόμα και ημέρες για να δώσουν αποτελέσματα, ενώ η εξάπλωση της νόσου απαιτεί ταχύτερη αντίδραση. Ένας αυτόματος τρόπος διάγνωσης που μπορεί να χρησιμοποιηθεί από το σπίτι επιτρέπει στους ανθρώπους να αποφύγουν την έκθεση σε πιθανά μολυσμένα περιβάλλοντα, όπως τα κέντρα διαγνωστικών εξετάσεων.

Ένας τέτοιος τρόπος διάγνωσης δεν απαιτεί δείγμα από τον ασθενή, καθιστώντας τον μη επεμβατικό. Επίσης, μπορεί να είναι προσβάσιμος σε άτομα που δεν έχουν πρόσβαση σε κλινικές. Η διάγνωση μέσω κινητού τηλεφώνου μπορεί να εξοικονομήσει πολύτιμο χρόνο και πόρους, που αλλιώς θα απαιτούνταν για εργαστηριακές εξετάσεις. Ενώ οι επαγγελματίες υγείας πρέπει να έρχονται σε επαφή με ασθενείς για τη διάγνωση, ένας αυτόματος τρόπος διάγνωσης θα τους προστάτευε από τον κίνδυνο μόλυνσης.

Συνολικά, η ανάπτυξη ενός γρήγορου, αυτόματου και προσβάσιμου τρόπου διάγνωσης της Covid-19 μέσω κινητού τηλεφώνου είναι απαραίτητη για τον έλεγχο της πανδημίας και την παροχή αξιόπιστης και άμεσης ιατρικής φροντίδας σε όσους το χρειάζονται, προστατεύοντας παράλληλα τους επαγγελματίες υγείας και το ευρύ κοινό από περαιτέρω εξάπλωση του ιού.

Η εργασία αυτή αποσκοπεί στην ανάπτυξη μη επεμβατικών τρόπων διάγνωσης για τον Covid-19 και την εφαρμογή τεχνικών διάγνωσης που βασίζονται σε ηχογραφήσεις αναπνοής ασθενών που πάσχουν από Covid-19, αλλά και υγιών ατόμων. Η διαδικασία περιλαμβάνει διάφορα στάδια, ξεκινώντας από τη συλλογή δεδομένων. Καταγράφονται ηχητικά δεδομένα τόσο στο νοσοκομείο όσο και μέσω crowdsourcing εφαρμογών.

Οι ηχογραφήσεις αναπνοής υπόκεινται σε επεξεργασία σήματος, με στόχο την εξαγωγή σημαντικών ποιοτικών χαρακτηριστικών, όπως ο μέσος χρόνος εισπνοής, εκπνοής και αναπνοής. Η επεξεργασία σήματος βοηθά στην ανίχνευση προτύπων και διακυμάνσεων στις ηχογραφήσεις.

Η τεχνολογία της βαθιάς μηχανικής μάθησης εφαρμόζεται για την ανάλυση των ηχητικών δεδομένων. Αλγόριθμοι μηχανικής μάθησης εκπαιδεύονται να αναγνωρίζουν μοτίβα στις ηχογραφήσεις που σχετίζονται με την Covid-19. Με βάση τα αποτελέσματα της ανάλυσης, προκύπτουν δείκτες που μπορούν να χρησιμοποιηθούν για τον προσδιορισμό της παρουσίας της νόσου. Οι τεχνικές αυτές επιτρέπουν μια γρήγορη και αυτόματη διάγνωση της Covid-19, ενθαρρύνοντας την αποτελεσματική παρακολούθηση και αντιμετώπιση της νόσου. Παράλληλα, αυτή η τεχνολογία μπορεί να χρησιμοποιηθεί για την εξέταση της αναπνοής γενικά και την αξιολόγηση της υγείας του αναπνευστικού συστήματος.

Α.Θεωρητικό Μέρος

Κεφάλαιο 1

Covid 19

Η ασθένεια COVID-19 είναι μια νέα εμφανιζόμενη ασθένεια που προκαλείται από ιό που ανήκει στην οικογένεια των κορονοϊών. Οι κοροναϊοί είναι μονόκλωνοι ιοί RNA. Κάτω από ένα ηλεκτρονικό μικροσκόπιο, είναι σχεδόν σφαιρικοί ή πλουμοειδής. Ως ένας από τους μεγαλύτερους ιούς RNA, το μέγεθος του γονιδιώματος του κορονοϊού είναι μεταξύ 26 και 32 kb. Ο Covid-19 εντοπίστηκε για πρώτη φορά στην πόλη Wuhan, στην επαρχία Hubei της Κίνας. Τον Δεκέμβριο του 2019, ο Παγκόσμιος Οργανισμός Υγείας (ΠΟΥ) ανακοίνωσε ότι ο ιός μπορεί να προκαλέσει αναπνευστική νόσο προκαλώντας συμπτώματα βήχα, πυρετού και πνευμονίας ενώ άλλα συχνά συμπτώματα είναι η δύσπνοια και η υποξία[1].

Η ασθένεια έχει εξαπλωθεί στην Κίνα αλλά και στον υπόλοιπο κόσμο καθιστώντας την πανδημική από τον WHO στις 30 Ιανουαρίου του 2020. Ο SARS-CoV-2 και οι υπόλοιποι ανθρώπινοι κοροναϊοί συγκαταλέγονται στην οικογένεια των Coronaviridae. Αυτοί οι ιοί προκαλούν στους ανθρώπους το αναπνευστικό σύνδρομο της Μέσης Ανατολής (MERS) ή Σοβαρό Οξύ Αναπνευστικό Σύνδρομο (SARS). Η SARS είναι επίσης μια ιογενής αναπνευστική νόσος που προκαλείται από τον SARS- κοροναϊό (SARS-CoV), ο οποίος αναφέρθηκε για πρώτη φορά το 2003 στη Νότια Κίνα και εξαπλώθηκε σε πολλές χώρες Παγκοσμίως. Επιπλέον, ανακοινώθηκαν για πρώτη φορά κρούσματα του ιού MERS στη Σαουδική Αραβία όπου προκλήθηκε ο θάνατος 858 ανθρώπων[2].

Από όσα είναι γνωστά για τον COVID-19, ένα μεγάλο ποσοστό των μολυσμένων ατόμων εμφανίζουν ελαφριά ή και καθόλου συμπτώματα, συμβάλλοντας σημαντικά στη μη σκόπιμη εξάπλωση της νόσου. Αυτή η κατάσταση απαιτεί άμεση και ακριβή διάγνωση της νόσου μέσω συχνών και εκτεταμένων δοκιμών για την πρόληψη εστιών. Ο Παγκόσμιος Οργανισμός Υγείας (WHO) έχει εντοπίσει αρκετά συμπτώματα της COVID-19 ασθένειας, όπως υψηλή θερμοκρασία, βήχας και αναπνευστικές δυσκολίες τα οποία έχει γνωστοποιήσει. Ωστόσο, αυτά τα συμπτώματα είναι κοινά για πολλά

αναπνευστικά νοσήματα και όχι απαραίτητα μοναδικά στον COVID-19, καθιστώντας δύσκολη την αυτοαξιολόγηση των ασθενών. Η συχνότερη μέθοδος για τη διάγνωση του SARS-CoV-2 χρησιμοποιεί αντίστροφη μεταγραφή-αλυσιδωτή αντίδραση πολυμεράσης (rRT-PCR) σε ρινοφαρυγγικά (NP) επιχρίσματα. Ωστόσο, η συλλογή δείγματος με το επίχρισμα NP είναι μια επεμβατική μέθοδος και δεν είναι ιδανική για προληπτικούς ελέγχους, προγνωστικούς σκοπούς και συνεχή παρακολούθηση, δεδομένου ότι απαιτεί στενή επαφή μεταξύ των παρόχων υγειονομικής περίθαλψης και των ασθενών. Αυτή η επαφή εισάγει σημαντικό κίνδυνο μετάδοσης του ιού σε παρόχους υγειονομικής περίθαλψης και άλλους ασθενείς και επιβαρύνει τα συστήματα υγειονομικής περίθαλψης[3].

Γι' αυτό τον λόγο κρίνεται απαραίτητη η εύρεση ενός πιο προσβάσιμου, άμεσου και ασφαλέστερου τρόπου διάγνωσης. Η διαδικασία αυτή ξεκινάει με την κατανόηση των συμπτωμάτων του κορωνοϊού, των μηχανισμών μόλυνσης του, της παθοβιολογίας του αλλά και των μεταβολών που επιφέρει στο ανθρώπινο σώμα και πιο συγκεκριμένα στο αναπνευστικό σύστημα.

1.1 Παθολογία και διάγνωση

Όταν ο ιός εισέρχεται στο σώμα, έρχεται σε επαφή με τους βλεννογόνους της μύτης, το στόμα και τα μάτια. Προσκολλάται στους υποδοχείς στα επιφανειακά κύτταρα του κυττάρου ξενιστή μέσω προσκολλήσεων ιικής γλυκοπρωτεΐνης. Όταν εισέρχεται σε ένα υγιές κύτταρο το χρησιμοποιεί για να δημιουργήσει νέα μέρη του ιού. Με αυτόν τον τρόπο πολλαπλασιάζονται και με τη σειρά τους οι νέοι ιοί και μολύνουν τα κοντινότερα κύτταρα.

Το αναπνευστικό σύστημα αποτελείται από την τραχεία η οποία διακλαδώνεται μέσα στους πνεύμονες και καταλήγει σε μικροσκοπικούς σάκους αερίων που ονομάζονται κυψελίδες. Οι κυψελίδες είναι υπεύθυνες για την ανταλλαγή αερίων στους πνεύμονες δηλαδή για την είσοδο οξυγόνου στην ροή του αίματος και η έξοδος του διοξειδίου του άνθρακα[4]. Ο SARS-Covid-2 έχει τη δυνατότητα να μολύνει τα επιθηλιακά κύτταρα του άνω ή του κάτω μέρους της αναπνευστικής οδού ενώ σε ορισμένες περιπτώσεις, η μόλυνση μπορεί να φτάσει μέχρι και τις κυψελίδες[5]. Πιο συγκεκριμένα η πρωτεΐνη του SARS-CoV-2 συνδέεται με το ACE2 υποδοχέα στην επιφάνεια των επιθηλιακών

κυττάρων του πνεύμονα, στα κυψελιδικά κύτταρα τύπου 2, και τα ενδοθηλιακά κύτταρα.

Μια ανθρώπινη διαμεμβράνη τύπου 2 η πρωτεάση της σερίνης (TMPRSS2) που είναι η πρωταρχική γλυκοπρωτεΐνη S, διευκολύνει την είσοδο του ιού στα κύτταρα ξενιστές. Ο ιός εισέρχεται στο κυτταρόπλασμα και απελευθερώνει το RNA του για αντιγραφή χρησιμοποιώντας τον μηχανισμό του κυττάρου ξενιστή, που έχει ως αποτέλεσμα τον πολλαπλασιασμό του ιού που εξαπλώνεται σε άλλα κύτταρα ξενιστές. Επομένως, τόσο το έμφυτο όσο και το προσαρμοστικό ανοσοποιητικό σύστημα ενεργοποιούνται και παράγονται οι προφλεγμονώδεις κυτοκίνες και οι χημειοκίνες[6]. Η επιφανειακή πεπλομερής πρωτεΐνη (πρωτεΐνη ακίδας) SARS-CoV-2, S γλυκοπρωτεΐνη, μπορεί να δράσει με δύο τρόπους για να εισέλθει στα κύτταρα-ξενιστές, τα οποία συνήθως είναι τα επιθηλιακά κύτταρα του πνεύμονα, τα ενδοθηλιακά κύτταρα και τα κυψελιδικά κύτταρα τύπου 2. Στη συνέχεια, ο SARS-CoV-2 αντιγράφει το RNA του, συνθέτει πρωτεΐνες και συναρμολογείται, με αποτέλεσμα τον πολλαπλασιασμό των ιών που εξαπλώνονται σε άλλα κύτταρα-ξενιστές[5]. Η κλινική στατιστική ανάλυση έδειξε ότι η χρόνια αποφρακτική πνευμονοπάθεια (ΧΑΠ) και το κάπνισμα αποτελούν παράγοντες κινδύνου μόλυνσης από SARS-CoV-2, καθώς μπορούν να προκαλέσουν τα επιθηλιακά κύτταρα και τα κυψελιδικά κύτταρα τύπου 2 να εκφράζουν περισσότερους υποδοχείς ACE2 σε επίπεδα γονιδίου, mRNA και πρωτεΐνης.

Η παραδοσιακή έμφυτη απόκριση του ανοσοποιητικού συστήματος στην μόλυνση από τον ιό περιλαμβάνει μακροφάγα, μονοκύτταρα, δενδριτικά κύτταρα και ουδετερόφιλα. Αυτά τα κύτταρα του ανοσοποιητικού συστήματος εκφράζουν μια ποικιλία υποδοχέων αναγνώρισης προτύπων (PRRs) που μπορούν να ανιχνεύσουν μοριακά μοτίβα επαγόμενα από παθογόνους ιούς (PAMPs). Οι υποδοχείς μετατροπής ενζύμου της αγγειοτενσίνης 2 (ACE2), στους οποίους συνδέεται ο κορονοϊός, βρίσκονται στις κυψελίδες τύπου 2, στα βλεφαροειδή και στα κύλικά κύτταρα στους αεραγωγούς.[7]

Με βάση τα κύτταρα που είναι πιθανότερο να προσβληθούν και μολυνθούν, ο SARS-CoV-2 μπορεί να χωριστεί σε τρεις φάσεις καθεμία από τις οποίες αντιστοιχεί σε διαφορετικό κλινικό στάδιο της νόσου:

- 1ο Στάδιο, Ασυμπτωματική φάση(1-2 μέρες μόλυνσης):

Ο εισπνεόμενος ιός SARS-CoV-2 συνδέεται με τα επιθηλιακά κύτταρα της ρινικής κοιλότητας και αρχίζει να αναπαράγεται. Όπως προαναφέραμε ο ACE2 είναι ο κύριος υποδοχέας τόσο για τον SARS-CoV 2 όσο και για τον SARS-CoV in vitro. Δεδομένα που αφορούν τον SARS-CoV υποδηλώνουν ότι τα βλεφαροειδή κύτταρα

είναι πρωτογενή μολυσμένα κύτταρα στους αγωγίμους αεραγωγούς. Ωστόσο, αυτή η έννοια μπορεί να χρειάζεται κάποια αναθεώρηση, καθώς το μονοκυτταρικό RNA υποδεικνύει χαμηλό επίπεδο έκφρασης ACE2 στα αγωγή κύτταρα αεραγωγού και καμία προφανή προτίμηση κυτταρικού τύπου. Σε αυτό το στάδιο υπάρχει τοπική εξάπλωση του ιού αλλά περιορισμένη έμφυτη ανοσολογική απόκριση και ο ιός μπορεί να ανιχνευθεί με ρινικά επιχρίσματα. Ακόμη και αν το ιικό φορτίο μπορεί να είναι χαμηλό, αυτά τα άτομα είναι μολυσματικά.

- Στάδιο 2, Ανώτερο Αναπνευστικό Σύστημα(επόμενες ημέρες)

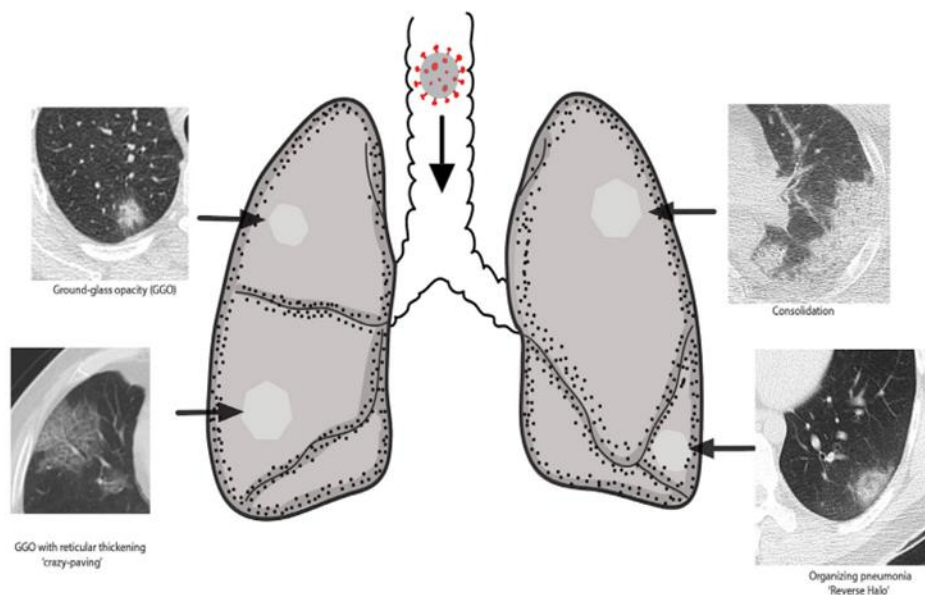
Ο ιός διαδίδεται και μεταναστεύει στην αναπνευστική οδό κατά μήκος των αγωγίμων αεραγωγών με αποτέλεσμα να ενεργοποιείται μια πιο ισχυρή έμφυτη ανοσολογική απόκριση. Σε αυτό το σημείο, η ασθένεια COVID-19 έχει εκδηλωθεί κλινικά. Τα μολυσμένα από ιούς επιθηλιακά κύτταρα αποτελούν κύρια πηγή βήτα και λάμδα ιντερφερονών. Το CXCL10 έχει επίσης σημειωθεί ως χρήσιμος δείκτης για την νόσο SARS. Για περίπου το 80% των μολυσμένων ασθενών, η νόσος θα είναι ήπια και θα περιορίζεται κυρίως στο άνω αναπνευστικό σύστημα. Αυτά τα άτομα μπορούν να παρακολουθούνται στο σπίτι με την αντιμετώπιση των συμπτωμάτων τους.

- Στάδιο 3,Υποξία, αλλοιώσεις θολής υαλού και εξέλιξη σε ARDS:

Δυστυχώς, περίπου το 20% των μολυσμένων ασθενών θα εξελιχθεί στο στάδιο 3 της νόσου και θα αναπτύξει πνευμονικές διηθήσεις και ορισμένοι από αυτούς θα αναπτύξουν πολύ σοβαρή νόσο. Οι αρχικές εκτιμήσεις για το ποσοστό θνησιμότητας είναι περίπου 2%, αλλά αυτό ποικίλλει σημαντικά με την ηλικία. Τα ποσοστά θνησιμότητας και νοσηρότητας μπορούν να αναθεωρηθούν όταν οριστεί καλύτερα ο επιπολασμός των ήπιων και ασυμπτωματικών περιπτώσεων. Ο ιός φτάνει τώρα στις μονάδες ανταλλαγής αερίων του πνεύμονα και μολύνει τα κυψελιδικά κύτταρα τύπου II. Τόσο ο SARS-CoV όσο και η γρίπη μολύνουν κατά προτίμηση τον κυτταρικό τύπο II σε σχέση με κύτταρα τύπου I. Οι μολυσμένες κυψελιδικές μονάδες τείνουν να είναι περιφεριακές. Ο SARS-CoV διαδίδεται μέσα στα κύτταρα τύπου II, απελευθερώνεται μεγάλος αριθμός ιικών σωματιδίων και τα κύτταρα υφίστανται απόπτωση και πεθαίνουν. Καθώς απελευθερώνονται ιικά σωματίδια που μολύνουν κύτταρα τύπου II σε γειτονικές μονάδες είναι πιθανόν να (αυτο)παραχθεί μία πνευμονική τοξίνη. Κανονικά, τα κύτταρα τύπου II είναι τα πρόδρομα κύτταρα για τα κύτταρα τύπου I. Αυτή η υποτιθέμενη ακολουθία

γεγονότων έχει παρουσιαστεί στο μοντέλο ποντικών της πνευμονίας από γρίπη. Το παθολογικό αποτέλεσμα του SARS και της Covid-19 είναι η διάχυτη κυψελιδική βλάβη με υαλώδεις μεμβράνες πλούσιες σε ινώδες και μερικά γιγαντιαία πολυπύρρηνα κύτταρα. Η ανώμαλη επούλωση του τραύματος μπορεί να οδηγήσει σε πιο σοβαρές ουλές και ίνωση από άλλες μορφές ARDS [8].

Γενικά, η εξέλιξη της νόσου σε σοβαρό τραυματισμό του πνεύμονα και σε σύνδρομο οξείας αναπνευστικής δυσχέρειας (ARDS) εξαρτάται από παράγοντες που δεν είναι σαφώς κατανοητοί αλλά περιλαμβάνουν το συστατικό της κυτταροτοξικότητας, της λεμφοπενίας και την επακόλουθη φλεγμονώδη απόκριση.



Εικόνα 1: Ένα ενδεικτικό διάγραμμα που δείχνει τυπικές ή κοινές απεικονιστικές εκδηλώσεις τραυματισμού του πνεύμονα COVID-19. GGO[8]

1.2 Απεικονιστικές εκδηλώσεις της Covid-19

Τα απεικονιστικά ευρήματα της Covid-19 μπορούν να φανούν τόσο στην ακτινογραφία θώρακος (CXR) όσο και στην αξονική τομογραφία θώρακα. Οι αλλοιώσεις του τύπου θολής υάλου (GGO) είναι η πιο συχνή απεικονιστική εκδήλωση.[9] Σε μια από τις μεγαλύτερες μελέτες σύγκρισης της διαγνωστικής ικανότητας του CT μεταξύ 1014 ασθενών, η αξονική τομογραφία θώρακα ήταν μη φυσιολογική στο 88% ασθενών με COVID-19 κατά την παρουσίαση με ευαισθησία 97% σε σύγκριση με RT-PCR.[10] Η

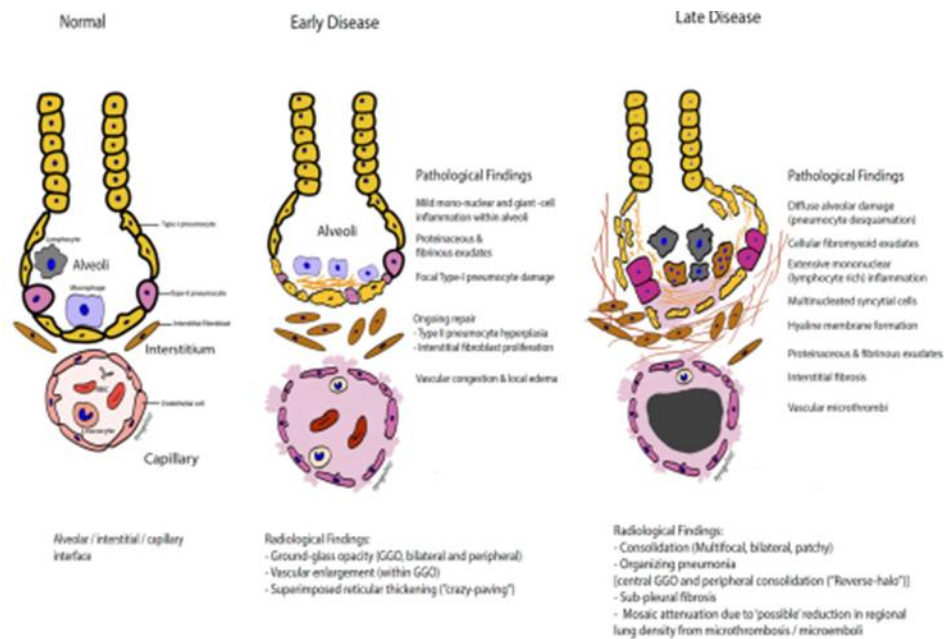
πολυλοβιακή προσβολή των πνευμόνων είναι συχνή [17] ενώ ο πιο συχνά προσβεβλημένος λοβός είναι ο κάτω δεξιός[11]. Καθώς ο παθογόνος εισπνέεται συνήθως μέσω των αναπνευστικών σταγονιδίων η νόσος είναι συνήθως σε βρογχοκεντρική κατανομή, δηλαδή αυτά τα μοτίβα είναι τυπικά κατά μήκος της κατανομής της αναπνευστικής οδού[9].

1.2.1 Κοινά απεικονιστικά χαρακτηριστικά

Οι αλλοιώσεις του τύπου θολής υάλου είναι η πιο κοινή αναφερόμενη απεικονιστική εκδήλωση COVID-19. Μια GGO ορίζεται ως αυξημένη εξασθένηση στην αξονική τομογραφία, η οποία δεν αποκρύπτει τις βρογχοαγγειακές δομές. Οι αμφίπλευρες αλλοιώσεις του τύπου θολής υάλου, κυρίως σε περιφερειακή κατανομή με επικράτηση κατώτερου λοβού, είναι ένα από τα τυπικά πρότυπα μόλυνσης από την COVID-19. Αυτές οι θολότητες τείνουν να είναι στρογγυλές στη μορφολογία, είναι συνήθως βρογχοκεντρικές ενώ εμπλέκονται κυρίως οι οπίσθιοι πνεύμονες[10]. Η δικτυωτή διάμεση πάχυνση στην COVID-19 είναι συνήθως ομαλή, περιφερειακή και μπορεί να παρατηρηθεί σε περιοχές με αλλοιώσεις του τύπου θολής υάλου [9]. Αυτό το μοτίβο συνήθως εμφανίζεται στην οξεία έως και τη χρόνια φάση της νόσου και πιθανότατα αντιπροσωπεύει τη στρατολόγηση διάμεσων φλεγμονωδών κυττάρων.

1.2.2 Ενοποίηση

Η ενοποίηση ορίζεται ως μια ομοιογενής αύξηση της εξασθένησης του πνευμονικού παρεγχύματος που έχει ως αποτέλεσμα την απόκρυψη των υποκείμενων αγγείων και βρόγχων. Η ενοποίηση είναι ένα άλλο κοινό χαρακτηριστικό μοτίβο πνευμονίας COVID-19 και είναι ενδεικτικό της εξέλιξης της νόσου. Παρόμοια με τις αλλοιώσεις του τύπου θολής υάλου, οι ενοποιήσεις είναι κατά κύριο λόγο περιφερικές και οπίσθιες με τη συμμετοχή των κατώτερων πνευμόνων.[12] Η σπηλαίωση είναι σπάνια και εάν υπάρχει, θα πρέπει να εξεταστεί το ενδεχόμενο βακτηριακής επιμόλυνσης.



Εικόνα 2 Ένα επεξηγηματικό διάγραμμα που δείχνει τη φυσιολογική κυψελιδική-διάμεση-τριχοειδική διεπαφή με παθολογικά ευρήματα που σχετίζονται με πρόιμη και όιμη πνευμονική λοίμωξη COVID-19 και σχετικές ακτινολογικές εκδηλώσεις. GGO. [20]

1.2.3 Σύγκριση με SARS και MERS

Πριν από μερικά χρόνια, είχαν κάνει την εμφάνιση τους δύο παρόμοιοι ιοί από την οικογένεια του κορονοϊού – το σοβαρό οξύ αναπνευστικό σύνδρομο (SARS) το 2003 και το αναπνευστικό σύνδρομο της Μέσης Ανατολής (MERS) το 2013. Τα παθολογικά ευρήματα των πνευμονικών λοιμώξεων από κορονοϊό (SARS και MERS) επικαλύπτονται με την τρέχουσα πανδημία COVID-19. Η κλινική εικόνα, ο τρόπος μετάδοσης και τα ευρήματα απεικόνισης της Covid-19 είναι αλληλεπικαλυπτόμενα με το SARS και το MERS. Οι αλλοιώσεις θολής υάλου σε περιφερικό μοτίβο με μεγαλύτερη επικράτηση στον κάτω λοβό είναι το πιο κοινό απεικονιστικό χαρακτηριστικό και στις τρεις ασθένειες. Ωστόσο η μονόπλευρη προσβολή των πνευμόνων, ειδικά στο πρώιμο στάδιο της νόσου, φαίνεται να είναι πιο συχνή σε SARS και MERS, ενώ η αμφοτερόπλευρη εμπλοκή των πνευμόνων παρατηρείται στην πλειονότητα των περιπτώσεων COVID-19. Επίσης, οι αρχικές ακτινογραφίες θώρακος φαίνεται να είναι θετικές σε ελαφρώς χαμηλότερο ποσοστό ασθενών με COVID-19 (55–60%) σε σύγκριση με τον SARS και τον MERS όπου το ποσοστό ανέρχεται περίπου στο 80%[13].

1.2.4 Εύρηματα από ιατροδικαστική αυτοψία

Μια 31χρονη γυναίκα βρέθηκε αναισθητη στο διαμέρισμά της κατά τη διάρκεια της πανδημίας COVID-19. Η μεταθανάτια αξονική τομογραφία (PMCT) αποκάλυψε διάχυτες, αμφοτερόπλευρες αλλοιώσεις θολής υάλου (bilateral GGO) που σχετίζονται με πανλοβικές ενοποιήσεις και βρογχογραφήματα. Τμήματα των πνευμόνων αποκάλυψαν ετερογενείς περιοχές υπόλευκης ενοποίησης χωρίς πυώδη έκκριση. Η ιστολογία των πνευμόνων έδειξε κυψελιδική βλάβη με γεωγραφικά ετερογενές μοτίβο, με το κεντρικό τμήμα των πνευμόνων να έχει επηρεαστεί περισσότερο. Στα ευρήματα παρατηρήθηκε κυρίως οίδημα, διάχυτη κυψελιδική βλάβη (DAD) στην εξιδρωματική φάση με την παρουσία υαλικών μεμβρανών σε μερικές κυψελίδες. Συνολικά το μοτίβο της βλάβης ήταν γεωγραφικά ετερογενές επηρεάζοντας κυρίως κεντρικά μέρη των πνευμόνων και διατηρώντας τα περιφερειακά μέρη ανέπαφα [14].

Σε ένα άλλο άρθρο αναφέρονται τα ευρήματα αυτοψίας 21 ασθενών με COVID-19 που νοσηλεύτηκαν στο Πανεπιστημιακό Νοσοκομείο της Βασιλείας και στο Cantonal Hospital Baselland στην Ελβετία. Η κύρια αιτία θανάτου ήταν η αναπνευστική ανεπάρκεια με εξιδρωματική διάχυτη κυψελιδική βλάβη και μαζική τριχοειδική συμφόρηση, που συχνά συνοδεύεται από μικροθρόμβους παρά την αντιπηκτική δράση. Δέκα περιπτώσεις έδειξαν υπερτιθέμενη βρογχοπνευμονία. Περαιτέρω ευρήματα περιελάμβαναν πνευμονική εμβολή (n=4), κυψελιδική αιμορραγία (n=3) και αγγειίτιδα (n=1). Όλοι οι ασθενείς διαγνώστηκαν με COVID-19 μέσω της χρήσης τεστ αλυσιδωτής αντίδρασης πολυμεράσης (PCR) για SARS-CoV-2 που πραγματοποιήθηκε σε ανασοφαρυγικό επίχρισμα, βρογχοκυψελιδικό υγρό πλύσης. Συχνότερα, οι ασθενείς παρουσίασαν ξηρό βήχα και πυρετό (n=16 και n=12, αντίστοιχα). Το ένα τρίτο των ασθενών παρουσίασαν σοβαρή τραχειοβρογχίτιδα. Τα μακροσκοπικά ευρήματα στους πνεύμονες ήταν ετερογενή, που κυμαίνονταν από αποσπασματικές έως διάχυτες περιοχές ενοποίησης έως σοβαρές και εκτεταμένες πυώδεις βρογχοπνευμονικές διηθήσεις. Σε όλες τις περιπτώσεις, το πνευμονικό παρέγχυμα ήταν βαρύ και συμπαγές με ανομοιόμορφο μπλε-κόκκινο χρώμα και ίχνη σοβαρής συμφόρησης. Το πιο σημαντικό ιστολογικό εύρημα ήταν η σοβαρή τριχοειδική συμφόρηση συνοδευόμενη από υαλώδεις μεμβράνες. [15].

1.2.5 Επίδραση Covid-19 στον βήχα

Όπως έχει προαναφερθεί ένα από τα πιο χαρακτηριστικά συμπτώματα της νόσου Covid-19 είναι ο ξηρός βήχας. Σε σύγκριση με άλλες αναπνευστικές ασθένειες, όπως βακτηριακές και ιογενείς πνευμονίες, η εκδήλωση της Covid-19 από πρώιμο στάδιο περιλαμβάνει φλεγμονή που περιορίζεται στις υποπεξωκοτικές ή περιβρογχαγγειακές περιοχές των πνευμόνων, εμφανίζοντας αλλοιώσεις θολής υάλου (GGOs) που συνήθως σχετίζονται με αγγειακή διαστολή.[10] Η εξέλιξη της Covid-19 σε μεταγενέστερα στάδια στους πνεύμονες μοιάζει με οξεία βρογχίτιδα καθώς ο SARS-CoV-2 μπορεί να εξαπλωθεί στο κάτω μέρος της αναπνευστικής οδού και να προκαλέσει φλεγμονή στις κυψελίδες ανοίγοντας έτσι το δρόμο προς δευτερογενείς καταστάσεις όπως η πνευμονία και το οξύ αναπνευστικό σύνδρομο (ARDS) [16]. Αυτό εμφανίζεται ως ακόμη ένα συχνό σύμπτωμα, αυτό της δύσπνοιας που οφείλεται στην συσσώρευση βλέννας και υγρού στις κυψελίδες με αποτέλεσμα να εμποδίζει την ανταλλαγή διοξειδίου του άνθρακα με οξυγόνο στον οργανισμό.

Κεφάλαιο 2

Βήχας και Δείκτης T_i/T_{tot} (χρόνος εισπνοής/συνολικός χρόνος αναπνοής)

2.1 Ανάλυση βήχα

Ο ήχος του βήχα αποτελείται από τρεις φάσεις: Φάση 1- Αρχική έκρηξη, Φάση 2- Θορυβώδης ροή αέρα και Φάση 3- Κλείσιμο γλωττίδας. Παρατηρήθηκε στο δείγμα βήχα υγιών ατόμων ότι η φάση 3 τελειώνει με τη δραστηριότητα των φωνητικών χορδών. Ακόμη παρατηρήθηκε ότι μετά τη Φάση 1, δηλαδή στην αρχική έκρηξη, τα επίπεδα ενέργειας είναι υψηλά σε υψηλότερες συχνότητες[17]. Από την ανάλυση κορυφών στο ενεργειακό περίβλημα ενός υγιούς βήχα προέκυψε ότι το δείγμα υγιούς βήχα είχε τον ελάχιστο αριθμό κορυφών σε σύγκριση με τους άλλους τρεις τύπους βήχα(βρογχίτιδα, άσθμα, Covid-19) όπως φαίνεται και στην εικόνα. Ξεκινά με μια κορυφή που ακολουθείται από μια περιοχή επίπλευσης με δύο μικρές κορυφές και πέφτει ακολουθώντας μια μεγάλη κορυφή[18].

