ΔΙΕΘΝΕΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΗΣ ΕΛΛΑΔΟΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΕΛΛΗΝΙΚΩΝ ΕΙΔΩΝ ΜΟΥΣΙΚΗΣ ΜΕΣΩ ΗΧΗΤΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΚΑΙ ΜΟΝΤΕΛΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ»

Των φοιτητών:
Καρδακάρη Χριστίνας-Ελένης
Αρ. Μητρώου: 514056
Στέλιου Νταϊλάκη
Αρ. Μητρώου:144230

Επιβλέπων
Ρήγας Κωτσάκης
Καθηγητής

Ημερομηνία 22/6/2022

Τίτλος Δ.Ε.: ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΕΛΛΗΝΙΚΩΝ ΕΙΔΩΝ ΜΟΥΣΙΚΗΣ ΜΕΣΩ ΗΧΗΤΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΚΑΙ ΜΟΝΤΕΛΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Κωδικός Δ.Ε. 21344

Ονοματεπώνυμο φοιτητών: Καρδακάρη Χρηστίνα-Ελένη, Στέλιος Νταϊλάκης

Ονοματεπώνυμο εισηγητή Κωτσάκης Ρήγας

Ημερομηνία ανάληψης Δ.Ε. 13-10-2021

Ημερομηνία περάτωσης Δ.Ε. 22-6-2022

Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Πρόλογος

Η παρούσα διπλωματική εργασία επιλέχθει για να αναδείξει μια εφαρμογή η οποία ενώ χρησιμοποιείται πολύ συχνά στην καθημερινότητα μας με ποικίλους τρόπους, είναι ακόμη άγνωστο πεδίο για πολλούς. Ακόμα και για τους ερευνητές είναι ένα θέμα το οποίο έχει πολλά περιθώρια βελτίωσης και εξέλιξης. Οπωσδήποτε είναι μια εφαρμογή η οποία τείνει να κάνει την ζωή μας πολύ πιο εύκολη όπως επίσης να βοηθήσει πολύ ανθρώπους με αισθητηριακές αναπηρίες.

Μέσα από αυτή την έρευνα καταφέραμε να αντιληφθούμε τις πολλαπλές λειτουργίες και εφαρμογές της κατηγοριοποίησης του ήχου εφιστώντας την προσοχή μας στην μουσική. Επίσης, καταλάβαμε την τεχνολογία πίσω από αυτό κατανοώντας φυσικά ότι το πεδίο έρευνας και μελέτης του αντικειμένου είναι πολύ μεγαλύτερο.

Περίληψη

Η συγκεκριμένη διπλωματική εργασία έχει ως θέμα την κατηγοριοποίηση της μουσικής. Αρχικά, αναλύονται τα προβλήματα κατηγοριοποίησης στον ήχο αναλύοντας κάποιες βασικές έννοιες, ενώ στη συνέχεια εξετάζεται συνοπτικά το πρόβλημα, στο πλαίσιο της ανάκτησης πληροφοριών μουσικής (Mir). Παρουσιάζονται και περιγράφονται τα χαρακτηριστικά του ήχου, ενώ ταυτόχρονα πραγματοποιούνται πειράματα με τα γνωστά αλγοριθμικά μοντέλα ταξινόμησης ηχητικού περιεχομένου, για την εξαγωγή χρήσιμων συμπερασμάτων.

Τα σημαντικότερα σημεία που αναδείχθηκαν μέσα από αυτή την έρευνα είναι το πόσο χρήσιμο κομμάτι την καθημερινότητας μας είναι η κατηγοριοποίηση του ήχου. Η κατάτμηση της μουσικής είναι κάτι το οποίο ήδη ευρέως χρησιμοποιείται και έχει διευκολύνει στην αναζήτηση συγκεκριμένων κομματιών, με τη χρησιμοποίηση κατάλληλων μηχανισμών μεταδεδομένων.

Είναι σαφές ότι υπάρχουν ακόμα πολλές εφαρμογές της τεχνολογίας αυτής κάποιες από τις οποίες είναι ήδη σε ευρεία χρήση και κάποιες οι οποίες είναι ακόμα στο ερευνητικό στάδιο. Υπάρχουν ακόμη πολλά προβλήματα τα οποία αναμένουν λύσεις. Σε ότι αφορά την κατηγοριοποίηση της μουσικής είναι ίσως η πιο διαδεδομένη εφαρμογή αυτής της τεχνολογίας.

Τέλος, παραθέτουμε την δική μας έρευνα η οποία αφορά την ελληνική μουσική. Τα αποτελέσματα που εξήχθησαν μέσα από αυτή την έρευνα φέρουν ικανοποιητική απόδοση ενώ ταυτόχρονα πραγματοποιήθηκε σύγκριση με πειράματα ταξινόμησης ξένης μουσικής. Αξίζει να αναφερθεί πως η κατηγοριοποίηση μουσικής με βάση το είδος φέρει αυξημένο ερευνητικό ενδιαφέρον, ενώ στην παρούσα εργασία παρουσιάζονται πρώιμα πειράματα και μοντέλα ταξινόμησης.

Segmentation of Greek music genre through sound features and machine learning models.

Kardakari Christina-Helen

Ntailakis Stelios

Abstract

The subject of this thesis is music segmentation. First, the problems of sound segmentation are analyzed by some basic concepts, and then the problem is briefly examined, in the context of retrieving music information (Mir). The characteristics of sound are presented and described, while at the same time experiments are performed with the known algorithmic models for classification of audio content, in order to draw useful conclusions.

The most important points that emerged through this research is what an important part of our daily life is sound segmentation. Music segmentation is already widely used and has made it easier to search for specific tracks, using appropriate metadata mechanisms.

It is clear that there are still many applications of this technology some of which are already in use and some are still in research. Many problems are waiting for solutions to be found. When it comes to music segmentation it is probably the most wildspread applications of this technology.

Finally, we adduce out own research that is about Greek music. The results obtained through this research are satisfactory while at the same time a comparison was made with foreign music classification experiments. It is worth noting that the categorization of music by genre has increased research interest, while in the present paper early experiments and classification models are presented.

Ευχαριστίες

Στον καθηγητή μας κύριο Ρ. Κωτσάκη για την βοήθεια και την καθοδήγηση του και στις οικογένειές μας.

Περιεχόμενα

Κατάλογος Εικόνων

Εισαγωγή

Σκοπός αυτής της πτυχιακής εργασίας είναι η μελέτη και κατανόηση της ταξινόμησης του ήχου και συγκεκριμένα της μουσικής καθώς και η παρουσίαση της δικής μας έρευνας πάνω σε αυτό.

Σε γενικές γραμμές η ταξινόμηση ήχου έχει πάρα πολλές δυνατότητες και εφαρμογές. Η δική μας έρευνα στόχευσε στην κατηγοριοποίηση ελληνικών μουσικών κομματιών σε τρία συγκεκριμένα είδη.

Στο πρώτο κεφάλαιο δίνεται ο ορισμός του ήχου και της κατηγοριοποίησης ηχητικού περιεχομένου. Επίσης, γίνεται αναφορά στην χρησιμότητα και στην εξέλιξη της κατηγοριοποίησης, στην εξέταση και τις πιθανές λύσεις του προβλήματος. Ακόμα, γίνεται εισαγωγή στην ανάκτηση πληροφοριών μουσικής και στις μελλοντικές προοπτικές.

Στο δεύτερο κεφάλαιο εξηγούνται τα χαρακτηριστικά των ήχων και χωρίζονται στις επιμέρους βασικές κατηγορίες τους.

Στο τρίτο κεφάλαιο αναφέρονται τα είδη μουσικής που επιλέχθηκαν για την έρευνα μας και ο λόγος για τον οποίο επιλέχθηκαν. Επίσης, εξηγούνται πιο αναλυτικά τα νευρωνικά δίκτυα.

Στο τέταρτο κεφάλαιο περιγράφεται με σαφήνεια το πρόβλημα της κατηγοριοποίησης και παρουσιάζονται οι εργασίες διάφορων ερευνητών από όλο τον κόσμο που έχουν ασχοληθεί με το θέμα.

Στο πέμπτο κεφάλαιο ξεκινάει η περιγραφή της διαδικασίας της δικής μας έρευνας. Ξεκινώντας με την συλλογή και προεπεξεργασία των δεδομένων και στην συνέχεια το στάδιο της κατάτμησης.

Στο έκτο κεφάλαιο αναλύεται το δεύτερο βήμα που είναι η εξαγωγή των ηχητικών χαρακτηριστικών και η κατάτμηση τους.

Στο έβδομο κεφάλαιο αναλύονται τα πειράματα που εκπονήθηκαν, οι τρόποι επικύρωσης και οι αλγόριθμοι εκμάθησης μοντέλων.

Έπειτα, στο όγδοο κεφάλαιο φαίνονται τα αποτελέσματα των πειραμάτων μέσα από πίνακες και ραβδογράμματα που δείχνουν αναλυτικά όλα τα επιμέρους χαρακτηριστικά.

Στο ένατο κεφάλαιο γίνεται αξιολόγηση των χαρακτηριστικών που βρέθηκαν παραπάνω δημιουργώντας έτσι μια σειρά κατάταξης.

Τέλος, στο δέκατο κεφάλαιο παραθέτονται τα συμπεράσματα που βγήκαν από το πείραμα και οι μελλοντικές κατευθύνσεις του πεδίου αυτού.

ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ

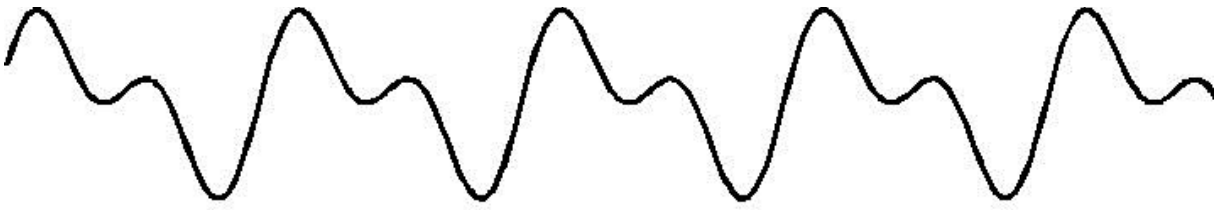
Κεφάλαιο 1: Προβλήματα κατηγοριοποίησης ήχων

1.1 Εισαγωγή

Στο πρώτο κεφάλαιο δίνονται ορισμοί για την κατάτμηση του ήχου και εξηγούμε γιατί είναι σημαντική, ενώ παράλληλα αναφέρουμε κάποια από τα προβλήματα που δημιουργούνται. Στη συνέχεια, παρουσιάζονται κάποια από τα προγράμματα και τους αλγόριθμους οι οποίοι τείνουν να τα λύσουν.

1.2 Η έννοια του ήχου

Το ηχητικό σήμα παράγεται από τις διακυμάνσεις της πίεσης του αέρα. Επαναλαμβάνονται συχνά σε συγκεκριμένες περιόδους χρόνου, έτσι ώστε να δημιουργούνται κύματα με το ίδιο σχήμα. Το ύψος του κύματος, γνωστό και ως πλάτος, δείχνει την ένταση του ήχου. Η πλειοψηφία των ήχων που συναντάμε δεν είναι αρμονικές και περιοδικές κυματομορφές. Τα σήματα διαφορετικών συχνοτήτων όταν προστίθενται μαζί, δημιουργούν πολύπλοκα επαναλαμβανόμενα μοτίβα. Τέτοιοι ήχοι είναι, για παράδειγμα, η ανθρώπινη φωνή ή τα μουσικά όργανα. Το ανθρώπινο αυτί έχει την ικανότητα να διακρίνει τους διαφορετικούς ήχους και να τους κατατάξει σε κατηγορίες [1].



Εικόνα 1. 1 Μουσική κυματομορφή με σύνθετο επαναλαμβανόμενο σήμα

1.3 Ορισμός κατηγοριοποίησης

Η κατηγοριοποίηση βασίζεται στο πλάτος του σήματος Root Mean Square (RMS). Το πρωτότυπο αρχείο διασπάται σε διαστήματα του ενός δευτερολέπτου για κάθε ένα από τα οποία υπάρχουν πενήντα τιμές RMS. Στη συνέχεια, υπολογίζεται η μέση τιμή και η διασπορά αυτών των τιμών. Στη πρώτη φάση του αλγορίθμου εντοπίζεται η χρονική στιγμή στην οποία γίνεται η αλλαγή και στη δεύτερη φάση εντοπίζεται η αλλαγή που έχει γίνει με ακρίβεια 20msec [2].

1.4 Χρησιμότητα και εξέλιξη

Ο άνθρωπος από την φύση του είναι ικανός να κάνει κατηγοριοποίηση του ήχου. Για παράδειγμα, μπορεί από τον ήχο να ξεχωρίσει τι ζώο ακούει, να ξεχωρίσει τις φωνές των ανθρώπων που γνωρίζει ή να ξεχωρίσει εάν χτυπάει το κουδούνι ή το τηλέφωνο. Υπάρχουν όμως περιπτώσεις όπου ο ήχος είναι πολύ σιγανός, αδιευκρίνιστος ή έχει πολλές παρεμβολές. Λόγου χάρη, εάν ακούει τη φωνή μέσα από

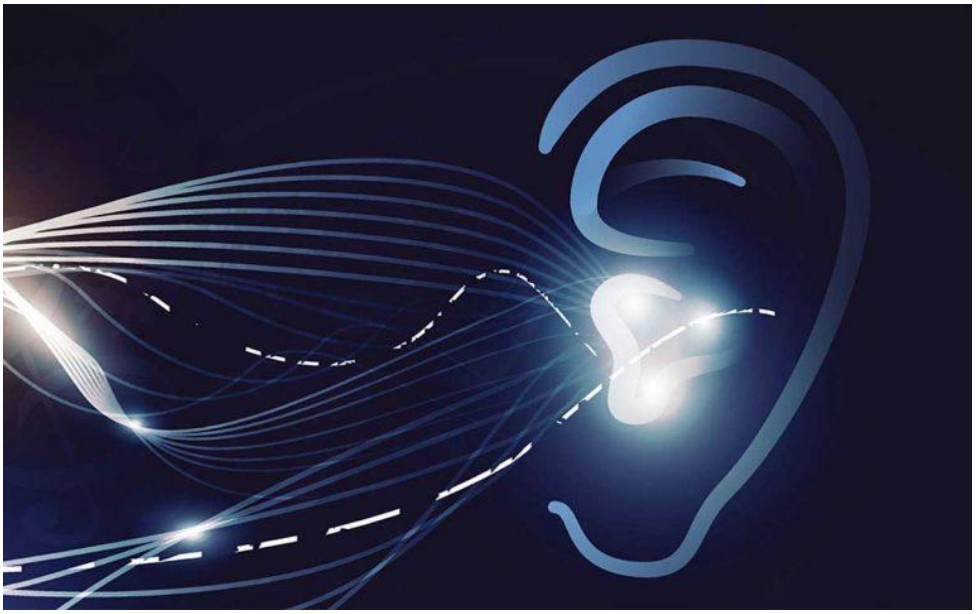
το τηλέφωνο.

Η αντίληψη του πώς γίνεται αυτή η διαδικασία θα μπορούσε να είναι πολύ διδακτική. Αρχικά, θα βοηθούσε στην σωστότερη διάγνωση και θεραπεία ακουστικών παθήσεων. Επίσης, θα ήταν πολύ βοηθητικό στην επιστήμη να υπάρχει πρόγραμμα το οποίο κατηγοριοποιεί τους ήχους με τον ίδιο τρόπο που το κάνει ο άνθρωπος. Παραδείγματος χάρη, συχνά οι γιατροί χρησιμοποιούν την ακοή τους για να εντοπίσουν αναπνευστικά προβλήματα ή καρδιακές ανωμαλίες. Είναι σαφές λοιπόν, ότι ένα τέτοιο μηχάνημα θα ήταν σωτήριο σε περιπτώσεις που η πρόσβαση του γιατρού στον ασθενή είναι δύσκολη ή χρονοβόρα.