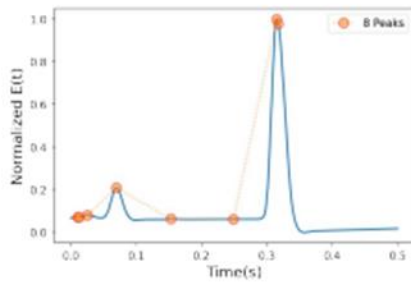


Figure 10: Peaks in Healthy Cough

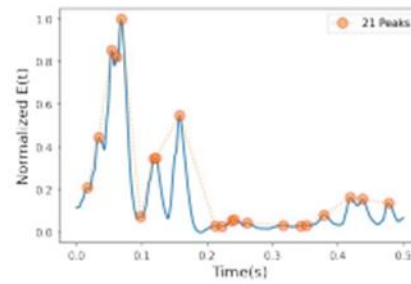


Figure 11: Peaks in Ashtma Cough

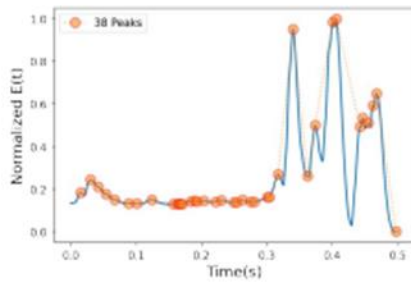


Figure 12: Peaks in Bronchitis Cough

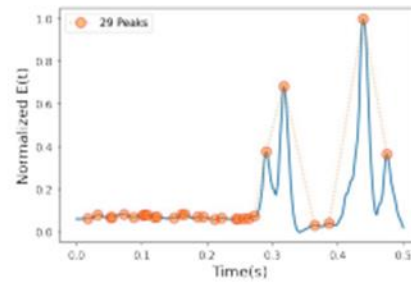


Figure 13: Peaks in COVID-19 Cough

Εικόνα 3 Ανάλυση κορυφών στα ενεργειακά περιβλήματα ηχητικών βήχα: από υγιή άτομο, με άσθμα, με βρογχίτιδα και Covid-19[18]

Το άσθμα και η βρογχίτιδα συγκαταλέγονται στην κατηγορία του υγρού βήχα (μεταφέρει βλέννα και τα πτύελα που προκαλούνται από βακτήρια ή ιούς, έκκριση στο κατώτερο αναπνευστικό σύστημα). Συγκεκριμένα, η δραστηριότητα των φωνητικών χορδών φαίνεται τυχαία, και η ενέργεια επεκτείνεται σε μια ευρύτερη ζώνη συχνοτήτων. Στην εικόνα το βρογχικό Ασθμα και η Βρογχίτιδα φαίνεται να επιδεικνύουν αυτά τα χαρακτηριστικά και τα επίπεδα ενέργειας. Στην εικόνα γίνονται εμφανείς ακόμη οι κορυφές στο περίγραμμα της ενέργειας του βήχα άσθματος και της βρογχίτιδας αντίστοιχα. Τα δείγματα βήχα της βρογχίτιδας εμφανίζουν τον μέγιστο αριθμό κορυφών. Ξεκινάει με μικρές κορυφές ενώ στο τέλος παρουσιάζει ψηλές κορυφές. Ενδιάμεσα στη φάση 2 παρατηρήθηκαν πολλαπλές μικρές κορυφές. Ωστόσο, τα δείγματα βήχα του άσθματος εμφανίζουν λιγότερες κορυφές από αυτά της βρογχίτιδας αλλά περισσότερες κορυφές από τους υγιείς και αυτούς με COVID-19, ξεκινώντας με πολλές μεγάλες κορυφές στη φάση 1, και τελειώνοντας με μικρές κορυφές στη φάση 3. Παρατηρείται ότι ο βήχας COVID-19 είναι συνεχής. Η ενέργεια κατανέμεται σε πολλές συχνότητες. Από την ανάλυση της κατανομής ενέργειας

πολλαπλών ήχων βήχα COVID-19, διαπιστώθηκε ότι η κατανομή ενέργειας ήταν υψηλή στη Φάση 2 και στη Φάση 3. Η ανώμαλη ταλαντωτική κίνηση στις φωνητικές χορδές μπορεί να παράγεται από την αλλοιωμένη αεροδυναμική πάνω από τη γλωττίδα λόγω αναπνευστικού ερεθισμού, τα οποία μπορούμε να δούμε στην εικόνα 3 που ακολουθεί[18].

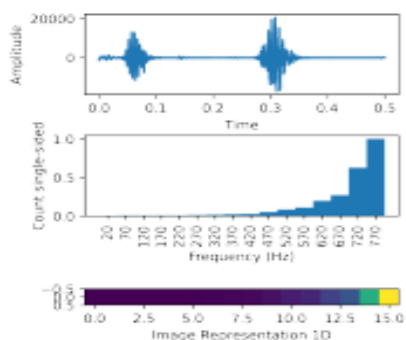


Figure 2: Healthy Cough

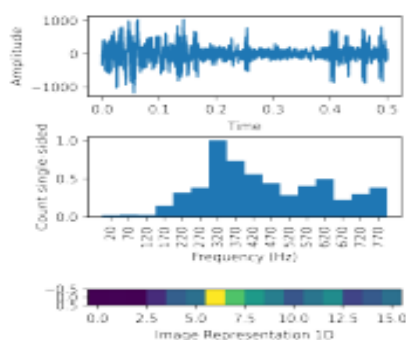


Figure 3: Asthma Cough

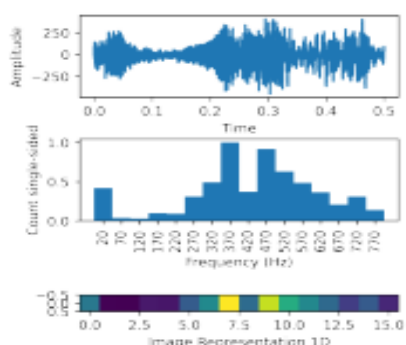


Figure 4: Bronchitis Cough

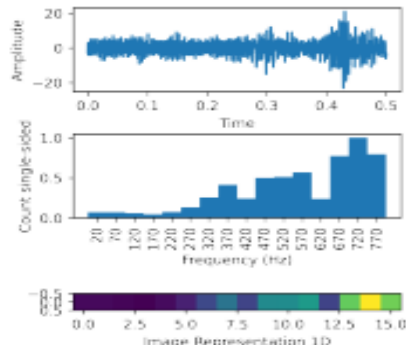


Figure 5: COVID-19 Cough

Εικόνα 4 Φάσμα ενεργειών ήχων βήχα[18]

2.2 Η Κλινική Σημασία του Κλασσικού Δείκτη T_i/T_{tot}

Ο δείκτης T_i/T_{tot} , ή αλλιώς ο λόγος του χρόνου της εισπνοής προς τον συνολικό χρόνο του αναπνευστικού κύκλου, είναι ένα μέτρο που χρησιμοποιείται για να αξιολογηθούν ανωμαλίες στην αναπνοή.

- T_i (Διάρκεια Εισπνοής): Το χρονικό διάστημα εισπνοής.

- T_{tot} (Συνολικός Χρόνος Αναπνοής): Ο συνολικός χρόνος ενός αναπνευστικού κύκλου, συμπεριλαμβανομένης της εισπνοής και της εκπνοής καθώς και το διάστημα μεταξύ τους.
- T_e (Διάρκεια Εκπνοής): Το χρονικό διάστημα εκπνοής.

Ο δείκτης T_i/T_{tot} αναφέρεται στον λόγο της διάρκειας της εισπνοής προς τον συνολικό χρόνο αναπνοής και έχει χρησιμοποιηθεί ως "κύκλος εργασίας" της αναπνευστικής συστοιχίας[19]. Αυτός ο δείκτης αντικατοπτρίζει το ποσοστό της διάρκειας κάθε κίνησης της αναπνοής που αφιερώνεται στην ανάπτυξη δύναμης από τους αναπνευστικούς μύες.

Επιπλέον, αφού αντανακλά την σχέση μεταξύ εισπνοής και εκπνοής, αυτός ο δείκτης παρέχει μια εκτίμηση του βαθμού απόφραξης των αεραγωγών και του επιπέδου stress που ασκείται στους αναπνευστικούς μύες[19]. Ο δείκτης T_i/T_{tot} συνδυάζεται συχνά με τον λόγο της μέσης πίεσης του διαφράγματος κατά τη διάρκεια μιας αναπνοής προς τη μέγιστη στατική πίεση του διαφράγματος, γνωστής ως P_{di}/P_{dimax} . Το προϊόν αυτών των δύο δεικτών ονομάζεται δείκτης τάσης-χρόνου του διαφράγματος (TT_{di}), με μια τιμή 0,15 να θεωρείται κρίσιμη για την κόπωση του διαφράγματος[20].

Ο δείκτης T_i/T_{tot} είναι σημαντικό εργαλείο για την αξιολόγηση των αναπνευστικών μοτίβων και την κατανόηση της αναπνευστικής λειτουργίας, ιδίως στο πλαίσιο της εκτίμησης της εργασίας των αναπνευστικών μυών και της πίεσης στους αεραγωγούς.

Σε μία άλλη μελέτη πραγματοποιείται σύγκριση της αναπνοής κατά την διάρκεια άσκησης μεταξύ ασθενών με διάμεση πνευμονοπάθεια και υγιών ατόμων. Το δείγμα αποτελούν 41 άτομα, όπου τα μισά είναι ασθενείς με διάμεση πνευμονοπάθεια και τα υπόλοιπα υγιή άτομα. Οι παρακολουθούμενοι παράμετροι περιλαμβάνουν τη Βιταλική Χωρητικότητα (VC), τη Μέγιστη Ισχύ (PO_{max}), την Αρτηριακή Οξυγόνωση (SaO_2), τον Πενταδιασταλτικό Όγκο (VT), τον Συνολικό Χρόνο Αναπνοής (T_{tot}), τον Χρόνο Εισπνοής (T_i) και το δείκτη T_i/T_{tot} .

Τα αποτελέσματα δείχνουν ότι οι τιμές T_i και T_{tot} μειώνονται σταδιακά κατά τη διάρκεια της άσκησης τόσο στους ασθενείς όσο και στα υγιή άτομα. Η μείωση του T_i συντονίζεται με αυτή του T_{tot} , με αποτέλεσμα να διατηρείται σταθερός ο λόγος T_i/T_{tot} στους ασθενείς (0,35 στην ανάπαυση, 0,34 στο PO_{max}), ενώ στα υγιή άτομα ο T_i μειώνεται λιγότερο από το T_{tot} , με αποτέλεσμα ο λόγος T_i/T_{tot} να αυξάνεται. Ο λόγος T_i/T_{tot} κατά τη διάρκεια της άσκησης σχετίζεται με τον Συνολικό Αναπνευστικό Όγκο άνω των 40 l/min και παρατηρείται ότι είναι χαμηλότερος στους πάσχοντες σε σχέση

με τους υγιείς. Συμπεραίνεται ότι η σύντομη διάρκεια της εισπνοής και ο υψηλός ρυθμός αναπνοής (σύντομο T_{tot}) εντοπίζονται ως αντίδραση στο ελαστικό φορτίο που ασκείται από τους πνεύμονες[21].

Επιπλέον με τη μελέτη τους στοχεύουν να διερευνήσουν την αναπνοή σε ασθενείς με μακροχρόνια συμπτώματα μετά από την νόσηση με Covid-19, οι οποίοι αντιμετωπίζουν ανεξήγητη δύσπνοια και εμφανίζουν φυσιολογικά αποτελέσματα στις μετρήσεις της πνευμονικής λειτουργίας, τόσο στην κατάσταση ανάπαυσης όσο και κατά τη διάρκεια της άσκησης. Οι παράμετροι εκτίμησης περιλαμβάνουν τον Χρόνο Εισπνοής (T_i) κατά τη διάρκεια της παροχής Αναπνευστικού Όγκου (VT), τον λόγο T_i/T_{tot} (Χρόνος Εισπνοής προς συνολικό Χρόνο Αναπνοής) καθώς και την εισπνεόμενη ροή, δηλαδή τον λόγο του VT (Παροχή Αναπνοής) προς τον T_i .

Τα αποτελέσματα δείχνουν ότι οι ασθενείς με μακροχρόνια συμπτώματα Covid-19 δεν διαφέρουν από τους υγιείς(control group) ως προς τον ρυθμό αναπνοής (RR) τόσο στην κατάσταση ανάπαυσης όσο και στο αποκορύφωμα της άσκησης, αλλά εμφανίζουν χαμηλότερες τιμές στην Παροχή Αναπνοής (VT) κατά το αποκορύφωμα. Επιπλέον, οι ασθενείς αυτοί εμφάνισαν σημαντικά υψηλότερες τιμές για τον λόγο T_i/T_{tot} τόσο στην κατάσταση ανάπαυσης όσο και κατά τη διάρκεια της άσκησης, ενώ είχαν χαμηλότερες τιμές για τον λόγο VT/VI κατά το αποκορύφωμα. Επιπλέον, στους ασθενείς με μακροχρόνια συμπτώματα Covid-19, ο λόγος T_i/T_{tot} κατά τη διάρκεια της άσκησης συσχετίζεται σημαντικά με τις τιμές της Ενδο-Τελικής Πίεσης του CO₂ (PETCO₂).

Συμπεραίνεται ότι οι ασθενείς με μακροχρόνια συμπτώματα Covid-19 έχουν ένα πρότυπο αναπνοής που είναι πιο επιρρεπές στη διαφραγματική κόπωση και λιγότερο αποτελεσματικό από αυτό των υγιών υποκειμένων. Η πλειονότητα της εργασίας της αναπνοής επιτυγχάνεται από το διάφραγμα. Μετά από μια ασθένεια, ιδίως εάν απαιτεί μηχανική υποστήριξη της αναπνοής ή σε συνθήκες γενικής φυσικής αποκατάστασης, το διάφραγμα φαίνεται να υποστηρίζει μεγάλο μέρος του έργου της αναπνοής[22].

Επιπλέον ο λόγος I:E (Inspiration to Expiration Ratio) αναφέρεται στο χρονικό διάστημα που κάθε φάση της αναπνοής καταλαμβάνει και παρέχει χρήσιμες πληροφορίες σχετικά με τον λόγο Χρόνου Εισπνοής (TI) προς Χρόνο Εκπνοής (TE), μπορώντας να χρησιμοποιηθεί για την περιγραφή διαταραχών της αναπνοής, καθώς μπορεί να ελλοχεύουν διαταραχές στην αναπνευστική ροή στο ανώτερο αναπνευστικό σύστημα[23].

Η φυσιολογική αναπνοή συνήθως παρουσιάζει λόγο I:E 1:2 ή 1:3 σε κατάσταση ανάπαυσης. Ενδεχομένως, η παρουσία αποφρακτικών προβλημάτων μπορεί να

οδηγήσει σε επιμήκυνση της εκπνοής ή της εισπνοής, οδηγώντας σε ανορθόδοξο λόγο I:E[24].

Σε ασθενείς με Χρόνια Αποφρακτική Πνευμονοπάθεια (ΧΑΠ), συνήθως παρουσιάζεται παρατεταμένη εκπνοή. Μια μελέτη που περιλαμβάνει 38 ασθενείς με ΧΑΠ και 43 υγιή υποκείμενα χρησιμοποίησε ένα αναλυτή πολυκαναλικού ήχου πνεύμονα (Stethographics) για να μετρήσει τον λόγο της διάρκειας της εισπνοής προς την εκπνοή (R1:I/E). Όπως ήταν αναμενόμενο, η παρατεταμένη εκπνοή ήταν πιο συνηθισμένη στους ασθενείς με ΧΑΠ από ό,τι στους υγιείς. Τιμές R1 μικρότερες από 0,6 ήταν συνήθεις στους ασθενείς με ΧΑΠ, αλλά χαμηλές τιμές R1 παρατηρήθηκαν και σε μερικούς υγιείς, ενώ σε μερικούς ασθενείς με ΧΑΠ παρατηρήθηκαν σχετικά υψηλές τιμές[25].

Μια άλλη μελέτη αναφέρει ότι ο λόγος I:E σε ασθενείς με ΧΑΠ είναι σημαντικά χαμηλότερος σε εκείνους με μέτρια σοβαρότητα σε σύγκριση με την ελαφριά σοβαρότητα, υποδεικνύοντας ότι η παρατεταμένη εκπνοή είναι χαρακτηριστική της μέτριας ΧΑΠ [26].

Ακόμη παρατηρούμε πως τα παραπάνω μετρούμενα μεγέθη έχουν κάποια συσχέτιση με μεγέθη που μετριοούνται και υπολογίζονται σε μία γνωστή διαγνωστική μέθοδο, την σπιρομέτρηση. Πιο συγκεκριμένα η σπιρομέτρηση είναι μια εξέταση που βοηθά στην διάγνωση πολλών προβλημάτων του αναπνευστικού όπως η χρόνια αποφρακτική πνευμονοπάθεια και ακόμη βοηθά στην παρακολούθηση της πορείας του ασθενή και της ανταπόκρισης αυτού στην θεραπεία του. Είναι ένα τεστ που εξετάζει κατά πόσο λειτουργούν σωστά οι πνεύμονες, άρα πόσο καλά πραγματοποιείται η διαδικασία της αναπνοής. Η λειτουργία της εισπνοής και της εκπνοής επηρεάζονται πάρα πολύ από τα ποικίλα αναπνευστικά προβλήματα, όπως το άσθμα, η χρόνια αποφρακτική πνευμονοπάθεια, η ιδιοπαθής πνευμονική ίνωση. Η μέτρηση αυτή γίνεται με ένα συγκεκριμένο όργανο που ονομάζεται σπιρόμετρο. Για να είναι όμως χρήσιμες οι μετρήσεις χρειάζονται πρώτα το φύλο, το ύψος και το βάρος του ασθενούς ώστε να μπορεί να συγκριθεί με άλλα ανθρωπομετρικά στοιχεία και να καθοριστούν οι φυσιολογικές τιμές για τους διάφορους δείκτες που εξετάζονται στην σπιρομέτρηση. Ο ασθενής εφόσον καθίσει, βάζει στο στόμα του ένα επιστόμιο που συνδέεται με το μηχάνημα, το οποίο καταγράφει κάθε δευτερόλεπτο τον όγκο, την ροή τον χρόνο του εισπνεόμενου και του εκπνεόμενου αέρα. Τα μεγέθη όπως που υπολογίζονται είναι εκπνεόμενος όγκος αέρα στο 1ο δευτερόλεπτο της εκπνοής, την βίαιη εκπνεόμενη χωρητικότητα και το πηλικο αυτών των δύο καθώς και τις εκπνευστικές ροές, είναι

αυτά που συνήθως μας οδηγούν στην διάγνωση[27]. Οι μετρικές αυτές είναι πιθανό να συσχετίζονται με την διάρκεια εισπνοής, την διάρκεια εκπνοής και τη συνολική διάρκεια αναπνοής που βρίσκεται και υπολογίζεται παρακάτω ώστε να τροφοδοτηθεί το XGBoost μοντέλο. Επιπλέον είναι δυνατόν οι όγκοι και χωρητικότητες εισπνοών και εκπνοών να μεταφράζονται με κάποιον τρόπο στις συχνότητες που υπάρχουν στα ηχητικά και κατ'επέκταση να εκλαμβάνονται σαν σύνθετα μοτίβα στα φασματογράμματα από ένα νευρωνικό δίκτυο. Στην παρακάτω ενότητα έγινε μια προσπάθεια εφαρμογής των γνώσεων που αφορούν τους αναπνευστικούς ρυθμούς.

Κεφάλαιο 3

Επεξεργασία Δεδομένων

Η επεξεργασία σήματος είναι ένα θεμελιώδες πεδίο μελέτης που παίζει κρίσιμο ρόλο σε ένα ευρύ φάσμα τομέων, συμπεριλαμβανομένων των τηλεπικοινωνιών, της επεξεργασίας ήχου και βίντεο, της βιοϊατρικής μηχανικής και της ανάλυσης δεδομένων. Η σημασία της, πηγάζει από την ικανότητά του να εξάγει πολύτιμες πληροφορίες από σήματα και να επιτρέπει την αποτελεσματική ανάλυση και ερμηνεία.

Στο πλαίσιο αυτής της εργασίας, η επεξεργασία σήματος έχει τεράστια σημασία για την κατανόηση και την εξαγωγή ουσιαστικών γνώσεων από πολύπλοκα σύνολα δεδομένων. Η μελέτη αυτή επικεντρώνεται στην εξαγωγή χρήσιμων χαρακτηριστικών που θα βοηθήσουν στην καλύτερη κατανόηση των μοτίβων της φυσιολογικής και της μη φυσιολογικής αναπνοής και κατ'επέκταση στην μη επεμβατική και ταχύτερη διάγνωση της Covid-19. Τα σήματα που συλλέχθηκαν παρουσιάζουν περίπλοκα μοτίβα και χαρακτηριστικά που απαιτούν περίπλοκες τεχνικές ανάλυσης και ερμηνείας.

Οι πρωταρχικοί στόχοι και στόχοι της επεξεργασίας σήματος στη μελέτη αυτή είναι διπλοί. Πρώτον, στοχεύει στην προεπεξεργασία των ακατέργαστων δεδομένων μειώνοντας το θόρυβο, βελτιώνοντας τα σχετικά χαρακτηριστικά και εξαλείφοντας τα τεχνουργήματα που θα μπορούσαν να επηρεάσουν την επακόλουθη ανάλυση. Αυτό το βήμα προεπεξεργασίας είναι ζωτικής σημασίας για τη διασφάλιση της ακρίβειας και της αξιοπιστίας των αποτελεσμάτων που προκύπτουν από περαιτέρω ανάλυση.

Δεύτερον, η επεξεργασία σήματος στην έρευνά στοχεύει στην εξαγωγή σχετικών πληροφοριών και στην αποκάλυψη κρυμμένων μοτίβων μέσα στα δεδομένα. Εφαρμόζοντας προηγμένους αλγόριθμους και τεχνικές, όπως υπερπαραμετρικό φίλτράρισμα, λογαριθμική συμπίεση, ανάλυση Fourier, επιδιώκεται η αποκάλυψη σημαντικών δομών, συσχετισμών ή ανωμαλιών που υπάρχουν στα σήματα. Αυτό επιτρέπει μια βαθύτερη κατανόηση των υποκείμενων φαινομένων και διευκολύνει τη λήψη αποφάσεων ή τη μοντελοποίηση πρόβλεψης σε αυτήν την ερευνητική περιοχή.

Αξιοποιώντας τεχνικές επεξεργασίας σήματος, στοχεύεται η άντληση πολύτιμων γνώσεων από τα πολύπλοκα σύνολα δεδομένων των ηχογραφήσεων που υπάρχουν στη διάθεσή μας. Η επιτυχής εφαρμογή αυτών των τεχνικών θα συμβάλει στην καλύτερη κατανόηση των επιπτώσεων της νόσου Covid-19 στο αναπνευστικό σύστημα και κατ'επέκταση στην αποτελεσματικότερη διάγνωσή του.

Στη διατριβή, θα γίνει εμβάθυνση στις διάφορες τεχνικές επεξεργασίας σήματος που χρησιμοποιούνται, την εφαρμογή τους και τα αποτελέσματά τους στο πλαίσιο της έρευνάς που θα πραγματοποιηθεί. Συνδυάζοντας αυτές τις τεχνικές με αλγορίθμους ομαδοποίησης, στοχεύουμε να συνεισφέρουμε στην απόκτηση γνώσεων σχετικά με τον τρόπο που επηρεάζει την φυσιολογική αναπνοή η νόσος covid-19.

3.1 Φιλτράρισμα

Στον τομέα της επεξεργασίας σήματος ήχου, μία από τις κύριες προκλήσεις είναι η βελτίωση της ποιότητας των σημάτων ήχου μειώνοντας τον ανεπιθύμητο θόρυβο. Ο θόρυβος χαμηλής συχνότητας, που χαρακτηρίζεται από ανεπιθύμητους ήχους βουητού, μπορεί να υποβαθμίσει σημαντικά την εμπειρία ακρόασης. Για να μετριάσουν αυτό το πρόβλημα, οι μηχανικοί χρησιμοποιούν μια ισχυρή τεχνική γνωστή ως υψιπερατό φιλτράρισμα. Με την επιλεκτική εξασθένηση των χαμηλών συχνοτήτων, τα υψιπερατά φίλτρα αφαιρούν αποτελεσματικά τον ανεπιθύμητο θόρυβο και συμβάλλουν στην παραγωγή καθαρότερων, πιο ευκρινών σημάτων ήχου. Παρακάτω θα αναφερθούμε στην έννοια και τη σημασία του φιλτραρίσματος υψηλής διέλευσης, τις βασικές αρχές του και τις πρακτικές του εφαρμογές στον τομέα της επεξεργασίας σήματος ήχου.

3.1.1 Υψιπερατό φιλτράρισμα

Το υψιπερατό φιλτράρισμα είναι μια τεχνική επεξεργασίας σήματος που έχει σχεδιαστεί για να επιτρέπει σε υψηλότερες συχνότητες να περνούν μέσα από ένα φίλτρο ενώ εξασθενεί ή εξαλείφει τις χαμηλότερες. Στο πλαίσιο των ηχητικών σημάτων, ο θόρυβος χαμηλής συχνότητας προέρχεται συχνά από πηγές όπως ηλεκτρικές παρεμβολές, βρόχους γείωσης, θόρυβος χειρισμού μικροφώνου ή περιβαλλοντικούς παράγοντες. Αυτά τα ανεπιθύμητα στοιχεία συνήθως εμπίπτουν στο χαμηλότερο εύρος του φάσματος ήχου, εμποδίζοντας την ακρίβεια και την καθαρότητα του περιεχομένου ήχου.

Τα υπερυπερατά φίλτρα λειτουργούν με βάση τις αρχές της απόκρισης συχνότητας και της συχνότητας αποκοπής. Η απόκριση συχνότητας ενός φίλτρου αντιπροσωπεύει τη συμπεριφορά του σε απόκριση σε διαφορετικές συχνότητες. Για τα υπερυπερατά φίλτρα, η απόκριση είναι σχεδιασμένη να είναι σχετικά επίπεδη για υψηλότερες συχνότητες, επιτρέποντάς τους να περνούν με ελάχιστη εξασθένηση. Ωστόσο, καθώς η συχνότητα μειώνεται κάτω από μια καθορισμένη συχνότητα αποκοπής, το φίλτρο εξασθενεί προοδευτικά το σήμα. Η συχνότητα αποκοπής αντιπροσωπεύει το σημείο στο οποίο το φίλτρο αρχίζει να εξασθενεί τις χαμηλότερες συχνότητες.

Τα υπερυπερατά φίλτρα μπορούν να εφαρμοστούν χρησιμοποιώντας διάφορες τεχνικές, όπως αναλογικά ηλεκτρονικά κυκλώματα, αλγόριθμους επεξεργασίας ψηφιακών σημάτων ή συνδυασμό και των δύο. Στα αναλογικά κυκλώματα, παθητικά εξαρτήματα όπως αντιστάσεις, πυκνωτές και επαγωγείς χρησιμοποιούνται για τη δημιουργία δικτύων φίλτρων που επιτρέπουν τη διέλευση υψηλότερων συχνοτήτων. Από την άλλη πλευρά, τα ψηφιακά υπερυπερατά φίλτρα υλοποιούνται χρησιμοποιώντας αλγόριθμους που χειρίζονται τα διακριτά δείγματα ενός ψηφιακού ηχητικού σήματος. Αυτοί οι αλγόριθμοι μπορούν να εφαρμοστούν σε πραγματικό χρόνο ή κατά τα στάδια μετά την επεξεργασία.

Η εφαρμογή του φιλτραρίσματος υψηλής διέλευσης προσφέρει πολυάριθμα πλεονεκτήματα και βρίσκει εφαρμογές ευρείας εμβέλειας στην επεξεργασία σήματος ήχου. Μερικά βασικά πλεονεκτήματα και περιπτώσεις χρήσης του περιλαμβάνουν:

- Μείωση θορύβου: Τα υπερυπερατά φίλτρα μειώνουν αποτελεσματικά τον θόρυβο χαμηλής συχνότητας, όπως βουητό, ή περιβαλλοντικούς ήχους, με αποτέλεσμα καθαρότερα ηχητικά σήματα. Αυτό είναι ιδιαίτερα ωφέλιμο στην επαγγελματική παραγωγή ήχου, την εγγραφή μουσικής και τη μετάδοση, όπου η ποιότητα και η σαφήνεια του περιεχομένου είναι πρωταρχικής σημασίας. Το υπερυπερατό φιλτράρισμα μπορεί ακόμη να βελτιώσει την ευκρίνεια της ομιλίας μειώνοντας το θόρυβο χαμηλής συχνότητας, καθιστώντας το ιδιαίτερα πολύτιμο σε εφαρμογές όπως συστήματα φωνητικής επικοινωνίας, τηλεδιάσκεψη και αλγόριθμοι αναγνώρισης ομιλίας.
- Ισοστάθμιση: Το υπερυπερατό φιλτράρισμα είναι επίσης ένα κρίσιμο στοιχείο των τεχνικών εξισορρόπησης. Μειώνοντας επιλεκτικά συγκεκριμένες περιοχές συχνοτήτων, οι μηχανικοί ήχου μπορούν να διαμορφώσουν την απόκριση

συχνότητας ενός σήματος, αντισταθμίζοντας τα χαρακτηριστικά των περιβαλλόντων εγγραφής ή δίνοντας έμφαση σε ορισμένες τονικές ιδιότητες.

Ένα μεγάλο α σημαίνει ότι η έξοδος θα μειωθεί πολύ αργά, αλλά θα επηρεαστεί επίσης έντονα ακόμη και από μικρές αλλαγές στην είσοδο. Από τη σχέση μεταξύ παραμέτρου α και χρονικής σταθεράς RC παραπάνω προκύπτει ότι ένα μεγάλο α αντιστοιχεί σε μεγάλο RC και επομένως σε χαμηλή γωνιακή συχνότητα του φίλτρου. Ως εκ τούτου, αυτή η περίπτωση αντιστοιχεί σε ένα υπερατά φίλτρο με πολύ στενή ζώνη διακοπής. Επειδή επηρεάζεται από μικρές αλλαγές και τείνει να διατηρεί τις προηγούμενες τιμές εξόδου του για μεγάλο χρονικό διάστημα, μπορεί να περάσει σχετικά χαμηλές συχνότητες. Ωστόσο, μια σταθερή είσοδος θα μειώνεται πάντα στο μηδέν, όπως θα ήταν αναμενόμενο με ένα υπερατά φίλτρο με μεγάλο RC.

Ένα μικρό α σημαίνει ότι η έξοδος θα εξασθενήσει γρήγορα και θα απαιτήσει μεγάλες αλλαγές στην είσοδο (δηλαδή, το $(x[i] - x[i-1])$ είναι μεγάλο) για να προκαλέσει μεγάλη αλλαγή στην έξοδο. Με τη σχέση μεταξύ παραμέτρου α και χρονικής σταθεράς RC παραπάνω, ένα μικρό α αντιστοιχεί σε ένα μικρό RC και επομένως υψηλή γωνιακή συχνότητα του φίλτρου[28]. Ως εκ τούτου, αυτή η περίπτωση αντιστοιχεί σε ένα υπερατά φίλτρο με πολύ μεγάλη ζώνη διακοπής. Επειδή απαιτεί μεγάλες (δηλαδή γρήγορες) αλλαγές και τείνει να ξεχνάει γρήγορα τις προηγούμενες τιμές εξόδου του, μπορεί να περάσει μόνο σχετικά υψηλές συχνότητες, όπως θα ήταν αναμενόμενο με ένα φίλτρο υψηλής διέλευσης με μικρό RC.

3.2 Λογαριθμική Συμπίεση

Στην επεξεργασία σήματος ήχου, η λογαριθμική συμπίεση είναι μια ισχυρή τεχνική που χρησιμοποιείται για τον χειρισμό του δυναμικού εύρους των σημάτων ήχου. Με τη συμπίεση του πλάτους της κυματομορφής ήχου, η λογαριθμική συμπίεση βοηθά στην εξισορρόπηση των επιπέδων έντασης, στη βελτίωση της καθαρότητας του ήχου και στη βελτίωση της συνολικής εμπειρίας ακρόασης.

Η λογαριθμική συμπίεση λειτουργεί με βάση την αρχή ότι η ανθρώπινη αντίληψη της έντασης είναι μη γραμμική. Αντί να κλιμακώνεται γραμμικά το πλάτος ενός ακουστικού σήματος, η λογαριθμική συμπίεση εφαρμόζει έναν μαθηματικό μετασχηματισμό που εξασθενεί τα υψηλότερα πλάτη περισσότερο από τα χαμηλότερα πλάτη. Αυτή η μη γραμμική χαρτογράφηση διασφαλίζει ότι διατηρούνται ήχοι χαμηλής

έντασης, ενώ μειώνεται το δυναμικό εύρος των πιο δυνατών ήχων. Ως αποτέλεσμα, η αντιληπτή ένταση είναι πιο ισορροπημένη και οι λεπτές λεπτομέρειες ήχου γίνονται πιο ηχηρές.

Η λογαριθμική συμπίεση μπορεί να εφαρμοστεί χρησιμοποιώντας διάφορους αλγόριθμους και τεχνικές. Μια κοινή προσέγγιση περιλαμβάνει την εφαρμογή μιας λογαριθμικής συνάρτησης στις τιμές πλάτους του ηχητικού σήματος. Η λογαριθμική συνάρτηση συμπίεζει τις τιμές πλάτους, δίνοντας έμφαση στα χαμηλότερα πλάτη ενώ μειώνει τα μεγαλύτερα. Αυτή η τεχνική φέρνει αποτελεσματικά το δυναμικό εύρος του σήματος ήχου μέσα σε ένα επιθυμητό εύρος, αποφεύγοντας την παραμόρφωση ή το ψαλίδισμα του σήματος.

Στην επεξεργασία ηχητικού σήματος, η λογαριθμική συμπίεση χρησιμοποιείται για να διορθωθεί το πρόβλημα του μεγάλου δυναμικού εύρους των τιμών σε ένα φασματόγραμμα. Συχνά, οι μικρές τιμές συναφούς σημασίας καλύπτονται ή κυριαρχούνται από μεγάλες τιμές λόγω της απόκλισης του δυναμικού εύρους[29]. Για να αντιμετωπιστεί αυτό το πρόβλημα, συνήθως χρησιμοποιείται μια κλίμακα ντεσιμπέλ. Η ιδέα είναι να μειωθεί η διαφορά μεταξύ μεγάλων και μικρών τιμών, ενισχύοντας τις μικρές τιμές. Συνήθως, χρησιμοποιούνται συναρτήσεις που βασίζονται σε λογάριθμους για να επιτευχθεί αυτό, όπως η λογαριθμική συμπίεση που εφαρμόστηκε στην περίπτωση μας. Η ένταση της συμπίεσης επηρεάζεται από την παράμετρο γάμμα. Όσο υψηλότερη είναι η τιμή του γάμμα, τόσο πιο έντονη είναι η συμπίεση. Η εξίσωση της λογαριθμικής συμπίεσης έχει ως εξής: $\Gamma(v) := \log(1 + \gamma \cdot v)$ [30].