Επίσης, οι μηχανικοί αυτοκινήτων συχνά μπορούν από τον ήχο που κάνει η μηχανή κατά τη λειτουργία της να εντοπίσουν το πρόβλημα που έχει προκύψει. Είναι, λοιπόν, σαφές ότι και αυτή η επιστήμη μπορεί να εξελιχθεί σημαντικά από την ύπαρξη ενός τέτοιου μηχανήματος.

Δεν υπάρχει αμφιβολία ότι ένα τέτοιο μηχάνημα, σαφέστατα, θα προσέφερε μεγαλύτερη ακρίβεια με τον ίδιο τρόπο που το κάνει ένα μικροσκόπιο. Επιπλέον, θα εξυπηρετούσε στο να μπορούμε να ηχογραφούμε τους ήχους ούτως ώστε να μπορούμε να το ξανακούσουμε, να τους απομονώσουμε ή να αφαιρέσουμε τις παρεμβολές.

Επιπροσθέτως, θα μπορούσε να χρησιμοποιηθεί στην αυτόματη καταγραφή του ήχου σε κείμενο, κώδικα Morse (μέθοδος για μετάδοση πληροφορίας με παλμούς μικρής και μεγάλης διάρκειας (τελείες παύλες) [3].



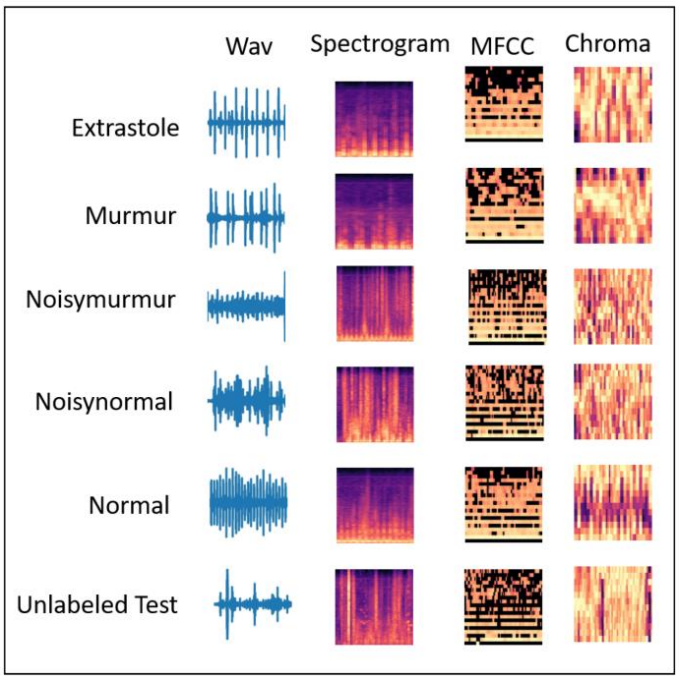
Εικόνα 1. 2 Έρευνα MIT πως ξεχωρίζουμε τις φωνές

1.5 Εξέταση του προβλήματος

Η κατηγοριοποίηση ήχου είναι η πιο σημαντική προϋπόθεση για την ανάλυση περιεχομένου πολυμέσων. Η ταξινόμηση του ήχου συνίσταται από τα φυσικά και αντιληπτά χαρακτηριστικά του. Η χρήση των κοινών χαρακτηριστικών θα πρέπει να οδηγεί στην ταξινόμηση των ήχων στις κατάλληλες κατηγορίες [4]. Ωστόσο, η ταξινόμηση βάση περιεχομένου μπορεί να δημιουργήσει σημαντικά προβλήματα όταν δεν αναφερόμαστε σε έναν ενιαίο τύπο όπως η μουσική ή ομιλία αλλά σε μεικτούς ήχους, για παράδειγμα ομιλία με μουσικό υπόβαθρο ή ήχους περιβάλλοντος [5]. Τέτοιου είδους αρχεία ήχου παρουσιάζουν δυσκολίες, όχι μόνο στην ταξινόμηση, αλλά και την επεξεργασία, στην αναγνώριση και στην καταγραφή τους [6].

1.6 Επίλυση του προβλήματος

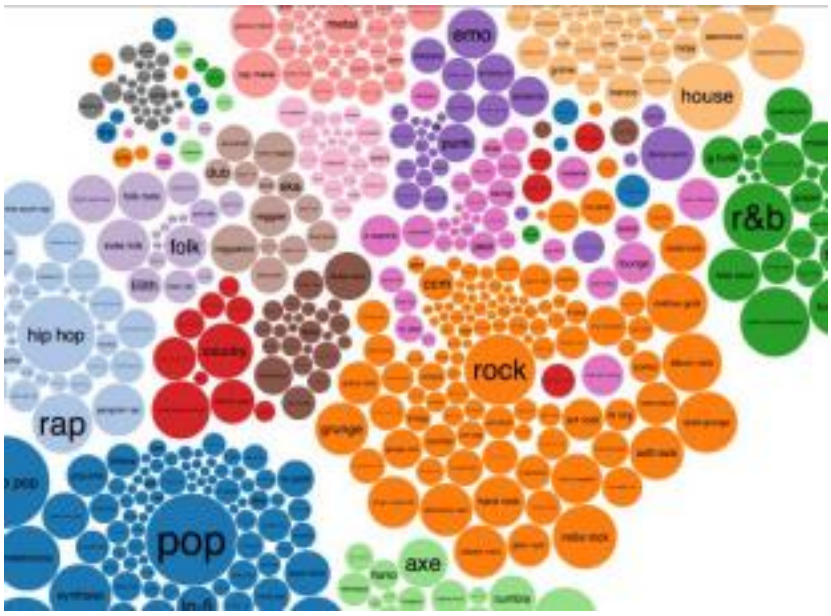
Προκειμένου να λυθούν τέτοια προβλήματα δημιουργήθηκαν διάφορα προγράμματα και αλγόριθμοι. Κοινό τους στοιχείο είναι το στάδιο της εκπαίδευσης και της δοκιμής. Η απόδοσή τους εξαρτάται από την πολυπλοκότητα, την ακρίβεια και την επιλογή των χαρακτηριστικών του σήματος [6]. Ένα από αυτά είναι ο ταξινομητής ήχου Support Vector Machine (SVM) που μπορεί να ταξινομήσει τα δεδομένα σε πέντε τύπους: μουσική, ομιλία, ήχοι περιβάλλοντος, ομιλία συνδυασμένη με μουσική και μουσική συνδυασμένη με ήχους περιβάλλοντος [5]. Άλλες μέθοδοι είναι οι συντελεστές κυψελίδας συχνότητας τήξης Mel-Frequency Cepstrum Coefficients (MFCC), φασματικές συχνότητες γραμμής Line Spectral Frequencies (LSF) και ενέργεια βραχείας διάρκειας Short Time Energy, ορθογώνια αναζήτηση αντιστοίχισης Orthogonal Matching Pursuit [6]. Μια από τις πιο σημαντικές τεχνολογίες αναγνώρισης και ταξινόμησης ήχου είναι τα συνελικτικά νευρωνικά δίκτυα Convolutional Neural Networks (CNN) η οποία συνδυάζει διαφορετικούς τύπους χαρακτηριστικών και στην συνέχεια αξιολογούνται, συγκρίνονται και συγχωνεύονται με στόχο την καλύτερη δυνατή ακρίβεια. Η μεγαλύτερη μελέτη σχετικά με τα CNN έγινε στην ταξινόμηση ήχων ζώων, όπου συγκεντρώθηκαν σύνολο δεδομένων ήχου πτηνών, νυχτερίδων και φαλαινών, έπειτα εκτελέστηκε ένας μεγάλος αριθμός πειραμάτων και συνδυάστηκε με τα τελειοποιημένα CNN [7].



Εικόνα 1. 3 Πνευμονικό ηχητικό σύστημα

1.7 Ανάκτηση πληροφοριών μουσικής MIR

Η ανάκτηση πληροφοριών μουσικής Music Information Retrieval (MIR) είναι ένας τομέας που λαμβάνει όλο και περισσότερο έδαφος τόσο στην ερευνητική κοινότητα όσο και στην μουσική βιομηχανία Αντιμετωπίζει το πρόβλημα της εύρεσης και της ανάκτησης τύπων μουσικής από ένα μεγάλο σύνολο δεδομένων μουσικής. Το MIR μπορεί να ταξινομήσει αυτόματα τη μουσική ανάλογα με το είδος, τη διάθεση, τον καλλιτέχνη, τα όργανα κ.λπ. Παρόλα αυτά, η ταξινόμηση εξακολουθεί να είναι ένα σημαντικό πρόβλημα στον MIR [8].



Εικόνα 1. 4 Ανάκτηση πληροφοριών μουσικής με τη χρήση Deep Learning και Modification

1.8 Μελλοντικές προοπτικές

Λόγω της μεγάλης σημασίας που έχει η ταξινόμηση οι ερευνητές βρίσκονται συνέχεια στην αναζήτηση νέων μεθόδων και έτσι υπάρχει ταχεία ανάπτυξη και συνεχής αναθεώρηση [8].
Μία ολοκληρωμένη μέθοδος ταξινόμησης, πέραν από αυτά τα προβλήματα, πρέπει να περιλαμβάνει την αναγνώριση της διαφοράς ομιλίας και μουσικής, την ανίχνευση του φύλου του ομιλητή και αναγνώριση ειδικών εφέ, ενώ οι περισσότερες τεχνικές ανάλυσης του ήχου επικεντρώνονται στη λύση συγκεκριμένων προβλημάτων, ο στόχος είναι η δημιουργία ενός γενικού πλαισίου [9]. Σκοπεύοντας να αξιολογηθεί η αποτελεσματικότητα της εκάστοτε τεχνικής, διεξήχθησαν πειράματα με όλα αυτά τα προβλήματα ήχου καθώς επίσης με την ανίχνευση στιγμιότυπων σε αθλητικά βίντεο και αναγνώριση μουσικού είδους [10].

1.9 Επίλογος

Στο πρώτο κεφάλαιο δόθηκε ένας ορισμός της κατηγοριοποίησης ως προς τον τρόπο που γίνεται κα την τεχνολογία στην οποία βασίζεται.
Στη συνέχεια, αναλύθηκαν οι πολλαπλοί τρόποι που μπορεί να είναι χρήσιμη αυτή η τεχνολογία σε επιστήμες όπως η ιατρική ή η μηχανολογία. Έγιναν ξεκάθαρα τα οφέλη καθώς και το γεγονός ότι μπορεί να αποβούν σωτήρια.

Αξίζει να σημειωθεί ότι παρόλα αυτά υπάρχουν ακόμη αρκετά προβλήματα και εμπόδια. Το κυριότερο είναι ότι η τεχνολογία δεν έχει εξελιχθεί αρκετά ώστε να μπορεί να ταξινομήσει ήχους οι οποίοι δεν είναι «καθαροί».
Κατά συνέπεια, είναι ανάγκη να υπάρξει περισσότερη μελέτη και εξέλιξη, τόσο στις ήδη υπάρχουσες τεχνολογίες, όπως το SVM, το MFCC, το LCF, το STE, το OMP και το CNN.
Μία τεχνολογία η οποία φαίνεται να κερδίζει όλο και περισσότερη προσοχή και ενδιαφέρον είναι η ανάκτηση πληροφοριών MIR η οποία ακόμα όμως, έχει ακόμα πολλά περιθώρια εξέλιξης στο θέμα της κατηγοριοποίησης.

Είναι γεγονός, ότι εξαιτίας των πολλαπλών πλεονεκτημάτων της τεχνολογίας αυτής, η ερευνητική κοινότητα θα αφιερώνει όλο και περισσότερο χρόνο, ενέργεια και χρηματοδότηση για την όσο το δυνατόν πιο στοχευόμενη αντιμετώπιση των προαναφερθέντων προβλημάτων και την πρόοδό της.

Κεφάλαιο 2: Χαρακτηριστικά ήχων

2.1 Εισαγωγή

Στο δεύτερο κεφάλαιο αναλύονται τα βασικά μουσικά χαρακτηριστικά ανάλογα με τις κατηγορίες τους. Πιο συγκεκριμένα, περιγράφονται οι πέντε κατηγορίες δυναμική, ρυθμός, τέμπο, χροιά και τόνος μέσα από τις ιδιότητες καθεμίας από αυτές. Τέλος, καταγράφονται και τα υψηλού επιπέδου χαρακτηριστικά τα οποία συνοψίζονται στις κατηγορίες δομή και μορφή, προβλέψεις, ομοιότητα και ανάκτηση και εξαγωγή.

2.2 Χαρακτηριστικοί εξαγωγείς

2.2.1 Δυναμικά χαρακτηριστικά

Η συνολική ενέργεια του σήματος

$$\boldsymbol{x_{rms}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} = \sqrt{\frac{x_1^2+x_2^2+\cdots+x_n^2}{n}} \qquad (2.1)$$

❖ **Κατάτμηση (segment)**

Κατάτμηση θεωρείται η ανίχνευση σημείων αλλαγής. Πιο συγκεκριμένα συνίσταται στον εντοπισμό των χρονικών ορίων των τμημάτων, π.χ. εισαγωγή, ρεφρέν και στίχο.

❖ **Χαμηλή ενέργεια (low energy)**

Ο τρόπος για να υπολογίσουμε πόσα κομμάτια του σήματος εμφανίζουν χαμηλή ενέργεια δηλαδή μικρότερη από τον μέσο όρο.

2.2.2 Ρυθμικά χαρακτηριστικά

Η εκτίμηση της ρυθμικότητας στο ηχητικό σήμα μπορεί να παρουσιαστεί χρησιμοποιώντας τους τελεστές που θα παρουσιαστούν παρακάτω.

❖ **Διακύμανση (fluctuation)**

Το σύνολο του φάσματος σε όλες τις ζώνες μας δίνει την περιοδικότητα του σήματος.

❖ **Φάσμα παλμών (beatspectrum)**

Είναι ο τρόπος με τον οποίο αντιλαμβανόμαστε την επανάληψη σε άπειρες κλίμακες.

❖ **Θέσεις έναρξης (onsets)**

Όταν υπάρχουν διαδοχικές εκρήξεις ενέργειας λόγω διαδοχικών παλμών, ο ρυθμός προσδιορίζεται στην καμπύλη έναρξης.

- **Μέση συχνότητα συμβάντων (event density)**

Υπολογίζει το πόσες φορές ξεκινάει μια νότα ανά δευτερόλεπτο.

- **Τέμπο σε παλμούς ανά λεπτό (tempo)**

Μετράει τις περιόδους από την καμπύλη έναρξης και υπολογίζει τον ρυθμό.

- **Μετρική ανάλυση (metre)**

Λεπτομερής περιγραφή του μέτρου σε ένα σύνολο μετρικών επιπέδων.

- **Μετρικό κέντρο και δύναμη (metroid)**

  - **Δυναμικό μετρικό κέντρο:**
    Από την μετρική κεντροειδή καμπύλη μπορούμε να υπολογίσουμε την εξέλιξη της μετρικής δραστηριότητας σε σχέση με τον χρόνο και να βρούμε το κέντρο των επιλεγμένων μετρικών επιπέδων**.**

  - **Δυναμική μετρική δύναμη:**
    Προσδιορίζει αν ο παλμός είναι «καθαρός» και δυνατός ή «αδύναμος» και ασαφής ή μία σύνθεση αυτών**.**

- **Ρυθμική διαύγεια (pulseclarity)**

Υπολογίζει την «καθαρότητα» του ρυθμού υποδεικνύοντας την ισχύ του.

2.2.3 Τέμπο χαρακτηριστικά

Η ποιότητα ενός μουσικού ήχου ή φωνής που ξεχωρίζει ανάλογα με την χροιά και την ένταση ήχου.