Συνολικά η λογαριθμική συμπίεση είναι μια θεμελιώδης τεχνική στην επεξεργασία σημάτων ήχου που εξισορροπεί αποτελεσματικά το δυναμικό εύρος των σημάτων ήχου. Μειώνοντας τα υψηλότερα πλάτη περισσότερο από τα χαμηλότερα πλάτη, η λογαριθμική συμπίεση ενισχύει τη δυναμική του ήχου, βελτιώνει την ευκρίνεια και διατηρεί σταθερά επίπεδα έντασης. Οι εφαρμογές του στην παραγωγή μουσικής, τη μετάδοση και τη βελτίωση ομιλίας υπογραμμίζουν τη συνάφεια και τη σημασία του για την παροχή εμπειριών ήχου υψηλής ποιότητας.

3.3 Μετασχηματισμός Fourier (STFT) και φασματόγραμμα (spectrogram)

Στη σφαίρα της επεξεργασίας του ακουστικού σήματος, το φασματόγραμμα αποτελεί βασικό εργαλείο, ξετυλίγοντας τις περίπλοκες λεπτομέρειες συχνότητας των σημάτων καθώς αυτά ξεδιπλώνονται δυναμικά με την πάροδο του χρόνου. Αυτή η γραφική αναπαράσταση διαδραματίζει κεντρικό ρόλο σε πολλές εφαρμογές, παρέχοντας μια ολοκληρωμένη εικόνα για τις μεταβαλλόμενες συνιστώσες συχνότητας μέσα σε ένα σήμα. Κεντρική θέση στη δημιουργία φασματογραμμάτων διαθέτει η ιδέα του Βραχυχρόνιου Μετασχηματισμού Φουριέ (STFT), μιας μαθηματικής τεχνικής που επιτρέπει την ανάλυση των διακυμάνσεων της συχνότητας σήματος εντός εντοπισμένων χρονικών τμημάτων. Παρακάτω θα εξερευνηθεί η ουσία του υπολογισμού του φασματογράμματος, υπογραμμίζοντας τη σημασία του σε διάφορους τομείς και θα πέσει φως στο θεμελιώδες πλαίσιο Βραχυχρόνιου Μετασχηματισμού Φουριέ (STFT).

Η σφαίρα της επεξεργασίας ήχου, που κυμαίνεται από την ανάλυση ομιλίας έως τη σύνθεση μουσικής, ακόμη και τα ιατρικά διαγνωστικά, αποκομίζει τα οφέλη του υπολογισμού του φασματογράμματος. Ένα φασματόγραμμα χρησιμεύει ως οπτική πύλη προς το ηχητικό σύμπαν, προσφέροντας μια εικονογραφική αναπαράσταση του τρόπου με τον οποίο οι συνιστώσες συχνότητας ενός σήματος μεταμορφώνονται μέσα από σύντομα χρονικά διαστήματα. Η πολύπλευρη σημασία του είναι ιδιαίτερα εμφανής στην αποκρυπτογράφηση των αποχρώσεων των προτύπων ομιλίας, στην αποκάλυψη μουσικών νοτών και δομών και στην αποκάλυψη κρυμμένων ανωμαλιών στη φυσιολογική ακουστική. Ως εκ τούτου, η κατανόηση της μηχανικής του υπολογισμού του φασματογράμματος αποκτά ύψιστη σημασία, γιατί ξεκλειδώνει μια σφαίρα ακουστικών ιδεών.

Στον πυρήνα του υπολογισμού του φασματογράμματος βρίσκεται ο βραχυχρόνιος μετασχηματισμός Fourier (STFT), μια μαθηματική πράξη που ενσωματώνει την ουσία της ανάλυσης εντοπισμένης συχνότητας. Σε αντίθεση με τον συμβατικό μετασχηματισμό Fourier που ξεδιπλώνει το φάσμα συχνοτήτων ενός ολόκληρου σήματος, το STFT εμβαθύνει σε τμήματα σήματος, επιτρέποντας μια λεπτομερή κατανόηση της κατανομής συχνότητας με την πάροδο του χρόνου. Αυτή η τμηματοποίηση περιλαμβάνει τη διάσπαση του σήματος σε επικαλυπτόμενα

παράθυρα, με τον μετασχηματισμό Fourier να εκτελείται σε κάθε τμήμα παραθύρου. Το αποτέλεσμα είναι μια δισδιάστατη αναπαράσταση, όπου ένας άξονας απεικονίζει τη χρονολογική πρόοδο, ένας άλλος αντιπροσωπεύει διακύμανση συχνότητας και το χρώμα ή η ένταση ενσωματώνει το πλάτος ή την ισχύ των αντίστοιχων στοιχείων συχνότητας.

Η αποτελεσματικότητα του STFT βασίζεται όχι μόνο στις μαθηματικές του αρχές αλλά και στην πραγματιστική εκτέλεσή του. Η διαδικασία περιλαμβάνει την κατανομή του σήματος σε συμπαγή παράθυρα, ακολουθούμενη συνήθως από πολλαπλασιασμό με μια συνάρτηση παραθύρου όπως Gaussian ή Hamming. Αυτό περιορίζει το φαινόμενο της φασματικής διαρροής, διασφαλίζοντας ότι τα μεμονωμένα στοιχεία συχνότητας παραμένουν διακριτά. Η επιλογή του τύπου παραθύρου εξισορροπεί περίπλοκα την ακρίβεια της ανάλυσης συχνότητας και χρόνου. Η πράξη της επικάλυψης αυτών των παραθύρων εμπλουτίζει περαιτέρω την ανάλυση δημιουργώντας μια απρόσκοπτη μετάβαση από το ένα παράθυρο στο άλλο, καταγράφοντας έτσι τη χρονική εξέλιξη του σήματος[31].

Η υλοποίηση του βραχυχρόνιου μετασχηματισμού Fourier(STFT) περιλαμβάνει μια σειρά διαδοχικών λειτουργιών που αποκαλύπτουν συλλογικά τη φασματική δυναμική ενός ηχητικού σήματος. Στη συνέχεια γίνεται εμβάθυνση στη διαδικασία βήμα προς βήμα του υπολογισμού του βραχυχρόνιου μετασχηματισμού Fourier, ρίχνοντας φως στα βασικά στοιχεία του και τη σημασία κάθε σταδίου.

Η διαδικασία του του βραχυχρόνιου μετασχηματισμού Fourier ξεκινά με την κατάτμηση του ηχητικού σήματος σε μικρότερα, διαχειρίσιμα τμήματα. Αυτά τα τμήματα, που συχνά αναφέρονται ως πλαίσια ή παράθυρα, χρησιμεύουν ως βάση για ανάλυση εντοπισμένης συχνότητας. Για να μετριαστεί η απότομη περικοπή των σημάτων στις άκρες του πλαισίου, χρησιμοποιούνται λειτουργίες παραθύρου όπως Hamming, Hanning ή Gaussian. Αυτές οι λειτουργίες λεπταίνουν απαλά τις άκρες κάθε παραθύρου, μειώνοντας τη φασματική διαρροή και διασφαλίζουν την ομαλή μετάβαση από το ένα πλαίσιο στο άλλο. Μόλις το ηχητικό σήμα διαιρεθεί αποτελεσματικά σε παράθυρα, ο μετασχηματισμός Fourier εφαρμόζεται σε κάθε παράθυρο ανεξάρτητα[32]. Αυτή η μαθηματική πράξη ξετυλίγει τη σύνθετη κυματομορφή στα ημιτονοειδή συστατικά της, αποκαλύπτοντας την παρουσία διαφόρων συχνοτήτων εντός του εντοπισμένου τμήματος.

Το αποτέλεσμα του Μετασχηματισμού Fourier είναι ένα σύνθετο φάσμα που αρθρώνει το πλάτος και τη φάση κάθε συστατικού συχνότητας. Αναλύοντας το σύνθετο φάσμα που λαμβάνεται από τον Μετασχηματισμό Fourier, διακρίνονται δύο βασικές πληροφορίες: το μέγεθος και η φάση. Το φάσμα μεγέθους απεικονίζει την ένταση των διαφορετικών συνιστωσών συχνότητας, μεταφέροντας την ισχύ που μεταφέρει το καθένα. Από την άλλη πλευρά, το φάσμα φάσης ξετυλίγει τη χρονική ευθυγράμμιση αυτών των στοιχείων μέσα στο παράθυρο. Μαζί, αυτές οι πτυχές παρέχουν μια ολοκληρωμένη κατανόηση του περιεχομένου συχνότητας του ηχητικού σήματος εντός του συγκεκριμένου χρονικού τμήματος.

Για την καταγραφή της παροδικής φύσης των σημάτων ήχου, τα καρέ επικαλύπτονται κατά τη διαδικασία STFT. Αυτή η επικάλυψη εγγυάται τη συνέχεια των πληροφοριών μεταξύ διαδοχικών παραθύρων, επιτρέποντας την απεικόνιση των εξελισσόμενων χαρακτηριστικών συχνότητας. Συνήθως, χρησιμοποιείται μια επικάλυψη 50%, αν και αυτή η τιμή μπορεί να προσαρμοστεί για να επιτευχθεί μια ισορροπία μεταξύ της χρονικής ακρίβειας και της υπολογιστικής απόδοσης. Καθώς η διαδικασία STFT ξεδιπλώνεται σε όλο το ηχητικό σήμα, δημιουργείται ένας καταρράκτης φασμάτων μεγέθους για κάθε καρέ. Η συλλογή αυτών των φασμάτων με την πάροδο του χρόνου σχηματίζει ένα δισδιάστατο πίνακα, κοινώς γνωστό ως φασματογράμμα. Αυτός ο πίνακας κρατά το χρόνο στον έναν άξονα, τη συχνότητα στον άλλο και το μέγεθος των συνιστωσών της συχνότητας αναπαρίσταται μέσω χρώματος ή έντασης. Αυτή η οπτική αναπαράσταση χρησιμεύει ως μια πλούσια πηγή πληροφοριών, επιτρέποντας στο ανθρώπινο μάτι να διακρίνει γρήγορα μοτίβα, τάσεις και αλλαγές στην κατανομή συχνότητας του σήματος ήχου. Στην περίπτωση συνεχούς χρόνου, η συνάρτηση που θα μετασχηματιστεί πολλαπλασιάζεται με μια συνάρτηση παραθύρου η οποία είναι μη μηδενική μόνο για ένα μικρό χρονικό διάστημα. Λαμβάνεται ο μετασχηματισμός Fourier (μια μονοδιάστατη συνάρτηση) του προκύπτοντος σήματος και, στη συνέχεια, το παράθυρο ολισθαίνει κατά μήκος του άξονα του χρόνου μέχρι το τέλος, καταλήγοντας σε μια δισδιάστατη αναπαράσταση του σήματος. Μαθηματικά, αυτό γράφεται ως: $F(x) = \int_{-\infty}^{+\infty} f(x)e^{-2\pi ix\xi} dx$, για κάθε πραγματικό αριθμό ξ .

Η θεωρητική βάση του βραχυχρόνιου Μετασχηματισμού Φουριέ (STFT) έθεσε τις βάσεις για μια βαθιά κατανόηση της ανάλυσης σημάτων ήχου. [33] Μεταφράζοντας αυτή την κατανόηση σε πρακτικές εφαρμογές, η βιβλιοθήκη `scipy` της `Python` παρέχει

ένα ισχυρό σύνολο εργαλείων για τον αποτελεσματικό υπολογισμό των φασματογραμμμάτων. Στη συνέχεια διερευνώνται οι λεπτομέρειες υλοποίησης του υπολογισμού του φασματογράμματος χρησιμοποιώντας τη συνάρτηση `scipy.signal.spectrogram`, επισημαίνοντας τις παραμέτρους, τα πλεονεκτήματα και τους περιορισμούς της.

Η συνάρτηση `scipy.signal.spectrogram` ενσωματώνει την περίπλοκη διαδικασία υπολογισμού του βραχυχρόνιου μετασχηματισμού Fourier μέσα σε λίγες γραμμές κώδικα. Αφαιρώντας τα πολύπλοκα μαθηματικά και τις περίπλοκες πράξεις, δίνει τη δυνατότητα στους ερευνητές και τους επαγγελματίες να επικεντρωθούν στην ανάλυση των φασματογραμμμάτων αντί να χαθούν στις πολυπλοκότητες της επεξεργασίας σήματος. Η συνάρτηση δέχεται μια σειρά παραμέτρων που προσαρμόζουν τον υπολογισμό στις συγκεκριμένες ανάγκες της ανάλυσης.

3.4 Περίγραμμα ισχύος- Power Envelope

Το power envelope αποτελεί μια θεμελιώδη μέθοδο, αποκαλύπτοντας κρίσιμες πληροφορίες για τα δυναμικά χαρακτηριστικά του ήχου. Αυτή η ενότητα εμβαθύνει στην εξαγωγή φακέλου ισχύος, ρίχνοντας φως στη σημασία της, τις υποκείμενες μεθοδολογίες και εφαρμογές της. Το περίγραμμα ισχύος, που συχνά αναφέρεται ως φάκελος πλάτους, ενσωματώνει τη μεταβαλλόμενη ένταση ή ενέργεια ενός ηχητικού σήματος με την πάροδο του χρόνου. Σε αντίθεση με το στιγμιαίο πλάτος που μπορεί να υπόκειται σε γρήγορες διακυμάνσεις, το περίγραμμα ισχύος παρέχει μια ομαλή αναπαράσταση της ισχύος του σήματος. Αυτή η έννοια βρίσκει τις ρίζες της στην κατανόηση ότι η ανθρώπινη αντίληψη της έντασης είναι λογαριθμική και όχι γραμμική. Έτσι, το περίγραμμα ισχύος ευθυγραμμίζεται πιο στενά με τον τρόπο με τον οποίο οι άνθρωποι αντιλαμβάνονται την ένταση του ήχου.

Για την κατανόηση της εξαγωγής του περιγράμματος ισχύος, είναι επιτακτική η διερεύνηση της φασματικής πυκνότητας ισχύος και των ζωνών συχνότητας. Η φασματική πυκνότητα ισχύος (PSD) περιγράφει την κατανομή της ισχύος ενός σήματος σε διάφορες συνιστώσες συχνότητας. Η εξαγωγή του PSD περιλαμβάνει την κατανομή του σήματος στις συχνότητες που το αποτελούν και την ποσοτικοποίηση της ισχύος σε κάθε συχνότητα.

Γι'αυτό εισάγεται η έννοια των ζωνών συχνοτήτων. Οι ζώνες συχνοτήτων ομαδοποιούν μια σειρά συχνοτήτων μαζί, επιτρέποντας μια πιο γενικευμένη ανάλυση. Αυτές οι ζώνες συχνά απέχουν λογαριθμικά, ευθυγραμμισμένες με την απόκριση του ανθρώπινου ακουστικού συστήματος σε διαφορετικές περιοχές συχνοτήτων.

Η εξαγωγή ενός φακέλου ισχύος περιλαμβάνει την άθροιση της ισχύος σε συγκεκριμένες ζώνες συχνοτήτων με την πάροδο του χρόνου. Αυτή η διαδικασία συγχωνεύει το περιεχόμενο ισχύος ενός σήματος σε μια πιο κατανοητή αναπαράσταση. Το φιλτράρισμα στις ζώνες συχνοτήτων όχι μόνο μειώνει την υπολογιστική πολυπλοκότητα της ανάλυσης αλλά διευκολύνει επίσης την εστίαση σε συγκεκριμένα ακουστικά στοιχεία του σήματος. Πιο συγκεκριμένα η διαδικασία περιγράφεται στη συνέχεια. Το ηχητικό σήμα χωρίζεται σε μικρότερα τμήματα, χρησιμοποιώντας συχνά τεχνικές όπως ο μετασχηματισμός Fourier βραχείας διάρκειας (STFT). Αυτά τα τμήματα καταγράφουν τη δυναμική του σήματος σε μικρότερα χρονικά διαστήματα. Για κάθε τμήμα, υπολογίζεται η φασματική πυκνότητα ισχύος. Αυτό συνεπάγεται τη μετατροπή του σήματος στην περιοχή συχνότητας και τον τετραγωνισμό του μεγέθους των μιγαδικών τιμών που προκύπτουν για να ληφθούν τιμές ισχύος. Αθροίζονται οι τιμές ισχύος εντός προκαθορισμένων ζωνών συχνοτήτων. Αυτές οι ζώνες αντιστοιχούν συνήθως σε συγκεκριμένα εύρη ακουστικών συχνοτήτων, όπως τα μπάσα, τα μεσαία και τα πρίμα. Οι αθροιστικές τιμές ισχύος συχνά εξομαλύνονται για να μειωθούν οι απότομες αλλαγές και να τονιστεί η συνολική τάση του περιβλήματος ισχύος.

3.4.1 Εντοπισμός κορυφών σε περίγραμμα ισχύος

Η ανίχνευση κορυφών εντός ενός φακέλου ισχύος είναι μια εξίσου σημαντική τεχνική στην επεξεργασία σήματος. Οι κορυφές, ως τοπικά μέγιστα, παρέχουν βασικές πληροφορίες σχετικά με την προβολή ορισμένων χαρακτηριστικών σε ένα ηχητικό σήμα. Αυτή η τεχνική βρίσκει εκτεταμένη χρήση σε διάφορες εφαρμογές, όπως η ανίχνευση συμβάντων ήχου, η ανάλυση μουσικής και η αναγνώριση ομιλίας. Αυτή η εξερεύνηση εμβαθύνει στον μηχανισμό ανίχνευσης αιχμής εντός των φακέλων ισχύος και στον επακόλουθο υπολογισμό των πλατών κορυφών. Η ανίχνευση αιχμής περιλαμβάνει τον εντοπισμό των υψηλότερων σημείων ή των τοπικών μέγιστων εντός ενός σήματος. Στο πλαίσιο των φακέλων ισχύος, οι κορυφές αντιπροσωπεύουν συχνά σημαντικές συγκεντρώσεις ενέργειας, υποδεικνύοντας συμβάντα ή μεταβάσεις στο ηχητικό σήμα. Η ανίχνευση αυτών των κορυφών είναι ζωτικής σημασίας για την

καταγραφή σημαντικών γεγονότων και τον χαρακτηρισμό των ηχητικών σημάτων από την άποψη της έντασης και του χρόνου.

Η συνάρτηση `find_peaks` από τη μονάδα `scipy.signal` εμφανίζεται ως ένα ισχυρό εργαλείο για την ανίχνευση κορυφών μέσα σε ένα σήμα. Αξιοποιώντας αυτή τη λειτουργία, είναι δυνατός ο εντοπισμός τοπικών μεγίστων με βάση διάφορες παραμέτρους, όπως η προεξοχή και το ύψος. Η προεξοχή αναφέρεται στην κατακόρυφη απόσταση μεταξύ της κορυφής και της κατώτερης γραμμής περιγράμματός της. Η συνάρτηση `find_peaks` προσφέρει την ευελιξία να φιλτραρίσματος κορυφών που δεν πληρούν συγκεκριμένα κριτήρια, επιτρέποντας την εκλεπτυσμένη αναγνώριση κορυφών.

Η επιτυχής εφαρμογή της συνάρτησης `find_peaks` απαιτεί προσεκτική εξέταση των παραμέτρων της:

- **Height Threshold:** Αυτή η παράμετρος καθορίζει το ελάχιστο ύψος που πρέπει να ληφθεί υπόψη μια κορυφή. Οι κορυφές κάτω από αυτό το όριο δεν λαμβάνονται υπόψη. Η προσαρμογή αυτού του ορίου μπορεί να επηρεάσει την ευαισθησία της ανίχνευσης αιχμής.
- **Prominence:** Οι κορυφές με τιμή προεξοχής κάτω από ένα συγκεκριμένο όριο μπορούν να θεωρηθούν θόρυβος ή μικρές διακυμάνσεις και μπορούν να εξαιρεθούν από τη διαδικασία ανίχνευσης.
- **Distance:** Οι κορυφές που είναι πιο κοντά από μια καθορισμένη απόσταση θεωρούνται συχνά ως μια ενιαία κορυφή. Αυτή η παράμετρος είναι κρίσιμη σε σενάρια όπου οι κορυφές σε κοντινή απόσταση πρέπει να αντιμετωπίζονται ως ξεχωριστά γεγονότα.

Πέρα από την ανίχνευση κορυφών, η κατανόηση του χρονικού εύρους των κορυφών είναι εξίσου απαραίτητη. Το πλάτος μιας κορυφής παρέχει πληροφορίες σχετικά με τη διάρκεια ενός συμβάντος ή ενός ηχητικού περιστατικού. Για να υπολογιστούν τα πλάτη των κορυφών, χρησιμοποιείται η συνάρτηση `peak_widths` από τη μονάδα `scipy.signal`. Αυτή η συνάρτηση αναλύει το σχήμα των κορυφών και επιστρέφει τα πλάτη των κορυφών, προσφέροντας πολύτιμες πληροφορίες σχετικά με τη χρονική έκταση των γεγονότων.

Οι τεχνικές ανίχνευσης αιχμής και ο μετέπειτα υπολογισμός του πλάτους εντός των φακέλων ισχύος συμβάλλουν σημαντικά στην κατανόηση της χρονικής και ενεργειακής δυναμικής των ηχητικών σημάτων. Οι λειτουργίες `find_peaks` και `peak_widths` από τη μονάδα `scipy.signal` εξουσιοδοτούν τους ερευνητές και τους επαγγελματίες να αναγνωρίζουν αποτελεσματικά τα κύρια χαρακτηριστικά στα ηχητικά σήματα.

3.5 Συσταδοποίηση

Η συσταδοποίηση (`clustering`), μια θεμελιώδης έννοια στον τομέα της ανάλυσης δεδομένων και της μηχανικής μάθησης, χρησιμεύει ως απαραίτητο εργαλείο για την αποκάλυψη κρυφών μοτίβων σε σύνολα δεδομένων. Παρακάτω γίνεται εμβάθυνση, διευκρινίζοντας τη γενική φύση, τις ξεχωριστές κατηγορίες και τις αξιοσημείωτες εφαρμογές του.

Στην ουσία, η συσταδοποίηση περιλαμβάνει την τμηματοποίηση ενός συνόλου δεδομένων σε υποσύνολα ή συστάδες, όπου τα σημεία δεδομένων εντός του ίδιου συμπλέγματος μοιράζονται ομοιότητες, ενώ εκείνα σε διαφορετικά συμπλέγματα παρουσιάζουν διαφορές. Αυτή η διαδικασία επιδιώκει να αποκαλύψει την υποκείμενη δομή των δεδομένων, επιτρέποντας γνώσεις που μπορεί να είναι αόριστες με άλλα μέσα. Η συσταδοποίηση λειτουργεί με την υπόθεση ότι τα σημεία δεδομένων που ανήκουν στην ίδια συστάδα σχετίζονται περισσότερο μεταξύ τους παρά με εκείνα σε άλλες συστάδες.

Η συσταδοποίηση είναι ένα από τα κυριότερα μελήματα της διερευνητικής ανάλυσης δεδομένων και μια κοινή τεχνική για στατιστική ανάλυση δεδομένων, που χρησιμοποιείται σε πολλούς τομείς, όπως η αναγνώριση προτύπων, η ανάλυση εικόνας, η ανάκτηση πληροφοριών και η βιοπληροφορική.

Η ίδια η ανάλυση συστάδων δεν είναι ένας μεμονωμένος αλγόριθμος, αλλά μια γενική εργασία που πρέπει να επιλυθεί. Μπορεί να επιτευχθεί με διάφορους αλγόριθμους που διαφέρουν σημαντικά ως προς την κατανόησή τους για το τι αποτελεί ένα σύμπλεγμα και τον τρόπο αποτελεσματικής εύρεσης τους. Οι δημοφιλείς έννοιες των συστάδων περιλαμβάνουν ομάδες με μικρές αποστάσεις μεταξύ των μελών του συμπλέγματος,

πυκνές περιοχές του χώρου δεδομένων, διαστήματα ή συγκεκριμένες στατιστικές κατανομές. Η συσταδοποίηση μπορεί επομένως να διατυπωθεί ως ένα πρόβλημα βελτιστοποίησης πολλαπλών στόχων. Ο κατάλληλος αλγόριθμος συσταδοποίησης και οι ρυθμίσεις παραμέτρων (συμπεριλαμβανομένων παραμέτρων όπως η συνάρτηση απόστασης προς χρήση, το όριο πυκνότητας ή ο αριθμός των αναμενόμενων συμπλεγμάτων) εξαρτώνται από το μεμονωμένο σύνολο δεδομένων και την προβλεπόμενη χρήση των αποτελεσμάτων. Η ανάλυση συστάδων δεν είναι μια αυτόματη εργασία, αλλά μια επαναληπτική διαδικασία ανακάλυψης γνώσης ή διαδραστικής βελτιστοποίησης πολλαπλών στόχων που περιλαμβάνει δοκιμή και αποτυχία. Συχνά είναι απαραίτητο να τροποποιηθεί η προεπεξεργασία δεδομένων και τις παραμέτρους του μοντέλου μέχρι το αποτέλεσμα να επιτύχει τις επιθυμητές ιδιότητες.

Οι τεχνικές συσταδοποίησης μπορούν να κατηγοριοποιηθούν ευρέως σε διάφορες ομάδες με βάση τους στόχους, τις μεθοδολογίες και τα χαρακτηριστικά τους:

3.5.1 Μέθοδοι κατάτμησης (Partitioning Methods)

Αυτές οι μέθοδοι χωρίζουν το σύνολο δεδομένων σε διακριτά μη επικαλυπτόμενα συμπλέγματα. Ο αλγόριθμος K-Means, ένα κλασικό παράδειγμα, εκχωρεί σημεία δεδομένων σε συστάδες ελαχιστοποιώντας το άθροισμα των τετραγωνικών αποστάσεων μεταξύ των σημείων και των κεντροειδών των συστάδων που τους έχουν εκχωρηθεί.[34, p. 1] Ο αλγόριθμος K-Means είναι αποτελεσματικός και χρησιμοποιείται ευρέως, αλλά η αποτελεσματικότητά του βασίζεται στην αρχική τοποθέτηση των κεντροειδών. Το πρόβλημα βελτιστοποίησης του K-means είναι NP-Hard αλλά υπάρχει ένας ευρέως γνωστός αλγόριθμος βελτιστοποίησης ο αλγόριθμος του Lloyd, που συχνά αναφέρεται απλώς ως "αλγόριθμος k-means". Ωστόσο, βρίσκει μόνο ένα τοπικό βέλτιστο και συνήθως εκτελείται πολλές φορές με διαφορετικές τυχαίες αρχικοποιήσεις. Οι παραλλαγές του k-means συχνά περιλαμβάνουν βελτιστοποιήσεις όπως η επιλογή του καλύτερου από πολλαπλές εκτελέσεις, αλλά και ο περιορισμός των κεντροειδών στα μέλη του συνόλου δεδομένων (k-medoids), η επιλογή διαμέσου (k-medians clustering), η επιλογή των αρχικών κέντρων λιγότερο τυχαία (k-means++) ή επιτρέποντας μια ασαφή ανάθεση συμπλέγματος (fuzzy c-means).

Οι περισσότεροι αλγόριθμοι τύπου k-means απαιτούν να καθοριστεί εκ των προτέρων ο αριθμός των συστάδων (k), το οποίο θεωρείται ένα από τα μεγαλύτερα μειονεκτήματα αυτών των αλγορίθμων. Επιπλέον, οι αλγόριθμοι προτιμούν συστάδες περίπου παρόμοιου μεγέθους, καθώς θα εκχωρούν πάντα ένα αντικείμενο στο πλησιέστερο κέντρο[35]. Αυτό συχνά οδηγεί σε λανθασμένα κομμένα περιγράμματα συμπλεγμάτων (κάτι που δεν προκαλεί έκπληξη, καθώς ο αλγόριθμος βελτιστοποιεί τα κέντρα των συστάδων και όχι τα σύνορα των συμπλεγμάτων).

Το K-means έχει μια σειρά από ενδιαφέρουσες θεωρητικές ιδιότητες. Αρχικά, χωρίζει το χώρο δεδομένων σε μια δομή γνωστή ως διάγραμμα Voronoi. Ακόμη, είναι εννοιολογικά κοντά στην ταξινόμηση του πλησιέστερου γείτονα και ως εκ τούτου είναι δημοφιλής στη μηχανική μάθηση.

3.5.2 Ιεραρχικές Μέθοδοι(Hierarchical Methods)

Αυτές οι μέθοδοι δημιουργούν μια ιεραρχία συστάδων, που συχνά αντιπροσωπεύονται ως δομή που μοιάζει με δέντρο, γνωστή ως δενδρόγραμμα. Η αθροιστική και η διαιρετική είναι δύο κοινές προσεγγίσεις σε αυτήν την κατηγορία. Οι συγκεντρωτικές μέθοδοι ξεκινούν με μεμονωμένα σημεία δεδομένων και τα συγχωνεύουν επαναληπτικά σε συστάδες, ενώ οι μέθοδοι διαίρεσης ξεκινούν με ολόκληρο το σύνολο δεδομένων και το χωρίζουν αναδρομικά σε μικρότερα συμπλέγματα. Η συσταδοποίηση που βασίζεται στη συνδεσιμότητα βασίζεται στην βασική ιδέα ότι τα αντικείμενα σχετίζονται περισσότερο με κοντινά αντικείμενα παρά με αντικείμενα πιο μακριά. Αυτοί οι αλγόριθμοι συνδέουν «αντικείμενα» για να σχηματίσουν «συστάδες» με βάση την απόστασή τους. Ένα σύμπλεγμα μπορεί να περιγραφεί σε μεγάλο βαθμό από τη μέγιστη απόσταση που απαιτείται για τη σύνδεση τμημάτων του συμπλέγματος[34].

3.5.3 Μέθοδοι με βάση την πυκνότητα(Density Based Methods)

Αυτές οι μέθοδοι προσδιορίζουν τα συμπλέγματα ως πυκνές περιοχές που χωρίζονται από αραιότερες περιοχές στο χώρο δεδομένων. Το DBSCAN (Density-Based Spatial Clustering of Applications with Noise) είναι μια εξέχουσα τεχνική σε αυτή την κατηγορία. Προσδιορίζει βασικά σημεία που έχουν έναν ελάχιστο αριθμό γειτόνων σε μια καθορισμένη απόσταση και επεκτείνει τα συμπλέγματα συνδέοντας προσβάσιμα σημεία. Ωστόσο, συνδέει μόνο σημεία που ικανοποιούν ένα κριτήριο πυκνότητας, στην

αρχική παραλλαγή που ορίζεται ως ένας ελάχιστος αριθμός άλλων αντικειμένων εντός αυτής της ακτίνας. Μια άλλη ενδιαφέρουσα ιδιότητα του DBSCAN είναι ότι η πολυπλοκότητά του είναι αρκετά χαμηλή. Απαιτεί δηλαδή γραμμικό αριθμό ερωτημάτων εύρους και ότι θα ανακαλύψει ουσιαστικά τα ίδια αποτελέσματα σε κάθε εκτέλεση. Επομένως δεν χρειάζεται να εκτελεστεί πολλές φορές[36].

3.5.4 Μέθοδοι που βασίζονται σε μοντέλα κατανομής(Model Based Methods)

Η ομαδοποίηση βάσει μοντέλου λαμβάνει δεδομένο ότι τα σημεία δεδομένων δημιουργούνται από ένα μείγμα κατανομών πιθανοτήτων. Τα μοντέλα διανομής χρησιμοποιούνται περισσότερο με μοντέλα ομαδοποίησης που σχετίζονται με τις στατιστικές. Οι συστάδες μπορούν στη συνέχεια να οριστούν εύκολα ως αντικείμενα που ανήκουν κατά πάσα πιθανότητα στην ίδια κατανομή. Μια βολική ιδιότητα αυτής της προσέγγισης είναι ότι μοιάζει πολύ με τον τρόπο που δημιουργούνται τεχνητά σύνολα δεδομένων: με δειγματοληψία τυχαίων αντικειμένων από μια διανομή. Ο αλγόριθμος Προσδοκίας-Μεγιστοποίησης (EM) χρησιμοποιείται συχνά σε αυτήν την κατηγορία με τη μέθοδο της γκαουσιανής σύνθεσης. Το σύνολο δεδομένων μοντελοποιείται συνήθως με έναν σταθερό αριθμό κατανομών Gauss ώστε να αποφευχθεί η υπερβολική προσαρμογή. Αυτά αρχικοποιούνται τυχαία και οι παράμετροι τους βελτιστοποιούνται επαναληπτικά για να ταιριάζουν καλύτερα στο σύνολο δεδομένων. Με αυτόν τον τρόπο υπολογίζει επαναληπτικά τις παραμέτρους των υποκείμενων κατανομών, συγκλίνει σε ένα τοπικό βέλτιστο και εκχωρεί σημεία δεδομένων σε συστάδες ανάλογα. Επομένως πολλές διαφορετικές εκτελέσεις είναι πιθανό να παράγουν διαφορετικά αποτελέσματα. Προκειμένου να επιτευχθεί μια σκληρή ομαδοποίηση, τα αντικείμενα συχνά αντιστοιχίζονται στη Gaussian κατανομή στην οποία πιθανότατα ανήκουν.

3.5.5 Ασαφής ομαδοποίηση(Fuzzy Clustering)

Η ασαφής ομαδοποίηση επιτρέπει στα σημεία δεδομένων να ανήκουν σε πολλαπλά συμπλέγματα με διαφορετικούς βαθμούς συμμετοχής. Σε αντίθεση με τη σκληρή ομαδοποίηση, όπου ένα σημείο ανήκει σε ένα μόνο σύμπλεγμα, η ασαφής ομαδοποίηση εξετάζει την πιθανότητα ένα σημείο να έχει μερική συμμετοχή σε πολλαπλές συστάδες[37].