- **Διάρκεια επιθέσεων (attacktime)**

Η φάση της επίθεσης αναγνωρίζει ορισμένους χαρακτηρισμούς ηχοχρώματος. Η φάση επίθεσης μπορεί να περιγραφεί πολύ απλά εκτιμώντας τη χρονική διάρκεια της.

- **Μέση κλίση επιθέσεων (attackslope)**

Είναι μία εναλλακτική περιγραφή της φάσης επίθεσης. Οι ήχοι εμφανίζονται στην ίδια κλίμακα με το αρχικό σήμα αλλά εκφρασμένοι σε δευτερόλεπτα.

- **Άλμα επίθεσης (attackleap)**

Υπολογίζεται η διαφορά πλάτους μεταξύ αρχής και τέλους της φάσης επίθεσης.

- **Μείωση της κλίσης (decreaseslope)**

Η φάση απελευθέρωσης αναζητάει το τοπικό ελάχιστο πριν και μετά από κάθε κορυφή μίας νότας. Η περιγραφή της φάσης αυτής έχει να κάνει με την μείωση της κλίσης της.

- **Διάρκεια (duration)**

Μετράται σε δευτερόλεπτα από την φάση επίθεσης μέχρι την φάση απελευθέρωσης.

❖ **Αναλογία πρόσημου-μεταβολών (zerocross)**

Ο θόρυβος ενός σήματος μετράται από το πόσες φορές διασχίζει τον άξονα των x, δηλαδή αλλάζει πρόσημο.

❖ **Συχνότητα αποκοπής (rollof)**

Ενέργεια υψηλής συχνότητας (I)

Ως προεπιλογή έχει οριστεί η αναλογία 0,85. Αναλόγως την ποσότητα σήματος που ανιχνεύεται πάνω από αυτή την αναλογία υπολογίζεται η ποσότητα της υψηλής συχνότητας στο σήμα.

❖ **Φωτεινότητα brightness**

Ενέργεια υψηλής συχνότητας (II)

Σε αυτή την μέθοδο βρίσκουμε την συχνότητα αποκοπής και υπολογίζουμε την ποσότητα της ενέργειας πάνω από αυτή την συχνότητα.

❖ **Κεντροειδές, διακύμανση, λοξότητα, κύρτωση, επιπεδότητα, εντροπία (centroid, spread, skewness, kurtosis, flatness, entropy)**

Αυτοί είναι οι όροι που περιγράφουν την φασματική κατανομή που μπορεί να περιγραφεί με στατικές ροπές.

❖ **Συντελεστές φάσματος συχνότητας (mfcc)**

Το mfcc προσφέρει μία περιγραφή του φασματικού σχήματος του ήχου.

❖ **Τραχύτητα (roughness)**

Αισθητηριακή παραφωνία: μετρώντας τις κορυφές όλων των ζευγών των ημιτονοειδών που είναι κοντά σε συχνότητα μπορούμε να υπολογίσουμε την αισθητηριακή ασυμφωνία ή τραχύτητα.

❖ **Παρατυπία φάσματος (regularity)**

Ο βαθμός παραλλαγής μεταξύ συνεχόμενων κορυφών στο φάσμα.

2.2.4 Χαρακτηριστικά χροιάς

Η ποιότητα ενός ήχου διέπεται από τον ρυθμό των δονήσεων που τον παράγουν. Δηλαδή, κατά πόσο ο τόνος είναι υψηλός ή χαμηλός.

❖ **Εκτίμηση χροιάς (pitch)**

Η χροιά εξάγεται είτε συνεχόμενα είτε ως ξεχωριστά κομμάτια.

❖ **Ψηφιακή διασύνδεση μουσικών οργάνων (midi)**

Κατάτμηση του ήχου ανάλογα με την χροιά που μεταδίδει και μετατροπή του σε αναπαράσταση MIDI.

❖ **Δυσαρμονικότητα (inharmonicity)**

Όσα κομμάτια δεν είναι πολλαπλάσια της βασικής συχνότητας δημιουργούν δυσαρμονικότητα καθώς παρουσιάζεται ενέργεια εκτός του ιδανικού αρμονικού.

2.2.5 Τονικά χαρακτηριστικά

Ο χαρακτήρας ενός μουσικού κομματιού όπως καθορίζεται από την νότα στην οποία παίζεται ή την σχέση των νοτών μίας κλίμακας.

❖ **Χρωμόγραμμα (Chromagram)**

Εναλλακτικά ονομάζεται «προφίλ κλάσης αρμονικής χροιάς». Με αυτό, μπορούμε να δούμε τον τρόπο με τον οποίο κατανέμεται η ενέργεια στην διάρκεια της χροιάς.

❖ **Ισχύς τόνου μουσικής (keystrength)**

Υπολογίζει την ισχύ η οποία πρέπει να είναι μεταξύ -1 και 1 συνδυάζοντας κάθε νότα.

❖ **Μουσικό κλειδί (key)**

Δείχνει μια γενική εκτίμηση του τονικού κέντρου και της σαφήνειας του.

❖ **Λειτουργία (mode)**

Αναγνωρίζει τις εναλλαγές μεταξύ μείζων και ελάσσονος με τιμές μεταξύ -1 και 1.Όσο πλησιάζει το 1 αυξάνονται οι πιθανότητες να είναι μείζων ενώ όσο πλησιάζει το -1 αυξάνονται οι πιθανότητες να είναι ελάσσονα.

❖ **Οπτικοποίηση (keysom)**

Εμφανίζει ένα ψευδο-χρωματικό χάρτη.

❖ **Τονικό κέντρο (tonalcentroid)**

Προβάλει το τονικό κέντρο των έξι διαστάσεων. Σχετίζει τις συγχορδίες σε κύκλους των πέμπτων, των ελασσόνων τρίτων και των μειζόνων τρίτων.

❖ **Λειτουργία ανίχνευσης αρμονικής αλλαγής (hcdf)**

Είναι η ροή του τονικού κέντρου.

2.3 Υψηλού επιπέδου χαρακτηριστικά ήχων

2.3.1 Δομή και μορφή

Χρησιμοποιούνται πιο περίτεχνα εργαλεία ούτως ώστε να γίνονται υψηλότερου επιπέδου αναλύσεις. Πιο συγκεκριμένα, τα ηχητικά αρχεία μπορούν αυτόματα να κατηγοριοποιηθούν σε όμοιες κατηγορίες μέσω της εκτίμησης των χρονικών ασυνεχειών.

❖ **Πίνακας ομοιότητας (similarity matrix)**

Παρουσιάζονται οι ομοιότητες ανάμεσα σε όλα τα ζεύγη των εισαγόμενων δεδομένων.

- **Καμπύλη καινοτομίας (novelty)**

Υποδεικνύει τις πιθανότητες να υπάρχουν μεταβάσεις ανάμεσα σε διαδοχικές καταστάσεις κατά τη διάρκεια του χρόνου. Αυτό φαίνεται από τα ύψη των κορυφών.

- **Κατάτμηση καινοτομίας (segment…"Novelty")**

Δείχνει την χρονική στιγμή που συμβαίνει ασυνέχεια χαρακτηριστικών έτσι ώστε να γίνεται σωστότερη κατάτμηση του ήχου.

2.3.2 Στατιστική

Παρέχει λειτουργίες κα εφαρμογές για να περιγράψουμε, να αναλύσουμε και να υποδείξουμε δεδομένα.

- **Μέσος όρος (mean)**

Δίνει το μέσο όρο των πλαισίων της χαρακτηριστικής συχνότητας. Αν αυτό χωριστεί σε τμήματα το αποτέλεσμα θα είναι μία σειρά σε διαδοχικά τμήματα.

- **Τυπική απόκλιση (standard deviation)**

Επιστρέφει την τυπική απόκλιση στα πλαίσια της χαρακτηριστικής συχνότητας. Η συχνότητα είναι ένας δομημένος πίνακας. Τα δεδομένα εξόδων θα είναι δομημένα με τον ίδιο τρόπο.

- **Διάμεσος (median)**

Επιστρέφει την διάμεσο στα πλαίσια της χαρακτηριστικής συχνότητας.

- **Στατιστικά στοιχεία (stat)**

Μπορεί να χρησιμοποιηθεί σε οποιοδήποτε αντικείμενο και να δώσει τα στατιστικά του στοιχεία σε δομημένη μορφή.

- **Ιστόγραμμα (histo)**

Χρησιμοποιείται σε οποιοδήποτε στοιχείο και δίνει το κατάλληλο ιστόγραμμα. Τα δεδομένα αποθηκεύονται σε θέσεις ίσων αποστάσεων.

- **Αναλογία προσήμου-μεταβολών (zerocross)**

Όπως αναφέρθηκε και στην ενότητα του τέμπο παραπάνω, μετράει τις φορές που το σήμα περνάει τον άξονα x, δηλαδή αλλάζει πρόσημο, και είναι ένδειξη θορύβου.

- **Κεντροειδές (centroid)**

Το κεντροειδές αναφέρεται στο γεωμετρικό κέντρο ενός σχήματος. Αυτή η επεξεργασία μας δίνει το κεντροειδές ενός σήματος.

- **Διάδοση (spread)**

Επαναφέρει την τυπική απόκλιση των δεδομένων.

❖ **Λοξότητα (skewness)**

Απαντάει με τον συντελεστή λοξότητας. Μετράει την συμμετρία της κατανομής. Αν έχει θετική τιμή θεωρείται θετικά λοξή με τιμές μεγαλύτερες από τη μέση ενώ, αν έχει περισσότερες αρνητικές τιμές, είναι αρνητικά λοξή. Μία απόλυτα συμμετρική κατανομή έχει μηδενική λοξότητα.

❖ **Κύρτωση (kurtosis)**

Επιστρέφει την κύρτωση δεδομένων. Δηλαδή, το πόσο συγκεντρωμένα είναι τα δεδομένα γύρω από τη μέση τιμή.

❖ **Επιπεδότητα (flatness)**

Προσδιορίζει αν η κατανομή θα είναι ομαλή ή θα έχει αιχμές.

❖ **Εντροπία (entropy)**

Η εντροπία Shannon παρουσιάζει το κατά πόσο υπάρχουν κυρίαρχες κορυφές σε μια κατανομή. Για παράδειγμα, εάν η καμπύλη παρουσιάζει επιπεδότητα η κατάσταση παρουσιάζει αβεβαιότητα στην έξοδο και η εντροπία είναι μέγιστη. Αντίθετα, εάν εμφανίζεται μια πολύ αιχμηρή κορυφή η κατάσταση εμφανίζει ελάχιστη αβεβαιότητα και ελάχιστη εντροπία.

❖ **Χαρακτηριστικά (features)**

Το σύνολο των χαρακτηριστικών αποθηκεύεται σε έναν πίνακα δομών και το αποτέλεσμα επιστρέφεται μέσω μίας μεταβλητής.

❖ **Στατιστική χαρτογράφηση (map)**

Χαρτογραφεί τη συσχέτιση των ηχητικών εγγραφών που βρίσκονται στο ίδιο σύνολο.

2.3.3 Προβλέψεις

Προβλέπει τα αποτελέσματα της γραμμικής παλινδρόμησης.

❖ **Συναισθήματα (emotion)**

Τα συναισθήματα που προκαλούνται από τη μουσική συνοψίζονται σε δύο κατηγορίες.

Ι.Στα βασικά συναισθήματα τα οποία χωρίζονται σε πέντε τάξεις: χαρά, λύπη, τρυφερότητα, θυμός και φόβος.

ΙΙ. Σε τρείς διαστάσεις: την πρόκληση ενέργειας, την εναλλαγή μεταξύ ευχαρίστησης και δυσφορίας κα την συναισθηματική ένταση.

❖ **Ταξινόμηση (classify)**

Η ταξινόμηση μπορεί να γίνει με τη διασταύρωση είτε ενός χαρακτηριστικού είτε ενός συνόλου χαρακτηριστικών.

❖ **Σύμπλεγμα (cluster)**

Χρησιμοποιεί κοινά στοιχεία διαφόρων ξεχωριστών ακουστικών σημάτων για να δημιουργήσει ένα καινούργιο ενιαίο σήμα.

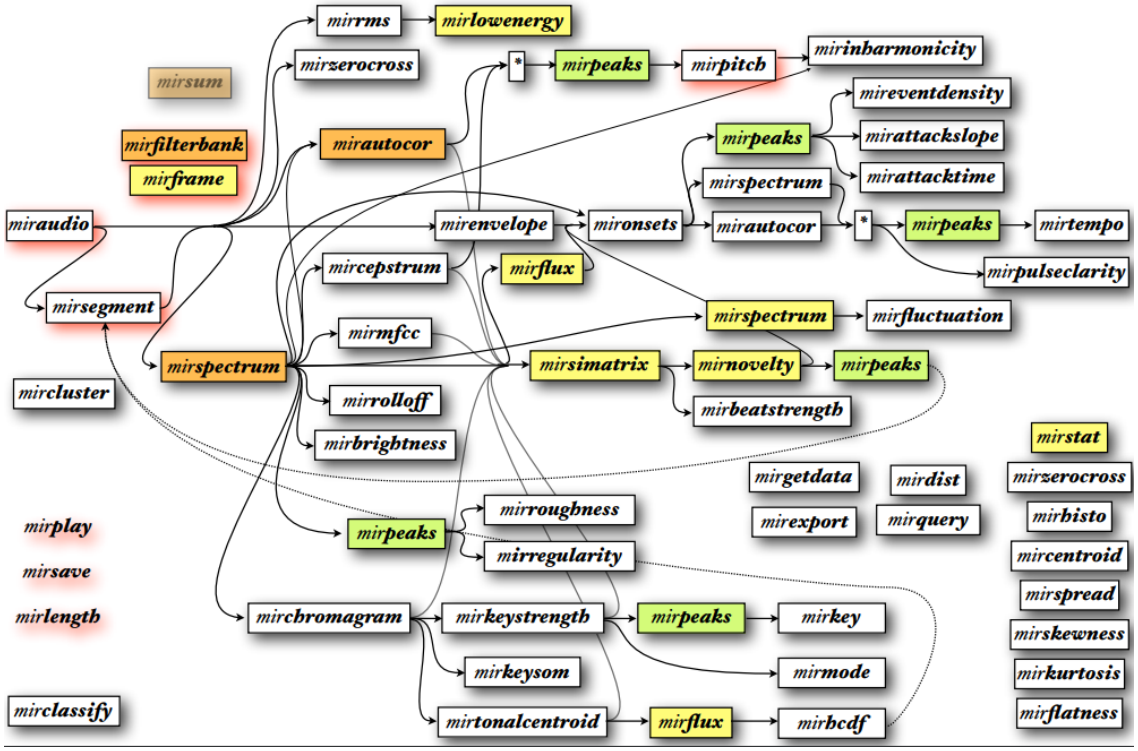
2.3.4 Εξαγωγή

❖ **Επιστροφή αποτελέσματος (getdata)**

Επιστρέφει τα δεδομένα εισόδου σε μία δομή η οποία μπορεί να χρησιμοποιηθεί για υπολογισμούς.

❖ **Εξαγωγή δεδομένων (export)**

Εξάγει όλα τα δεδομένα σε ένα αρχείο κειμένου.



Εικόνα 2. 1 MIRtoolbox Features

2.4 Επίλογος

Στο δεύτερο κεφάλαιο δόθηκαν ορισμοί για τα χαρακτηριστικά ήχων αφού πρώτα κατηγοριοποιήθηκαν κατάλληλα σε μία από τις βασικές κατηγορίες δυναμική, ρυθμός, τέμπο, χροιά και τόνος. Μέσα από αυτή την ανάλυση, μπορέσαμε να εξοικειωθούμε με ορισμούς και έννοιες οι οποίοι θα χρησιμοποιηθούν στα επόμενα κεφάλαια και είναι απαραίτητοι κυρίως στο πρακτικό μέρος αυτής της εργασίας.