Κεφάλαιο 4

Μηχανική Μάθηση

Η μηχανική μάθηση είναι ένας κλάδος της τεχνητής νοημοσύνης (AI) και της επιστήμης των υπολογιστών που εστιάζει στη χρήση δεδομένων και αλγορίθμων για τη μίμηση του τρόπου με τον οποίο μαθαίνουν οι άνθρωποι, βελτιώνοντας σταδιακά την ακρίβειά της. Η μηχανική μάθηση είναι ένα σημαντικό συστατικό του αναπτυσσόμενου τομέα της επιστήμης των δεδομένων. Μέσω της χρήσης στατιστικών μεθόδων, οι αλγόριθμοι εκπαιδεύονται να κάνουν ταξινομήσεις ή προβλέψεις και να αποκαλύπτουν βασικές γνώσεις σε έργα εξόρυξης δεδομένων. Αυτές οι πληροφορίες στη συνέχεια συμβάλλουν στη λήψη αποφάσεων εντός των εφαρμογών και των επιχειρήσεων, επηρεάζοντας τις βασικές μετρήσεις ανάπτυξης. Ένα μοντέλο μηχανικής μάθησης αντί να προγραμματίζεται με ακριβείς οδηγίες, εκπαιδεύεται χρησιμοποιώντας τεράστιες ποσότητες δεδομένων. Αυτά τα δεδομένα συνήθως περιλαμβάνουν χαρακτηριστικά εισαγωγής (όπως εικόνες, κείμενο ή αριθμητικές τιμές) και τις αντίστοιχες εξόδους ή ετικέτες-στόχους τους. Τα μοντέλα μηχανικής μάθησης στοχεύουν όχι μόνο στην απομνημόνευση των δεδομένων στα οποία εκπαιδεύονται, αλλά και στη γενίκευση από αυτά. Γενίκευση σημαίνει να μπορείς να κάνεις ακριβείς προβλέψεις ή αποφάσεις για νέα, πρωτοεμφανιζόμενα δεδομένα. Αυτή είναι μια βασική διάκριση. Η πραγματική δοκιμή ενός μοντέλου μηχανικής μάθησης δεν είναι η απόδοσή του στα δεδομένα εκπαίδευσης αλλά η ικανότητά του να διαχειρίζεται αποτελεσματικά νέα δεδομένα από τον πραγματικό κόσμο.

4.1 Εποπτευόμενη μάθηση

Στην εποπτευόμενη μάθηση, ο αλγόριθμος μαθαίνει να αντιστοιχίζει εισόδους σε αντίστοιχες εξόδους με βάση τα δεδομένα εκπαίδευσης με ετικέτα. Καθώς τα δεδομένα εισόδου τροφοδοτούνται στο μοντέλο, το μοντέλο προσαρμόζει τα βάρη του μέχρι να τοποθετηθεί κατάλληλα. Αυτό συμβαίνει ως μέρος της διαδικασίας πολλαπλής επικύρωσης για να διασφαλιστεί ότι το μοντέλο αποφεύγει την υπερπροσαρμογή. Η

εποπτευόμενη μάθηση συμβάλλει στην επίλυση ποικίλων προβλημάτων του πραγματικού κόσμου σε κλίμακα. Οι τύποι αλγορίθμων εποπτευόμενης μάθησης περιλαμβάνουν την ενεργή μάθηση, την ταξινόμηση και την παλινδρόμηση. Οι αλγόριθμοι ταξινόμησης χρησιμοποιούνται όταν οι έξοδοι περιορίζονται σε ένα περιορισμένο σύνολο τιμών και οι αλγόριθμοι παλινδρόμησης χρησιμοποιούνται όταν οι έξοδοι ενδέχεται να έχουν οποιαδήποτε αριθμητική τιμή εντός ενός φάσματος.

4.2 Μη εποπτευόμενη μάθηση

Η μάθηση χωρίς επίβλεψη περιλαμβάνει την εύρεση προτύπων ή δομών σε δεδομένα χωρίς ετικέτα. Αυτοί οι αλγόριθμοι ανακαλύπτουν κρυφά μοτίβα ή ομαδοποιήσεις δεδομένων χωρίς την ανάγκη ανθρώπινης παρέμβασης. Η ικανότητα αυτής της μεθόδου να ανακαλύπτει ομοιότητες και διαφορές στις πληροφορίες την καθιστά ιδανική για διερευνητική ανάλυση δεδομένων, στρατηγικές cross-selling, τμηματοποίηση πελατών και αναγνώριση εικόνας και προτύπων. Χρησιμοποιείται επίσης για τη μείωση του αριθμού των χαρακτηριστικών σε ένα μοντέλο μέσω της διαδικασίας μείωσης διαστάσεων. Οι αλγόριθμοι μάθησης χωρίς επίβλεψη, αντί να ανταποκρίνονται στην ανατροφοδότηση εντοπίζουν κοινά σημεία στα δεδομένα και αντιδρούν με βάση την παρουσία ή την απουσία τέτοιων κοινών στοιχείων σε κάθε νέο κομμάτι δεδομένων. Μια κοινή εφαρμογή είναι η ομαδοποίηση, όπου τα σημεία δεδομένων ομαδοποιούνται βάσει ομοιοτήτων. Οι κατηγοριοποιήσεις της μη επιβλεπόμενης μηχανικής μάθησης περιλαμβάνουν: τη συσταδοποίηση, τη μείωση διαστάσεων, εκμάθηση κανόνων συσχέτισης, ανίχνευση ανωμαλιών με τον προσδιορισμό ακραίων τιμών, μοντέλα παραγωγής

4.3 Ενισχυτική μάθηση

Αυτός ο τύπος μάθησης αφορά τους πράκτορες που μαθαίνουν πώς να λαμβάνουν αλληλουχίες αποφάσεων λαμβάνοντας ανατροφοδότηση με τη μορφή ανταμοιβών ή ποινών. Αυτό το μοντέλο μαθαίνει καθώς προχωρά χρησιμοποιώντας δοκιμή και σφάλμα. Μια σειρά επιτυχών αποτελεσμάτων θα ενισχυθεί για να αναπτυχθεί η καλύτερη σύσταση ή πολιτική για ένα δεδομένο πρόβλημα.

Σε αυτήν την ενότητα, θα πραγματοποιηθεί εμβάθυνση στη σφαίρα των τεχνικών μηχανικής μάθησης, διερευνώντας τόσο εποπτευόμενους όσο και μη εποπτευόμενους αλγόριθμους μάθησης καθώς και μοντέλα βαθιάς μάθησης. Αυτές οι τεχνικές είναι

θεμελιώδεις για την αξιοποίηση της δύναμης της μηχανικής μάθησης για εργασίες επεξεργασίας σήματος.

4.4 Βαθιά Μάθηση

Η βαθιά μηχανική μάθηση υπάγεται στην μηχανική μάθηση και ξεχωρίζει από αυτή με την χρήση πολλαπλών επιπέδων στον δίκτυο. Επομένως μπορεί να είναι είτε επιβλεπόμενη είτε μη επιβλεπόμενη. Τα μοντέλα βαθιάς μηχανικής μάθησης έχουν σχεδιαστεί για να μαθαίνουν αυτόματα σχετικά χαρακτηριστικά από δεδομένα, εξαλείφοντας μεγάλο μέρος της ανάγκης για χειροποίητη μηχανική χαρακτηριστικών. Αυτή η πτυχή είναι ιδιαίτερα πλεονεκτική στην επεξεργασία σήματος, όπου τα ακατέργαστα δεδομένα μπορεί να είναι πολύπλοκα και υψηλών διαστάσεων.

Ένα από τα βασικά πλεονεκτήματα της βαθιάς μάθησης είναι η ικανότητά της να εκτελεί εκμάθηση χαρακτηριστικών. Στην παραδοσιακή μηχανική εκμάθηση, οι ειδικοί συχνά σχεδιάζουν χειροκίνητα χαρακτηριστικά που βοηθούν τους αλγόριθμους να κατανοούν τα δεδομένα. Η βαθιά μάθηση, από την άλλη πλευρά, μπορεί να μάθει αυτόματα αυτά τα χαρακτηριστικά από τα δεδομένα, καθιστώντας την εξαιρετικά προσαρμόσιμη σε διαφορετικές εργασίες επεξεργασίας σήματος.

Τα βαθιά νευρωνικά δίκτυα επεξεργάζονται τις πληροφορίες ιεραρχικά. Μαθαίνουν χαρακτηριστικά χαμηλού επιπέδου, όπως άκρες και υφές στα αρχικά επίπεδα και σταδιακά δημιουργούν περίπλοκες, υψηλού επιπέδου αναπαραστάσεις στα επόμενα επίπεδα. Αυτή η ιεραρχική προσέγγιση είναι ανεκτίμητη όταν ασχολούμαστε με σήματα όπου υπάρχουν μοτίβα σε πολλαπλές κλίμακες.

Είναι επίσης πολύ επεκτάσιμα. Μπορούν να χειριστούν τεράστια σύνολα δεδομένων και πολύπλοκα προβλήματα, καθιστώντας τα καταλληλά για την αποτελεσματική επεξεργασία μεγάλων όγκων δεδομένων σήματος.

Στη βαθιά μάθηση, κάθε επίπεδο μαθαίνει να μετατρέπει τα δεδομένα εισόδου του σε μια ελαφρώς πιο αφηρημένη και σύνθετη αναπαράσταση. Σε μια εφαρμογή αναγνώρισης εικόνας, η ακατέργαστη είσοδος μπορεί να είναι μια μήτρα εικονοστοιχείων. Το πρώτο αναπαραστατικό επίπεδο μπορεί να αφαιρέσει τα εικονοστοιχεία και να κωδικοποιήσει τις άκρες. Το δεύτερο στρώμα μπορεί να συνθέτει και να κωδικοποιεί διατάξεις ακμών. Το τρίτο στρώμα μπορεί να κωδικοποιεί τη μύτη

και τα μάτια. και το τέταρτο στρώμα μπορεί να αναγνωρίσει ότι η εικόνα περιέχει ένα πρόσωπο[38]. Είναι σημαντικό ότι μια διαδικασία βαθιάς μάθησης μπορεί να μάθει από μόνη της ποια χαρακτηριστικά θα τοποθετήσει βέλτιστα σε ποιο επίπεδο. Αυτό δεν εξαλείφει την ανάγκη για συντονισμό με το χέρι. Για παράδειγμα, διαφορετικοί αριθμοί επιπέδων και μεγεθών επιπέδων μπορούν να παρέχουν διαφορετικούς βαθμούς αφαίρεσης.

Η λέξη "Βαθιά" στη βαθιά μηχανική μάθηση αναφέρεται στον αριθμό των επιπέδων μέσω των οποίων μετασχηματίζονται τα δεδομένα. Πιο συγκεκριμένα, τα συστήματα βαθιάς μάθησης έχουν πολλές μετατροπές μεταξύ εισόδου και εξόδου (CAP). Η CAP είναι η αλυσίδα των μετασχηματισμών από την είσοδο στην έξοδο. Τα CAP περιγράφουν πιθανές αιτιώδεις συνδέσεις μεταξύ εισόδου και εξόδου. Για ένα νευρωνικό δίκτυο τροφοδοσίας, το βάθος των CAP είναι αυτό του δικτύου και είναι ο αριθμός των κρυφών επιπέδων συν ένα (καθώς το επίπεδο εξόδου είναι επίσης παραμετροποιημένο). Για επαναλαμβανόμενα νευρωνικά δίκτυα, στα οποία ένα σήμα μπορεί να διαδοθεί μέσω ενός στρώματος περισσότερες από μία φορές, το βάθος CAP είναι δυνητικά απεριόριστο[39]. Κανένα καθολικά αποδεκτό όριο βάρους δεν χωρίζει τη ρηχή μάθηση από τη βαθιά μάθηση, αλλά οι περισσότεροι ερευνητές συμφωνούν ότι η βαθιά μάθηση περιλαμβάνει βάθος CAP μεγαλύτερο από 2. Το CAP του βάρους 2 έχει αποδειχθεί ότι είναι μια καθολική προσέγγιση με την έννοια ότι μπορεί να μιμηθεί οποιαδήποτε συνάρτηση. Πέρα από αυτό, περισσότερα επίπεδα δεν προσθέτουν στην ικανότητα προσέγγισης συναρτήσεων του δικτύου. Τα μοντέλα βαθιάς μάθησης ($CAP > 2$) μπορούν να εξάγουν καλύτερα χαρακτηριστικά από τα ρηχά μοντέλα και ως εκ τούτου, τα επιπλέον επίπεδα βοηθούν στην αποτελεσματική εκμάθηση των χαρακτηριστικών.

Η βαθιά μάθηση θα είναι ανεκτίμητη στο πλαίσιο μεγάλων δεδομένων, καθώς εξάγει υψηλού επιπέδου πληροφορίες από πολύ μεγάλους όγκους δεδομένων[40].

Η έρευνα της βαθιάς μάθησης επιχειρεί να μοντελοποιήσει δεδομένα μεγάλης κλίμακας με πολλαπλά στρώματα επεξεργασίας με πολύπλοκες δομές. Έτσι, σε αντίθεση με τις αρχιτεκτονικές των ρηχών μοντέλων μηχανικής μάθησης, οι αρχιτεκτονικές βαθιάς μάθησης αποτελούνται από πολλαπλές μη γραμμικές μεταμορφώσεις. Υπάρχει ένας τεράστιος αριθμός παραλλαγών αρχιτεκτονικών σε βάθος και διαφορετικές αρχιτεκτονικές μπορούν να χρησιμοποιηθούν για την

αναπαράσταση διαφορετικών πηγών δεδομένων. Για παράδειγμα, το συνελκτικό νευρωνικό δίκτυο είναι η πιο δημοφιλής αρχιτεκτονική για την αναγνώριση εικόνων, και το επαναλαμβανόμενο νευρωνικό δίκτυο είναι περισσότερο εφαρμόσιμο σε διαδοχικές εργασίες όπως η αναγνώριση γραπτού και προφορικού λόγου[41]. Στη συνέχεια, θα αναφερθούν αυτές οι αρχιτεκτονικές και θα ερευνηθεί η καταλληλότητα τους για χειρισμό διαφορετικών δεδομένων.

4.4.1 Συνελκτικά νευρωνικά δίκτυα (CNN)

Τα συνελκτικά νευρωνικά δίκτυα (CNN) είναι μια εξειδικευμένη κατηγορία βαθιών νευρωνικών δικτύων που έχουν σχεδιαστεί για την επεξεργασία δεδομένων τύπου πλέγματος, όπως εικόνες και βίντεο. Έχουν φέρει επανάσταση σε πολλούς τομείς, συμπεριλαμβανομένης της όρασης υπολογιστών, της αναγνώρισης ομιλίας και της επεξεργασίας σήματος, λόγω της ικανότητάς τους να καταγράφουν αποτελεσματικά τις χωρικές ιεραρχίες. Είναι ένας τύπος τεχνητών νευρικών δικτύων εμπρόσθιας τροφοδοσίας (feed-forward) που χρησιμοποιούν συνέλιξη σε τουλάχιστον ένα από τα επίπεδα τους. Είναι εμπνευσμένα από βιολογικά νευρωνικά δίκτυα. Το CNN συνδυάζει τεχνητά νευρωνικά δίκτυα και διακριτή συνέλιξη για την επεξεργασία εικόνας που μπορεί να χρησιμοποιηθεί για την αυτόματη εξαγωγή χαρακτηριστικών. Συνεπώς είναι ειδικά σχεδιασμένο για την αναγνώριση δισδιάστατων δεδομένων, όπως π.χ εικόνες και βίντεο. Οι εικόνες μπορούν να χρησιμοποιηθούν απευθείας ως είσοδος του δικτύου, το οποίο αποφεύγει τη σύνθετη διαδικασία εξαγωγής χαρακτηριστικών και ανακατασκευής δεδομένων όπως στους παραδοσιακούς αλγόριθμους αναγνώρισης εικόνας.

Τα συνελκτικά νευρωνικά δίκτυα μοντελοποιούνται ως συλλογές νευρώνων που συνδέονται σε ένα ακυκλικό γράφημα. Κάθε επίπεδο αυτού του δικτύου παίρνει τις εξόδους του προηγούμενου στρώματος ως εισόδους. Επομένως, οι κύκλοι δεν επιτρέπονται στην αρχιτεκτονική των CNN. Συνήθως, οι ξεχωριστοί νευρώνες μεταξύ δύο γειτονικών στρωμάτων είναι πλήρως συνδεδεμένοι, αλλά οι νευρώνες σε ένα μόνο στρώμα δεν μοιράζονται καμία σύνδεση. Ένα βαθύ νευρωνικό συνελκτικό δίκτυο περιλαμβάνει περισσότερα από 2 επίπεδα. Μπορεί να υπάρχουν 20 ή περισσότερα στρώματα και το κάθε στρώμα μπορεί να είναι λειτουργικά διαφορετικό. Τα νευρωνικά συνελκτικά δίκτυα περιλαμβάνουν τις εξής κατηγορίες επιπέδων:

4.4.1.1 Συνελικτικά Επίπεδα

Τα CNN αντλούν το όνομά τους από την εκτεταμένη χρήση συνελικτικών επιπέδων. Αυτά τα επίπεδα εφαρμόζουν συνελικτικές λειτουργίες χρησιμοποιώντας φίλτρα ή πυρήνες με δυνατότητα εκμάθησης για τη σάρωση των δεδομένων εισόδου. Συνδυάζοντας τα δεδομένα, τα συνελικτικά επίπεδα μπορούν να ανιχνεύσουν βασικά χαρακτηριστικά όπως άκρες, υφές και τοπικά μοτίβα. Αυτό το χωρικό φιλτράρισμα τους επιτρέπει να συλλαμβάνουν σχετικές πληροφορίες, ενώ μειώνουν τη διάσταση των δεδομένων, καθιστώντας τα υπολογιστικά διαχειρίσιμα.

4.4.1.2 Στρώματα συγκέντρωσης(pooling)

Μετά από συνελικτικές λειτουργίες, τα συνελικτικά νευρωνικά δίκτυα συχνά ενσωματώνουν στρώματα συγκέντρωσης. Η ομαδοποίηση είναι μια τεχνική downsampling που μειώνει τις χωρικές διαστάσεις των δεδομένων. Βοηθά στη διατήρηση της μεταφραστικής μεταβλητότητας του δικτύου, επιτρέποντάς του να αναγνωρίζει χαρακτηριστικά ανεξάρτητα από την ακριβή τους θέση. Η συγκέντρωση συμβάλλει επίσης στη στιβαρότητα και την αποτελεσματικότητα του μοντέλου. Το επίπεδο συγκέντρωσης λαμβάνει ένα μέρος του δείγματος από τα δεδομένα εισόδου. Το στρώμα συγκέντρωσης διαιρεί την είσοδο σε πολλές μικρές περιοχές και εφαρμόζει ορισμένες συναρτήσεις όπως η εύρεση του μέσου όρου ή τού μεγιστού σε κάθε περιοχή. Σε αντίθεση με το συνελικτικά επίπεδα, δεν υπάρχει επικάλυψη στον υπολογισμό των στρωμάτων συγκέντρωσης.

4.4.1.3 Πλήρως συνδεδεμένα επίπεδα

Προς το τέλος μιας αρχιτεκτονικής CNN, τα πλήρως συνδεδεμένα επίπεδα συγκεντρώνουν χαρακτηριστικά υψηλού επιπέδου και συνήθως ακολουθούνται από συναρτήσεις ενεργοποίησης. Αυτά τα επίπεδα επιτρέπουν στο δίκτυο να κάνει προβλέψεις με βάση τις αναπαραστάσεις που έχει μάθει το μοντέλο. Το βάθος και το πλάτος αυτών των πλήρως συνδεδεμένων στρωμάτων μπορούν να ρυθμιστούν ώστε να ταιριάζουν στην πολυπλοκότητα της εργασίας. Το πλήρως συνδεδεμένο στρώμα παίρνει όλους τους νευρώνες από το προηγούμενο στρώμα και συνδέεται σε κάθε νευρώνα που έχει. Τα πλήρως συνδεδεμένα στρώματα δεν βρίσκονται χωρικά πια

(μπορούν να οπτικοποιηθούν ότι είναι μονοδιάστατα), επομένως δεν μπορούν να υπάρχουν συνελκτικά στρώματα μετά από ένα πλήρως συνδεδεμένο στρώμα[42].

4.4.1.4 Χρήση Νευρωνικών συνελκτικών Δικτύων

Συμπερασματικά τα CNN χρησιμοποιούνται σε πληθώρα εργασιών στην επεξεργασία σήματος όπως είναι η ταξινόμηση εικόνων, η ανίχνευση αντικειμένων και η τμηματοποίηση εικόνας. Τα συνελκτικά νευρωνικά δίκτυα υπερέχουν στην ταξινόμηση εικόνων, όπου ο στόχος είναι να αποδοθεί μια ετικέτα σε μια εικόνα με βάση το περιεχόμενό της. Μοντέλα όπως τα AlexNet, VGG και ResNet αποτελούν σημεία αναφοράς στην ακρίβεια ταξινόμησης εικόνων. Πλέον τα συνελκτικά νευρωνικά δίκτυα κρίνονται απαραίτητα για εργασίες που αφορούν τον εντοπισμό και την ταξινόμηση αντικειμένων εντός εικόνων ή καρέ βίντεο. Προσεγγίσεις όπως το R-CNN αξιοποιούν τα συνελκτικά νευρωνικά δίκτυα για να επιτύχουν αποτελέσματα τεχνολογιών αιχμής. Στην ιατρική απεικόνιση και την όραση υπολογιστών, η τμηματοποίηση εικόνων διαχωρίζει τις εικόνες σε σημαντικές περιοχές ή αντικείμενα. Τα συνελκτικά νευρωνικά δίκτυα έχουν επιδείξει εξαιρετικές επιδόσεις σε αυτόν τον τομέα. Πέρα από την όραση υπολογιστών, τα συνελκτικά νευρωνικά δίκτυα έχουν επίσης εφαρμοστεί σε εργασίες επεξεργασίας σήματος, συμπεριλαμβανομένης της αναγνώρισης ομιλίας, της ανάλυσης ήχου και της πρόβλεψης χρονοσειρών. Αυτό δείχνει την ευελιξία τους και τη δυνατότητα για μεταφορά γνώσης μεταξύ τομέων.

4.4.2 Extreme Gradient Boosting (XGBoost)

Το Extreme Gradient Boosting, που συχνά συντομεύεται ως XGBoost, είναι ένας ισχυρός και ευέλικτος αλγόριθμος μηχανικής εκμάθησης που έχει αποκτήσει τεράστια δημοτικότητα τόσο στις παραδοσιακές εφαρμογές μηχανικής εκμάθησης όσο και στις εφαρμογές επεξεργασίας σήματος. Το XGBoost ανήκει στην οικογένεια εκμάθησης συνόλου, γνωστή για την εξαιρετική απόδοση πρόβλεψης και τη στιβαρότητά του. Το XGBoost βασίζεται στην έννοια της εκμάθησης συνόλου, όπου πολλαπλά μοντέλα, συχνά δέντρα αποφάσεων, συνδυάζονται για να σχηματίσουν ένα ισχυρότερο, πιο ακριβές μοντέλο. Αυτά τα δέντρα αποφάσεων είναι γνωστά ως αδύναμα μοντέλα εκμάθησης και είναι συνήθως ρηγά για να αποφευχθεί η υπερβολική προσαρμογή. Το XGBoost χρησιμοποιεί ένα σύνολο τέτοιων δέντρων, όπου κάθε επόμενο δέντρο διορθώνει τα λάθη που έγιναν από τα προηγούμενα. Το XGBoost προσφέρει πολλά

πλεονεκτήματα όταν εφαρμόζεται σε εργασίες επεξεργασίας σήματος. Μερικά από αυτά είναι ότι:

- Μπορεί να χειριστεί αποτελεσματικά θορυβώδη δεδομένα και ακραίες τιμές, κάτι που είναι απαραίτητο για την επεξεργασία σήματος όπου η ποιότητα των δεδομένων μπορεί να ποικίλει.
- Παρέχει πληροφορίες σχετικά με τη σημασία των χαρακτηριστικών, επιτρέποντάς να προσδιοριστούν ποια σήματα ή χαρακτηριστικά συμβάλλουν περισσότερο στις προβλέψεις του μοντέλου. Αυτό μπορεί να βοηθήσει στην επιλογή χαρακτηριστικών και στην ανάλυση σήματος.
- Είναι επεκτάσιμο και μπορεί να φιλοξενήσει μεγάλα σύνολα δεδομένων, καθιστώντας το κατάλληλο για την επεξεργασία εκτεταμένων δεδομένων σήματος.
- Υποστηρίζει ταξινόμηση πολλαπλών κλάσεων και πολλαπλών ετικετών, επιτρέποντας την ταξινόμηση των σημάτων σε πολλαπλές κατηγορίες ή ετικέτες.

Κεφάλαιο 5

Μετρικές Αξιολόγησης

- **Ακρίβεια (Accuracy):**
Η ακρίβεια μιας μετρικής αξιολόγησης αναφέρεται στο πόσο συχνά ένα μοντέλο καταφέρνει να προβλέψει σωστά την κλάση ενός δείγματος. Υπολογίζεται ως τον λόγο των σωστών προβλέψεων προς τον συνολικό αριθμό των προβλέψεων.
- **Ακρίβεια (Precision):**
Η ακρίβεια μιας μετρικής αξιολόγησης υπολογίζει το πόσο ακριβείς είναι οι θετικές προβλέψεις σε σχέση με το σύνολο των προβλέψεων που έγιναν για μια συγκεκριμένη κλάση. Υπολογίζεται ως τον λόγο των σωστών θετικών προβλέψεων προς τον συνολικό αριθμό των προβλέψεων που έγιναν για την αντίστοιχη κλάση.
- **Ανάκληση (Recall):**
Η ανάκληση μιας μετρικής αξιολόγησης μετρά το πόσα από τα πραγματικά θετικά δείγματα κατάφερε να προβλέψει σωστά το μοντέλο. Υπολογίζεται ως ο λόγος των σωστών θετικών προβλέψεων προς τον συνολικό αριθμό των πραγματικών θετικών δειγμάτων.
- **F1 Σκορ (F1 Score):**
Το F1 σκορ είναι ένας μέσος όρος της ακρίβειας και της ανάκλησης. Υπολογίζεται ως ο αντίστροφος του ανακρατημένου αριθμού της μέσης τιμής των αντίστοιχων αντίστοιχων τιμών. Αποτελεί ένα καλό μέτρο όταν οι κλάσεις δεδομένων είναι μη ισορροπημένες.
- **Μετρική Τιμή Silhouette**
Για να υπολογιστεί η βαθμολογία Silhouette για κάθε παρατήρηση/σημείο δεδομένων, πρέπει να βρεθούν οι ακόλουθες αποστάσεις για όλα τα σημεία: Αρχικά μετριέται η μέση απόσταση ανάμεσα στην παρατήρηση και στα υπόλοιπα δεδομένα του ίδιου συμπλέγματος. Η απόσταση αυτή μπορεί επίσης να ονομαστεί ως μέση απόσταση εντός συστάδας και συμβολίζεται με το γράμμα a . Στη συνέχεια υπολογίζεται η μέση απόσταση μεταξύ της

παρατήρησης και όλων των άλλων σημείων του επόμενου πλησιέστερου συμπλέγματος. Αυτή η απόσταση μπορεί επίσης να ονομαστεί ως μέση απόσταση πλησιέστερης συστάδας και συμβολίζεται με το γράμμα b . Η βαθμολογία S για κάθε παρατήρηση υπολογίζεται από την εξίσωση. Η τιμή κυμαίνεται από -1 έως 1 . Εάν η βαθμολογία είναι 1 , το σύμπλεγμα είναι πυκνό και καλά διαχωρισμένο από άλλα συμπλέγματα. Μια τιμή κοντά στο 0 αντιπροσωπεύει επικαλυπτόμενες συστάδες με δείγματα πολύ κοντά στο όριο απόφασης των γειτονικών συστάδων. Μια αρνητική βαθμολογία $[-1, 0]$ υποδηλώνει ότι τα δείγματα μπορεί να έχουν εκχωρηθεί σε λάθος ομάδες.

Κεφάλαιο 6

Σχετική Έρευνα

Αυτό το κεφάλαιο είναι αφιερωμένο στην ανασκόπηση της υπάρχουσας έρευνας στον τομέα της ανίχνευσης της νόσου Covid-19 από ηχογραφήσεις αναπνοής, βήχα και ομιλίας. Πρόκειται για έναν σχετικά νέο τομέα έρευνας που προέκυψε ως απάντηση στην πανδημία της Covid-19, καθώς οι ερευνητές επιδιώκουν να αναπτύξουν μη επεμβατικές και χαμηλού κόστους μεθόδους για τη διάγνωση της νόσου. Το κεφάλαιο εξετάζει διαφορετικές προσεγγίσεις και τεχνικές που έχουν χρησιμοποιηθεί για την ανάλυση ηχογραφήσεων, συμπεριλαμβανομένων αλγορίθμων μηχανικής μάθησης, ακουστικής ανάλυσης και εμβιομηχανικών μοντέλων. Εξετάζονται τα πλεονεκτήματα και οι περιορισμοί κάθε προσέγγισης και συζητούνται οι δυνατότητές τους για κλινική χρήση. Το κεφάλαιο βασίζεται σε μια σειρά ερευνητικών άρθρων από διάφορους κλάδους, συμπεριλαμβανομένης της ιατρικής, της μηχανικής και της επιστήμης των υπολογιστών, παρέχοντας μια ολοκληρωμένη επισκόπηση της τελευταίας τεχνολογίας σε αυτόν τον τομέα. Τα ευρήματα αυτής της ανασκόπησης έχουν επιπτώσεις για την ανάπτυξη νέων διαγνωστικών εργαλείων και την ανάπτυξη υπαρχουσών τεχνολογιών για την καταπολέμηση της Covid-19.

Η εργασία με τίτλο "Detection of COVID-19 in smartphone-based breathing recordings: A pre-screening deep learning tool" των Mohanad Alkhodari και Ahsan H. Khandoker παρουσιάζει μια μελέτη που στοχεύει στη διερεύνηση της σκοπιμότητας χρήσης αναπνευστικών ήχων που βασίζονται στη χρήση smartphone και βαθιάς μάθησης για τη διάγνωση ασθενών με COVID-19, συμπεριλαμβανομένων των ασυμπτωματικών περιπτώσεων. Η μελέτη χρησιμοποίησε μια δημόσια διαθέσιμη βάση δεδομένων που ονομάζεται Coswara, η οποία παρείχε 480 ήχους αναπνοής που καταγράφηκαν από 120 άτομα με COVID-19 και 120 υγιή άτομα μέσω μικροφώνου smartphone μέσω μιας εφαρμογής ιστότοπου.

Το πλαίσιο βαθιάς μάθησης που προτείνεται στη μελέτη βασίζεται σε έναν συνδυασμό χειροποίητων χαρακτηριστικών και βαθιά ενεργοποιημένων χαρακτηριστικών που μαθαίνονται μέσω μιας αρχιτεκτονικής νευρωνικών δικτύων που αποτελείται από ένα συνελκτικό νευρωνικό δίκτυο (CNN) και αμφίδρομες μονάδες βραχυπρόθεσμης

μνήμης (BiLSTM). Οι συγγραφείς διεξήγαγαν μια στατιστική ανάλυση των προφίλ των ασθενών, αποκαλύπτοντας μια σημαντική διαφορά στην ισχαιμική καρδιακή νόσο μεταξύ ατόμων με COVID-19 και υγιών ατόμων. Επιπρόσθετα, ανέλυσαν την κατανομή των τιμών των συνδυασμένων εγκεφαλικών συντελεστών συχνότητας Mel (MFCC) και πρότειναν μια προσέγγιση βαθιάς μάθησης που πέτυχε υψηλή ακρίβεια διάγνωσης τόσο για ρηχές όσο και για βαθιές αναπνοές.

Η εργασία εξετάζει επίσης τα βήματα προετοιμασίας δεδομένων, συμπεριλαμβανομένων τεχνικών αύξησης δεδομένων, όπως ο έλεγχος έντασης και η χρονική μετατόπιση, η επιλογή χαρακτηριστικών με τη χρήση τεστ χ -τετράγωνου (chi-square test). Το μοντέλο βαθιάς μάθησης βελτιστοποιήθηκε χρησιμοποιώντας τον ADAM solver με συγκεκριμένες παραμέτρους διαμόρφωσης και η απόδοσή του αξιολογήθηκε χρησιμοποιώντας διάφορες μετρήσεις, όπως ακρίβεια, ευαισθησία, ειδικότητα, ακρίβεια, βαθμολογία F1 και περιοχή κάτω από τις καμπύλες χαρακτηριστικών λειτουργίας δέκτη (AUROC).

Τα ευρήματα της μελέτης υποδηλώνουν ότι το προτεινόμενο μοντέλο βαθιάς μάθησης μπορεί να κάνει αποτελεσματικά τη διάκριση μεταξύ πασχόντων της νόσου Covid-19 και υγιών ατόμων, με πολλά υποσχόμενα αποτελέσματα για σκοπούς προκαταρκτικού ελέγχου. Αυτή η έρευνα έχει τη δυνατότητα να προσφέρει μια πρώιμη, γρήγορη και οικονομικά αποδοτική μέθοδο για την ανίχνευση της νόσου Covid-19, συμπληρώνοντας την υπάρχουσα ανάλυση αλυσιδωτής αντίδρασης πολυμεράσης ανάστροφης μεταγραφής (RT-PCR), ειδικά κατά τη διάρκεια της πανδημίας COVID-19[43].

Η ερευνητική μελέτη από τους Madhurananda Pahar, Marisa Klopper, Robin Warren και Thomas Niesler με τίτλο "COVID-19 detection in cough, breath and speech using deep transfer learning and bottleneck features" παρουσιάζει μια πειραματική έρευνα για την αποτελεσματικότητα της μεταφοράς μάθησης και της εξαγωγής χαρακτηριστικών στον εντοπισμό της COVID-19 από ηχογραφήσεις βήχα, αναπνοής και ομιλίας. Ο στόχος είναι η δημιουργία μιας μεθόδου ανίχνευσης COVID-19 που είναι μη επαφική, προσβάσιμη και οικονομικά εφικτή και μπορεί να χρησιμοποιηθεί σε φθηνές συσκευές καταναλωτών, όπως ένα smartphone.