Στην κατηγορία της δυναμικής ανήκουν η κατάτμηση και η χαμηλή ενέργεια. Έπειτα, ο ρυθμός περιλαμβάνει την διακύμανση, φάσμα παλμών, τις θέσεις έναρξης, τη μέση συχνότητα συμβάντων, το τέμπο σε παλμούς ανά λεπτό, τη μετρική ανάλυση, το μετρικό κέντρο και δύναμη και τέλος την ρυθμική διαύγεια. Σε ότι αφορά το τέμπο περιγράψαμε την διάρκεια επιθέσεων, την μέση κλίση επιθέσεων, το άλμα επίθεσης, την μείωση της κλίσης, τη διάρκεια, την αναλογία προσήμου-

μεταβολών, την κύλιση, την φωτεινότητα, το κεντροειδές, την εξάπλωση, τη λοξότητα, την κύρτωση, την επιπεδότητα, την εντροπία, τους συντελεστές φάσματος συχνότητας, την τραχύτητα και την παρατυπία φάσματος. Επίσης, στην χροιά περιλαμβάνονται η εκτίμηση χροιάς, η ψηφιακή διασύνδεση μουσικών οργάνων και η δυσαρμονικότητα. Τέλος, τα χαρακτηριστικά που περιλαμβάνει ο τόνος είναι το χρωμόγραμμα, η ισχύς τόνου μουσικής, το μουσικό κλειδί, η λειτουργία, η οπτικοποίηση, το τονικό κέντρο, η λειτουργία ανίχνευσης αρμονικής αλλαγής και η κατάτμηση με την τεχνική HCDF.

Επιπρόσθετα, αναλύθηκαν τα υψηλού επιπέδου χαρακτηριστικά ήχων. Ξεκινώντας από τη δομή και μορφή η οποία εμπεριέχει τον πίνακα ομοιότητας, την καμπύλη καινοτομίας και την κατάτμηση καινοτομίας. Επιπλέον, η στατιστική εκφράζεται μέσω του μέσου όρου, της τυπικής απόκλισης, την διαμέσου, των στατιστικών στοιχείων, του ιστογράμματος, της αναλογίας πρόσημου μεταβολών, του κεντροειδούς, της διάδοσης, της λοξότητας, της αιχμής, της επιπεδότητας, της εντροπίας, των χαρακτηριστικών και της στατιστικής χαρτογράφησης.

Κατόπιν, σημαντικό ρόλο παίζει το κομμάτι των προβλέψεων. Λειτουργεί με βάση τις δύο κατηγορίες συναισθημάτων, τα βασικά που είναι η χαρά, η λύπη, η τρυφερότητα, ο θυμός και ο φόβος και των τριών διαστάσεων που είναι η πρόκληση ενέργειας, η εναλλαγή μεταξύ ευχαρίστησης και δυσφορίας και η συναισθηματική ένταση. Ακόμη, αξίζει να αναφερθούν η ταξινόμηση και το σύμπλεγμα. Η ομοιότητα και η ανάκτηση είναι το επόμενο στάδιο. Εκεί, συναντώνται η απόσταση και το ερώτημα. Το τελευταίο στάδιο είναι η εξαγωγή όπου εκεί γίνεται η επιστροφή αποτελέσματος και η εξαγωγή δεδομένων.

Κεφάλαιο 3: Μοντέλα-αλγόριθμοι κατηγοριοποίησης

3.1 Εισαγωγή

Στο τρίτο κεφάλαιο δίνεται αρχικά ένας ορισμός της μουσικής και μια σύντομη αναφορά στην σημασία της από την αρχαιότητα. Στη συνέχεια, αναφέρονται και αναλύονται τα βασικά είδη μουσικής και περιγράφονται με βάση τα γνωρίσματα που τα διέπουν. Επίσης, διατυπώνεται ο τρόπος με τον οποίο λειτουργούν οι αλγόριθμοι και αναπτύσσονται κάποια από τα πιο δημοφιλή νευρωνικά δίκτυα.

3.2 Ορισμός της μουσικής

Ως μουσική ορίζεται η τέχνη που βασίζεται στην οργάνωση ήχων και σκοπός της είναι η σύνθεση, εκτέλεση και ακρόαση / λήψη ενός έργου. Το όνομα της προέρχεται από τις Μούσες της αρχαίας ελληνικής μυθολογίας. Στην αρχαία Ελλάδα η μουσική ήταν άρρηκτα συνδεδεμένη με την ποίηση, την μελωδία και τον χορό ως μέλη της τέχνης του θεάτρου. Μέχρι και σήμερα η μουσική είναι η τέχνη που ολοκληρώνει τις σκέψεις, τα συναισθήματα και τις ψυχικές καταστάσεις του ανθρώπου [11].

3.3 Είδη μουσικής

Μία εφαρμογή της ταξινόμησης η οποία είναι πολύ δημοφιλής αλλά και αρκετά δύσκολη είναι αυτή της κατηγοριοποίησης της μουσικής. Όπως, και με όλες τις κατηγορίες ήχου που έχουν αναφερθεί σε αυτή την εργασία, έτσι και για την μουσική γίνεται ανάλυση περιεχομένου ώστε να ταξινομηθεί κατάλληλα. Το πρόβλημα προκύπτει διότι πολλά μουσικά κομμάτια έχουν στοιχεία από πολλά είδη μουσικής.

3.3.1 Παραδοσιακή μουσική

Γνωστή και ως «δημοτική» μουσική αναφέρεται σε τοπικές συνθέσεις ελλαδικών περιοχών. Οι δημιουργοί τους είναι άγνωστοι καθώς η ιστορία τους ξεκινάει πάνω από έναν αιώνα πριν. Στην πραγματικότητα είναι ένα είδος το οποίο περιλαμβάνει πολλά άλλα είδη καθώς κάθε περιοχή χαρακτηρίζεται από τους δικούς της ρυθμούς, μελωδίες και μουσικά όργανα. Επίσης ξεχωρίζουν και βάση του περιεχομένου για παράδειγμα του γάμου ή της ξενιτιάς [12].

Εικόνα 3. 1 Η ελληνική παραδοσιακή μουσική

3.3.2 Λαϊκή

Λαϊκό ορίζεται το είδος μουσικής που διαδέχτηκε το ρεμπέτικο τις δεκαετίες 1950 έως 1970. Σε αντίθεση με το δημοτικό τραγούδι που εμφανίζει τοπική ποικιλομορφία, το λαϊκό έχει πανελλήνια ομοιομορφία. Ενώ αρχικά ήταν δημοφιλές μόνο στην εργατική τάξη πολύ σύντομα μεταφέρθηκε και στις τάξεις των πλουσίων. Έτσι χωρίστηκε σε κατηγορίες όπως Το ελαφρολαϊκό και βαρύ λαϊκό. Ωστόσο για πολλά χρόνια, όπως και το ρεμπέτικο, αντιμετωπίστηκε με καχυποψία, εχθρότητα και συκοφαντία [13].

Εικόνα 3. 2 Ελληνική λαϊκή μουσική τις δεκαετίες ΄50-΄60

3.3.3 Ροκ

Το ρόκ στην Ελλάδα ξεκίνησε από παιδιά και νέους που ήθελαν να ξεφύγουν από το πολιτικό τραγούδι και να μοιάσουν στους αγαπημένους τους ρόκ σταρ του εξωτερικού όπως ο Morrison, ο Hendrix και η Joplin. Μέσα από αυτά τα τραγούδια εξέφραζαν τα κοινωνικά προβλήματα και τις ανησυχίες της δικής τους γενιάς. Ο μεγαλύτερος εκπρόσωπος του είδους στην Ελλάδα, είναι ο Παύλος Σιδηρόπουλος του οποίου οι στίχοι εξερευνούν το πρόβλημα των ναρκωτικών και η μετάδοση του δίσκου του στο ραδιόφωνο απαγορεύεται [12].

Εικόνα 3. 3 Παύλος Σιδηρόπουλος, Έλληνας ρόκερ

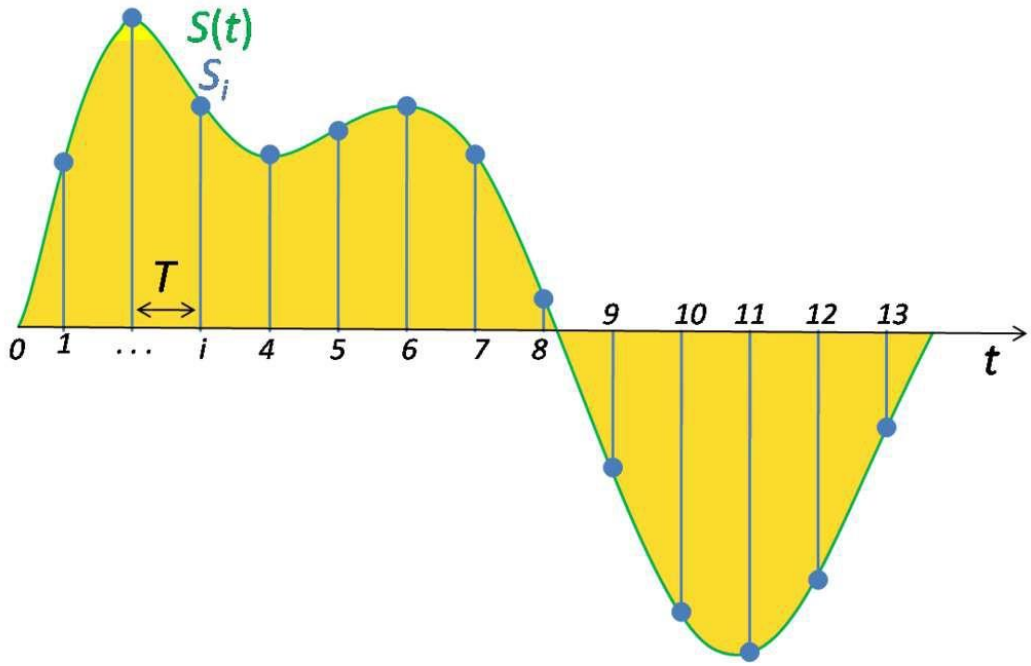
3.3.4 Ραπ

Το ελληνικό ραπ επηρεασμένο από την αμερικάνικη κουλτούρα είναι η εξέλιξη του χιπ χοπ. Περιλαμβάνει πολλές υποκατηγορίες όπως DJing (**D**isc **J**ockey) και το MCing (**M**aster of **C**eremony) [13]. Το όνομά της προέρχεται από τις λέξεις **R**hythmic **A**merica **P**oetry και περιγράφει τον τρόπο ρυθμικής ομιλίας με τον οποίο εκφράζεται αυτό το είδος μουσικής [14].

Εικόνα 3. 4 Active Member, ελληνικό χιπ-χοπ συγκρότημα

3.4 Ψηφιοποίηση του ήχου

Το πρώτο βήμα για την ταξινόμηση του ήχου είναι η ψηφιοποίησή του. Για να το καταφέρουμε αυτό πρέπει να μετατρέψουμε το σήμα σε ψηφιακό, ώστε να μπορεί να επεξεργαστεί από τους αλγόριθμους. Αυτό το καταφέρνουμε μετρώντας το πλάτος του ήχου σε σταθερά χρονικά διαστήματα. Αυτές οι μετρήσεις ονομάζονται δείγματα. Ένας συνηθισμένος ρυθμός δειγματοληψίας είναι 44.100 δείγματα ανά δευτερόλεπτα, το οποίο σημαίνει ότι από ένα ηχητικό αρχείο δέκα δευτερολέπτων μπορούμε να πάρουμε 441.000 δείγματα [1].



Εικόνα 3. 5 Δείγμα μετρήσεων σε τακτά χρονικά διαστήματα

3.5 Αλγόριθμοι

Κάθε καρέ ενός τραγουδιού μπορεί να ανήκει σε μια από τις τρείς κατηγορίες μη φωνητική, φωνητική ή σιωπή. Αυτές οι κατηγορίες βοηθούν στο να χωριστεί ένα μουσικό κομμάτι σε ενότητες. Για παράδειγμα, η εισαγωγή και το τέλος ενός μουσικού κομματιού είναι συνήθως μη φωνητικά, ενώ το χρονικό διάστημα που τραγουδιούνται οι στίχοι είναι φωνητικά. Η σιωπή, αποτελεί ένα μέρος του κομματιού που μπορεί να αγνοηθεί κατά την κατηγοριοποίηση και συνήθως αφορά την αρχή ή το τέλος του αρχείου ήχου.

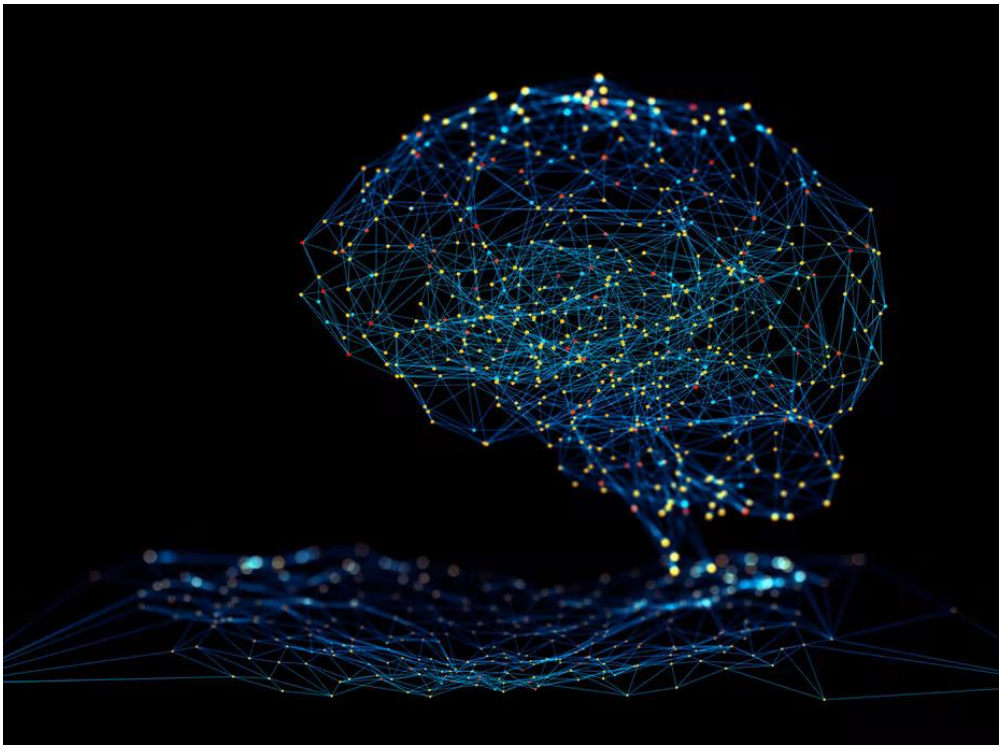
Οι αλγόριθμοι που υπάρχουν μέχρι τώρα, υπολογίζουν τα χαρακτηριστικά με βάση το φάσμα συχνοτήτων. Τα αποτελέσματα τους όμως δεν είναι ικανοποιητικά, πιθανότατα, διότι δίνεται ίση σημασία σε όλα τα κομμάτια του φάσματος. Επίσης, έχει παρατηρηθεί, ότι η μέγιστη φωνητική συχνότητα κατά την ερμηνεία ενός τραγουδιού είναι 3000Hz. Επομένως, έγινε μια προσπάθεια με το να ενισχυθεί το μουσικό εύρος από 300 έως 3000Hz. Ωστόσο, ούτε αυτή η προσέγγιση λειτούργησε καθώς επηρέασαν οι υψηλότερες συχνότητες του φάσματος. Ως εκ τούτου, εφαρμόστηκε ένα φίλτρο διέλευσης ζώνης με χαμηλότερη συχνότητα αποκοπής στα 400Hz και διαφορετικές υψηλότερες συχνότητες αποκοπής όπως 6000Hz, 3000Hz και 2000Hz. Ο κύριος λόγος που επιλέχθηκε χαμηλότερη συχνότητα αποκοπής 400Hz είναι για να αποφευχθεί η χρήση βασικών συχνοτήτων που εμφανίζονται ομοιόμορφα σε όλο το φάσμα. Η βέλτιστη υψηλότερη συχνότητα αποκοπής είναι τα 3000Hz [15].