Οι ερευνητές χρησιμοποιούν σύνολα δεδομένων που περιλαμβάνουν διάφορους ήχους, συμπεριλαμβανομένων βήχα, φτέρνισμα, ομιλία και άλλους θορύβους, αλλά δεν περιέχουν ετικέτες COVID-19. Προεκπαιδεύουν τρία βαθιά νευρωνικά δίκτυα: ένα CNN, ένα LSTM και ένα Resnet50. Αυτά τα προεκπαιδευμένα δίκτυα είτε προσαρμόζονται μέσω μιας διαδικασίας μάθησης, χρησιμοποιώντας μικρότερα σύνολα δεδομένων που περιέχουν ήχους βήχα με ετικέτες COVID-19, είτε χρησιμοποιούνται ως εξαγωγείς χαρακτηριστικών. Διάφορα χαρακτηριστικά εξάγονται από ηχητικά σήματα στον τομέα του χρόνου, συμπεριλαμβανομένων των συντελεστών συχνότητας Mel (MFCCs), των ενεργειών της τράπεζας φίλτρων καταγραφής σε γραμμική απόσταση, των συντελεστών ταχύτητας και επιτάχυνσης των MFCC, του ρυθμού μηδενικής διέλευσης (ZCR) και της κύρτωσης. Τα επικαλυπτόμενα πλαίσια χρησιμοποιούνται για να εξασφαλίσουν μια σταθερή διάσταση εισόδου για ταξινόμηση, διατηρώντας παράλληλα τη χρονική δομή του ήχου. Οι υπερπαραμέτροι εξαγωγής χαρακτηριστικών περιλαμβάνουν το μήκος πλαισίου (\mathcal{F}), τον αριθμό πλαισίων (\mathcal{S}), τον αριθμό των MFCC κατώτερης τάξης (\mathcal{M}) και τον αριθμό των φίλτρων γραμμικής απόστασης (\mathcal{B}). Ο ταξινομητής Resnet50, εκπαιδευμένος μέσω αυτής της διαδικασίας μεταφοράς μάθησης, επιτυγχάνει εξαιρετική απόδοση, με περιοχές κάτω από τη χαρακτηριστική καμπύλη λειτουργίας (ROC AUC) των 0,98, 0,94 και 0,92 αντίστοιχα για βήχα, αναπνοή και ομιλία. Ο βήχας φέρει το ισχυρότερο χαρακτηριστικό της COVID-19, ακολουθεί η αναπνοή και η ομιλία. Η εφαρμογή της μεταφοράς μάθησης και η εξαγωγή χαρακτηριστικών μπουκαλιού χρησιμοποιώντας μεγαλύτερα σύνολα δεδομένων χωρίς ετικέτες COVID-19 οδηγεί όχι μόνο σε βελτιωμένη απόδοση αλλά και σε σημαντική μείωση του προτύπου απόκλισης των AUC του ταξινομητή, προσφέροντας καλύτερη γενίκευση[44].

Στη μελέτη με τίτλο «Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data» οι Chloë Brown et.al περιγράφουν την ανάλυσή τους επί μιας μεγάλης συλλογής δεδομένων αναπνοής που συγκεντρώθηκε από το κοινό με σκοπό την υποστήριξη στη διάγνωση της Covid-19. Οι συγγραφείς εξετάζουν την δυνατότητα ανίχνευσης των ήχων που σχετίζονται με την COVID-19 από άλλους ήχους αναπνοής, όπως αυτούς που προκαλούνται από την άσθμα ή από υγιή άτομα. Τα αποτελέσματα της μελέτης δείχνουν ότι ακόμη και ένας απλός ταξινομητής μηχανικής μάθησης είναι σε θέση να διακρίνει σωστά τους ήχους που σχετίζονται με την υγεία από αυτούς που σχετίζονται με την COVID-19. Επιπλέον,

δείχνουν πώς μπορούν να διακρίνουν έναν χρήστη που έχει θετικό τεστ για την COVID-19 και πάσχει από βήχα από έναν υγιή χρήστη με βήχα, καθώς και από χρήστες με θετικό τεστ COVID-19 και βήχα από χρήστες που πάσχουν από άσθμα και έχουν βήχα. Τα μοντέλα τους επιτυγχάνουν ένα AUC άνω του 80% σε όλες τις δοκιμές. Στην έρευνα αυτή, χρησιμοποιούνται διάφορα χαρακτηριστικά ήχου για τον αυτόματο εντοπισμό διαφορετικών πτυχών της ηχογραφημένης αναπνοής. Αυτά τα χαρακτηριστικά περιλαμβάνουν: **Tempo, Period, RMS Energy, Spectral Centroid, Roll-off Frequency, Zero-crossing, MFCC (Mel-Frequency Cepstral Coefficients)**.

Επιπρόσθετα, εξάγονται διάφορα στατιστικά χαρακτηριστικά για τα χαρακτηριστικά που παράγουν χρονοσειρές, προκειμένου να αποτυπώσουν τις κατανομές πέρα από τη μέση τιμή. Ακόμη, χρησιμοποιούνται χαρακτηριστικά από μάθηση μεταφοράς μέσω του VGGish, ένα νευρωνικό δίκτυο που χρησιμοποιείται για να εξάγει αυτόματα χαρακτηριστικά από τον ήχο. Συνολικά, δημιουργούνται διανύσματα χαρακτηριστικών με διάσταση έως και 733 για κάθε μονάδα (βήχας, αναπνοή) και αυτά τα χαρακτηριστικά υποβάλλονται σε μείωση διαστάσεων μέσω της Ανάλυσης Κύριων Συνιστωσών (PCA) για να διατηρηθεί ένα τμήμα της αρχικής εξηγημένης διακύμανσης.

Οι συγγραφείς επικεντρώνονται σε τρεις κλινικά σημαντικές δυαδικές εργασίες ταξινόμησης, βασισμένες στη συλλογή δεδομένων ήχου αναπνοής και βήχα:

Εργασία 1: Διάκριση μεταξύ χρηστών που έχουν δηλώσει ότι διαγνώστηκαν θετικοί για COVID-19 (COVID-θετικοί) και χρηστών που δεν έχουν δηλώσει θετικό τεστ για COVID-19, διαθέτουν άδειο ιατρικό ιστορικό, δεν έχουν ποτέ καπνίσει, δεν έχουν συμπτώματα και βρίσκονται σε χώρες όπου τότε δεν είχε εξαπλωθεί η COVID-19 (μη-COVID).

Εργασία 2: Διάκριση μεταξύ χρηστών που έχουν δηλώσει ότι διαγνώστηκαν θετικοί για COVID-19 και έχουν δηλώσει βήχα ως σύμπτωμα (ένα συχνό σύμπτωμα σε ασθενείς με COVID) (COVID-θετικοί με βήχα) από χρήστες που δήλωσαν ότι δεν έχουν δοκιμάσει θετικά για COVID-19, έχουν άδειο ιατρικό ιστορικό, δεν έχουν ποτέ καπνίσει, βρίσκονται σε χώρες όπου τότε δεν είχε εξαπλωθεί η COVID-19 και έχουν βήχα ως σύμπτωμα (μη-COVID με βήχα).

Εργασία 3: Διάκριση μεταξύ χρηστών που έχουν δηλώσει ότι διαγνώστηκαν θετικοί για COVID-19 και έχουν δηλώσει βήχα ως σύμπτωμα (COVID-θετικοί με βήχα) από χρήστες που δήλωσαν ότι δεν έχουν δοκιμάσει θετικά για COVID-19, προέρχονται από χώρες όπου τότε δεν είχε εξαπλωθεί η COVID-19, έχουν αναφέρει το άσθμα στο ιατρικό τους ιστορικό και έχουν βήχα ως σύμπτωμα (μη-COVID με βήχα).[45].

Στην μελέτη με τίτλο ‘Deep Learning Audio Spectrograms Processing to the Early COVID-19 Detection’ οι *Ciro Rodriguez Rodriguez et al.* επικεντρώνονται στην ανάπτυξη ενός μοντέλου βαθιάς μάθησης για την ανίχνευση περιπτώσεων COVID-19 μέσω ηχητικών φασματογραμμμάτων βήχα, φτάρνισματος και άλλων ηχητικών αναπνοής από ανθρώπους που έχουν μολυνθεί. Η μεθοδολογία βασίστηκε σε τεχνικές βαθιάς μάθησης, χρησιμοποιώντας ένα σύνολο δεδομένων με ηχητικά δείγματα από άτομα που είναι άρρωστα και άλλα που δεν είναι. Χρησιμοποιήθηκε η αρχιτεκτονική Xception του ImageNet για την εκπαίδευση του μοντέλου με την τεχνική Fine-Tuning. Τα αποτελέσματα έδειξαν μια ακρίβεια περίπου από 0,75 έως 0,80, η οποία επηρεάζεται δραστικά από την ποιότητα του διαθέσιμου συνόλου δεδομένων. Παρ' όλα αυτά, τα αποτελέσματα δείχνουν ότι το μοντέλο μπορεί να παρουσιάσει πολύ καλύτερα αποτελέσματα εάν εργαστεί με ένα σύνολο δεδομένων που περιλαμβάνει μόνο ηχητικά δείγματα αναπνοής από περιπτώσεις COVID-19 υψηλής ποιότητας.

Επιπλέον, αναφέρεται ότι η μελέτη επικεντρώνεται στην ανάπτυξη μιας εφαρμογής κινητού τηλεφώνου με την ονομασία "AI4COVID-19", η οποία βασίζεται στην τεχνητή νοημοσύνη και έχει ως στόχο την ταχεία ανίχνευση της Covid-19 για να καλύψει την έλλειψη ενός τεστ που θα καλύπτει μεγάλη κλίμακα του πληθυσμού. Συμπερασματικά, το μοντέλο μπορεί να χρησιμοποιηθεί για την ανίχνευση της Covid-19 μέσω των ηχητικών αναπνοής, αλλά η ποιότητα του συνόλου δεδομένων παίζει καθοριστικό ρόλο στην ακρίβειά του[46].

Στην *Diagnosis_of_pneumonia_from_sounds_collected_using_low_cost_cell_phones* από τους *Insu Song et al.*, αναπτύχθηκε ένα χαμηλού κόστους σύστημα βασισμένο σε κινητό τηλέφωνο για τη γρήγορη διάγνωση προβλημάτων υγείας του αναπνευστικού. Συλλέχθηκαν συνολικά 367 ηχητικά δείγματα αναπνοής από παιδικά νοσοκομεία για την ανάπτυξη ακριβών μοντέλων διάγνωσης. Πραγματοποιήθηκε λεπτομερής ανάλυση των ηχητικών δεδομένων, και επιλέχθηκαν σημαντικά στατιστικά χαρακτηριστικά για κάθε ηλικιακή ομάδα. Στη συνέχεια, αναπτύχθηκε ένα μοντέλο διάγνωσης που

αξιολογήθηκε σε ένα δυαδικό πρόβλημα ταξινόμησης: πνευμονία έναντι μη πνευμονίας. Η μεθοδολογία περιλαμβάνει τη συλλογή ήχου αναπνοής, τη μείωση θορύβου, την εξαγωγή χαρακτηριστικών από τον ήχο αναπνοής και τη στατιστική ανάλυση και επιλογή σημαντικών χαρακτηριστικών για την ταξινόμηση. Τέλος, χρησιμοποιήθηκαν δύο ταξινομητές, το K-Nearest Neighbor (K-NN) και το Support Vector Machine (SVM), για την ταξινόμηση ανάλογα με τις διάφορες ηλικιακές ομάδες. Τα αποτελέσματα έδειξαν ότι η προτεινόμενη μέθοδος μπορούσε να ταξινομήσει αποτελεσματικά την πνευμονία, ακόμη και σε περιβαλλοντικούς θορύβους, με ακρίβεια 91,98%, ευαισθησία 92,06% και ειδικότητα 90,68%. Τα αποτελέσματα υποδεικνύουν ότι οι ηχητικές αναπνοές που καταγράφονται με χαμηλού κόστους κινητές συσκευές μπορούν να χρησιμοποιηθούν για την αποτελεσματική ανίχνευση της πνευμονίας[47].

Η μελέτη με τίτλο " Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory ", πραγματοποιήθηκε από τους συγγραφείς M. Fraiwan, L. Fraiwan, M. Alkhodari, και O. Hassanin. Στόχος της μελέτης ήταν να εξεταστεί η ικανότητα της βαθιάς μάθησης στην αναγνώριση πνευμονικών νοσημάτων από ηχογραφημένους ήχους του πνεύμονα.

Για τη μελέτη, χρησιμοποιήθηκε ένα σύνολο δεδομένων που περιλάμβανε συνολικά 103 ασθενείς, οι οποίοι προήλθαν από ηχογραφήματα του στηθοσκοπίου που καταγράφηκαν στο Νοσοκομείο King Abdullah University της Ιορδανίας. Επιπλέον, προστέθηκαν δεδομένα από 110 ασθενείς από ένα δημόσια διαθέσιμο αποθετήριο δεδομένων του Συνεδρίου Βιοϊατρικής Υγείας.

Η αρχιτεκτονική του δικτύου βαθιάς μάθησης περιλάμβανε δύο στάδια: συνελκτικά νευρωνικά δίκτυα (CNN) και διπλές μονάδες μακροπρόθεσμης μνήμης (BDLSTM). Το μοντέλο εκπαιδεύτηκε χρησιμοποιώντας ένα σχήμα διασταυρούμενης επικύρωσης k-fold με δέκα διασταυρούμενες επικυρώσεις και αξιολογήθηκε χρησιμοποιώντας αρκετές μετρικές απόδοσης, συμπεριλαμβανομένου του Cohen's kappa, της ακρίβειας, της ευαισθησίας, της ειδικότητας, της ακρίβειας και του F1-score.

Ο αλγόριθμος πέτυχε μέση ακρίβεια(accuracy) 99,62% και σαφήνεια(precision) 98,85% στην κατηγοριοποίηση των ασθενών βάσει των τύπων των πνευμονικών νοσημάτων χρησιμοποιώντας το CNN + BDLSTM. Επιπλέον, επιτεύχθηκε συνολική

συμφωνία 98,26% μεταξύ των προβλέψεων και των αρχικών κατηγοριών εντός της διαδικασίας εκπαίδευσης.[48].

Η εργασία αυτή, με τίτλο "Respiratory Sound Analysis for Detection of Pulmonary Diseases" από τον Vipul Jindal, εξετάζει τη χρήση ανάλυσης των ήχων της αναπνοής ως διαγνωστικό εργαλείο για τα πνευμονικά νοσήματα. Συγκεκριμένα, εξετάζει τη χρήση της μεθόδου Welch για την ανάλυση των ήχων της αναπνοής. Χρησιμοποιώντας αυτήν τη μέθοδο, πραγματοποιείται η ταξινόμηση των κυματομορφών του φιλτραρισμένου ήχου της αναπνοής για διάφορες ηλικιακές ομάδες ανδρών και γυναικών. Αυτό γίνεται για να εντοπιστούν διάφορα νοσήματα, όπως άσθμα, βρογχίτιδα, έμφυση, φυματίωση και πνευμονία, με βάση τις κυματομορφές των ήχων της αναπνοής σε διάφορα εύρη φασματικής ισχύος.

Η μέθοδος Welch διακρίνει διάφορα είδη ήχων της αναπνοής, όπως σφυγμοί, ράλλοι, και βρογχοκήλες, και συνδυάζει αυτήν την ανάλυση με την δημιουργία φασματικών διαγραμμάτων ισχύος για την αναγνώριση νοσημάτων. Η έρευνα αυτή έχει σημαντική συνεισφορά στον τομέα της διάγνωσης πνευμονικών νοσημάτων μέσω της ανάλυσης των ήχων της αναπνοής[49].

Οι συγγραφείς Hui-Hui Wang, Jia-Ming Liu, Mingyu You* and Guo-Zheng Li στη μελέτη με τίτλο "Audio Signals Encoding for Cough Classification Using Convolutional Neural Networks: A Comparative Study" εξερευνούν τη σημασία της ανίχνευσης του βήχα, η οποία έχει σημαντική κλινική αξία καθώς παρέχει αντικειμενικά κριτήρια για την αξιολόγηση και διάγνωση των αναπνευστικών νοσημάτων. Κινητοποιημένοι από τα μεγάλα επιτεύγματα των συνελκτικών νευρωνικών δικτύων (CNNs) τα τελευταία χρόνια, εξετάζουν 5 διαφορετικούς τρόπους κωδικοποίησης των ηχητικών σημάτων ως εικόνες και τα χειρίζονται ως είσοδο στα CNNs, ώστε η τεχνολογία επεξεργασίας εικόνας να μπορεί να εφαρμοστεί για την ανάλυση των ηχητικών σημάτων.

Προκειμένου να εξερευνήσουν την βέλτιστη μέθοδο κωδικοποίησης των ηχητικών σημάτων για την κατηγοριοποίηση του βήχα, πραγματοποίησαν συγκριτικές εξετάσεις σε ιατρικά δεδομένα που περιλαμβάνουν 70.000 ηχητικά τμήματα από 26 ασθενείς. Τα πειραματικά αποτελέσματα δείχνουν ότι το φάσμα RASTA-PLP είναι η καλύτερη μέθοδος κωδικοποίησης των ηχητικών σημάτων ως εικόνες για την κατηγοριοποίηση

βήχα, παρέχοντας μέση ακρίβεια 0,9965 μετά από 200 επαναλήψεις σε δοκιμαστικά δείγματα και F1-σκορ 0,9768 σε δείγματα που επαναδειγματοληπτήθηκαν από το σύνολο δοκιμής. Επομένως, η μέθοδος βασισμένη στην επεξεργασία εικόνας φαίνεται ότι είναι μια ελπιδοφόρα επιλογή για την επεξεργασία των ηχητικών σημάτων. Για να χρησιμοποιηθούν τα CNNs, τα δεδομένα εισόδου οργανώνονται ως ένα σύνολο χαρακτηριστικών. Στη συνέχεια, εφαρμόζονται τα συνελκτικά επίπεδα και τα επίπεδα συγκέντρωσης(pooling). Τα CNNs έχουν την ιδιότητα της ανεξαρτησίας από μετατοπίσεις για πολλές εργασίες επεξεργασίας εικόνας, κάτι που είναι ευνοϊκό για τη βελτίωση της απόδοσης στην κατηγοριοποίηση βήχα[50].

Στη μελέτη με τίτλο ‘Artificial intelligence enabled preliminary diagnosis for COVID-19 from voice cues and questionnaires’ αναπτύσσονται δυνητικές περιπτώσεις χρήσης ευφυούς ανάλυσης ομιλίας για τον εντοπισμό της Covid-19. Με την ανάλυση ηχητικών εγγραφών από ασθενείς με θετικά και αρνητικά αποτελέσματα για την COVID-19, κατασκευάστηκαν μοντέλα βάσει της ομιλίας και των συμπτωμάτων για την αυτόματη κατηγοριοποίηση της υγειονομικής κατάστασης των ασθενών, εάν είναι θετικοί ή όχι στην COVID-19. Για το σκοπό αυτό, δημιουργήθηκαν πολλά ακουστικά χαρακτηριστικά και χρησιμοποιήθηκαν διάφοροι αλγόριθμοι μηχανικής μάθησης. Τα πειράματα δείχνουν ότι προέκυψε μια μέση ακρίβεια 80% για τον εκτιμητή της νόσου Covid-19 θετικού ή αρνητικού, με βάση πολλές ηχητικές εγγραφές βήχα και φωνής /α/, και μια μέση ακρίβεια 83% για την διάγνωση COVID-19 με βάση έξι ερωτήσεις που αφορούν τα συμπτώματα.

Σε αυτήν τη μελέτη λαμβάνονται υπόψη τρία σετ ακουστικών χαρακτηριστικών. Πρώτον, το σύνολο χαρακτηριστικών Computational Paralinguistics Challenge (ComPARE) όπου εξάγεται με το εργαλείο openSMILE. Δεύτερον, ένα σύνολο ακουστικών χαρακτηριστικών που εξάγονται από δωρεάν διαθέσιμο script και βιβλιοθήκες με το ανοικτού κώδικα λογισμικό PRAAT και LIBROSA. Το τρίτο σετ ακουστικών χαρακτηριστικών είναι ένα διάνυσμα χαρακτηριστικών 1024, που εξάγεται χρησιμοποιώντας ένα βαθύ συνελκτικό νευρωνικό δίκτυο (D-CNN) που είχε εκπαιδευτεί προηγουμένως για αδύναμα ετικετοποιημένες ηχητικές εγγραφές από το σύνολο δεδομένων ESC-50.

Το σύνολο χαρακτηριστικών, το οποίο εξάγεται χρησιμοποιώντας το D-CNN για ηχητικά γεγονότα και σκηνές, επέδειξε καλύτερη απόδοση από τα άλλα σύνολα

χαρακτηριστικών. Σημειώστε ότι ο ταξινομητής RF απέδωσε καλύτερα από τον ταξινομητή SVM, επιτυγχάνοντας ακρίβεια 0,73, περιοχή κάτω από την καμπύλη (AUC) 0,69 και F1-σκορ 0,81. Η καλύτερη απόδοση επιτεύχθηκε εκεί όπου επιτεύχθηκε η υψηλότερη AUC[51].

Στην εργασία με τίτλο ‘An ensemble learning approach to digital coronavirus preliminary screening from cough sounds’ οι EmadA. Mohammed1 et al. αναπτύσσουν ένα αξιόπιστο ταξινομητή για ένα προ-ταξινομητικό μοντέλο COVID-19 από δεδομένα ήχου βήχα που συγκεντρώνονται από το κοινό. Οι ηχογραφήσεις βήχα που συγκεντρώθηκαν από το κοινό περιέχουν ένα μεταβλητό αριθμό βηχών, με ορισμένα αρχεία ήχου να είναι πιο ενδιαφέροντα από άλλα. Η ακριβής ανίχνευση της νόσου Covid-19 από τα δεδομένα ήχου απαιτεί την αντιμετώπιση δύο κύριων προκλήσεων: (i) το μεταβλητό αριθμό βήχα σε κάθε ηχογράφιση και (ii) το χαμηλό αριθμό θετικών περιπτώσεων COVID σε σύγκριση με τους υγιείς στα δεδομένα.

Η εργασία χρησιμοποιεί δύο open και crowdsourced συλλογές δεδομένων ηχογραφήσεων βήχα και τέμνει κάθε ηχογράφιση βήχα σε μη-επικαλυπτόμενα τμήματα βήχα. Η διαίρεση των ηχητικών αρχείων επιτρέπει την εμπλοτύιση των αρχικών δεδομένων χωρίς υπερδειγματοληψία, διαιρώντας τα αρχικά αρχεία ήχου βήχα σε μη-επικαλυπτόμενα τμήματα.

Κάθε τμήμα ήχου βήχα μετατρέπεται σε έξι αναπαραστάσεις εικόνας για περαιτέρω ανάλυση. Διεξάγονται εκτενείς πειράματα με μηχανική μάθηση(shallow), Συνελκτικά Νευρωνικά Δίκτυα (CNN) και προεκπαιδευμένα μοντέλα CNN. Τα αποτελέσματα των μοντέλων συγκρίνονται με άλλες πρόσφατες εργασίες που εφαρμόζουν τη μηχανική μάθηση σε δεδομένα ήχου βήχα για τον εντοπισμό της COVID-19. Η μέθοδος επέδειξε υψηλή απόδοση χρησιμοποιώντας ένα σύνολο μοντέλων στο σύνολο δεδομένων ελέγχου, με περιοχή κάτω από την καμπύλη διακριτικών λειτουργιών ακρίβεια = 0.77, ακρίβεια = 0.80, ανάκληση = 0.71, μέτρο F1 = 0.75 και Κάππα = 0.53. Τα αποτελέσματα δείχνουν μια βελτίωση στην ακρίβεια πρόβλεψης του μοντέλου προ-ταξινομητή COVID-19 σε σύγκριση με τα άλλα μοντέλα.[3].

Στην δημοσίευση με τίτλο ‘Pneumonia Diagnosis based on CNN-LSTM-BERT Model’ απο τους Yingxue Yu et al., εφαρμόστηκε ένα υπάρχον μοντέλο CNN-LSTM-BERT, που είχε χρησιμοποιηθεί για τη διάγνωση COVID-19, στη διάγνωση της πνευμονίας.

Σκοπός ήταν να διερευνηθεί η αξιοπιστία του μοντέλου όταν απουσιάζουν δεδομένα από διάφορες θωρακικές θέσεις και να προσομοιωθεί ένα πραγματικό σενάριο. Εκπαιδεύτηκε το μοντέλο σε δύο σύνολα δεδομένων από τη Γενεύη και το Porto Alegre. Το μοντέλο αποδείχθηκε αξιόπιστο και ανθεκτικό στην απουσία δεδομένων, και παρουσίασε υψηλή ακρίβεια στη διάγνωση της πνευμονίας.

Για την εξαγωγή χαρακτηριστικών, οι ακουστικές εγγραφές μετατράπηκαν στην κλίμακα Mel και χρησιμοποιήθηκαν 32 τιμές MFCC σε αυτή την ανάλυση. Το μοντέλο περιλαμβάνει δύο μέρη: τον κωδικοποιητή ήχου CNN-LSTM και το μοντέλο εκπαίδευσης BERT. Ο κωδικοποιητής ήχου CNN-LSTM αποτελείται από δύο συνελκτικά επίπεδα CNN που εκτελούν μια διδιάστατη συνέλιξη κατά μήκος του άξονα της συχνότητας, ακολουθούμενα από συνοπτικές προσκηνώσεις. Στη συνέχεια, χρησιμοποιείται μοντέλο LSTM για να αντιμετωπίσει τη μεταβλητότητα του χρόνου. Το μοντέλο BERT λαμβάνει την ενισχυμένη ενσωμάτωση ήχου από τον κωδικοποιητή ήχου, προσθέτει την θέση των δεδομένων και εκπαιδεύεται για τη διάγνωση της πνευμονίας. Εν τέλει, το μοντέλο αποδείχθηκε αποτελεσματικό στη διάγνωση της πνευμονίας ακόμα και με απουσία δεδομένων[52].

Η έρευνα 'Analysis of COVID-19 Using Imaging and Audio Modalities' από τους Omar Alaaeldin' et al. στοχεύει να συμβάλει στην κατανόηση και ανάλυση των προτύπων της Covid-19, με σκοπό να συμβάλει στην ιατρική διάγνωση της νόσου και τη δημιουργία ενός προσιτού και γρήγορου εργαλείου στον τομέα των ελέγχων για τον COVID-19 χρησιμοποιώντας τεχνικές βαθιάς μάθησης. Η μελέτη εξετάζει οικονομικές μεθόδους διάγνωσης, όπως οι ακτινογραφίες και οι υπολογιστικές τομογραφίες των πνευμόνων, καθώς και ηχητικές εγγραφές βήχα, προκειμένου να επιτύχει ελπιδοφόρα αποτελέσματα. Δημιουργήθηκαν δύο μοντέλα Συνελκτικών Νευρικών Δικτύων (CNN) για αυτόν τον σκοπό.

Το πρώτο μοντέλο εκπαιδεύτηκε σε ένα σύνολο δεδομένων από 38.000 ακτινογραφίες και τομογραφίες των πνευμόνων για να διακρίνει αν η εικόνα απεικονίζει έναν υγιή άνθρωπο, έναν ασθενή με COVID-19 ή έναν ασθενή με πνευμονία. Αυτό το μοντέλο πέτυχε ένα εντυπωσιακό ποσοστό ακρίβειας 95,9%.

Το δεύτερο μοντέλο CNN εκπαιδεύτηκε σε περίπου 2.000 ηχητικές εγγραφές βήχα που μετατράπηκαν σε Μελ-φάσματα και χρησιμοποιήθηκαν για να αναγνωρίσουν αν τα

Μελ-φάσματα των ηχητικών εγγραφών ανήκουν σε ασθενείς με COVID-19 ή όχι. Αυτό το μοντέλο πέτυχε ένα ποσοστό ακρίβειας 82,1%[53].

Τέλος, σε μια μετάνανάλυση με τίτλο ‘Artificial intelligence in clinical care amidst COVID-19 pandemic: A systematic review’, οι Eleni S. Adamidi, Konstantinos Mitsis, Konstantina S. Nikita πραγματοποιούν συστημική βιβλιογραφική αναφορά σε μελέτες μηχανικής μάθησης για την πρόγνωση ή διάγνωση της COVID-19. Η μελέτη αναγνώρισε 14 μοντέλα τεχνητής νοημοσύνης (AI) που αναπτύχθηκαν για τη διαλογή COVID-19, 38 μοντέλα που στοχεύουν στη διάγνωση του COVID-19 και 50 προγνωστικών μοντέλων που προβλέπουν κρίσιμα αποτελέσματα, όπως η ανάγκη για εισαγωγή στη ΜΕΘ, η χρήση αναπνευστήρα, ο κίνδυνος θνησιμότητας, η σοβαρότητα της νόσου και η διάρκεια παραμονής στο νοσοκομείο.

Ακόμη συμπεράνθηκε ότι 43 μελέτες χρησιμοποίησαν δεδομένα απεικόνισης (π.χ. ακτινογραφίες, CT) για την ανάπτυξη μοντέλων AI και 58 μελέτες χρησιμοποίησαν κλινικά δεδομένα, συμπεριλαμβανομένων αποτελεσμάτων εργαστηρίων και δημογραφικών πληροφοριών.

Επιπλέον, πραγματοποιήθηκε ανάλυση RoB για την αξιολόγηση της εφαρμοσιμότητας και της αξιοπιστίας των περιλαμβανόμενων μελετών σε κλινικά περιβάλλοντα. Αυτό βοηθά στην αξιολόγηση των πιθανών μεροληψιών και της συνολικής αξιοπιστίας των μοντέλων AI.

Οι πιο συχνά χρησιμοποιούμενοι αλγόριθμοι AI για σκοπούς ταξινόμησης σε όλες τις μελέτες ήταν οι Random Forests (RF), Linear Regression (LR), Support Vector Machines (SVM), Convolutional Neural Networks (CNN), Decision Trees (DT) και XGBoost (XGB), οι οποίοι είναι γενικά έξι ευρέως χρησιμοποιούμενες μέθοδοι ταξινόμησης.

Επομένως, η συστηματική ανασκόπηση αναδεικνύει τον σημαντικό ρόλο της τεχνητής νοημοσύνης στη βελτίωση της διαλογής, της διάγνωσης και της πρόγνωσης του COVID-19. Ωστόσο, η μελέτη τονίζει επίσης την ανάγκη για προσεκτική επικύρωση και αξιολόγηση κινδύνων για να διασφαλιστεί η αξιοπιστία και η αποτελεσματικότητα των μοντέλων σε πραγματικά κλινικά περιβάλλοντα.[54]

B. Πρακτικό Μέρος

Κεφάλαιο 7

Τα δεδομένα

Τα δεδομένα που χρησιμοποιούνται χωρίζονται σε κλινικά και σε μη κλινικά δεδομένα. Τα κλινικά ηχογραφήθηκαν και καταχωρήθηκαν σε νοσοκομεία με την βοήθεια επαγγελματιών υγείας. Τα μη κλινικά δεδομένα συλλέχθηκαν μέσω μιας διαδικτυακής εφαρμογής (crowdsourced) που περιλάμβανε ερωτηματολόγιο χρήστη και έδινε την δυνατότητα ηχογράφησης των ζητούμενων, κατόπιν οδηγιών τα οποία αποτελούν το πληθοποριστικό σύνολο δεδομένων SMARTY4COVID. Το έργο smarty4covid φιλοδοξεί τη δημιουργία ενός ευφυούς πολυτροπικού πλαισίου για την αξιολόγηση του κινδύνου και την παρακολούθηση του COVID-19, βασισμένο σε Επεξηγηματική Βαθιά Μάθηση (Explainable Deep Learning). Μετά από τις απαραίτητες εγκρίσεις από την Επιτροπή Ηθικής της Έρευνας του Εθνικού Μετσόβιου Πολυτεχνείου, υλοποιήθηκε και δημοσιοποιήθηκε μια ευέλικτη διαδικτυακή εφαρμογή (www.smarty4covid.org) ως μέσο συλλογής δεδομένων. Το σύνολο δεδομένων smarty4covid περιέχει συνολικά 18,265 ηχητικές καταγραφές βήχα, αναπνοής (κανονικής, βαθιάς) και φωνής, που αντιστοιχούν σε 4,673 χρήστες (Έλληνες και Κύπριους πολίτες).[55] Στα δεδομένα περιλαμβάνονται οι ηχογραφήσεις αλλά και ένα json αρχείο που υπάρχει σαν καρτέλα ασθενή και περιλαμβάνει χρήσιμα στοιχεία για τον ασθενή (όπως ηλικία, φύλο, ύψος, βάρος), την κατάστασή του, την υγεία του, την ύπαρξη τυχόν υποκείμενων νοσημάτων, την κλινική του εικόνα, τα συμπτώματα που εμφάνισε, την αναπνευστική του συχνότητα, αν καπνίζει, αν έχει εμβολιαστεί, αν έχει εισαχθεί στο νοσοκομείο, την ημερομηνία εισαγωγής κ.α. Το σύνολο δεδομένων smarty4covid δημοσιεύτηκε σε μορφή βάσης γνώσεων διαδικτυακής οντολογίας γλώσσας (OWL), επιτρέποντας τη συγχώνευση δεδομένων από άλλα σχετικά σύνολα δεδομένων, την εκτέλεση πολύπλοκων ερωτημάτων και το συλλογιστικό σκέλος. Μέρος του συνόλου δεδομένων που συλλέχθηκε από το κοινό ήταν μη έγκυρο λόγω εσφαλμένων καταχωρημένων ηχογραφήσεων από τους χρήστες και την παρουσία παραμορφώσεων και υψηλού θορύβου στα δεδομένα. Η διαδικασία καθαρισμού των δεδομένων πραγματοποιήθηκε μέσω μιας καμπάνιας συλλογής δεδομένων με τη χρήση του εργαλείου ανοιχτού κώδικα Label Studio[55] Έπειτα από την συλλογή των

δεδομένων ακολούθησε η χειρωνακτική τοποθέτηση ταμπελών όπου αφορούσε την ποιότητα, την ύπαρξη θορύβου αλλά και την ερμηνεία των ηχητικών δεδομένων. Τα κλινικά δεδομένα περιλαμβάνουν 112 ηχητικά δεδομένα αναπνοών από τα οποία χρησιμοποιήθηκε ένα μέρος για τον εντοπισμό της αναπνευστικής συχνότητας. Η επιλογή των ηχογραφήσεων που κρίθηκαν ως καλά δείγματα έγινε με βάση την ποιότητα του ήχου και την ακρίβεια του περιεχομένου.