3.6 Νευρωνικά δίκτυα

Τα νευρωνικά δίκτυα είναι ένας τομέας ο οποίος κερδίζει όλο και περισσότερο έδαφος στην επιστημονική κοινότητα λόγω της ανάπτυξης υλικών και της υψηλής διαθεσιμότητας δεδομένων. Ένα από τα πιο γνωστά παραδείγματα, είναι η ακουστική μοντελοποίηση για αυτόματη αναγνώριση ομιλίας, Automated Speech Recognition (ASR).

Τα συνελικτικά νευρωνικά δίκτυα Convolutional Neural Networks (CNN) συνήθως συνδέονται με την αναγνώριση και ταξινόμηση εικόνας ωστόσο, εφαρμόζεται και στην ταξινόμηση ήχου. Στο CNN εφαρμόζονται εργασίες με ταξινόμηση ομιλίας και μουσικής σημαντικά βελτιωμένες από την μηχανή διανυσματικής υποστήριξης Support Vector Machine (SVM) και τη γενικευμένη μέθοδο στιγμών Generalized Method of Moments (GMM).

Τα επαναλαμβανόμενα νευρωνικά δίκτυα που ασχολούνται με τις χρονικές ακολουθίες πληροφοριών μπορούν και να μοντελοποιούν χρονικές εξαρτήσεις εισάγοντας τον βρόγχο ανάδρασης μεταξύ εισόδου κα εξόδου. Τα δίκτυα μακράς βραχύχρονης μνήμης Long Short-Term Memory (LSTM) είναι ένα είδος επαναλαμβανόμενου νευρωνικού δικτύου Recurrent Neural Network (RNN) που εισάγει την έννοια της μνήμης κυψέλης. Η κυψέλη μαθαίνει, κρατά και ξεχνά πληροφορίες. Έτσι, τα δίκτυα LSTM είναι ένα πολύ ισχυρό εργαλείο, το οποίο μπορεί να γίνει ακόμα πιο ισχυρό συνδυάζοντας δύο δίκτυα LSTM και δημιουργώντας ένα αμφίδρομο δίκτυο Bidirectional (BLSTM). Το ένα δίκτυο διεργάζεται την διαδρομή προς τα εμπρός και το άλλο την διαδρομή προς τα πίσω. Αυτά τα δίκτυα έχουν εφαρμοστεί επιτυχώς σε διάφορες εργασίες μοντελοποίησης [16].



Εικόνα 3. 6 Νευρωνικά δίκτυά

3.7 Επίλογος

Στο κεφάλαιο αυτό αναφερόμαστε στη μουσική κα στην σημασία της από την αρχαιότητα ως σήμερα. Έπειτα, περιγράφονται τα είδη της μουσικής που έχουν επιλεχθεί για τους σκοπούς της εργασίας αυτής. Τα είδη αυτά είναι η παραδοσιακή ή αλλιώς δημοτική μουσική η οποία ξεχωρίζει αναλόγως με την περιοχή στην οποία γεννήθηκε. Στη συνέχεια, αναλύθηκε το λαϊκό τραγούδι το οποίο είναι και το πιο χαρακτηριστικό είδος ελληνικής μουσικής. Η ροκ μουσική παρόλο που δεν ξεκίνησε στην Ελλάδα έπαιξε και αυτή σημαντικό ρόλο στην ανατροφή μιας ολόκληρης γενιάς. Κάτι παρόμοιο συνέβη και με την ραπ μουσική, η οποία παραμένει ένα από δημοφιλέστερα είδη μουσικής.

Παρακάτω, περιγράφηκε συνοπτικά η διαδικασία με την οποία γίνεται η ψηφιοποίηση του ήχου. Επιπλέον, παρουσιάζεται ο τρόπος με τον οποίο λειτουργούν πολλοί αλγόριθμοι χωρίζοντας τα κομμάτια σε μη φωνητικά, φωνητικά και σιωπή. Ο υπολογισμός των χαρακτηριστικών βάση συχνοτήτων οδήγησε στο συμπέρασμα ότι η βέλτιστη συχνότητα αποκοπής είναι τα 3000Hz. Τέλος, επεξηγούνται τα νευρωνικά δίκτυα, τα οποία, ενώ συνήθως χρησιμοποιούνται στην ταξινόμηση εικόνας, μπορούν να εφαρμοστούν και στην ταξινόμηση ήχου, ομιλίας και μουσικής. Τα πιο διαδεδομένα νευρωνικά δίκτυα είναι το CNN, το SVM και το RNN.

Κεφάλαιο 4: Παρουσίαση του προβλήματος

4.1 Εισαγωγή

Στο παρακάτω κεφάλαιο αναλύονται διεξοδικά τα βήματα και κατ' επέκταση το αποτέλεσμα του εν λόγω πειράματος. Επίσης, περιγράφονται σε θεωρητικό επίπεδο η διαδικασία η οποία ακολουθήθηκε, τα βήματα που εκτελέστηκαν καθώς και η μεθοδολογία.

4.2 Το πρόβλημα της κατηγοριοποίησης

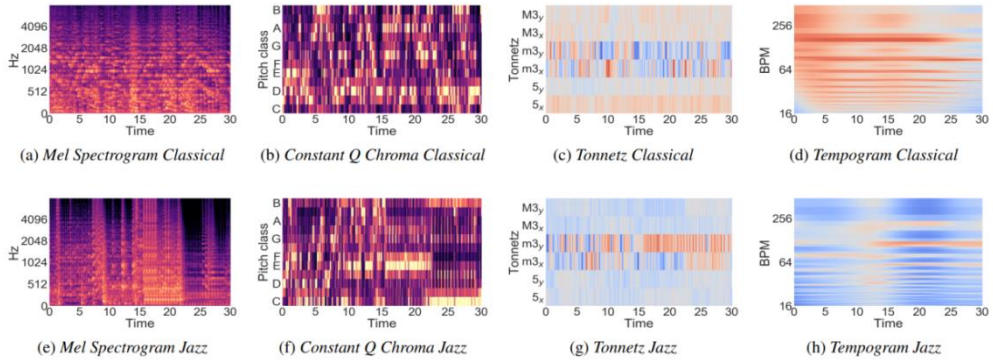
Η μουσική κατηγοριοποιείται βάσει ενός συνόλου κανόνων που αφορούν τον ήχο. Το είδος της μουσικής είναι υποκειμενικό από άνθρωπο σε άνθρωπο και μπορεί να είναι ασαφές. Επιπλέον, ένα μουσικό αρχείο μπορεί να αντιστοιχεί σε παραπάνω από ένα είδος καθώς θα χρησιμοποιεί στοιχεία από παραπάνω από μία κατηγορία. Επομένως, η κατηγοριοποίηση δεν μπορεί να είναι απόλυτα ακριβής αλλά μόνο μία προσέγγιση.

Η ταξινόμηση των μουσικών ειδών μπορεί να γίνει βάσει των ήδη καθορισμένων χαρακτηριστικών. Αυτό λέγεται εποπτευόμενη μάθηση. Μία άλλη προσέγγιση, είναι η μη εποπτευόμενη μάθηση. Με αυτή την προσέγγιση, αναλύονται τα τραγούδια εξετάζοντας τα χαρακτηριστικά τους και τα ταξινομεί σε ομάδες με βάση τις ομοιότητες τους [1].

Η ανάκτηση πληροφοριών μουσικής (MIR) είναι ένα πεδίο το οποίο αφορά την ανάλυση μουσικού περιεχομένου συνδυάζοντας την επεξεργασία σήματος, την μηχανική μάθηση και την θεωρία της μουσικής. Η μέθοδος αυτή δίνει την δυνατότητα στους αλγόριθμους να κατανοήσουν και να επεξεργαστούν τα μουσικά δεδομένα. Η αναγνώριση μουσικού είδους Music Genre Recognition (MGR) είναι ένα πολύ σημαντικό υποπεδίο της MIR. Το μουσικό είδος ορίζεται ως ένα εκφραστικό στυλ το οποίο περιλαμβάνει ορχηστρικούς και φωνητικούς τόνους με δομημένο τρόπο. Η αυτόματη αναγνώριση του είδους είναι ένα πολύ ενδιαφέρον πρόβλημα στα πλαίσια της MIR καθώς βοηθάει στην σύσταση προτεινόμενων τραγουδιών βάση προηγούμενων προτιμήσεων και στην οργάνωση μουσικών βάσεων δεδομένων [17].

4.3 Σχετικές εργασίες

Η πρώτη σημαντική εργασία στην αναγνώριση μουσικού είδους έγινε από τους Tzanetaki and Cook. Τα χαρακτηριστικά που βασίζονται στον ρυθμό κα την χροιά προτάθηκε να ταξινομηθούν χρησιμοποιώντας έναν συνδυασμό του αλγόριθμου Gaussian (GMM) και του K-πλησιέστερου γείτονα (KNN). Η χρήση μηχανών υποστήριξης διανυσμάτων προτάθηκε από τους Xu κ.ά. Οι Costa κ.ά. υπέδειξαν τα χαρακτηριστικά φασματογράφων ως προσέγγιση.



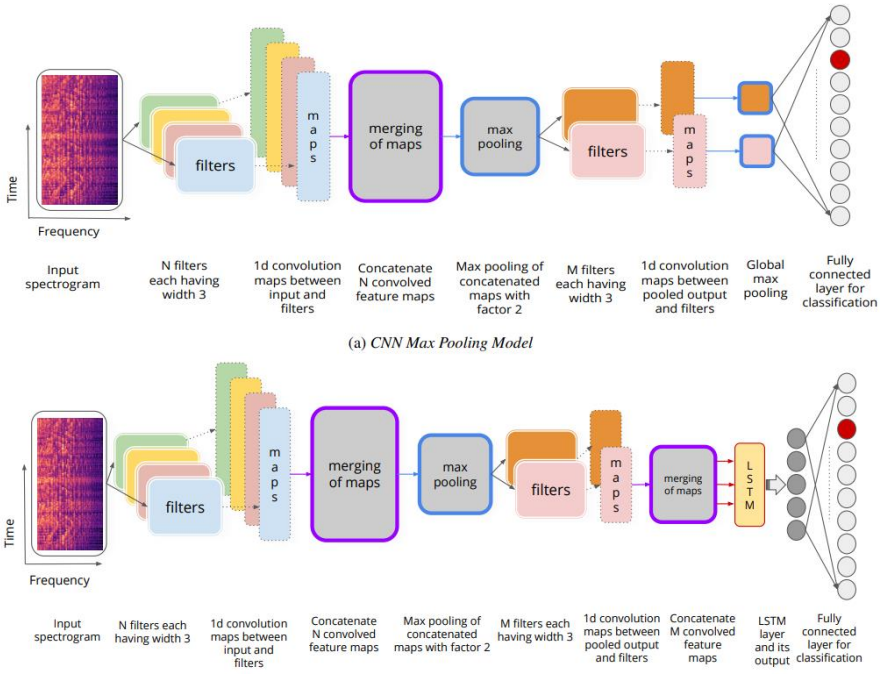
Εικόνα 4. 1 Ιδιαίτερα φασματικά και ρυθμικά χαρακτηριστικά δύο τραγουδιών που ανήκουν στο είδος «Κλασικά» και «Τζαζ». Φασματογράφημα Mel και το Constant Q Chroma είναι χαρακτηριστικά φασματικού τομέα, ενώ το Tonnetz και τοTempogram είναι χαρακτηριστικά τομέα ρυθμό

Ένα μεγάλο κομμάτι ανθρώπων που ασχολήθηκαν με το NGR χρησιμοποίησαν τεχνητά νευρωνικά δίκτυα. Η NGR εκμεταλλεύτηκε προς όφελος της το πεδίο της αναγνώρισης ομιλίας καθώς και τα δύο βασίζονται στην ανάλυση χαρακτηριστικών. Επίσης, υπάρχουν μέθοδοι που πρώτα ανακτώνται η μελωδία ή οι συγχορδίες και έπειτα αυτά τα δεδομένα χρησιμοποιούνται για την μηχανική εκμάθηση. Δεν μελετάται ο ήχος απευθείας αλλά η σύνθεση του. Ο JIANG και η ομάδα του έβγαλαν ορθά αποτελέσματα με την χρήση φασματικής αντίθεσης με βάση την οκτάβα Octave-based Spectral Contrast (OSC). Αργότερα αυτή η μέθοδος εξελίχθηκε σε φασματική αντίθεση διαμόρφωσης με βάση την οκτάβα Octave-based Modulation Spectral Contrast (OMSC) από τον C.-H.Lee και την ομάδα του [17].

4.3.1 Αναγνώριση είδους μουσικής με χρήση νευρωνικών δικτύων και εκμάθηση μεταφοράς

Μία εργασία από το Ινδικό Ινστιτούτο Τεχνολογίας στην Πάτνα εστίασε στην βάση δεδομένων GTZAN η οποία μελετάται στο πλαίσιο της NGR. Αυτή η βάση δεδομένων περιέχει δέκα διαφορετικά είδη τα οποία είναι blues, κλασσική, country, hip-hop, jazz, metal, pop, reggae και rock. Προκειμένου να αναγνωριστεί το είδος της μουσικής πρώτα εξασκούνται τα μοντέλα νευρωνικών δικτύων σε μία σειρά εξαγόμενων φασματικών και ρυθμικών χαρακτηριστικών. Επίσης, χρησιμοποιείται ένα σύστημα μεταφοράς μάθησης το οποίο εξάγει τα αξιοσημείωτα χαρακτηριστικά των τραγουδιών. Στη συνέχεια, ένα πολυστρωματικό δίκτυο εκπαιδεύεται στα χαρακτηριστικά αυτά για να προβλέψει τα είδη. Τέλος, οι προβλέψεις των διάφορων μοντέλων συνδυάζονται χρησιμοποιώντας πλειοψηφική ψηφοφορία.

Ένα ποικίλο σύνολο χαρακτηριστικών φασματικών και ρυθμικών τομέων εξάγονται από ακατέργαστα μουσικά Wav σήματα. Από το σύνολο των χαρακτηριστικών των NNETZ και TEMPOGRAM είναι ρυθμικά χαρακτηριστικά ενώ τα υπόλοιπα τα οποία είναι Mel Spectogram, Mel Cepstral, Delta and Double Delta Coefficients, Delta Coefficients, Double Delta Coefficients, Energy Normalized Chromagram, Constant Q Chromagram, Short Time Fourier Transform (STFT) Chromagram.Τα μουσικά δεδομένα στην βάση δεδομένων GTZAN δειγματίζονται στα 22.000 Hz και είναι περίπου διάρκειας 30 δευτερολέπτων το οποίο έχει ως αποτέλεσμα περίπου 661.500 δείγματα. Υπολογίζοντας τα χαρακτηριστικά για κάθε παράθυρο ολίσθησης με 2048 δείγματα με μετατόπιση 1024 δειγμάτων. Ενισχύουμε με τον κατάλληλο αριθμό μηδενικών στο τέλος έτσι ώστε να υπάρχει ένα σύνολο από 661.500/1024=646 παράθυρα και κάθε τραγούδι να αναπαριστάται από (646, k) διαστατικό πλέγμα χαρακτηριστικών. Η ακριβής επιλογή του k εξαρτάται από το χαρακτηριστικό που υπολογίζεται. Επίσης, στην συγκεκριμένη εργασία χρησιμοποιήθηκαν παραλλαγές των CNN και CNN-LSTM μοντέλων για την πρόβλεψη μουσικών ειδών. Χρησιμοποιήθηκαν μονοδιάστατες συνελίξεις στα μοντέλα τους. Εδώ, τα εξαγόμενα χαρακτηριστικά έχουν διαστάσεις των (646, k) και τα συνελικτικά φίλτρα έχουν διαστάσεις των (3, k). Η μονοδιάστατη λειτουργία συνέλιξης εκτελείται ολισθαίνοντας τα φίλτρα πάνω από τα 646 χρονικά βήματα με παράθυρο. Η λειτουργία αυτή συμβολίζεται ως συνέλιξη 1D επειδή τα συνελικτικά φίλτρα και τα χαρακτηριστικά έχουν ίδιο μήκος και ως εκ τούτου η ολίσθηση των φίλτρων εκτελείται μόνο κατά το πλάτος (χρονική διάσταση) των χαρακτηριστικών. Συνολικά εφαρμόστηκαν τέσσερα διαφορετικά μοντέλα CNN και CNN-LSTM σε όλα τα εξαγόμενα δισδιάστατα χαρακτηριστικά ξεχωριστά ώστε να προβλεφτεί το είδος της μουσικής. Η δομή αυτών των μοντέλων περιγράφεται στο σχήμα 2.1 .



Εικόνα 4. 2 α) CNN Max Pooling μοντέλο, β) CNN Max Pooling LSTM μοντέλο

Για τα δύο διαφορετικά είδη μονοδιάστατων διανυσμάτων χρησιμοποιούμε δύο ξεχωριστά πολυστρωματικά αντίληπτρα MultiLayer Perceptron (MLP) μοντέλα πρόβλεψης είδους.

Η βάση δεδομένων GTZAN περιέχει 1000 ηχητικά κομμάτια καθένα από τα οποία είναι διάρκειας 30 δευτερολέπτων. Όλα τα κομμάτια είναι μονοφωνικού ήχου 16-bit, 22.050 Hz αρχεία σε μορφή Wav. Περιέχει δέκα είδη τραγουδιών και κάθε είδος αντιπροσωπεύεται από εκατό κομμάτια. Τα μοντέλα αυτά αξιολογήθηκαν στο πλαίσιο ταξινόμησης δέκα κατηγοριών. Τα πειράματα εκτελέστηκαν δέκα φορές με ρύθμιση διασταυρωμένης επιβεβαίωσης. Προκειμένου να διατηρηθεί ομοιόμορφη κατανομή μουσικών ειδών τοποθετήθηκαν ογδόντα τραγούδια από κάθε είδος σε διαχωρισμό ακολουθίας και είκοσι τραγούδια από κάθε είδος σε διαχωρισμό επιβεβαίωσης. Το αποτέλεσμα μέσης ακρίβειας των μοντέλων αυτών αναφέρεται στο σχήμα 2.2 .

Features & Models	CNN Max Pooling	CNN Max Pooling LSTM	CNN Average Pooling	CNN Average Pooling LSTM	Multilayer Perceptron
Mel Spectrogram	**83.0**	73.6	**82.5**	75.7	-
Mel Coefficients	80.2	**79.0**	81.6	**80.5**	-
Delta Mel Coefficients	70.4	77.2	74.5	77.0	-
Double Delta Mel Coefficients	72.1	72.9	72.1	76.5	-
Energy Normalized Chromagram	45.7	34.5	43.0	36.2	-
Constant Q Chromagram	60.0	49.4	57.5	45.6	-
STFT Chromagram	62.8	52.5	63.4	53.7	-
Tonnetz Features	50.2	53.5	51.0	55.8	-
Tempogram Features	41.5	42.0	41.6	43.3	-
Averaged Signal Features	-	-	-	-	77.1
Transfer Learning Features	-	-	-	-	**85.5**

Εικόνα 4. 3 Μέση δεκαπλάσια βαθμολογία ακρίβειας διασταυρούμενης επικύρωσης για διαφορετικά χαρακτηριστικά και μοντέλα

Το πρώτο συμπέρασμα που παρατηρείται είναι ότι λαμβάνεται καλύτερο αποτέλεσμα από το NLP όταν χρησιμοποιούνται οι δυνατότητες εκμάθησης μεταφοράς μουσικής. Τα αποτελέσματα αυτά ήταν αναμενόμενα καθώς το πρωτότυπο σύστημα εκπαιδεύτηκε από την βάση δεδομένων χιλίων τραγουδιών (million song dataset) η οποία περιέχει ένα πολύ μεγάλο σύνολο ετικετών που αφορούν διάφορες πτυχές της μουσικής όπως η διάθεση, οι δεκαετίες, τα όργανα και, φυσικά, το είδος. Επιπλέον, πραγματοποιήθηκαν πειράματα με περισσότερη λεπτομέρεια, ώστε να παραχθούν καλύτερα αποτελέσματα. Παρατηρήθηκε ότι τα χαρακτηριστικά φασματογράμματος mel παράγουν καλύτερα αποτελέσματα στα μοντέλα CNN Max Pooling και CNN Average Pooling, ενώ οι συντελεστές mel παράγουν καλύτερα αποτελέσματα για τα μοντέλα CNN Max Pooling LSTM και CNN Average Pooling LST [17].

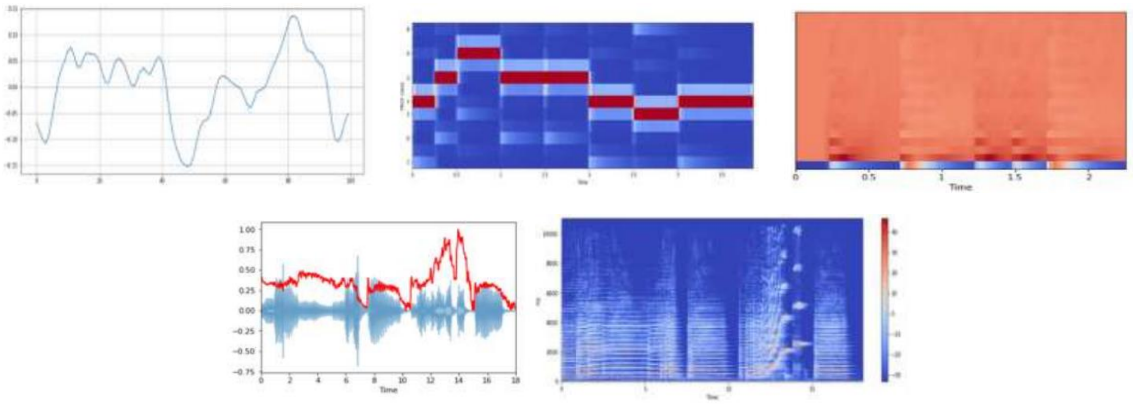
4.3.2 Αναγνώριση μουσικού είδους και ταξινόμηση

Ερευνητές του τμήματος μηχανικών υπολογιστών του Ινστιτούτου Μηχανικής Xavier της Ινδίας είχαν μια άλλη προσέγγιση. Προσπάθησαν να δημιουργήσουν ένα μοντέλο που όχι μόνο θα ταξινομούσε τα μουσικά είδη, αλλά θα έβγαζε ως αποτέλεσμα και το ποσοστό κάθε μουσικού είδους που περίεχε ένα τραγούδι. Παραδείγματος χάρη, εάν ένα τραγούδι χρησιμοποιεί ροκ και ποπ είδος, το αποτέλεσμα που θα εξαχθεί θα είναι αντίστοιχο ποσοστό των δύο ειδών ενώ για τα υπόλοιπα είδη θα δείχνει 0%. Αν ένα τραγούδι είναι αποκλειστικά του είδους blues το αποτέλεσμα θα είναι 100% το αντίστοιχο είδος. Παράλληλα, οι ερευνητές έθεσαν και τις εξής προϋποθέσεις για την επιτυχία της συνεργασίας τους. Καταρχάς, το έργο και το μοντέλο έπρεπε να δημιουργηθούν με μηδενική δαπάνη χρημάτων. Επίσης, το μοντέλο που χρησιμοποιήθηκε για την ταξινόμηση όφειλε να χρησιμοποιεί πολύ λίγη μνήμη και να λειτουργεί με την ελάχιστη δυνατή RAM και επεξεργαστική ισχύ. Τέλος, το μοντέλο έπρεπε να διαμορφωθεί και να δοκιμαστεί με βάση τις παραπάνω συνθήκες και η ακρίβεια του αποτελέσματος να είναι αποδεκτή.

Όπως αναφέρθηκε, το πρώτο βήμα ήταν να επιλεχθεί μια βάση δεδομένων, η οποία θα κατανάλωνε την ελάχιστη δυνατή μνήμη και θα ήταν κατάλληλη για το μοντέλο τους. Για αυτό τον σκοπό αυτό, αναλύθηκαν διάφορες βάσεις δεδομένων οι οποίες βρέθηκαν διαθέσιμες δωρεάν, όπως για παράδειγμα η βάση δεδομένων Free Music Archive (FMA), η βάση δεδομένων χιλίων τραγουδιών ή η βάση δεδομένων GTZAN, κ.λπ. Ταυτόχρονα, προσπάθησαν να δημιουργήσουν μια δική τους βάση δεδομένων ωστόσο η προσπάθεια αυτή απαιτούσε πολύ χρόνο και κόπο. Έτσι, κατέληξαν να χρησιμοποιήσουν την βάση δεδομένων GTZAN. Όπως αναφέρθηκε σε προηγούμενη ενότητα, η βάση δεδομένων αυτή περιέχει χίλια κομμάτια των 30 δευτερολέπτων με κάθε είδος να απαρτίζεται από εκατό τραγούδια το καθένα. Η βάση δεδομένων αυτή είναι περίπου 1,2 Gbps. Αυτό, την κατέστησε κατάλληλη για το συγκεκριμένο μοντέλο, διότι προσφέρει γρήγορες λειτουργίες και επίσης είναι δωρεάν.

Τα αρχεία ήχου στην βάση δεδομένων GTZAN είναι σε μορφή .eu, αλλά επειδή χρησιμοποιήθηκε η python για να δημιουργηθεί το μοντέλο, έπρεπε να γίνει μετατροπή σε μορφή Wav η οποία είναι συμβατή με την συγκεκριμένη γλώσσα προγραμματισμού. Παρόλο που υπάρχουν πολλοί διαφορετικοί τρόποι προσέγγισης για την κατηγοριοποίηση τραγουδιών βάση του είδους, προκειμένου να τηρηθούν οι προϋποθέσεις που τέθηκαν, θεωρήθηκε προτιμότερο να δουλέψουν με αριθμητικά και απτά δεδομένα παρά με αρχεία ήχου. Συνεπώς, το επόμενο βήμα ήταν η εξαγωγή των κατάλληλων δεδομένων από τα αρχεία ήχου και η αποθήκευσή τους σε αρχεία μορφής .csv.

Στην προεπεξεργασία δεδομένων έπρεπε να εξαχθούν μόνο τα απαραίτητα δεδομένα έτσι ώστε το τελικό αρχείο .csv να είναι όσο το δυνατό μικρότερο. Για αυτό τον λόγο, επιλέχθηκαν τα εξής χαρακτηριστικά Zero Crossing Rate, Spectral Centroid, Spectral Rolloff, Mel- Frequency Cepstral Coefficients, Chroma Frequencies, Roote Mean Square Errors (RMSE). Τα χαρακτηριστικά αυτά επιλέχτηκαν, καθώς κανένα από τα είδη δεν μοιράζονται τα ίδια ακριβώς χαρακτηριστικά, για παράδειγμα δυο διαφορετικά είδη δεν μοιράζονται το ίδιο Zero Crossing Rate ή το ίδιο Spectral Rolloff. Αυτά τα χαρακτηριστικά μπορούν να εξαχθούν μόνο από μια μορφή εικόνας αυτών των αρχείων ήχου και όχι απευθείας από τα αρχεία ήχου. Συνεπώς, πρέπει πρώτα να μετατραπεί κάθε αρχείο ήχου σε φασματογράφημα και να αποθηκευτεί ξεχωριστά. Μόλις ολοκληρωθεί αυτή η διαδικασία, πρέπει να εξαχθούν τα χαρακτηριστικά που αναφέρθηκαν προηγουμένως από τα αντίστοιχα φασματογράμματα. Η εξαγωγή των χαρακτηριστικών και η μετατροπή των αρχείων ήχου σε φασματογράμματα, μπορούν να επιτευχθούν, χρησιμοποιώντας το πακέτο Librosa της python.



Εικόνα 4. 4 α) Spectogram, β) Zero Crossing Rate, γ) Spectral Centroid, δ) Mel- Frequency Cepstral Coefficients, ε) Chroma Frequencies

Έπειτα, χρησιμοποιήθηκε το αρχείο .csv που δημιουργήθηκε μετά την εξαγωγή χαρακτηριστικών ως σύνολο δεδομένων.

❖ **Μοντέλα CNN/RNN**

Η πιο κλασική προσέγγιση ταξινόμησης των ειδών είναι το μοντέλο CNN, το οποίο επιτρέπει ευελιξία στην είσοδο δεδομένων και παράλληλα παρέχει αποδεκτά αποτελέσματα. Το CNN και το RNN είναι η πρώτη επιλογή, καθώς το CNN αναλύει απευθείας τις εικόνες και γίνεται εύκολη η εργασία στα φασματογράμματα, ενώ το RNN μπορεί να λειτουργήσει με δεδομένα και γραφήματα. Δοκιμάστηκε η εφαρμογή των CNN και RNN με τη χρήση των Tensorflow και Keras. Επίσης, πειραματίστηκαν με την παράλληλη προσέγγιση και με συνελικτική επαναλαμβανόμενη προσέγγιση. Για το μοντέλο συνελικτικής επαναλαμβανόμενης προσέγγισης, χρησιμοποίησαν αναφορές από την εργασία των Keunwoo Choi κ.α. Εκμεταλλεύτηκαν την ενεργοποίηση RELOU και το βελτιστοποιητή ADAM. Η συνάρτηση απώλειας (loss function) η οποία χρησιμοποιήθηκε ήταν κατηγορηματική διασταυρούμενη εντροπία (categorical cross entropy). Το μοντέλο εκπαιδεύτηκε για διάφορα σύνολα δεδομένων επικύρωσης. Το μοντέλο αυτό, παρείχε αποτελέσματα ακριβείας και εξακολουθεί να είναι αποτελεσματικό.

Για το παράλληλο μοντέλο που αναφέρθηκε παραπάνω, η εργασία έγινε από τους Lin Fen και Shenlen Liu. Σε αυτό το μοντέλο, διαπιστώθηκε ότι τα φασματογράμματα και τα δεδομένα εικόνας επεξεργάζονται παράλληλα από τα μοντέλα CNN και RNN. Αυτή η διεργασία, αύξησε έστω και ελάχιστα, την ακρίβεια του μοντέλου ωστόσο, αύξησε το χρόνο και τους πόρους η οποίοι απαιτούνται για την εκπαίδευση.

❖ **Προσέγγιση μηχανικής εκμάθησης**

Σε αυτή την προσέγγιση, χρησιμοποιήθηκαν οι βασικοί αλγόριθμοι μηχανικής εκμάθησης οι οποίοι είναι εύκολο να κατανοηθούν και να εφαρμοστούν με σκοπό τη δημιουργία ενός μοντέλου το οποίο στη συνέχεια θα εκπαιδευτεί και θα δοκιμαστεί. Η πρώτη προσέγγιση ήταν με το SVM. Σε αυτή την προσέγγιση, χρησιμοποιήθηκε η ενίσχυση ελαφριάς κλίσης Light Gradient Boost (LGB) για την

εξαγωγή χαρακτηριστικών για το SVM. Ορίστηκαν βασικές παράμετροι για το SVM και έγινε χρήση της κλίμακας Standard Scalar και για την επιλογή χαρακτηριστικών χρησιμοποιήθηκε ο ταξινομητής LGBM. Η δεύτερη προσέγγιση ήταν η αρχιτεκτονική XGBoost. Ορίστηκαν οι βασικές παράμετροι και το XGB χρησιμοποιήθηκε για την επιλογή χαρακτηριστικών, καθώς είναι από μόνο του ένα αρκετά ισχυρό εργαλείο. Το βιογραφικό αναζήτησης πλέγματος ήταν απαραίτητο και στις δύο προσεγγίσεις. Μετά την εκπαίδευση και τη δοκιμή και των δύο μοντέλων, βρέθηκε ότι το μοντέλο SVM απέδωσε καλύτερα από τα δύο, με ακρίβεια εξόδου περίπου 70%. Αυτό το μοντέλο είναι μία αποτελεσματική προσέγγιση στην ταξινόμηση ειδών μουσικής παρόλα αυτά, επιδέχεται περαιτέρω βελτίωση. Λαμβάνοντας ένα ακόμα μεγαλύτερο σύνολο δεδομένων με πιο λεπτομερείς πληροφορίες η ακρίβεια μπορεί να αυξηθεί σημαντικά. Τέλος, χρησιμοποιείται λιγότερη μνήμη RAM [18].