Κεφάλαιο 8

Μεθοδολογία

8.1 Προεπεξεργασία

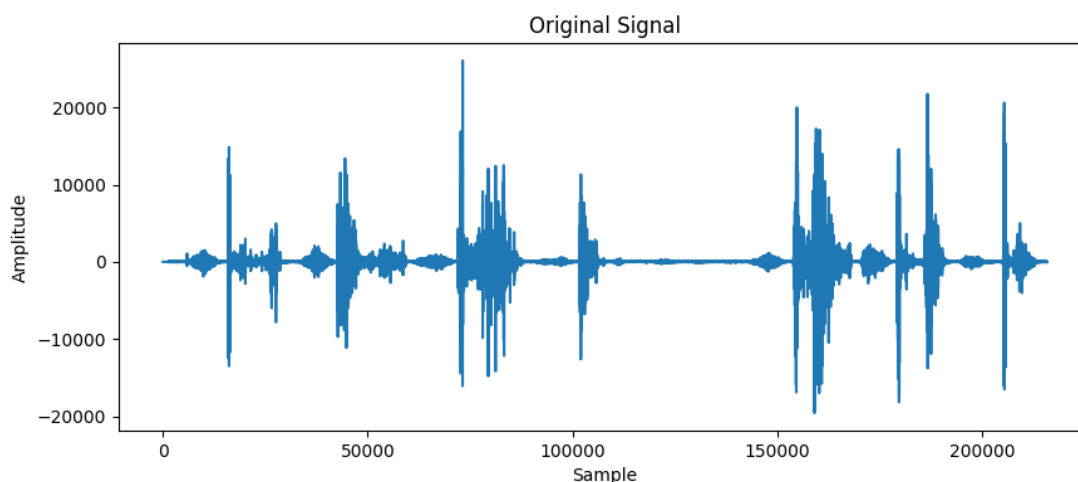
Η επεξεργασία των ηχητικών δεδομένων αναπνοής άρχισε βρίσκοντας τον αριθμό των αναπνοών που έχουν ηχογραφηθεί σε κάθε ηχογραφημένο αρχείο αναπνοής. Αρχικά η μέθοδος αυτή βασίστηκε στην εύρεση των κορυφών δηλαδή των μεγίστων στα γραφήματα του ήχου. Η διαδικασία που ακολουθήθηκε εν συντομία είναι η παρακάτω και παραδείγματα των σημάτων μετά από κάθε βήμα συναντώνται στις εικόνες 5-9 :

- Χρήση ζωνοπερατού φίλτρου για την απομάκρυνση θορύβου και λογαριθμική συμπίεση (Noise Removal + Log Compression).
- Υπολογισμός της έντασης του περιγράμματος από το φασματόγραμμα (Computation of Intensity envelope from spectrogram).
- Ευρεση των κορυφών πάνω στο Gaussian-smoothed περίγραμμα της έντασης των κορυφών (peak intensity envelope) (Left/Right Base of peaks also recorded).

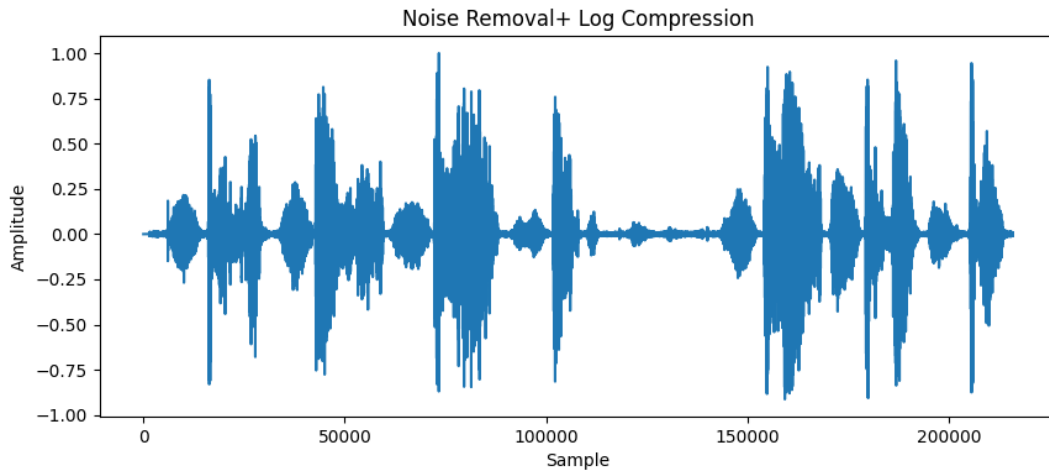
Όλα τα αρχεία ήχου επαναδειγματίζονται στα 7000Hz. Η προκύπτουσα συχνότητα Nyquist καλύπτει το εύρος των πνευμονικών ήχων (25-1500Hz) καθώς η μέγιστη τιμή των συχνοτήτων που περιλαμβάνουν οι πνευμονικοί ήχοι είναι 1500 και πολλαπλασιασμένο επί 2 είναι 3000 Hz <7000 ώστε να είναι δυνατή η ανακατασκευή του σήματος. Στη συνέχεια, η λογαριθμική συμπίεση χρησιμοποιείται για τη μείωση των διαφορών στο δυναμικό εύρος στις ηχητικές καταγραφές. Πιο συγκεκριμένα εφαρμόστηκε λογαριθμική συμπίεση καθώς στην επεξεργασία ηχητικού σήματος, το πρόβλημα με αναπαραστάσεις σε ένα φασματόγραμμα(όπου και θα χρησιμοποιηθεί στη συνέχεια) είναι ότι οι τιμές του διαθέτουν μεγάλο δυναμικό εύρος. Αυτό έχει σαν αποτέλεσμα πολλές φορές, οι μικρές, αλλά συναφείς τιμές να κυριαρχούνται ή να συγκαλύπτονται από μεγάλες τιμές. Επομένως, συνηθίζεται να χρησιμοποιείται συχνά μια κλίμακα ντεσιμπέλ, όπου η ιδέα είναι να εξισορροπηθεί αυτή η απόκλιση μειώνοντας τη διαφορά μεταξύ μεγάλων και μικρών τιμών έχοντας σαν αποτέλεσμα την ενίσχυση των μικρών τιμών. Ακόμη, για να αντιμετωπιστεί αυτό το πρόβλημα εφαρμόζονται συχνά άλλοι τύποι συναρτήσεων που βασίζονται σε λογάριθμους,

δηλαδή η εφαρμογή λογαριθμικής συμπίεσης όπου εφαρμόστηκε και στην συγκεκριμένη περίπτωση. Ο βαθμός της συμπίεσης επηρεάζεται από τη σταθερά γάμμα. Όσο υψηλότερη είναι η τιμή γάμμα τόσο πιο επιθετική και απότομη είναι η συμπίεση. Στη συνέχεια κανονικοποιούμε τα σήματα.

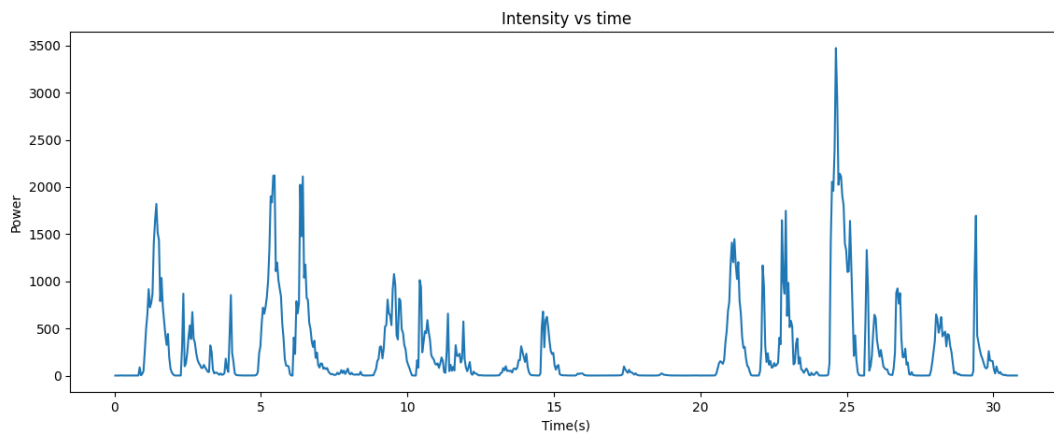
Τα φασματογράμματα παράγονται χρησιμοποιώντας ένα μέγεθος παραθύρου FFT 50ms. Οι κάδοι που προκύπτουν στο φασματόγραμμα πολλαπλασιάζονται με το τετράγωνο της συχνότητάς τους, καθώς η ισχύς ενός ηχητικού κύματος είναι ανάλογη του αρθροίσματος των τετραγώνων των τιμών αλλά και των συχνοτήτων από τις οποίες συντίθεται όπως προκύπτει από τον τύπο της ισχύος για ένα κύμα $I = 2\pi^2\rho f^2v\Delta$. Το φασματόγραμμα αποτελεί μια εικόνα στην οποία ο ένας άξονας περιλαμβάνει τον χρόνο, ο άλλος άξονας τις συχνότητες και για κάθε σημείο της εικόνας δηλαδή για κάθε ζεύγος χρόνου – συχνότητας υπάρχει μία τιμή έντασης που αναπαρίσταται με ένα χρώμα ανάλογα την τιμή που έχει.



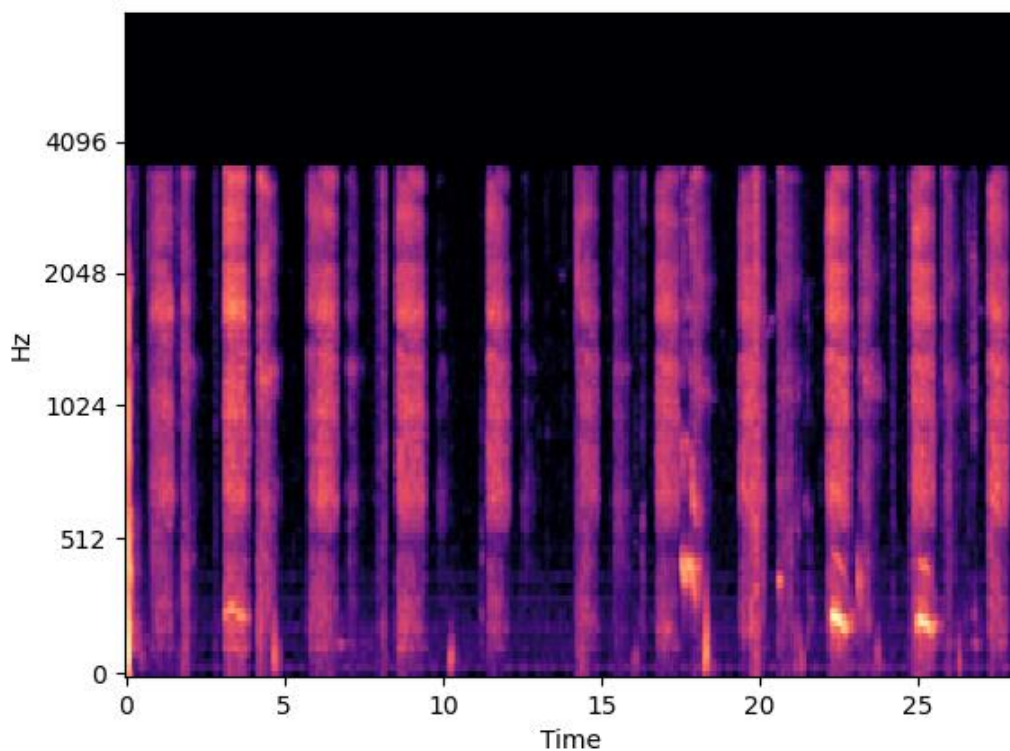
Εικόνα 5 Αρχικό Γράφημα Ήχου



Εικόνα 6 Γράφημα ήχου μετά την αποθροβοποίηση και την λογαριθμική συμπίεση

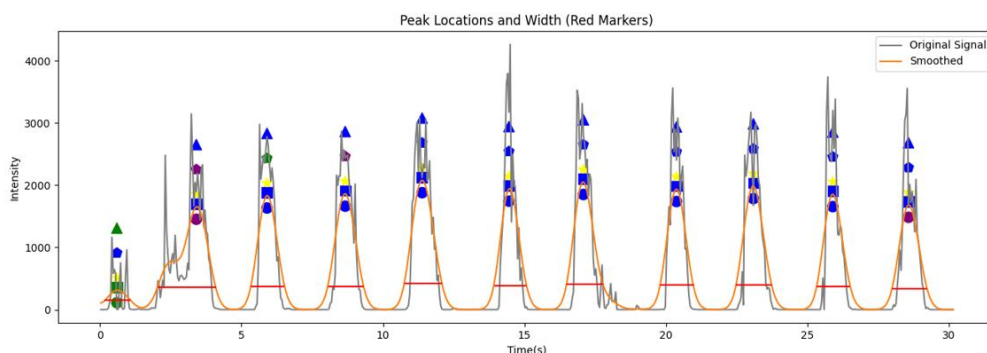


Εικόνα 7 Γράφημα με περίγραμμα Έντασης



Εικόνα 8 Φασματογράμμα

Στη συνέχεια, χρησιμοποιείται ένα Γκαουσιανό φίλτρο για την εξομάλυνση της καμπύλης πάνω στην οποία εκτελείται, η ανίχνευση κορυφών με την χρήση τη συνάρτησης `scipy.signal.find_peaks`. Η συνάρτηση αυτή διευκόλυνε όχι μόνο τον εντοπισμό των κορυφών δηλαδή των εισπνοών και των εκπνοών αλλά και τον υπολογισμό της χρονικής διάρκειας κάθε τέτοιου συμβάντος, πληροφορία απαραίτητη για να υπολογιστεί η εκτιμώμενη αναπνευστική συχνότητα. Στην από κάτω εικόνα παρουσιάζεται παράδειγμα εντοπισμού κορυφών:



Εικόνα 9 Γράφημα που παρουσιάζει την εύρεση κορυφών και τον υπολογισμό των πλατών τους

Στο συγκεκριμένο σημείο για να αξιολογηθεί η ακρίβεια της μεθόδου υπολογίστηκε το μέσο τετραγωνικό σφάλμα δηλαδή το mse το οποίο μετρά τον μέσο όρο των τετραγώνων του σφάλματος, δηλαδή τη μέση τετραγωνική διαφορά μεταξύ των εκτιμώμενων τιμών(αυτών που υπολογίσαμε) και των πραγματικών τιμών. Εν προκειμένω, οι πραγματικές τιμές προέρχονται από τη μέτρηση του αναπνευστικού ρυθμού των υποκειμένων μετά την ηχογράφιση της αναπνευστικής τους δραστηριότητας. Επομένως είναι πιθανό να υπάρχει εγγενής απόκλιση και στις δύο μετρήσεις. Το MSE είναι ένα μέτρο της ποιότητας ενός εκτιμητή. Καθώς προέρχεται από το τετράγωνο της διαφοράς, είναι πάντα μια θετική τιμή που μειώνεται καθώς το σφάλμα πλησιάζει το μηδέν. Η τιμή MSE που βρέθηκε δεν ήταν αμελητέα με αποτέλεσμα να χρειαστεί περαιτέρω διερεύνηση και επεξεργασία.

Εξετάστηκαν λοιπόν οι περιπτώσεις που κυμαίνονται σε τιμές μικρότερες από -5 η μεγαλύτερες από +5 κορυφές. Υπήρχαν συνολικά 27 τέτοιες περιπτώσεις για τις οποίες αναλύθηκαν τα γραφήματά τους, ακούστηκαν οι αντίστοιχες ηχογραφήσεις, πραγματοποιήθηκε η σύνδεση μεταξύ ήχων και κορυφών και καταγράφηκε η ποιότητα και οι πιθανοί λόγοι για τους οποίους η μέθοδος δεν έφερε τα επιθυμητά αποτελέσματα για τις περιπτώσεις αυτές. Από τις περιπτώσεις αυτές αφού εξερευνήθηκαν και κατηγοριοποιήθηκαν, εντοπίστηκαν εκείνες στις οποίες μπορούσε να υπάρξει σημαντική βελτίωση μέσω περαιτέρω επεξεργασίας. Η διαδικασία αυτή συμπεριλάμβανε συσταδοποίηση με στόχο να κατηγοριοποιηθούν οι εντοπισμένες κορυφές σε εισπνοές, εκπνοές και θόρυβος εφόσον ένα συχνό πρόβλημα στις εγγραφές ήταν η μη καταγραφή και επομένως εντοπισμός των εισπνοών.

Αυτό επιχειρείται στη συνέχεια με τη βοήθεια της μη επιβλεπόμενης μηχανικής μάθησης (unsupervised learning) και πιο συγκεκριμένα της συσταδοποίησης.

8.2 Συσταδοποίηση

Για την συσταδοποίηση τα χαρακτηριστικά που επιλέχθηκαν αρχικά είναι τα βασικά ποσοτικά χαρακτηριστικά που ορίζουν μια κορυφή ,δηλαδή το ύψος, το πλάτος και η προβολή (prominence). Η προβολή μιας κορυφής μετρά πόσο ξεχωρίζει η κορυφή λόγω του εγγενούς ύψους της και της θέσης της σε σχέση με άλλες κορυφές. Μια χαμηλή απομονωμένη κορυφή μπορεί να είναι πιο εμφανής από μια που είναι υψηλότερη, αλλά είναι σε γενικότερο πλαίσιο ένα μη αξιοσημείωτο μέλος μιας υψηλής περιοχής.

Έτσι λοιπόν επιλέχθηκαν και εφαρμόστηκαν 5 αλγόριθμοι συσταδοποίησης οι οποίοι είναι k-means, DBSCAN, Spectral Clustering και Hierarchical Clustering, Meanshift.

8.2.1 K-means Συσταδοποίηση

Η ομαδοποίηση **k-means** είναι μια μέθοδος ομαδοποίησης διανυσμάτων που στοχεύει να χωρίσει n παρατηρήσεις σε k συστάδες τα οποία κάθε παρατήρηση ανήκει στο σύμπλεγμα με τον πλησιέστερο μέσο όρο (κέντρα συστάδων ή κέντρο συστάδων), που χρησιμεύει ως πρωτότυπο το σύμπλεγμα. Αυτό έχει ως αποτέλεσμα την κατάτμηση του χώρου δεδομένων σε κελιά Voronoi. Η ομαδοποίηση k-means ελαχιστοποιεί τις διακυμάνσεις εντός τής κλάσης (τετράγωνες Ευκλείδειες αποστάσεις), αλλά όχι τις κανονικές Ευκλείδειες αποστάσεις, που θα ήταν το πιο δύσκολο πρόβλημα: ο μέσος όρος βελτιστοποιεί τα τετραγωνικά σφάλματα, ενώ μόνο η γεωμετρική διάμεσος ελαχιστοποιεί τις Ευκλείδειες αποστάσεις. Το πρόβλημα είναι υπολογιστικά δύσκολο (NP-hard). Ωστόσο, αποτελεσματικοί ευρετικοί αλγόριθμοι συγκλίνουν γρήγορα σε ένα τοπικό βέλτιστο.

8.2.2 DBSCAN Συσταδοποίηση

Η χωρική ομαδοποίηση εφαρμογών με θόρυβο βάσει πυκνότητας (**DBSCAN**) είναι ένας αλγόριθμος ομαδοποίησης δεδομένων που προτάθηκε από τους Martin Ester, Hans-Peter Kriegel, Jörg Sander και Xiaowei Xu το 1996. Είναι ένας μη παραμετρικός αλγόριθμος ομαδοποίησης που βασίζεται στην πυκνότητα σημείων σε κάποιο χώρο, δηλαδή ομαδοποιεί σημεία που έχουν μεγάλη πυκνότητα μεταξύ τους (σημεία με πολλούς κοντινούς γείτονες), επισημαίνοντας τα ως ακραία σημεία που βρίσκονται μόνα τους σε περιοχές χαμηλής πυκνότητας (των οποίων οι κοντινότεροι γείτονες είναι πολύ μακριά). Ο DBSCAN είναι ένας από τους πιο συνηθισμένους αλγορίθμους ομαδοποίησης και επίσης χρησιμοποιείται περισσότερο στην επιστημονική βιβλιογραφία. Ο DBSCAN χρειάζεται δύο παραμέτρους για να λειτουργήσει : 1. Την παράμετρο ϵ (eps) και 2. τον ελάχιστο αριθμό σημείων που απαιτούνται για να σχηματιστεί μια πυκνή περιοχή[a] (minPts). Ο αλγόριθμος αρχίζει με μια αυθαίρετη αφετηρία την οποία δεν έχει επισκεφθεί. Η ε-γειτονιά αυτού του σημείου ανακτάται και εάν περιέχει αρκετά πολλά σημεία, ξεκινά ένα σύμπλεγμα. Διαφορετικά, το σημείο χαρακτηρίζεται ως θόρυβος. Μια σημαντική λεπτομέρεια είναι ότι αυτό το σημείο μπορεί αργότερα να βρεθεί σε ένα επαρκούς μεγέθους ε-περιβάλλον ενός διαφορετικού σημείου με αποτέλεσμα να γίνει μέρος ενός συμπλέγματος. Ως εκ τούτου, όλα τα

σημεία που βρίσκονται εντός της ε-γειτονιάς προστίθενται, αλλά προστίθεται επίσης και η δική τους ε-γειτονιά όταν είναι πυκνή. Αυτή η διαδικασία συνεχίζεται μέχρι να βρεθεί πλήρως το σύμπλεγμα που συνδέεται με την πυκνότητα. Στη συνέχεια, ένα νέο μη επισκέψιμο σημείο επιλέγεται και υποβάλλεται σε επεξεργασία, οδηγώντας στην ανακάλυψη ενός περαιτέρω συμπλέγματος ή θορύβου. Το DBSCAN μπορεί να χρησιμοποιηθεί με οποιαδήποτε συνάρτηση απόστασης (καθώς και συναρτήσεις ομοιότητας ή άλλα κατηγορήματα). Η συνάρτηση απόστασης (dist) μπορεί επομένως να θεωρηθεί μια επιπλέον παράμετρος.

8.2.3 Φασματική Συσταδοποίηση

Οι τεχνικές φασματικής ομαδοποίησης (Spectral Συσταδοποίηση) χρησιμοποιούν το φάσμα (ιδιοτιμές) του πίνακα ομοιότητας των δεδομένων για να πραγματοποιήσουν μείωση διαστάσεων πριν από την ομαδοποίηση σε λιγότερες διαστάσεις. Ο πίνακας ομοιότητας δίνεται ως είσοδος και αποτελείται από μια ποσοτική αξιολόγηση της σχετικής ομοιότητας κάθε ζεύγους σημείων στο σύνολο δεδομένων. Ο βασικός Αλγόριθμος αποτελείται από τα εξής βήματα: Υπολογισμός του Λαπλασιανού L (ή το κανονικοποιημένο Λαπλασιανό), υπολογισμός των πρώτων k ιδιοδιανυσμάτων (τα ιδιοδιανύσματα που αντιστοιχούν στις k μικρότερες ιδιοτιμές του L), ορισμός του πίνακα που σχηματίζεται από τα πρώτα k ιδιοδιανύσματα, ομαδοποίηση των κόμβων του γραφήματος με βάση αυτά τα χαρακτηριστικά.

8.2.4 Ιεραρχική συσταδοποίηση

Η ιεραρχική ομαδοποίηση (ονομάζεται αλλιώς και ιεραρχική ανάλυση συστάδων ή HCA ή hierarchical συσταδοποίηση) είναι μια μέθοδος ανάλυσης συστάδων που επιδιώκει να δημιουργήσει μια ιεραρχία συστάδων. Οι στρατηγικές για την ιεραρχική ομαδοποίηση εμπίπτουν στις εξής δύο κατηγορίες:

Συγκεντρωτική (Agglomerative): Αυτή η προσέγγιση εκτελείται "από κάτω προς τα πάνω" δηλαδή κάθε παρατήρηση ξεκινά στο δικό της σύμπλεγμα και στη συνέχεια τα ζεύγη συστάδων συγχωνεύονται καθώς κάποιος κινείται προς τα πάνω στην ιεραρχία.

Διάσπαση (Divisive): Αυτή η προσέγγιση πραγματοποιείται "από πάνω προς τα κάτω" δηλαδή όλες οι παρατηρήσεις ξεκινούν σε ένα σύμπλεγμα και εκτελούνται αναδρομικά διαχωρισμοί καθώς κάποιος κινείται προς τα κάτω στην ιεραρχία.

Γενικά, οι συγχωνεύσεις και οι διασπάσεις καθορίζονται με άπληστο τρόπο. Τα αποτελέσματα της ιεραρχικής ομαδοποίησης παρουσιάζονται συνήθως σε ένα δενδρόγραμμα. Ο κλασικός αλγόριθμος για την ιεραρχική συσσωρευτική ομαδοποίηση (HAC) έχει χρονική πολυπλοκότητα $O(n^3)$ και απαιτεί μνήμη $\Omega(n^2)$, γεγονός που τον καθιστά πολύ αργό ακόμη και για μεσαία σύνολα δεδομένων. Στην περίπτωση μας χρησιμοποιήθηκε η Agglomerative προσέγγιση δηλαδή από κάτω προς τα πάνω.

Οι αλγόριθμοι αυτοί χρησιμοποιήθηκαν για να κατατάξουν σε κάθε διάγραμμα, που προέκυψε από κάθε ηχητικό αναπνοής, τις κορυφές που ανιχνεύθηκαν από τον προηγούμενο τρόπο σε ένα cluster(μια συστάδα) με σκοπό να κατηγοριοποιηθούν οι εισπνοές στην ίδια κατηγορία, όπως και οι εκπνοές μεταξύ τους και παράλληλα να χαρακτηριστούν ορισμένες κορυφές που δεν ανήκουν σε ήχο αναπνοής ως θόρυβος.

Για να εξακριβωθεί όμως η καταλληλότητα των χαρακτηριστικών που χρησιμοποιήθηκαν για τα συγκεκριμένα δεδομένα και τους συγκεκριμένους αλγόριθμους χρησιμοποιήθηκε μια μετρική προσδιορισμού καταλληλότητας για το σύνολο των αναπνευστικών δεδομένων μας. Επομένως χρησιμοποιήθηκε η μετρική τιμή Silhouette για την αξιολόγηση της ποιότητας των συμπλεγμάτων που δημιουργήθηκαν χρησιμοποιώντας τους παραπάνω αλγορίθμους ομαδοποίησης αξιολογώντας το πόσο καλά τα δείγματα συγκεντρώνονται με άλλα δείγματα που είναι παρόμοια μεταξύ τους. Η μετρική τιμή Silhouette υπολογίζεται για κάθε δείγμα διαφορετικών ομάδων.

Μετά τον υπολογισμό της διάρκειας κάθε κορυφής(είτε εισπνοής είτε εκπνοής) και τον διαχωρισμό των κορυφών σε εισπνοές, εκπνοές και outliers θα γίνει εφικτός ο υπολογισμός του μέσου χρόνου εισπνοής, εκπνοής και συνολικής αναπνοής για κάθε δείγμα. Τα μεγέθη αυτά ενέχουν σημαντικές πληροφορίες που συμβάλλουν στην σωστή κατηγοριοποίηση και διάγνωση.

Ύστερα από αυτήν την επεξεργασία και ανάλυση βρίσκονται στη διάθεσή μας μέγεθη και μετρικές που μπορούν να αξιοποιηθούν ώστε να δημιουργηθεί ένα χρήσιμο εργαλείο διάγνωσης που θα τροφοδοτείται από αυτά και θα αποφασίζει αν ο ασθενής πάσχει από Covid-19 ή όχι. Γι'αυτό τον λόγο αναπτύχθηκε ένα σύνθετο μοντέλο διάγνωσης που αποτελείται από δύο απλά μοντέλα: 1. Ένα CNN νευρωνικό, συνελκτικό δίκτυο που τροφοδοτείται από φασματογράμματα και ένα XGBOOST που

τροφοδοτείται από τις μέσες διάρκειες των εισπνοών, εκπνοών και συνολικών αναπνοών.

8.3 Συνελικτικό Νευρωνικό Δίκτυο

8.3.1 Η Εκπαίδευση Συνελικτικού Νευρωνικού Δικτύου τροφοδοτούμενο με φασματογράμματα

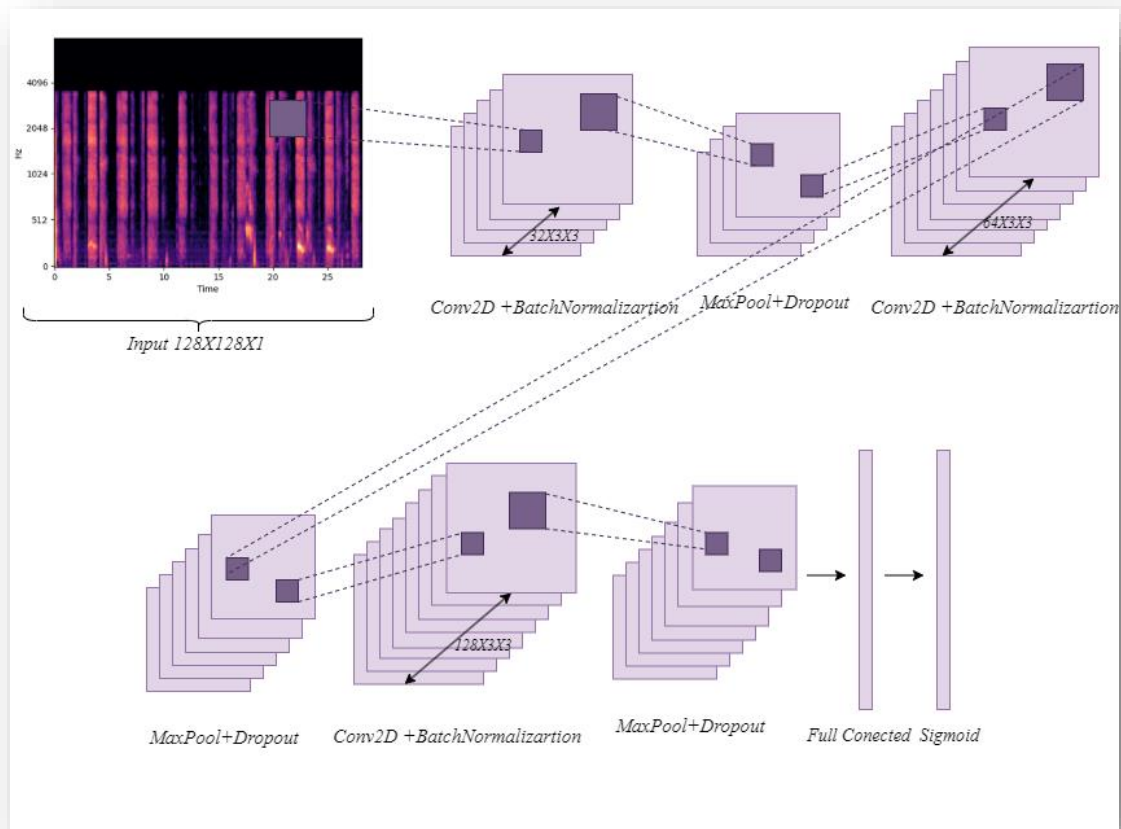
Στη συνέχεια με βάση σχετική βιβλιογραφία χτίσαμε αρχιτεκτονικές νευρωνικών συνελικτικών δικτύων που εκπαιδεύτηκαν με εικόνες φασματογραμμάτων, προερχόμενες από την επεξεργασία των ηχητικών ηχογραφήσεων της αναπνοής. Πιο συγκεκριμένα αφού φορτώθηκαν τα ηχογραφημένα δείγματα, με τη χρήση κώδικα και τη χρήση έτοιμων βιβλιοθηκών της Python όπως Librosa δημιουργήθηκαν τα φασματογράμματα. Στη συνέχεια ακολούθησε η επεξεργασία των δειγμάτων πριν την εκπαίδευση που περιλαμβάνει την κανονικοποίηση των φασματογραμμάτων και την επιλογή των μη μηδενικών ή κενών δειγμάτων ως προς την γνώση μας για τη νόσηση ή όχι από Covid-19.

Δημιουργήθηκε και εκπαιδεύτηκε πληθώρα μοντέλων με στόχο την επίτευξη της μέγιστης ακρίβειας και μείωσης της απώλειας. Σε κάθε εκπαίδευση παρατηρήθηκε η διακύμανση και η πρόοδος αυτών των μετρικών, τόσο στο σύνολο εκπαίδευσης αλλά και στο σύνολο αξιολόγησης. Στην εικόνα 7 παρατίθεται η αρχιτεκτονική του πρώτου μοντέλου και στην ενότητα αρχιτεκτονική μοντέλου αναλύεται η δομή του.

8.3.2 Αρχιτεκτονική Μοντέλου

Περιγράφεται η αρχιτεκτονική του αρχικού μοντέλου, όπως φαίνεται και στην εικόνα 7 η οποία περιλαμβάνει το επίπεδο εισόδου όπου καθορίζεται η μορφή των εισερχομένων δεδομένων, τρία επίπεδα συνελικτικών νευρωνικών επιπέδων max pooling επιπέδων και dropout επιπέδων, το Flatten επίπεδο, το πλήρως συνδεδεμένο επίπεδο και το επίπεδο της σιγμοειδούς συνάρτησης. Τα συνελικτικά επίπεδα περιλαμβάνουν το καθένα με τη σειρά 32, 64 και 128 φίλτρα. Στον πυρήνα (kernel) των Επιπέδων Συνέλιξης (Convolutional Layers) καθορίζονται δυναμικά σημαντικές συσχετίσεις από τα δεδομένα χρησιμοποιώντας φίλτρα. Ο αριθμός στην αρχή κάθε επιπέδου δηλώνει πόσα φίλτρα θα δημιουργηθούν. Το MaxPooling layer μειώνει την

έξοδο των προηγούμενων επιπέδων εκτελώντας έναν πυρήνα πάνω στα δεδομένα και επιλέγοντας τη μεγαλύτερη τιμή από ένα πεδίο 2×2 . Τα Dropout επίπεδα απενεργοποιούν τυχαία τον δεδομένο ποσοστό νευρώνων ανά εποχή. Με αυτήν τη μεθοδολογία, μπορεί να μειωθεί το φαινόμενο της υπερπροσαρμογής στα εκπαιδευτικά δεδομένα. Πριν το πλήρως συνδεδεμένο φίλτρο υπάρχει το Flatten επίπεδο όπου δημιουργεί ένα 1-διάστατον τανυστή/διάνυσμα από ένα n-διάστατο εισαγόμενο ώστε να μπορεί η πληροφορία να τροφοδοτηθεί στο επόμενο επίπεδο. Τέλος το μοντέλο περιλαμβάνει ένα Dense Layer με 256 νευρώνες και έναν πραγματικό ταξινομητή (ένα επίπεδο με 1 νευρώνα για κάθε κλάση). Μέσω της σιγμοειδούς συνάρτησης ενεργοποίησης, εκπέμπονται τιμές μεταξύ 0 και 1 που μπορούν να ερμηνευτούν ως η πιθανότητα για κάθε κλάση. Οι 4 παραλλάγες μοντέλων που υπάρχουν είναι η αρχική, η βαθειά που έχει μεγαλύτερο αριθμό επιπέδων από το αρχικό, η φαρδιά που έχει μεγαλύτερο αριθμό φίλτρων σε κάθε επίπεδο σε σχέση με το αρχικό, αυτή με τα περισσότερα φίλτρα που περιλαμβάνει επίπεδα με παρόμοιο αριθμό φίλτρων μεγαλύτερο σε σχέση με το αρχικό μοντέλο και η αντίστροφη αρχιτεκτονική που περιέχει επίπεδα με φθίνον αριθμό φίλτρων σε αντίθεση με το αρχικό που έχει αύξον.