4.3.3 Αναγνώριση μουσικού είδους

Με τον αριθμό των ηχητικών αρχείων να αυξάνεται συνεχώς αυξάνεται και η ανάγκη για ταξινόμηση και οργάνωσή τους. Η αυτόματη αναγνώριση μουσικού είδους είναι ένα υποπεδίο της ανάκτησης μουσικών πληροφοριών (MIR). Ο προβληματισμός σχετικά με το σε ποια κατηγορία ανήκει ένα μουσικό αρχείο είναι ζήτημα ταξινόμησης. Η μουσική ταξινομείται βάση χρόνου, γεωγραφικής προέλευσης, θέματος ή ένα οποιοδήποτε άλλο σύνολο κανόνων που αφορούν τον ήχο. Οι άνθρωποι ταξινομούν τη μουσική με βάση την αντίληψή τους ως προς το ηχητικό σήμα. Για αυτό το λόγο, το είδος μουσικής είναι υποκειμενικό από άτομο σε άτομο και μπορεί να είναι διφορούμενο. Επιπροσθέτως, ένα αρχείο μουσικής μπορεί να ανήκει σε παραπάνω από ένα είδος ή κατηγορία. Επομένως, δεν μπορεί να αποδοθεί αντικειμενικά ένα τραγούδι σε ένα είδος. Αυτή η παρατήρηση, ώθησε τη μελέτη των Vogler και Othman. Μελετώντας το θεωρητικό υπόβαθρο, εντόπισαν ότι υπάρχουν δύο σύνολα δεδομένων και άγνωστη μεταξύ τους σχέση. Το πρώτο σύνολο είναι αυτό των αρχείων της μουσικής και το δεύτερο σύνολο είναι αυτό των ειδών της μουσικής. Για να βρεθεί η σχέση μεταξύ των δύο συνόλων, χρησιμοποιήθηκε η μηχανική εκμάθηση.

Το σύνολο δεδομένων πρέπει να περιέχει και τα δύο σύνολα προκειμένου να βρεθεί η μεταξύ τους σχέση. Πολλοί από τους ερευνητές χρησιμοποίησαν την βάση δεδομένων GTZAN. Ωστόσο, για τους σκοπούς αυτής της μελέτης, θεωρήθηκε ότι υπήρχαν ελαττώματα. Έτσι, επέλεξαν να δημιουργήσουν τη δική τους βάση δεδομένων. Χρησιμοποιήθηκαν μουσικές συλλογές οι οποίες πωλούνται από διάφορες εμπορικές εταιρείες και περιλαμβάνουν 100 τραγούδια ανά είδος. Στη συνέχεια, καθένα από αυτά τα σύνολα δεδομένων, χωρίστηκαν σε σύνολα για εκπαίδευση και σε σύνολα για δοκιμή. Στο τέλος της μελέτης, παρατηρήθηκε ότι υπήρχαν αρχεία τα οποία δεν ταίριαζαν σε καμία κατηγορία.

Μία εργασία ταξινόμησης μπορεί να διαμορφωθεί ως εποπτευόμενη εργασία μηχανικής εκμάθησης, η οποία παρέχει τη δυνατότητα σε ένα πρόγραμμα να μάθει ένα μοτίβο χωρίς συγκεκριμένο προγραμματισμό των κανόνων ταξινόμησης. Η εποπτευόμενη μηχανική εκμάθηση αναφέρεται σε ένα υποπεδίο της μηχανικής εκμάθησης όπου ο αλγόριθμος παρέχεται με δεδομένα με ετικέτα. Ο αλγόριθμος εκπαιδεύεται από τα επισημασμένα δεδομένα, προκειμένου να είναι σε θέση να επηρεάσει δεδομένα που μέχρι εκείνη τη στιγμή ήταν κρυφά. Η μηχανική εκμάθηση τυπικά περιλαμβάνει τα βήματα εξαγωγής χαρακτηριστικών, επιλογής χαρακτηριστικών, μοντελοποίησης και αξιολόγησης αυτών των μοντέλων. Στον τομέα της μηχανικής εκμάθησης, υπάρχουν διαφορετικοί τύποι μοντέλων καθένα από αυτά με τα πλεονεκτήματα και τα μειονεκτήματά του. Οι ερευνητές εδώ χρησιμοποίησαν την προσέγγιση νευρωνικών δικτύων, καθώς παρέχει μεγαλύτερη ευελιξία όσο αναφορά τη ρύθμιση του μοντέλου και επειδή η προσέγγιση αυτή, παρουσιάζει μεγάλο ενδιαφέρον στη κοινότητα της μηχανικής εκμάθησης. Ένα νευρωνικό δίκτυο αποτελείται από απλές υπολογιστικές μονάδες, γνωστές

ως νευρώνες. Ένας νευρώνας δέχεται πολλαπλές εισόδους και εφαρμόζει μία λειτουργία ενεργοποίησης. Ο νευρώνας, στη συνέχεια, θα ενεργοποιήσει μία έξοδο σύμφωνα με τη συγκεκριμένη λειτουργία ενεργοποίησης. Κάθε είσοδος έχει τα αντίστοιχα βάρη. Διαισθητικά, τα βάρη αυτά, θα θεωρηθούν ως σύνδεση της ισχύος μεταξύ των νευρώνων. Ένα παράδειγμα της συνάρτησης ενεργοποίησης αποτελεί η σιγμοειδής συνάρτηση. Η λειτουργία ενεργοποίησης καθορίζει την έξοδο του νευρώνα. Αυτή η έξοδος, θα μεταδοθεί στον επόμενο συνδετικό νευρώνα ή σε περίπτωση που βρίσκεται στο επίπεδο της εξόδου, θα είναι η τελική έξοδος.

$$\boldsymbol{S(t) = \frac{1}{1+e^{-t}}} \qquad \textit{(4.1)}$$

Ένα δίκτυο νευρώνων σχηματίζεται με τη διασύνδεση των νευρώνων σε στρώματα. Το πιο εξωτερικό στρώμα, είναι το στρώμα εξόδου, ενώ το πιο εσωτερικό είναι το στρώμα εισόδου. Τα ενδιάμεσα στρώματα ονομάζονται κρυφά στρώματα. Τα νευρωνικά δίκτυα με περισσότερα από δύο κρυφά στρώματα ονομάζονται βαθιά νευρωνικά δίκτυα. Η διασύνδεση αυτών των νευρώνων, παρέχει μια κατανεμημένη αναπαράσταση σε όλο το πλαίσιο του δικτύου.

Ένα νευρωνικό δίκτυο, μπορεί να φανεί ως ένας καθολικός εκτιμητής συνάρτησης. Μπορούμε να χρησιμοποιήσουμε αλγόριθμους εκμάθησης για να εκπαιδεύσουμε ένα νευρωνικό δίκτυο. Για να ταιριάξει ένα νευρωνικό δίκτυο σε μία συγκεκριμένη λειτουργία το δίκτυο θα προσαρμόσει τα βάρη των νευρώνων, σύμφωνα με τον αλγόριθμο μάθησης. Ο αλγόριθμος μάθησης θα αποφασίσει πόσο δυνατά ή αδύναμα θα είναι τα βάρη του.

Τυπικά, ένας αλγόριθμος εκμάθησης λειτουργεί μεταδίδοντας την τιμή εισόδου από τα επίπεδο εισόδου στο επίπεδο εξόδου. Στη συνέχεια, οι τιμές που λαμβάνονται από τα επίπεδα εξόδου συγκρίνονται με την πραγματική τιμή ετικέτας. Η διαφορά μεταξύ πραγματικής τιμής και της τιμής εξόδου, ονομάζεται σφάλμα. Έπειτα, η τιμή αυτή μεταφέρεται προς τα πίσω από την έξοδο στην είσοδο. Χρησιμοποιώντας τεχνικές βελτιστοποίησης, τα βάρη μπορούν να ρυθμιστούν έτσι ώστε, να ελαχιστοποιηθεί το σφάλμα. Το βήμα μετάδοσης της τιμής εισόδου στο εξωτερικό στρώμα είναι γνωστό ως διάδοση προς τα εμπρός (Forward Propagation), ενώ το βήμα διάδοσης του σφάλματος προς το εσωτερικό στρώμα ονομάζεται διάδοση προς τα πίσω (Back Propagation).

Τα νευρωνικά δίκτυα είναι ισχυρά, αλλά έχουν μειονεκτήματα όσο αναφορά την ταχύτητα και την αποτελεσματικότητα. Όσο αυξάνεται ο αριθμός των νευρώνων και των στρωμάτων αυξάνονται εκθετικά και οι απαιτήσεις σε πόρους υπολογισμού. Αυτό μπορεί να καταστήσει την χρήση του μη πρακτική. Ωστόσο, η ισχύς των μονάδων γραφικής επεξεργασίας Graphical Processing Units (GPU) έχουν αυξηθεί. Η GPU είναι υλικό ειδικά σχεδιασμένο για να επιταχύνει δύσκολους υπολογισμούς στον τομέα των γραφικών. Το είδος των υπολογισμών που γίνονται στις GPU έχουν συνήθως την μορφή γραμμικής άλγεβρας. Στην ίδια κατηγορία υπολογισμών ανήκουν και τα νευρωνικά δίκτυα. Αυτό σημαίνει ότι είναι εφικτή η εκμετάλλευση των GPU για επιτάχυνση των υπολογισμών. Ένα από τα πλεονεκτήματα είναι η ανθεκτικότητα του στον θόρυβο. Στην περίπτωση της ταξινόμησης των ειδών τα δεδομένα θεωρούνται θορυβώδη καθώς ο ορισμός των ειδών είναι ασαφής. Υπάρχουν διάφορες κατηγορίες νευρωνικών δικτύων κατάλληλες για διαφορετικές εργασίες. Τα δύο είδη νευρωνικών δικτύων που χρησιμοποιήθηκαν στην συγκεκριμένη εργασία είναι το CNN και το RNN.

Όσο αναφορά την τεχνική εκτέλεση της μελέτης, εξετάστηκε η ταξινόμηση μουσικών ειδών με το λογισμικό «Neuroph Studio». Το λογισμικό αυτό προσφέρει διεπαφή γραφικού χρήστη Graphic User Interface (GUI). Στη συνέχεια, εξέτασαν την περίπτωση να παρακάμψουν το GUI και να

χρησιμοποιήσουν απευθείας το πλαίσιο Neuroph. Ωστόσο το Neuroph βασίζεται στην Java η οποία έχει σχεδιαστεί να λειτουργεί αποκλειστικά στην CPU. Η μηχανική εκμάθηση χρειάζεται υπολογιστές υψηλής απόδοσης για αυτό και χρησιμοποιήθηκε το πλαίσιο «keras» το οποίο παρέχει μαζική παραλληλοποίηση στις GPU χρησιμοποιώντας theano back-end. Προγραμματίζεται μέσω της γλώσσας προγραμματισμού python. Στην συνέχεια, σχεδίασαν και εφάρμοσαν ένα μέσο πληροφορίας που επιτρέπει ως είσοδο μια διεύθυνση URL του YouTube και δίνει ως έξοδο ένα από τα τρία είδη ροκ, ποπ ή χιπ-χοπ. Τα βήματα που εκτελέστηκαν με την σειρά είναι λήψη αρχείου μουσικής, η ανάλυση χαρακτηριστικών και η είσοδος των χαρακτηριστικών στο νευρωνικό δίκτυο. Έπειτα, έγιεν χρήση του Vamp5 το οποίο είναι ένα σύστημα επεξεργασίας ήχου για τα πρόσθετα που εξάγουν περιγραφικές πληροφορίες από τον ήχο. Για να είναι εφικτή η χρήση του Vamp χρειάζεται ένας host κατάλληλο για τον σκοπό αυτό. Ο Sonic Annotator επιτρέπει την χρήση πρόσθετων Vamp ομαδικά και λαμβάνοντας υπόψη ότι είναι ένα εργαλείο γραμμής εντολών μπορεί εύκολα να εκτελέσει ομαδική επεξεργασία μέσω της λειτουργίας Bash. Το μεγαλύτερο πλεονέκτημα της χρήσης Vamp είναι η τεράστια διαθεσιμότητα εφαρμογών και προσθηκών. Ο Sonic Annotator εξάγει τα χαρακτηριστικά ήχου και τα ορίζει ως έξοδο σε αρχεία csv. Κάθε γραμμή ενός αρχείου csv αποτελείται από πολλαπλές στήλες. Η πρώτη στήλη είναι η σφραγίδα του χρόνου σε δευτερόλεπτα. Η επόμενη στήλη θα περιέχει τα χαρακτηριστικά εντός της χρονικής αυτής της διάρκειας. Ο Sonic Annotator παρουσιάζει δυσκολία λόγω της ποσότητας της μνήμης που χρησιμοποιείται κατά της εξαγωγή. Παρατηρήθηκε ότι ο διαχωρισμός των αρχείων σε κλίπ των 30 δευτερολέπτων οδηγεί στην αποφυγή αυτού του σφάλματος. Για την διάσπαση των αρχείων χρησιμοποιήθηκε το λογισμικό ffmpeg [19].

4.4 Διαδικασία παρούσας εργασίας

Η μουσική, είναι μία τέχνη η οποία έχει μεγάλη ποικιλία σε είδη, αναλόγως το περιεχόμενο, των στίχων, το ρυθμό, τη μελωδία και τα μουσικά όργανα που χρησιμοποιούνται. Η ελληνική μουσική σκηνή έχει πολλά να προσφέρει σε κάθε έναν από αυτούς τους τομείς. Ανάμεσα στον πλούτο των ειδών της ελληνικής μουσικής ξεχωρίσαμε τρεις κατηγορίες. Αυτές είναι το λαϊκό, το ροκ και η ραπ. Τα κριτήρια σύμφωνα με τα οποία προτιμήθηκαν οι τρεις συγκεκριμένες κατηγορίες αναφέρονται παρακάτω.

Το λαϊκό είδος επιλέχθηκε καθώς αποτελεί ορόσημο της ελληνικής κουλτούρας και περιέχει μουσικά όργανα τα οποία κατά κύριο λόγο συναντώνται στην ελληνική μουσική. Κάποια από αυτά είναι το μπουζούκι, ο μπαγλαμάς, λύρα, λαούτο κλπ [20].

Η ροκ μουσική προήλθε από την Rock n' Roll. Χαρακτηριστικά της όργανα είναι η ηλεκτρική κιθάρα, τα ντραμς, το μπάσο και το ηλεκτρικό αρμόνιο. Ξεχωρίζει για τον έντονο ρυθμό και την μοναδική μελωδία φωνών [21].

Στη ραπ μουσική δίνεται έμφαση στους στίχους οι οποίοι είναι αυτοσχέδιοι, γραμμένοι σε καθομιλουμένη έκφραση και με πολλά στοιχεία αργκό. Η μουσική έχει στοιχεία από την Soul, την Jazz και ποικίλα άλλα μουσικά ρεύματα [22].