Εικόνα 10 Δομή βασικού μοντέλου CNN

8.3.3 Εκπαίδευση του Μοντέλου

Πριν από την εκπαίδευση των μοντέλων, οι ετικέτες των κλάσεων κωδικοποιούνται χρησιμοποιώντας τη μέθοδο One-Hot-Encoding, έτσι ώστε να μπορούν να χρησιμοποιηθούν από το μοντέλο. Στη συνέχεια τα δεδομένα διαιρούνται σε δύο σύνολα, ένα σύνολο εκπαίδευσης και ένα σύνολο ελέγχου. Το σύνολο εκπαίδευσης χρησιμοποιείται για την πραγματική εκπαίδευση του μοντέλου, ενώ το σύνολο ελέγχου χρησιμοποιείται για να εκτιμηθεί η απόδοση του μοντέλου σε δεδομένα που δεν έχει ξαναδεί. Ο βελτιστοποιητής Adam χρησιμοποιείται για την ελαχιστοποίηση της συνάρτησης απώλειας `binary_crossentropy`. Κατά τη διάρκεια της εκπαίδευσης χρησιμοποιούνται επίσης `Learning Rate Scheduler` για προσαρμογή του ρυθμού μάθησης κατά τη διάρκεια της εκπαίδευσης, `Early Stopping` για διακοπή νωρίς αν η απώλεια δεν βελτιώνεται για μια συγκεκριμένη περίοδο και `Reduce LR on Plateau` δηλαδή μείωση του ρυθμού μάθησης αν η απώλεια στο σύνολο ελέγχου δεν βελτιώνεται. Το μοντέλο δηλαδή εκπαιδεύεται στα δεδομένα εκπαίδευσης με διαχείριση του ρυθμού μάθησης και διακόπτεται νωρίς αν δεν υπάρχει βελτίωση στο

σύνολο ελέγχου. Το μοντέλο αξιολογείται στο σύνολο ελέγχου για να εκτιμηθεί η ακρίβεια του σε δεδομένα που δεν έχει δει κατά τη διάρκεια της εκπαίδευσης

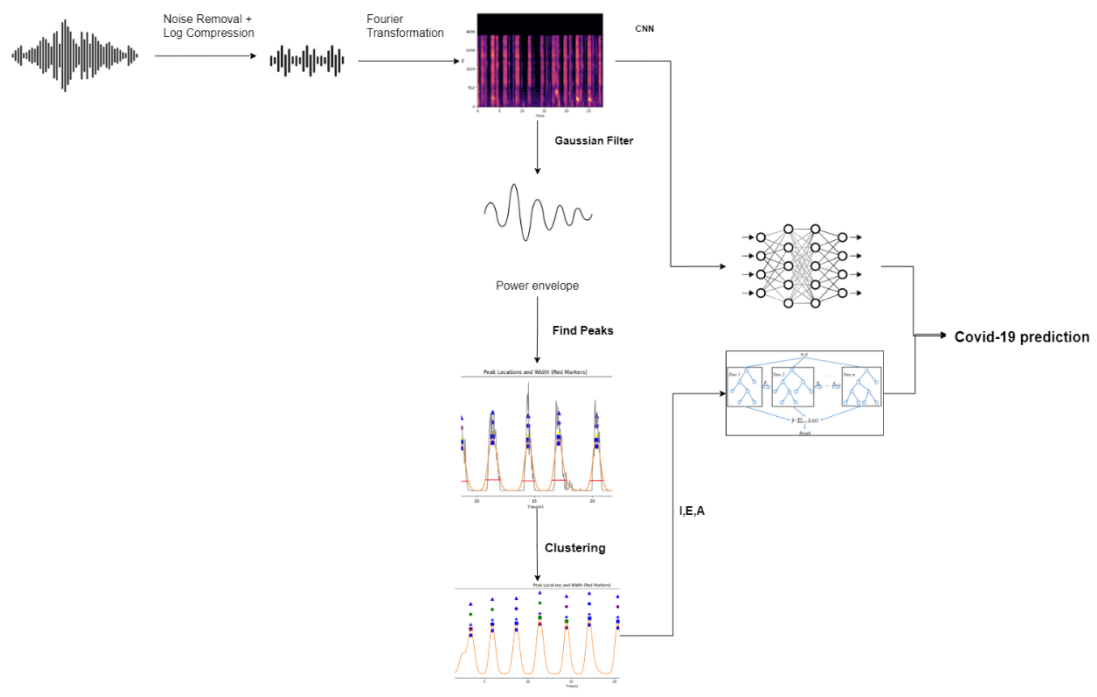
8.4 Εκπαίδευση του XGBoost Μοντέλου

Ωστόσο για μεγαλύτερη ακρίβεια χρησιμοποιήθηκε και μία άλλη μέθοδος, το μοντέλο XGBoost το οποίο τροφοδοτήθηκε με αριθμητικά δεδομένα, συγκεκριμένα με την διάρκεια εισπνοής(I), την διάρκεια εκπνοής(E) και τη συνολική διάρκεια αναπνοής(A) σε διαφορετικούς συνδυασμούς όπως I,E,A ή I,E, I/E. Οι μετρικές αυτές έχουν σημασία για την διάγνωση καθώς έχει επισημανθεί και πιο πάνω ότι η φυσιολογική αναπνοή συνήθως παρουσιάζει λόγο I:E 1:2 ή 1:3 σε κατάσταση ανάπαυσης και ενδεχομένως η παρουσία αποφρακτικών προβλημάτων μπορεί να οδηγήσει σε επιμήκυνση της εκπνοής ή της εισπνοής, οδηγώντας σε μη αναμενόμενο λόγο I:E. Γι' αυτόν τον λόγο χρησιμοποιήθηκαν αυτές οι μετρικές για την εκπαίδευση του XGBoost. Αρχικά ορίζονται τα σύνολα εκπαίδευσης και ελέγχου και στη συνέχεια υπολογίζεται το `scale_pos_weight` για την αντιμετώπιση της ανισορροπίας κλάσεων: $scale_pos_weight = 0.4 * \sum(y_train == 0) / \sum(y_train == 1)$ και δημιουργείται ένα XGBoost μοντέλο με τις κατάλληλες υπερπαραμέτρους:

```
scale_pos_weight=scale_pos_weight, learning_rate=0.0001, max_depth=9,  
n_estimators=7000, reg_alpha=1, reg_lambda=1.
```

Το μοντέλο εκπαιδεύεται στα δεδομένα εκπαίδευσης και πραγματοποιούνται προβλέψεις στα δεδομένα ελέγχου. Τέλος υπολογίζεται η ακρίβεια του μοντέλου.

Με αυτόν τον τρόπο εκπαιδεύεται ένα μοντέλο XGBoost χρησιμοποιώντας τα δεδομένα εκπαίδευσης, πραγματοποιούνται προβλέψεις στα δεδομένα ελέγχου και υπολογίζεται η ακρίβεια του μοντέλου. Ακόμη εφαρμόστηκε `parameter grid` για να βρεθούν οι καταλληλότερες παράμετροι για το μοντέλο καθώς είναι γνωστό πως για το XGBoost είναι αρκετά σημαντική η επιλογή τους. Οι δοκιμές με τον XGBoost πραγματοποιήθηκαν με τα δεδομένα εισόδου `inhale`, `exhale`, `full breath`. Στην παρακάτω εικόνα παρουσιάζεται η μεθοδολογία που ακολουθήθηκε:



Εικόνα 11 Μεθοδολογία

Κεφάλαιο 9

Αποτελέσματα

9.1 Εντοπισμός κορυφών

Για την εξακρίβωση της αποτελεσματικότητας του εντοπισμού κορυφών αλλά και της διαδικασίας που ακολουθήθηκε, χρησιμοποιήθηκε η μετρική MSE που όπως αναφέρεται πιο πάνω είναι ένα μέτρο της ποιότητας ενός εκτιμητή. Αναμένεται μια θετική τιμή που όσο υψηλότερη είναι τόσο μεγαλύτερη αποδεικνύεται ότι είναι η απόκλιση των μετρούμενων μεγεθών. Η τιμή MSE που βρέθηκε είναι ίση με 27.81 που μπορεί σίγουρα να χαρακτηριστεί ως μη αμελητέα. Αυτό φανερώνει ότι ο αριθμός των κορυφών σε αρκετές ηχογραφήσεις απέχει από τον αριθμό που εντοπίστηκαν και καταγράφηκαν από την προηγούμενη διαδικασία.

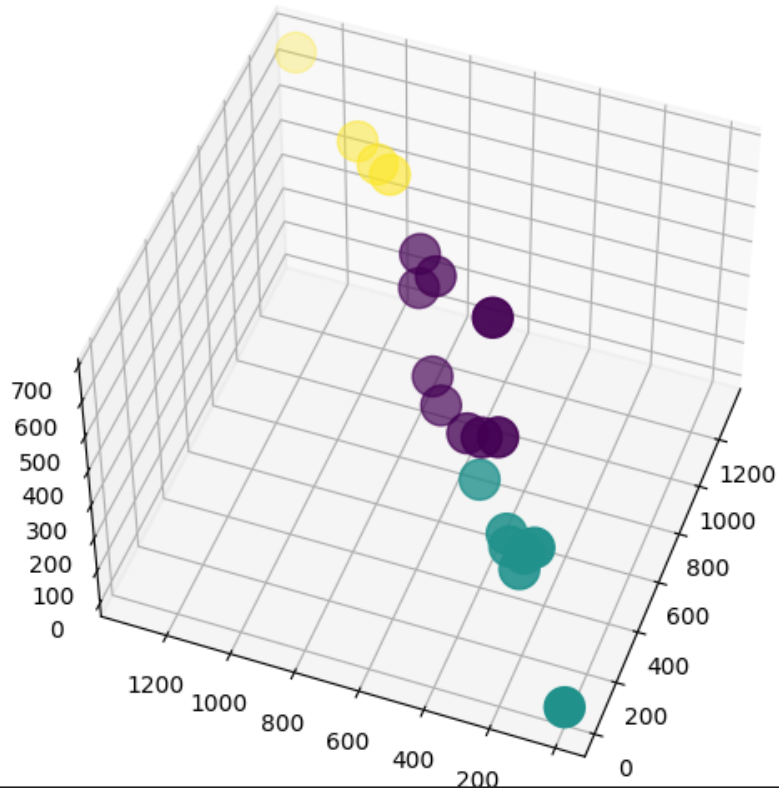
Έτσι κρίθηκε απαραίτητο να διερευνηθούν οι λόγοι για την απόκλιση αυτή. Εξετάστηκαν λοιπόν οι περιπτώσεις που κυμαίνονται σε τιμές μικρότερες από -5 η μεγαλύτερες από +5. Υπήρχαν συνολικά 27 τέτοιες περιπτώσεις για τις οποίες αναλύθηκαν τα γραφήματά τους, ακούστηκαν οι αντίστοιχες ηχογραφήσεις, πραγματοποιήθηκε η σύνδεση μεταξύ ήχων και κορυφών και καταγράφηκε η ποιότητα και οι πιθανοί λόγοι για τους οποίους η μέθοδος δεν έφερε τα επιθυμητά αποτελέσματα για τις περιπτώσεις αυτές. Συνολικά οι αιτίες κατατάσσονται σε τρεις κατηγορίες :

1. Σε ηχητικά που δεν υπάρχουν δηλαδή ούτε ακούγονται ούτε ανιχνεύονται καθόλου αναπνοες.
2. Σε ηχητικά που υπάρχουν καταγεγραμμένοι πολλοί εξωτερικοί θόρυβοι.
3. Σε ηχητικά στα οποία έχουν καταγραφεί μόνο οι εκπνοές σε αντίθεση με τις εισπνοές οι οποίες είναι πιο αδύναμες και σε πολλές περιπτώσεις ανύπαρκτες

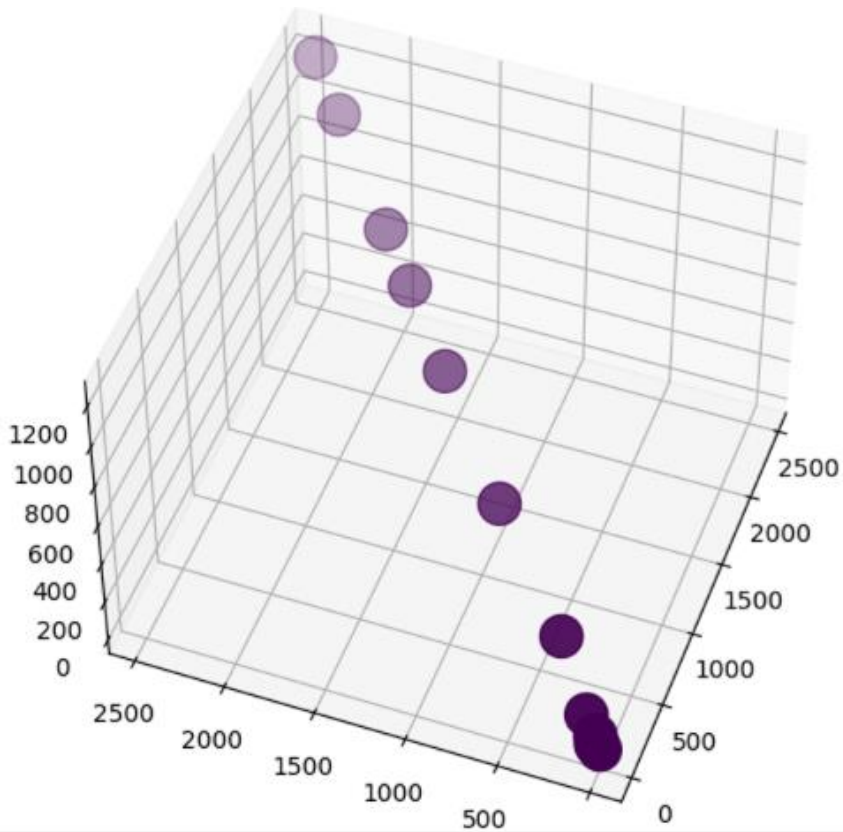
Η τρίτη κατηγορία επιδεχόταν βελτιώσεις καθώς με την κατηγοριοποίηση των κορυφών σε εισπνοές, εκπνοές και θόρυβος έγινε εφικτό να εντοπιστούν οι εγγραφές που έλειπαν οι εισπνοές και από τις εκπνοές μόνο να εντοπιστεί η διάρκεια της εκπνοής και του αναπνευστικού κύκλου. Αυτό επιχειρείται στη συνέχεια με τη βοήθεια της μη επιβλεπόμενης μηχανικής μάθησης (unsupervised learning) και πιο συγκεκριμένα της συσταδοποίησης.

9.2 Συσταδοποίηση

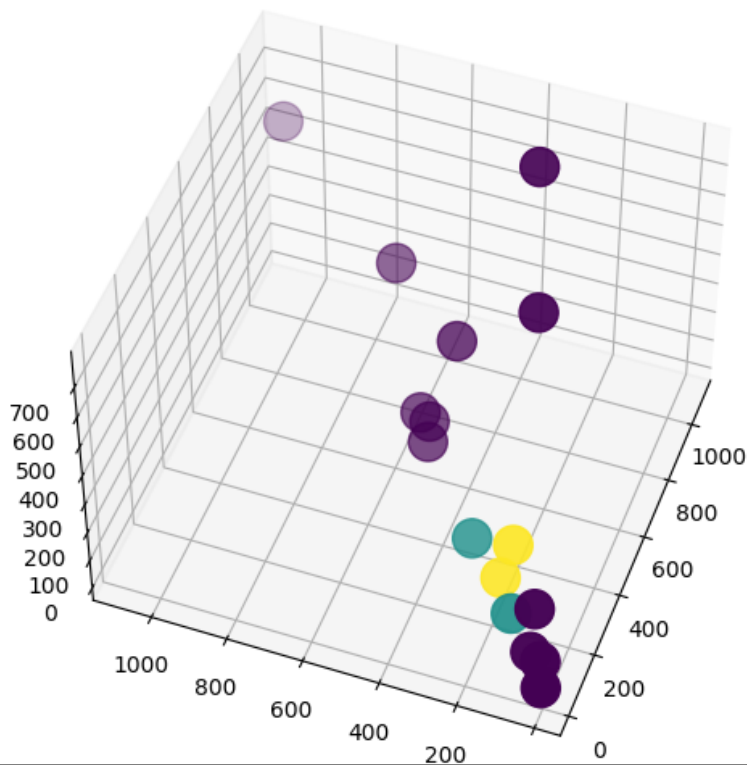
Οι αλγόριθμοι που χρησιμοποιήθηκαν είναι k-means, DBSCAN, Spectral Clustering και Hierarchical Clustering, Meanshift και παραδείγματα εφαρμογής τους φαίνονται στις παρακάτω εικόνες.



Εικόνα 12 Συσταδοποίηση με αλγόριθμο K-mean



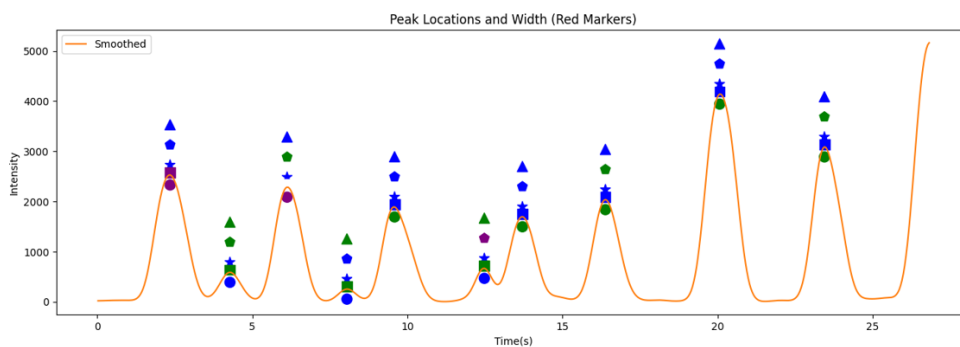
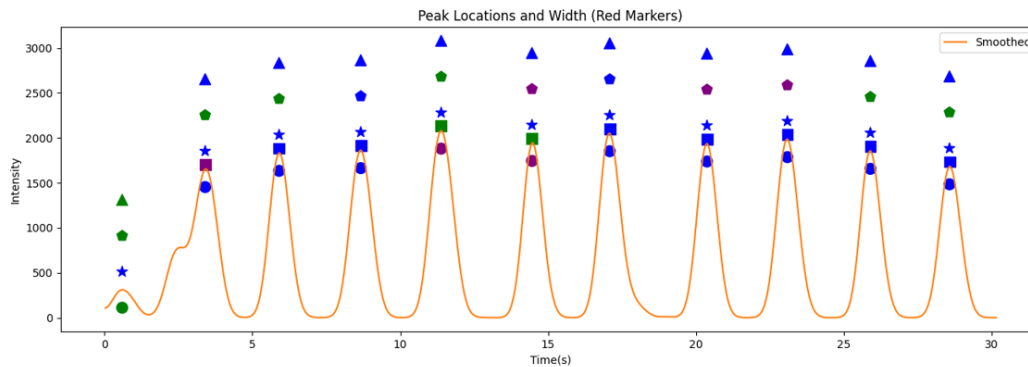
Εικόνα 13 Συσταδοποίηση με αλγόριθμο DBSCAN



Εικόνα 14 Συσταδοποίηση με αλγόριθμο Spectral

Οι αλγόριθμοι κατέταξαν σε κάθε διάγραμμα ,που προέκυψε από κάθε ηχητικό αναπνοής, τις κορυφές που ανιχνεύθηκαν από τον προηγούμενο τρόπο σε ένα

cluster(μια συστάδα) με σκοπό να κατηγοριοποιηθούν οι εισπνοές στην ίδια κατηγορία, όπως και οι εκπνοές μεταξύ τους και παράλληλα να χαρακτηριστούν ορισμένες κορυφές που δεν ανήκουν σε ήχο αναπνοής ως θόρυβος. Παρακάτω μπορούμε να δούμε την κατηγοριοποίηση αυτή πάνω σε ένα διάγραμμα.

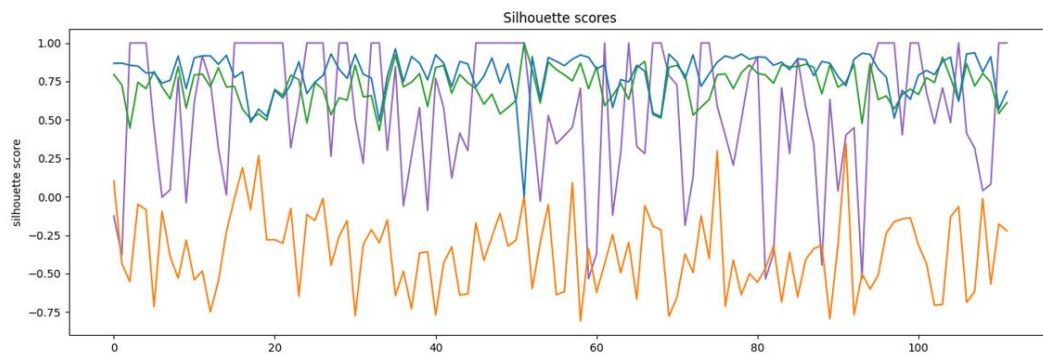


Εικόνα 15 Γραφήματα με κατηγοριοποίηση κορυφων με βάση τους αλγόριθμους συσταδοποίησης

Κάθε διαφορετικό σύμβολο αντιστοιχεί σε διαφορετικό αλγόριθμο συσταδοποίησης και κάθε διαφορετικό χρώμα αντιστοιχεί σε ένα διαφορετικό cluster που κατέταξε ο αλγόριθμος την κορυφή. Πιο συγκεκριμένα το αστεράκι είναι του αλγορίθμου DBSCAN, το κυκλάκι του K-Means, το πεντάγωνο του Spectral Clustering και το τρίγωνο για τον Hierarchical Clustering. Με βάση αυτά τα σύμβολα γίνεται κατανοητό σε πόσα και ποια clusters έχει κατατάξει ο κάθε αλγόριθμος την κάθε κορυφή. Συνολικά παρατηρούνται στα διαγράμματα ότι την χειρότερη επίδοση παρουσιάζει ο αλγόριθμος DBSCAN.

Για να εξακριβωθεί η καταλληλότητα των αλγορίθμων και να βρεθεί ο βέλτιστος υπολογίστηκε η μετρική silhouette. Στο παρακάτω διάγραμμα παρατίθενται μετρικές

τιμές silhouette για όλους τους αλγόριθμους και όλα τα διαγράμματα. Κατα τον άξονα x προχωράνε τα διαφορετικά διαγράμματα που προέκυψαν από τις ηχογραφήσεις αναπνοών και κατα μήκος του άξονα y απεικονίζεται η τιμή silhouette. Πιο συγκεκριμένα η γραφική με μωβ χρώμα αφορά τις τιμές για τον αλγόριθμο DBSCAN, η γραφική με το πράσινο αφορά τον αλγόριθμο K-means, η γραφική με το πορτοκαλί χρώμα αφορά τον αλγόριθμο Spectral και η γραφική με το μπλε χρώμα αφορά τον αλγόριθμο Hierarchical. Ακόμη για κάθε αλγόριθμο υπολογίστηκε η μέση τιμή από όλα τα διαγράμματα ώστε να γίνει εφικτό να εξάγει ένα γενικό συμπέρασμα. Οι τιμές που βρέθηκαν είναι 0.55 για τον αλγόριθμο DBSCAN, 0.72 για τον αλγόριθμο K-means, -0.35 για τον αλγόριθμο Spectral και 0.8 για τον αλγόριθμο Hierarchical.



Εικόνα 16 Γράφημα των μετρικών τιμών Silhouette

Από τον υπολογισμό της διάρκειας κάθε κορυφής(είτε εισπνοής είτε εκπνοής) και τον διαχωρισμό των κορυφών σε εισπνοές, εκπνοές και outliers έγινε εφικτός ο υπολογισμός του μέσου χρόνου εισπνοής, εκπνοής και συνολικής αναπνοής για κάθε δείγμα. Τα μεγέθη αυτά εμπεριέχουν σημαντικές πληροφορίες που συμβάλλουν στην σωστή κατηγοριοποίηση και διάγνωση.

9.3 Μοντέλο CNN

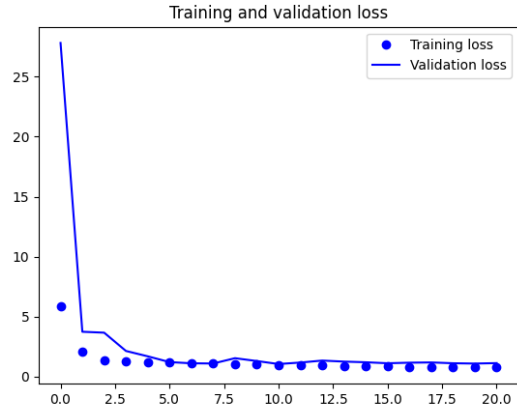
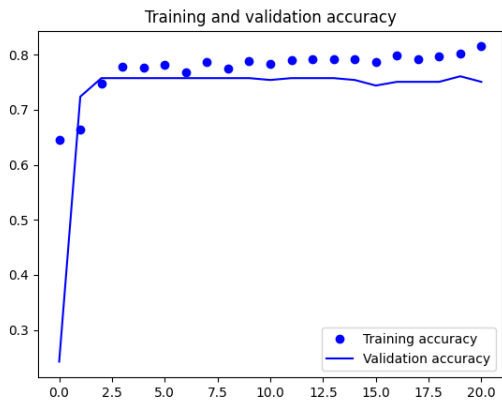
Το κύριο μοντέλο που δοκιμάστηκε είναι αυτό που απεικονίζεται στην εικόνα 12. Αξίζει να σημειωθεί ότι τα καλύτερα αποτελέσματα από αυτό το μοντέλο σημειώθηκαν με την χρήση learning rate scheduler και τη χρήση early stopping για την αποφυγή overfitting. Στον παρακάτω πίνακα συνοψίζονται όλες οι δοκιμές που πραγματοποιήθηκαν για το νευρωνικό μοντέλο:

Μοντέλο	Oversampling	Class Weights	Accuracy	Precision	Recall	F1score
Αρχική Αριτεκτονική (model1)	-	-	0.81	0.17	0.11	0.12
Αρχική Αριτεκτονική (model1)	Random Oversampler with strategy= “minority”	-	0.78	0.23	0.21	0.22
Αρχική Αριτεκτονική (model1)	SMOTE oversampler sampling strategy=auto και random state=42	-	0.74	0.24	0.25	0.24
Αρχική Αριτεκτονική (model1)	Random Oversampler με sampling strategy=0.5	-	0.76	0.25	0.11	0.12
Αρχική Αριτεκτονική (model1)	SMOTE oversampler με sampling strategy=auto και k_neighbors= 7	-	0.78	0.21	0.23	0.22
Φαρδιά αρχιτεκτονικ ή(wide_mode l)	-	-	0.79	0.2	0.19	0.21

Βαθιά αρχιτεκτονικ ή	-	-	0.76	0.18	0.2	0.19
Αρχιτεκτονικ ή με περισσότερα φίλτρα	-	-	0.81	0.23	0.21	0.22
Περισσότερα φίλτρα	Random Oversampler with strategy="mi nority"	-	0.78	0.24	0.23	0.22
Περισσότερα φίλτρα	-	Yes	0.81	0.23	0.25	0.21
Αντίστροφη αρχιτεκτονικ ή με φθίνον αριθμό φίλτρων σε κάθε επίπεδο	-	-	0.77	0.21	0.22	0.20

Παρακάτω συναντάμε αναλυτικά τις γραφικές παραστάσεις κάθε δοκιμής και ορισμένες παρατηρήσεις :

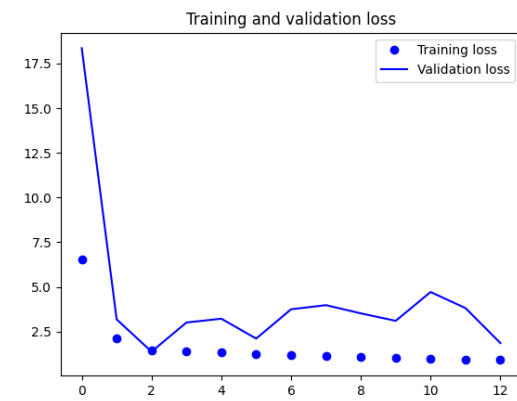
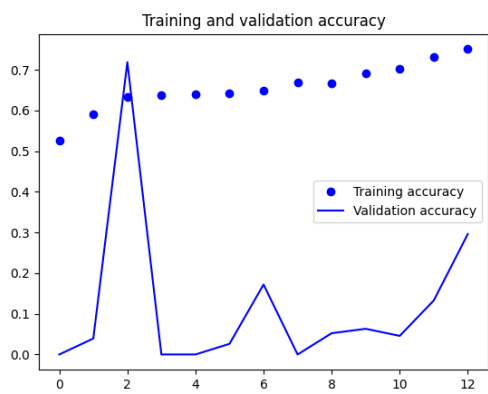
9.3.1 Αρχική εκπαίδευση μοντέλου χωρίς επαύξηση της μειονότητας κλάσης



Εικόνα 17 Αρχική εκπαίδευση

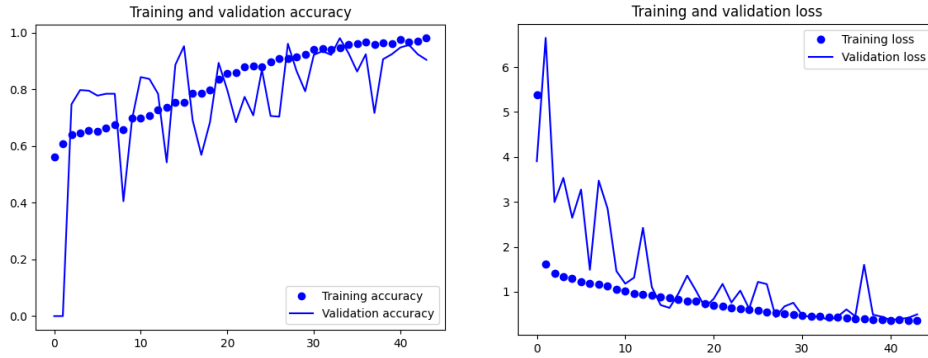
9.3.2 Υπερδειγματολόγηση

Random Oversampling with strategy= “minority”



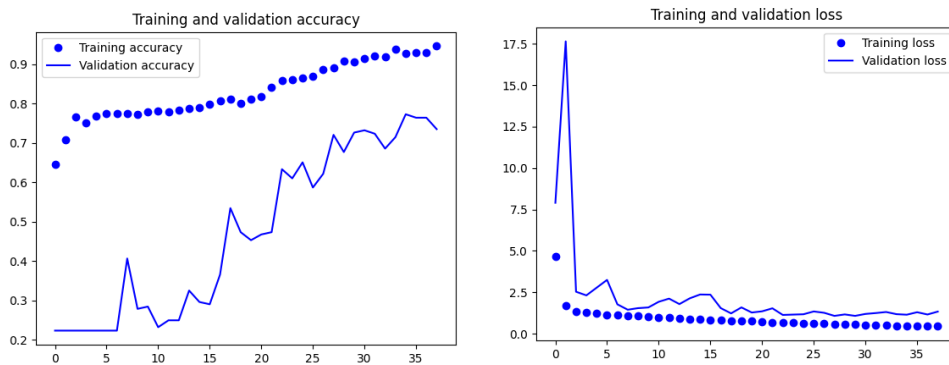
Εικόνα 14 Εκπαίδευση με υπερδειγματολόγηση της κλάσης μεινότητας

SMOTE oversampler με sampling strategy=auto και random state=42



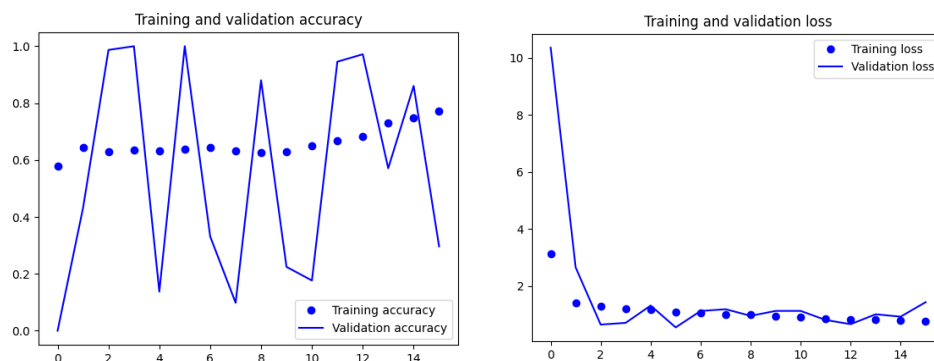
Εικόνα 15 Εκπαίδευση με Oversampler = SMOTE(sampling_strategy='auto')

Random Oversampler με sampling strategy=0.5



Εικόνα 16 RandomOverSampler(sampling_strategy=0.5)

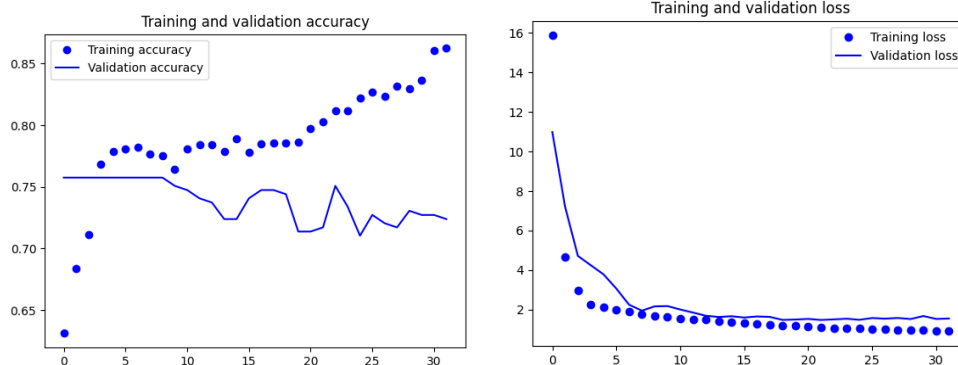
SMOTE oversampler με sampling strategy=auto και k_neighbors=7



Εικόνα 17 Εκπαίδευση με SMOTE k_neighbors=7

9.3.3 Φαρδιά Αρχιτεκτονική

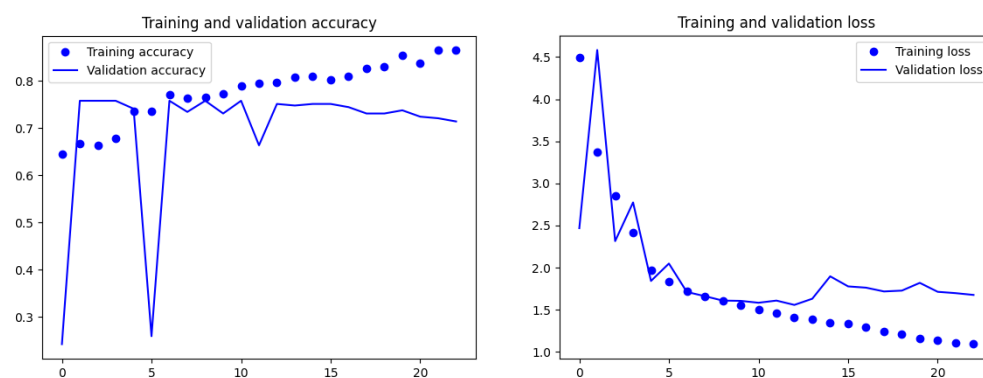
Πέτυχε σχεδόν τα ίδια αποτελέσματα με την βασική αρχιτεκτονική. Επομένως η απλούστερη αρχιτεκτονική είναι καλύτερη εφόσον πετυχαίνει παρόμοια ή και καλύτερη ακρίβεια με λιγότερο χρόνο και μικρότερη πολυπλοκότητα.



Εικόνα 18 Εκπαίδευση Φαρδιάς Αρχιτεκτονικής

Σε αυτό το διάγραμμα ακρίβειας επίσης παρατηρούμε ότι ενώ η ακρίβεια εκπαίδευσης βελτιώνεται με την πάροδο του χρόνου, η ακρίβεια εξέτασης χειροτερεύει κάτι το οποίο είναι ένδειξη υπερπροσαρμογής(overfitting).

9.3.4 Βαθεία αρχιτεκτονική



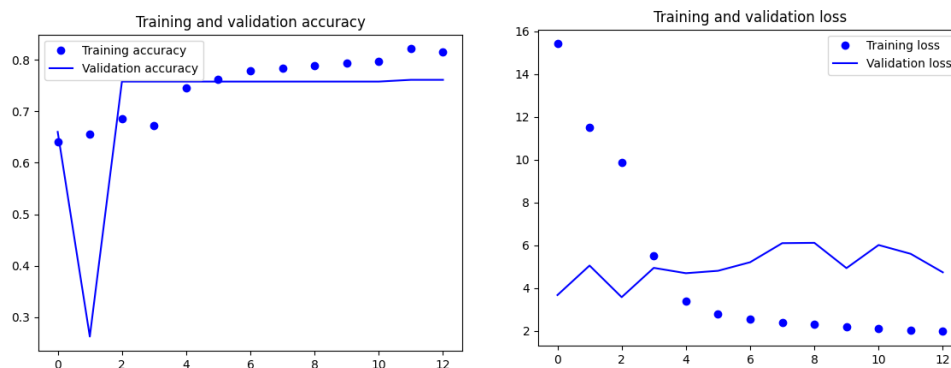
Εικόνα 19 Εκπαίδευση Βαθείας Αρχιτεκτονικής

Αυτή η αρχιτεκτονική κατάφερε λίγο μικρότερη ακρίβεια και παρατηρούμε ότι ενώ το training loss συνεχίζει να μειώνεται κατά τη διάρκεια της εκπαίδευσης το validation

loss δεν ακολουθεί την ίδια μείωση γεγονός που μας οδηγεί στο συμπέρασμα ότι προκαλείται υπερπροσαρμογή. Ωστόσο, το μοντέλο καταφέρνει να επιτύχει αρκετά υψηλή ακρίβεια καθώς χρησιμοποιούνται τεχνικές αποφυγής υπερπροσαρμογής όπως learning rate, scheduling, early stopping και τεχνικές κανονικοποίησης (regularization) όπως dropout και L2-regularizers. Εφόσον λοιπόν σε όλα τα μοντέλα που καταλήξαμε χρησιμοποιούνται αυτές τις τεχνικές δεν παρατηρείται μεγάλη διακύμανση στις ακρίβειες και τα αποτελέσματα.

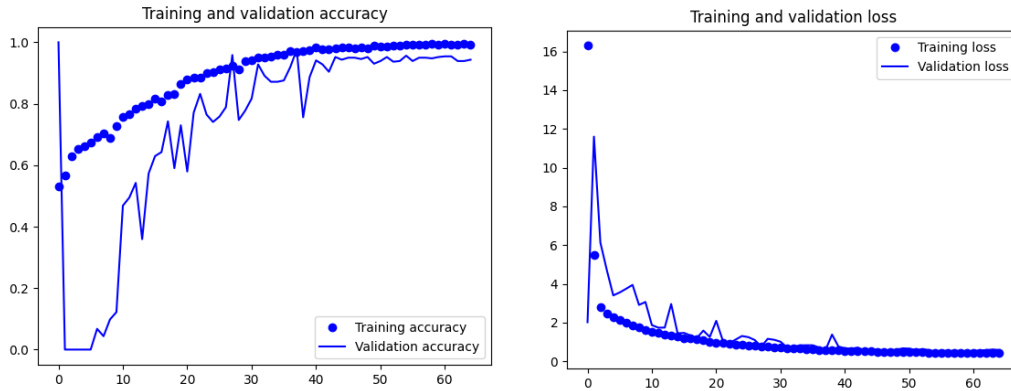
9.3.5 Αρχιτεκτονική με περισσότερα φίλτρα

Με αυτή την αρχιτεκτονική επιτυγχάνεται αρκετά υψηλή Ακρίβεια αλλά μεγαλύτερη απώλεια. Τα περισσότερα φίλτρα και η μεγαλύτερη πολυπλοκότητα καταφέρνουν να αιχμαλωτίσουν περισσότερες λεπτομέρειες και πιο σύνθετα μοτίβα στα φασματογράμματα.



Εικόνα 20 Εκπαίδευση Αρχιτεκτονικής με περισσότερα φίλτρα

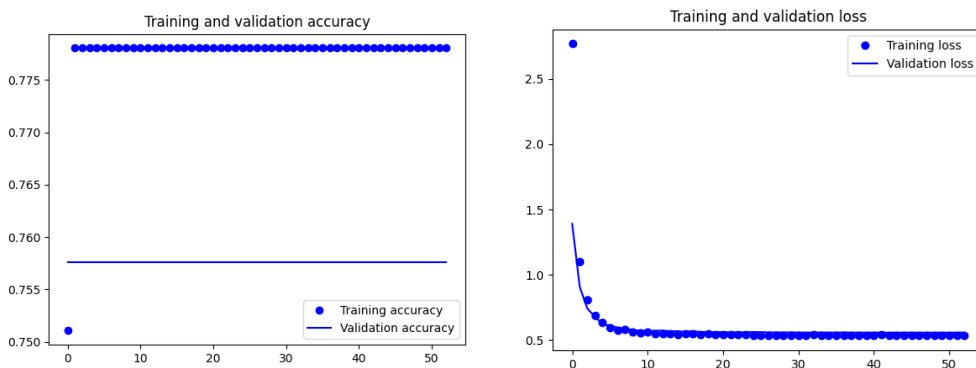
Περισσότερα φίλτρα και Random Oversampler with strategy="minority"



Εικόνα 21 Εκπαίδευση αρχιτεκτονικής με περισσότερα φίλτρα, με υπερδειγματοληψία της κλάσης μεινότητας

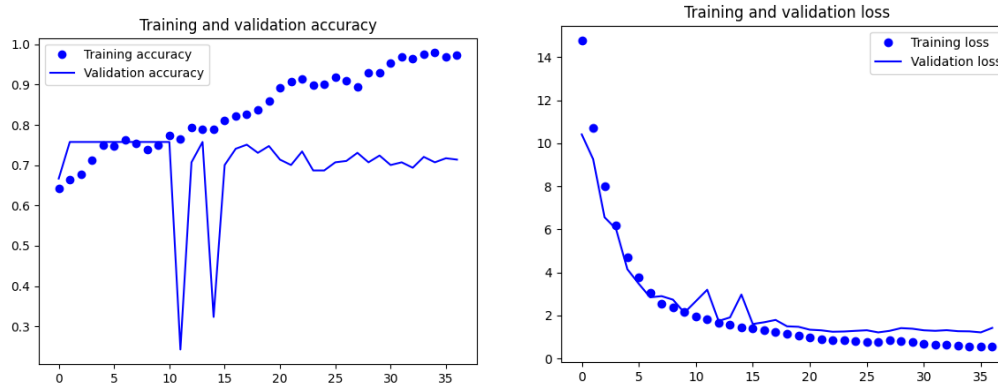
Στα γραφήματα παρατηρούμε πως κατά τη διάρκεια της εκπαίδευσης η ακρίβεια εκπαίδευσης αλλά και η ακρίβεια επικύρωσης έχουν σταθερή αύξηση και καταφέρνουν να πιτύχουν πολύ υψηλές τιμές κοντά στο 95%.

Περισσότερα φίλτρα και βάρη κλάσεων



Εικόνα 22 Εκπαίδευση με βάρη κλάσεων

Αντίστροφη αρχιτεκτονική με φθίνον αριθμό φίλτρων σε κάθε επίπεδο



Εικόνα 18 Εκπαίδευση αντίστροφης αρχιτεκτονικής

Ακολουθώντας την υπάρχουσα βιβλιογραφία εκπαιδεύτηκαν τα παραπάνω μοντέλα νευρωνικών συνελκτικών δικτύων με διάφορες παραλλαγές, εφαρμογή βαρών λόγω ανισοροπίας των δύο κλάσεων και oversampling της μειωηφούσας κλάσης. Παρατηρήθηκε πως η ακρίβεια κυμαίνεται απο 75-80% κάτι που είναι αρκετά υψηλό δεδομένων των περιορισμών που είχαμε. Συγκεκριμένα οι περιορισμοί που αντιμετωπίσαμε ήταν τα λίγα δεδομένα καθώς συνήθως για την εκπαίδευση ενός CNN χρειάζονται περισσότερα δεδομένα για να εκπαιδευτεί πλήρως ένα μοντέλο και να εντοπιστούν τα σύνθετα μοτίβα που αποκαλύπτουν συσχετίσεις μεταξύ των ηχητικών και της διάγνωσης Covid-19. Ακόμη ένα πρόβλημα που αντιμετωπίστηκε είναι η ανισοροπία των κλάσεων καθώς υπήρχαν πολλά περισσότερα δείγματα υγιών υποκειμένων σε σχέση με τους πάσχοντες με αποτέλεσμα το μοντέλο να προβλέπει σωστά τα δείγματα της πρώτης κλάσης και να αστοχεί αρκετά στην δεύτερη. Ένας επιπλέον παράγοντας που δυσκόλεψε την διαδικασία ήταν η μη καθαρότητα και ευκρίνεια των εγγραφών που πρόσθετε θόρυβο και πιθανώς δυσκόλευε την εύρεση σύνθετων μοτίβων.

9.4 Μοντέλο XGBoost

Οι καλύτερες παράμετροι που βρέθηκαν τη χρήση parameter grid ήταν: {'learning rate': 0.01, 'max_depth': 5, 'n_estimators': 1000, 'reg_alpha': 0.1, 'reg_lambda': 0.1, 'scale_pos_weight': 0.25} και πέτυχαν Ακρίβεια: 0.78. Δεδομένων των περιορισμών και των δυσκολιών που αντιμετωπίστηκαν θεωρούμε πως είναι ένα αρκετά επιτυχές

αποτέλεσμα. Η διαδικασία εύρεσης των μέσων όρων των εισπνοών και των εκπνοών συνοδεύτηκε με πολλές δυσκολίες καθώς οι εγγραφές παρουσιάζαν αρκετό θόρυβο και δεν ήταν ευδιάκριτα τα συμβάντα αναπνοής, εκπνοής. Γι'αυτό για τα μη κλινικά δεδομένα πραγματοποιήθηκε χειρωνακτικός διαχωρισμός και κατηγοριοποίηση των εγγραφών ώστε να επιλεγθούν μόνο οι χρήσιμες εγγραφές. Όπως αναφέρεται και πιο πάνω γίνεται χρήση συσταδοποίησης για να επισημανθεί ο τύπος του κάθε συμβάντος αλλά και να διαπιστωθεί η ύπαρξη θορύβου. Επιπλέον επειδή τα δεδομένα είναι περιορισμένης έκτασης και διαθέτουν περισσότερα δείγματα από την μία κλάση το μοντέλο που εκπαιδεύτηκε προβλέπει σε μεγάλο ποσοστό ότι τα δείγματα ανήκουν στην κλάση πλειοψηφίας. Ωστόσο το συμπέρασμά μας από την πρόβλεψη είναι πως οι χρόνοι εισπνοής και εκπνοής είναι ικανοποιητικοί δείκτες για την ένδειξη εμποδίου στην αναπνευστική οδό και κατ'επέκταση στην διάγνωση COVID-19.

9.5 Συνδυασμός των δύο μοντέλων

Στη συνέχεια, για να επιτευχθεί ένα πιο ανθεκτικό μοντέλο, συνδυάσαμε τις προβλέψεις των δύο επιμέρους μοντέλων. Η πρώτη δοκιμή που έγινε πέτυχε Combined Model Ακρίβεια: 0.79 που έγινε με τον συνδυασμό I,E, I/E, ενώ η δεύτερη δοκιμή ήταν Combined Model Ακρίβεια: 0.80 που έγινε με τον συνδυασμό I,E,A για το μοντέλο XGBoost. Συνολικά το υβριδικό μοντέλο πέτυχε αρκετά υψηλή ακρίβεια. Ο συνδυασμός των δύο μοντέλων επιτεύχθηκε βρίσκοντας το μέσο όρο τους.[56] Από τη μία το νευρωνικό δίκτυο που επέδειξε μεγάλη ικανότητα εντοπισμού σύνθετων μοτίβων στο επίπεδο και κατάφερε με την ιεραρχική τοποθέτηση των φίλτρων να συνθέσει πληροφορίες που οδηγούν στην σωστή κατάταξη των δειγμάτων. Και από την άλλη το μοντέλο XGBoost με τα δέντρα απόφασης του επέδειξε ικανοποιητική απόδοση στα αριθμητικά δεδομένα. Με αυτό το υβριδικό μοντέλο έγινε εφικτό να αξιοποιηθούν τα πλεονεκτήματα των δύο αυτών μοντέλων και να μετριαστούν οι αδυναμίες τους.

Κεφάλαιο 10

Συμπεράσματα

Στο πειραματικό κομμάτι που προηγήθηκε, πραγματοποιήθηκε μια προσπάθεια διάγνωσης της Covid-19 μέσω ηχητικών ηχογραφήσεων και σημάτων αναπνοής και της επεξεργασίας αυτών με τεχνικές ανάλυσης και επεξεργασίας σήματος αλλά και τεχνικές μηχανικής μάθησης. Πιο συγκεκριμένα, η πρακτική διαδικασία αρχίζει με την προεπεξεργασία των ηχητικών δειγμάτων αναπνοής. Ξεκινήσαμε την επεξεργασία των ηχητικών δεδομένων αναπνοής βρίσκοντας αρχικά τον αριθμό των αναπνοών που έχουν ηχογραφηθεί σε κάθε ηχογραφημένο αρχείο αναπνοής. Η μέθοδος βασίστηκε στην εύρεση των κορυφών δηλαδή των μεγίστων στα γραφήματα του ήχου. Η διαδικασία που ακολουθήθηκε περιλαμβάνει την χρήση ζωνοπερατού φίλτρου για την απομάκρυνση θορύβου, τη λογαριθμική συμπίεση τον υπολογισμό της έντασης του περιγράμματος από το φασματόγραμμα και την εύρεση των κορυφών πάνω στο Gaussian-smoothed περίγραμμα της έντασης των κορυφών. Στη συνέχεια επιλέχθηκαν και εφαρμόστηκαν 5 αλγόριθμοι συσταδοποίησης οι οποίοι είναι k-means, DBSCAN, Spectral Clustering και Hierarchical Clustering, Meanshift ώστε να βρεθούν οι πιο αξιοσημείωτες κορυφές και να χαρακτηριστούν ως εισπνοές, εκπνοές ή θόρυβος. Από τον εντοπισμό των κορυφών και τον υπολογισμό της διάρκειας κάθε κορυφής(εισπνοής ή εκπνοής) και την κατηγοριοποίηση των κορυφών σε εισπνοές, εκπνοές και outliers έγινε εφικτός ο υπολογισμός του μέσου χρόνου εισπνοής, εκπνοής και συνολικής αναπνοής για κάθε δείγμα. Στη συνέχεια αναπτύξαμε ένα υβριδικό μοντέλο διάγνωσης που αποτελείται από δύο απλά μοντέλα: 1. Ένα CNN νευρωνικό, συνελκτικό δίκτυο που τροφοδοτείται από φασματογράμματα και ένα XGBOOST που τροφοδοτείται από τις μέσες διάρκειες των εισπνοών, εκπνοών και συνολικών αναπνοών. Το CNN αποτελείται από διάφορα επίπεδα και φίλτρα που περιλαμβάνουν το πρώτο επίπεδο στο οποίο καθορίζεται η μορφή των εισερχομένων δεδομένων, τρία επίπεδα Conv2D & MaxPooling, επίπεδα Flatten & Dropout, πυκνά επίπεδα Dense Layer με 256 νευρώνες καθώς και τον πραγματικό ταξινομητή ένα επίπεδο με 1 νευρώνα για κάθε κλάση. Το μοντέλο XGBoost τροφοδοτήθηκε με τα αριθμητικά δεδομένα: διάρκεια εισπνοής(I), την διάρκεια εκπνοής(E) και τη συνολική διάρκεια αναπνοής(A) σε διαφορετικούς συνδυασμούς όπως I,E,A ή I,E, I/E. Ο συνδυασμός των

δύο δηλαδή το υβριδικό μοντέλο κατάφερε ακρίβεια 0.8 παρά τους περιορισμούς που αντιμετωπίστηκαν όπως αναλύονται παραπάνω.

Η παρούσα εργασία συγκεντρώνει και αναλύει την υπάρχουσα σχετική βιβλιογραφία ενώ ταυτόχρονα δοκιμάζει πρακτικά μερικές από τις μεθόδους. Επιπλέον εξερευνεί τη σημασία του λόγου των διαρκειών I/E(Εισπνοής/Εκπνοής) στην διάγνωση της Covid-19, κάτι το οποίο είναι πρωτοφανές για τη συγκεκριμένη θεματολογία. Ακόμη προσφέρει εκτενή εξήγηση της νόσησης από Covid-19 και τις επιδράσεις της στο αναπνευστικό σύστημα. Γίνεται αντιληπτό πως η εργασία προτείνει νέες μεθόδους εύρεσης της μέσης διάρκειας Εισπνοής, Εκπνοής και Αναπνοής αλλά και επεξεργασίας σήματος αναπνοών.

Μελλοντικά θα ήταν ωφέλιμο να επαναληφθεί το πείραμα για τον λόγο Εισπνοής/Εκπνοής σε μεγαλύτερη κλίμακα και με μεγαλύτερη ποιότητα δεδομένων καθώς αυτά ήταν τα κύρια εμπόδια που αντιμετωπίστηκαν κατά τη διάρκεια εκπόνησης της εργασίας ενώ ταυτόχρονα ευελπιστούμε πως οι εξελίξεις στην τεχνητή νοημοσύνη και στην επεξεργασία σήματος που αναμένονται θα βελτιώσουν την ακρίβεια και την αποδοτικότητα της ανάλυσης αναπνευστικών ήχων. Τέλος κρίνεται αναγκαίο να επισημανθεί η σημασία του εγχειρήματος αυτού η οποία έγκειται στην ευκολία και την αμεσότητα ενός τέτοιου τρόπου διάγνωσης. Είναι γνωστό πως οι έως τώρα τρόποι διάγνωσης(pcr, self-test) απαιτούν εξοπλισμό ή την παρουσία κάποιου φορέα υγείας. Κάτι τέτοιο καθιστά τη διάγνωση δύσκολη, χρονοβόρα, κοστοβόρα, μη οικολογική και επιβαρυντική για τους φορείς υγείας. Αντιθέτως ο αυτόματος τρόπος που αναλύεται παραπάνω δεν απαιτεί εξοπλισμό ή τη διαμεσολάβηση του ανθρώπου με αποτέλεσμα να καθιστά τη διαδικασία γρήγορη, φθηνή, διευκολυντική, οικολογική και προστατεύει από την έκθεση σε πιθανώς μολυσμένα περιβάλλοντα.

Βιβλιογραφία

- [1] J. Laguarda, F. Hueto, and B. Subirana, "COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings," *IEEE Open J. Eng. Med. Biol.*, vol. 1, pp. 275–281, 2020, doi: 10.1109/OJEMB.2020.3026928.
- [2] H. S. Maghded, K. Z. Ghafoor, A. S. Sadiq, K. Curran, D. B. Rawat, and K. Rabie, "A Novel AI-enabled Framework to Diagnose Coronavirus COVID-19 using Smartphone Embedded Sensors: Design Study," in *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, Las Vegas, NV, USA: IEEE, Aug. 2020, pp. 180–187. doi: 10.1109/IRI49571.2020.00033.
- [3] E. A. Mohammed, M. Keyhani, A. Sanati-Nezhad, S. H. Hejazi, and B. H. Far, "An ensemble learning approach to digital corona virus preliminary screening from cough sounds," *Sci. Rep.*, vol. 11, no. 1, p. 15404, Dec. 2021, doi: 10.1038/s41598-021-95042-2.
- [4] "Αναπνοή," *Βικιπαίδεια*. Jan. 22, 2024. Accessed: Feb. 05, 2024. [Online]. Available: <https://el.wikipedia.org/w/index.php?title=%CE%91%CE%BD%CE%B1%CF%80%CE%BD%CE%BF%CE%AE&oldid=10393407>
- [5] S. Beyerstedt, E. B. Casaro, and É. B. Rangel, "COVID-19: angiotensin-converting enzyme 2 (ACE2) expression and tissue susceptibility to SARS-CoV-2 infection," *Eur. J. Clin. Microbiol. Infect. Dis. Off. Publ. Eur. Soc. Clin. Microbiol.*, vol. 40, no. 5, pp. 905–919, May 2021, doi: 10.1007/s10096-020-04138-6.
- [6] A. Kumar *et al.*, "COVID-19 Mechanisms in the Human Body—What We Know So Far," *Front. Immunol.*, vol. 12, p. 693938, Nov. 2021, doi: 10.3389/fimmu.2021.693938.
- [7] W.-L. Su, K.-C. Lu, C.-Y. Chan, and Y.-C. Chao, "COVID-19 and the lungs: A review," *J. Infect. Public Health*, vol. 14, no. 11, pp. 1708–1714, Nov. 2021, doi: 10.1016/j.jiph.2021.09.024.
- [8] R. J. Mason, "Pathogenesis of COVID-19 from a cell biology perspective," *Eur. Respir. J.*, vol. 55, no. 4, p. 2000607, Apr. 2020, doi: 10.1183/13993003.00607-2020.
- [9] D. Cozzi *et al.*, "Ground-glass opacity (GGO): a review of the differential diagnosis in the era of COVID-19," *Jpn. J. Radiol.*, vol. 39, no. 8, pp. 721–732, 2021, doi: 10.1007/s11604-021-01120-w.
- [10] N. Malguria, L.-H. Yen, T. Lin, A. Hussein, and E. K. Fishman, "Role of Chest CT in COVID-19," *J. Clin. Imaging Sci.*, vol. 11, p. 30, 2021, doi: 10.25259/JCIS_138_2020.
- [11] H. S. Darwish, M. Y. Habash, and W. Y. Habash, "Chest computed tomography imaging features in patients with coronavirus disease 2019 (COVID-19)," *J. Int. Med. Res.*, vol. 49, no. 5, May 2021, doi: 10.1177/03000605211010631.
- [12] A. Bhatt, A. Ganatra, and K. Kotecha, "COVID-19 pulmonary consolidations detection in chest X-ray using progressive resizing and transfer learning techniques," *Heliyon*, vol. 7, no. 6, p. e07211, Jun. 2021, doi: 10.1016/j.heliyon.2021.e07211.
- [13] P. Nagpal *et al.*, "Imaging of COVID-19 pneumonia: Patterns, pathogenesis, and advances," *Br. J. Radiol.*, vol. 93, no. 1113, p. 20200538, Sep. 2020, doi: 10.1259/bjr.20200538.
- [14] D. Aguiar, J. A. Lobrinus, M. Schibler, T. Fracasso, and C. Lardi, "Inside the lungs of COVID-19 disease," *Int. J. Legal Med.*, vol. 134, no. 4, pp. 1271–1274, Jul. 2020, doi: 10.1007/s00414-020-02318-9.
- [15] T. Menter *et al.*, "Postmortem examination of COVID-19 patients reveals diffuse alveolar damage with severe capillary congestion and variegated findings in lungs and other organs suggesting vascular dysfunction," *Histopathology*, vol. 77, no. 2, pp. 198–209, 2020, doi: 10.1111/his.14134.

- [16] P. G. Gibson, L. Qin, and S. H. Pua, "COVID-19 acute respiratory distress syndrome (ARDS): clinical features and differences from typical pre-COVID-19 ARDS," *Med. J. Aust.*, vol. 213, no. 2, pp. 54-56.e1, Jul. 2020, doi: 10.5694/mja2.50674.
- [17] M. Polverino, F. Polverino, M. Fasolino, F. Andò, A. Alfieri, and F. De Blasio, "Anatomy and neuro-pathophysiology of the cough reflex arc," *Multidiscip. Respir. Med.*, vol. 7, no. 1, p. 5, Jun. 2012, doi: 10.1186/2049-6958-7-5.
- [18] A. Pal and M. Sankarasubbu, "Pay attention to the cough: early diagnosis of COVID-19 using interpretable symptoms embeddings with cough sound signal processing," in *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, Virtual Event Republic of Korea: ACM, Mar. 2021, pp. 620–628. doi: 10.1145/3412841.3441943.
- [19] M. J. Tobin, "Monitoring of pressure, flow, and volume during mechanical ventilation," *Respir. Care*, vol. 37, no. 9, pp. 1081–1096, Sep. 1992.
- [20] F. Bellemare and A. Grassino, "Effect of pressure and timing of contraction on human diaphragm fatigue," *J. Appl. Physiol.*, vol. 53, no. 5, pp. 1190–1195, Nov. 1982, doi: 10.1152/jappl.1982.53.5.1190.
- [21] J. G. Burdon, K. J. Killian, and N. L. Jones, "Pattern of breathing during exercise in patients with interstitial lung disease.," *Thorax*, vol. 38, no. 10, pp. 778–784, Oct. 1983, doi: 10.1136/thx.38.10.778.
- [22] C. Veldman, W. S. de Boer, H. A. M. Kerstjens, M. A. Edens, and J. W. K. van den Berg, "Sonographic follow-up of diaphragm function in COVID-19: an exploratory study," *ERJ Open Res.*, Jan. 2023, doi: 10.1183/23120541.00623-2022.
- [23] M. F. M. Shakhah, A. A. Wahab, and M. I. M. Salim, "Assessment of inspiration and expiration time using infrared thermal imaging modality," *Infrared Phys. Technol.*, vol. 99, pp. 129–139, Jun. 2019, doi: 10.1016/j.infrared.2019.04.012.
- [24] E. Sembroski, C. M. McDowell, and M. M. Mannion, "Pulseless Electrical Activity Cardiac Arrest," *J. Educ. Teach. Emerg. Med.*, vol. 5, no. 1, pp. S1–S25, Jan. 2020, doi: 10.21980/J8Z055.
- [25] J. J. Murphy, P. G. Allen, T. H. Stevens, and D. Weatherhead, "A Meta-analysis of Hypothetical Bias in Stated Preference Valuation," *Environ. Resour. Econ.*, vol. 30, no. 3, pp. 313–325, Mar. 2005, doi: 10.1007/s10640-004-3332-z.
- [26] "Full article: Dynamic change in respiratory resistance during inspiratory and expiratory phases of tidal breathing in patients with chronic obstructive pulmonary disease." Accessed: Sep. 26, 2023. [Online]. Available: <https://www.tandfonline.com/doi/full/10.2147/COPD.S30399>
- [27] J. M. Haynes, "Basic spirometry testing and interpretation for the primary care provider," *Can. J. Respir. Ther. CJRT Rev. Can. Thérapie Respir. RCTR*, vol. 54, no. 4, p. 10.29390/cjrt-2018-017, 2018, doi: 10.29390/cjrt-2018-017.
- [28] "High-pass filter," *Wikipedia*. Mar. 01, 2023. Accessed: Sep. 26, 2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=High-pass_filter&oldid=1142309305
- [29] J. Smith, "Signal Processing Libraries for Faust".
- [30] "C3S1_LogCompression." Accessed: Sep. 26, 2023. [Online]. Available: https://www.audiolabs-erlangen.de/resources/MIR/FMP/C3/C3S1_LogCompression.html
- [31] "Short-time Fourier transform - Wikipedia." Accessed: Sep. 26, 2023. [Online]. Available: https://en.wikipedia.org/wiki/Short-time_Fourier_transform
- [32] "The Wavelet Tutorial." Accessed: Sep. 29, 2023. [Online]. Available: <https://users.rowan.edu/~polikar/WTtutorial.html>
- [33] A. V. Oppenheim, R. W. Schaffer, and R. W. Schaffer, *Discrete-time signal processing*, Third edition, Pearson new international edition. in Always learning. Harlow: Pearson, 2014.

- [34] "2.3. Clustering — scikit-learn 1.3.1 documentation." Accessed: Sep. 28, 2023. [Online]. Available: <https://scikit-learn.org/stable/modules/clustering.html>
- [35] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982, doi: 10.1109/TIT.1982.1056489.
- [36] R. C. Tryon, *Cluster analysis [by] Robert C. Tryon [and] Daniel E. Bailey*. New York: McGraw-Hill, 1970.
- [37] E. Schubert, "Knowledge Discovery in Databases," *Clust. Anal.*.
- [38] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013, doi: 10.1109/TPAMI.2013.50.
- [39] D. Yang, H. Li, W. Yu, Y. Guo, and W. Liang, "CNN-Cap: Effective Convolutional Neural Network-based Capacitance Models for Interconnect Capacitance Extraction," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 28, no. 4, p. 56:1-56:22, May 2023, doi: 10.1145/3564931.
- [40] N. Rusk, "Deep learning," *Nat. Methods*, vol. 13, no. 1, Art. no. 1, Jan. 2016, doi: 10.1038/nmeth.3707.
- [41] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, Art. no. 7553, May 2015, doi: 10.1038/nature14539.
- [42] X. Hao, G. Zhang, and S. Ma, "Deep Learning," *Int. J. Semantic Comput.*, vol. 10, no. 03, pp. 417–439, Sep. 2016, doi: 10.1142/S1793351X16500045.
- [43] M. Alkhodari and A. H. Khandoker, "Detection of COVID-19 in smartphone-based breathing recordings: A pre-screening deep learning tool," *PLOS ONE*, vol. 17, no. 1, p. e0262448, Jan. 2022, doi: 10.1371/journal.pone.0262448.
- [44] M. Pahar, M. Klopper, R. Warren, and T. Niesler, "COVID-19 detection in cough, breath and speech using deep transfer learning and bottleneck features," *Comput. Biol. Med.*, vol. 141, p. 105153, Feb. 2022, doi: 10.1016/j.compbio.2021.105153.
- [45] C. Brown *et al.*, "Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Aug. 2020, pp. 3474–3484. doi: 10.1145/3394486.3412865.
- [46] C. R. Rodriguez, D. Angeles, R. Chafloque, F. Kaseng, and B. Pandey, "Deep Learning Audio Spectrograms Processing to the Early COVID-19 Detection," in *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, Bhimtal, India: IEEE, Sep. 2020, pp. 429–434. doi: 10.1109/CICN49253.2020.9242583.
- [47] I. Song, "Diagnosis of pneumonia from sounds collected using low cost cell phones," in *2015 International Joint Conference on Neural Networks (IJCNN)*, Killarney: IEEE, Jul. 2015, pp. 1–8. doi: 10.1109/IJCNN.2015.7280317.
- [48] L. Fraiwan, O. Hassanin, M. Fraiwan, B. Khassawneh, A. M. Ibnian, and M. Alkhodari, "Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers," *Biocybern. Biomed. Eng.*, vol. 41, no. 1, pp. 1–14, Jan. 2021, doi: 10.1016/j.bbe.2020.11.003.
- [49] V. Jindal, V. Agarwal, and S. Kalaivani, "Respiratory Sound Analysis for Detection of Pulmonary Diseases," in *2018 IEEE Applied Signal Processing Conference (ASPCON)*, Kolkata, India: IEEE, Dec. 2018, pp. 293–296. doi: 10.1109/ASPCON.2018.8748284.
- [50] Hui-Hui Wang, Jia-Ming Liu, Mingyu You, and Guo-Zheng Li, "Audio signals encoding for cough classification using convolutional neural networks: A comparative study," in *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Washington, DC, USA: IEEE, Nov. 2015, pp. 442–445. doi: 10.1109/BIBM.2015.7359724.

- [51] C. Shimon, G. Shafat, I. Dangoor, and A. Ben-Shitrit, "Artificial intelligence enabled preliminary diagnosis for COVID-19 from voice cues and questionnaires," *J. Acoust. Soc. Am.*, vol. 149, no. 2, pp. 1120–1124, Feb. 2021, doi: 10.1121/10.0003434.
- [52] Y. Yu, T. Zhang, and Q. Zheng, "Pneumonia Diagnosis based on CNN-LSTM-BERT Model," p. 4.
- [53] O. Alaaeldin, O. S. El Ahl, L. Elmahy, M. Ihab, and W. Gomaa, "Analysis of COVID-19 Using Imaging and Audio Modalities," in *2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, Seoul, Korea, Republic of: IEEE, Jan. 2022, pp. 1–10. doi: 10.1109/IMCOM53663.2022.9721730.
- [54] E. S. Adamidi, K. Mitsis, and K. S. Nikita, "Artificial intelligence in clinical care amidst COVID-19 pandemic: A systematic review," *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 2833–2850, 2021, doi: <https://doi.org/10.1016/j.csbj.2021.05.010>.
- [55] K. Zarkogianni *et al.*, "The smarty4covid dataset and knowledge base as a framework for interpretable physiological audio data analysis," *Sci. Data*, vol. 10, no. 1, p. 770, Nov. 2023, doi: 10.1038/s41597-023-02646-6.
- [56] M. Athanasiou, K. Zarkogianni, K. Karytsas, and K. S. Nikita, "An LSTM-based Approach Towards Automated Meal Detection from Continuous Glucose Monitoring in Type 1 Diabetes Mellitus," in *2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE)*, 2021, pp. 1–5. doi: 10.1109/BIBE52308.2021.9635246.