Εν συνεχεία, πραγματοποιήθηκε λήψη πενήντα τραγουδιών για κάθε είδος σε μορφή MP3 (MPEG-1 Audio Layer 3). Προϋπόθεση υπήρξε το αρχείο ήχου να είναι καθαρό από τυχόν παρεμβολές. Αυτό σημαίνει, ο ήχος να μην προέρχεται από ζωντανή μετάδοση καθώς είναι πιθανή η οχλαγωγία κοινού . Επίσης, αποφεύχθηκε η χρήση τραγουδιών από μουσικά βίντεο λόγω πιθανόν μεγάλων εισαγωγών ή ενδιάμεσων διακοπών. Ο κώδικας που χρησιμοποιήθηκε απαιτεί η μορφή αρχείου ήχου να είναι σε WAV (Waveform Audio File Format). Αυτός είναι ο λόγος που προχωρήσαμε στην κατάλληλη μετατροπή. Έπειτα, προσθέσαμε τον κώδικα στο φάκελο των τραγουδιών, τα εισάγαμε όλα μαζί στο

πρόγραμμα Matlab. Από το Matlab βγήκαν κάποιοι πίνακες Excel οι οποίοι περιείχαν τα εξαγόμενα χαρακτηριστικά που στην συνέχεια τα χρησιμοποιήσαμε ως είσοδο σε εφαρμογές ταξινόμησης.

4.5 Επίλογος

Σε αυτό το κεφάλαιο συζητιέται αρχικά το πρόβλημα της κατηγοριοποίησης. Στην συνέχεια, παρατίθενται σχετικές εργασίες που έγιναν για την αντιμετώπιση των προβλημάτων αυτών και την βελτίωση της τεχνολογίας. Μελετήθηκε η εργασία του Ινδικού Ινστιτούτου Τεχνολογίας η οποία χρησιμοποίησε την βάση δεδομένων GTZAN. Επίσης η μελέτη του Ινστιτούτου Μηχανικής Xavier την Ινδίας παρόλο που εξερεύνησε διάφορες δωρεάν βάσεις δεδομένων όπως η Free Music Archive (FMA) και την βάση δεδομένων χιλίων τραγουδιών κατέληξε και αυτή να αντλήσει πληροφορίες από την GTZAN. Αυτό που έγινε σαφές είναι ότι η πιο συνηθισμένες και ακριβής τεχνικές ταξινόμησης είναι τα μοντέλα CNN/RNN και η προσέγγιση μηχανικής εκμάθησης.

Το ζήτημα της ταξινόμησης μουσικής είναι ότι μπορεί να γίνει με βάση ένα μεγάλο σύνολο κριτηρίων που αφορούν τον ήχο και είναι πολύ διαφορετικά μεταξύ τους. Έτσι η ταξινόμηση της μουσικής θεωρείται γενικά υποκειμενική. Πάνω σε αυτό το πρόβλημα στήριξαν την μελέτη τους οι Vogler και Othman. Οι συγκεκριμένοι ερευνητές κατέληξαν στην χρήση νευρωνικών δικτύων. Τέλος περιγράφεται συνοπτικά η διαδικασία που ακολουθήθηκε στην παρούσα εργασία.

ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ

Πρόλογος

Στο μέρος αυτό παρουσιάζεται το πειραματικό μέρος της πτυχιακής εργασίας. Το πείραμα της εργασίας αυτής αφορά την κατηγοριοποίηση ελληνικών τραγουδιών σε είδη μουσικής με μεθόδους μηχανικής μάθησης. Στα κεφάλαια που ακλουθούν περιλαμβάνεται η διαδικασία και τα κριτήρια συλλογής ηχητικών δεδομένων και η ταξινόμηση τους σε είδη μουσικής καθώς και η προεπεξεργασία τους σε κατάλληλη μορφή για να διεξαχθούν οι μετρήσεις .Έπειτα αναλύονται τα χαρακτηριστικά τα οποία επιλέχθηκαν για εξαγωγή από το data-set μας με σκοπό να με σκοπό να χρησιμοποιηθούν για την εκπαίδευση των μοντέλων μηχανικής μάθησης τα οποία προβλέπουν σε ποιο είδος ελληνικής μουσικής ανήκει ένα ελληνικό τραγούδι. Ακολουθεί σχολιασμός των αποτελεσμάτων και των σφαλμάτων καθώς και η αξιολόγηση των χαρακτηριστικών όπου χρησιμοποιήθηκαν για την εξαγωγή.

Κεφάλαιο 5: Συλλογή και προεπεξεργασία δεδομένων

5.1 Εισαγωγή

Σε αυτό το κεφάλαιο αναλύεται το πρώτο μέρος του πειράματος, η συλλογή και επεξεργασία δεδομένων. Διατυπώνονται οι σκέψεις και οι επιλογές για την δημιουργία του συνόλου δεδομένων μας (data-set) καθώς και τι μορφή θα έχει, με ποιά κριτήρια επιλέχθηκε η μουσική και από ποιες πηγές. Έπειτα, ερευνάται σε ποια μορφή πρέπει να κωδικοποιηθούν ώστε να γίνει η κατάτμηση σε παράθυρα για να καθιστούν κατάλληλα για εξαγωγή χαρακτηριστικών.

5.2 Συλλογή δεδομένων

Τα είδη μουσικής που εν τέλει χρησιμοποιήθηκαν για την διεξαγωγή του πειράματος ήταν τρία. Το ελληνικό ροκ, το ελληνικό ραπ και το ελληνικό λαϊκό τραγούδι. Για την δημιουργία του data-set χρησιμοποιήθηκαν 150 τραγούδια τα οποία ταξινομήθηκαν ανά 50 στα τρία είδη μουσικής.

Έγινε προσπάθεια οι ηχογραφήσεις και οι εκτελέσεις των τραγουδιών να είναι στην καλύτερη διαθέσιμη ποιότητα, ώστε να μην επηρεαστούν οι μετρήσεις από ήχους, όπως το χαρακτηριστικό τριγμό του βινυλίου που συναντήσαμε αρκετά σε παλιές ηχογραφήσεις του ελληνικού λαϊκού τραγουδιού. Επίσης, δεν επιλέξαμε ζωντανές εκτελέσεις όπου υπάρχουν διάφορες παρεμβολές, όπως οι φωνές του κοινού.

5.3 Προεπεξεργασία δεδομένων

Τα αρχεία που επιλέχθηκαν είχαν θέματα διαφορετικής στάθμης, οπότε ακολούθησε κανονικοποίηση (normalization) στο πρόγραμμα Audacity, έτσι ώστε να έχουν όλα την ίδια στάθμη έντασης και να καθιστούν συγκρίσιμα. Επίσης, έγινε μετατροπή των τραγουδιών σε μορφή wav στα 44.100 Hz ρυθμός δειγματοληψίας και σε 16-bit κβάντιση προκειμένου όλα τα τραγούδια να έχουν το ίδιο format.

5.4 Κατάτμηση

Το τελευταίο στάδιο προεπεξεργασίας δεδομένων ήταν ο διαχωρισμός του κάθε ηχητικού αρχείου σε παράθυρα (segmentation). Με αυτόν τον τρόπο κάθε ηχητικό αρχείο διασπάστηκε σε διακριτά μέρη έτσι ώστε από το κάθε μέρος να μπορέσουμε να εξάγουμε τιμές από ανάλογα χαρακτηριστικά και να τα χρησιμοποιήσουμε ως είσοδο σε αλγόριθμους μηχανικής μάθησης. Σε παρόμοια προβλήματα τα τραγούδια κατατμήθηκαν σε παράθυρα 1 δευτερολέπτου [23], [24] και 3 δευτερολέπτων [25]. Έγινε επανάληψη του πειράματος με 0.5 δευτερόλεπτα, 1 δευτερόλεπτο και 2 δευτερόλεπτα για να εξερευνηθεί κατά πόσο ο μεγαλύτερος υπολογιστικός όγκος φέρνει πιο ακριβή αποτελέσματα. Σε αυτό το σημείο αξίζει να αναφερθεί ότι όσο μικρότερα είναι τα παράθυρα τόσο αυξάνεται ο όγκος δεδομένων. Γίνεται περεταίρω ανάλυση της διαδικασίας κατάτμησης στο κεφάλαιο 6.

Να σημειωθεί ότι σε παρόμοια πειράματα πριν την κατάτμηση σε παράθυρα χρησιμοποιήθηκε μικρότερη διάρκεια τραγουδιού όπως 30 ή 40 δευτερόλεπτα [25], [23]. Εξαιτίας της πρωτοτυπίας του εν λόγω πειράματος, που αφορά ελληνική μουσική, χρησιμοποιήθηκε η πλήρης διάρκεια του τραγουδιού για μεγαλύτερη ακρίβεια. Εφαρμογές όπως το Shazam είναι ικανά να ταυτοποιούν τραγούδια και είδη μουσικής μέσα σε κάποια ms διότι υπάρχει βάση αληθείας (database) και δεν υπάρχει ανάγκη εκπαίδευσης μοντέλου μηχανικής μάθησης.

5.5 Επίλογος

Σε αυτό το κεφάλαιο έγινε περιγραφή του πρώτου μέρους του πειράματος που αφορούσε την επιλογή ειδών μουσικής, την συλλογή τραγουδιών και την επιλογή κριτηρίων βάση των οποίων έγινε η ταξινόμηση. Επίσης, τέθηκαν κάποιοι κανόνες βάση των οποίων επιλέχθηκε το υλικό.

Μετά την κατάλληλη κωδικοποίηση των αρχείων ξεκίνησε η κατάτμηση σε παράθυρα. Αυτό που ξεχωρίζει την συγκεκριμένη εργασία, πέραν του ότι αφορά την ελληνική μουσική είναι ότι χρησιμοποιήθηκε όλη η διάρκεια αυτών των τραγουδιών.

Κεφάλαιο 6: Εξαγωγή ηχητικών χαρακτηριστικών

6.1 Εισαγωγή

Αυτό το κεφάλαιο πραγματεύεται την επιλογή και περιγραφή των ηχητικών παραμέτρων για εξαγωγή ταυτόχρονα με την διαδικασία της κατάτμησης, καταλήγοντας στην τελική κατασκευή του data-set που πρόκειται να εκπαιδευτεί μέσω μηχανικής μάθησης.

6.2 Ηχητικά Χαρακτηριστικά

Η εξαγωγή ηχητικών χαρακτηριστικών πραγματοποιήθηκε σε κάθε παράθυρο από αυτά που έχουν επιλεγεί, όπως περιγράφεται στο προηγούμενο κεφάλαιο. Για τον λόγο αυτό χρησιμοποιήθηκε μια εξειδικευμένη εργαλειοθήκη του Matlab που ονομάζεται MIRToolBox το οποίο επικεντρώνεται πάνω στην εξαγωγή, στην δημιουργία και στον υπολογισμό ηχητικών ιδιοτήτων.

Μελετώντας παρόμοια πειράματα ηχητικής αναγνώρισης όπου έχουν υλοποιηθεί, ξεχώρισαν κάποια χαρακτηριστικά τα οποία παρατηρήθηκαν ότι είχαν μεγάλη συχνότητα εμφάνισης. Στο [25, 23] παρατηρήθηκε το low energy, στο [25], [23], [26] to zero crossing rate, στο [25], [26], [27] το roll-off frequency, στο [26], [28] το Spectral Centroid, στο [26] to Spread, το Skewness και Kurtosis και τέλος το MFCCs σε όλα προαναφερθέντα papers.

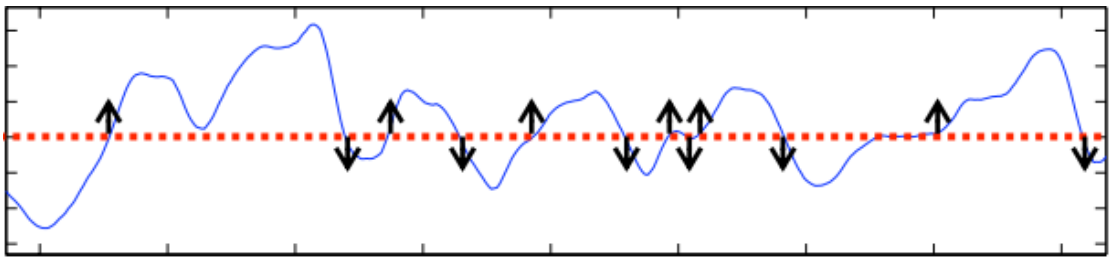
Στο πλαίσιο αυτό χρησιμοποιήθηκαν τα παρακάτω ηχητικά χαρακτηριστικά.

• **Low-Energy Feature**

Η Χαμηλή Ενέργεια είναι το ποσοστό παραθύρων που έχουν λιγότερη ενέργεια από τη μέση ενέργεια όλων των παραθύρων. Μουσική με φωνητικά, η όποια έχει παύσεις, θα έχει μεγάλη τιμή χαμηλής ενέργειας, ενώ μουσική με έγχορδα θα έχει χαμηλή αυτήν την τιμή [23].

•**Zero-Crossing Rate**

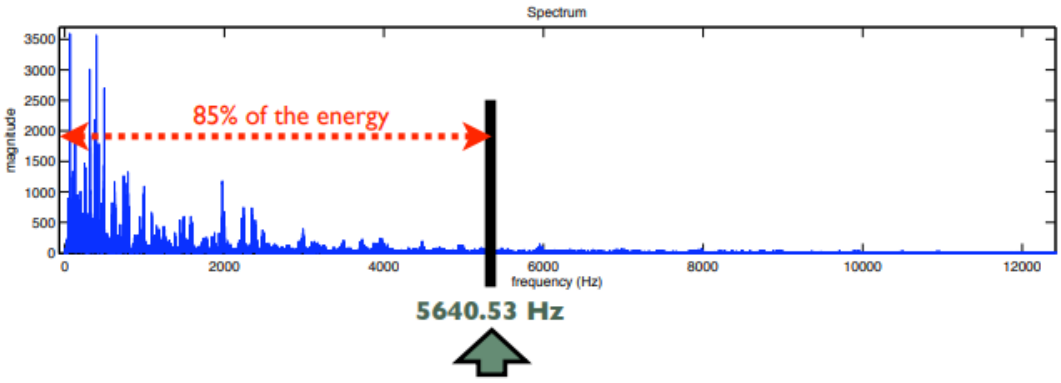
Ο Ρυθμός Μηδενικού Επιπέδου είναι ο ρυθμός με τον οποίο ένα σήμα διασχίζει τον άξονα χ δηλαδή μεταβαίνει από θετικές τιμές στο μηδέν, από το μηδέν στο αρνητικό ή το αντίστροφο. Η τιμές του έχουν χρησιμοποιηθεί πολύ σε εφαρμογές αναγνώρισης φωνής και σε εφαρμογές άντλησης μουσικών πληροφοριών για ταξινόμηση κρουστών ήχων. Εάν η τιμή είναι υψηλή σημαίνει ότι έχουμε θόρυβο.



Εικόνα 6. 1 Ρυθμός μηδενικού επιπέδου. Η οριζόντια πορτοκαλί γραμμή είναι το πλάτος=0 και η μπλε γραμμή είναι το ηχητικό σήμα

• **Spectral Roll-off**

Η ενεργειακή συχνότητα αποκοπής είναι η συχνότητα κάτω από την οποία βρίσκεται ένα συγκεκριμένο ποσοστό (συνήθως 85%) της συνολικής φασματικής ενέργειας. Έτσι, γίνεται σαφές εάν το σήμα περιέχει υψηλές συχνότητες. Εδώ έγινα πειράματα και με τα ποσοστά 30%, 50%, 70% και 90%.

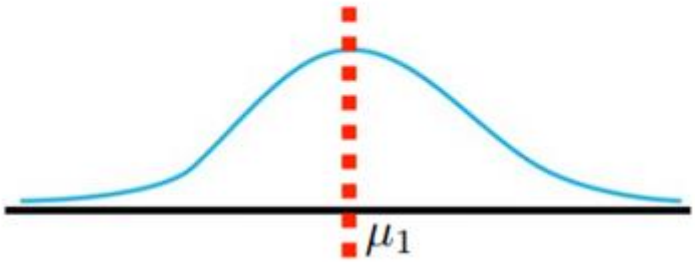


Εικόνα 6. 2 Το 85% της ενέργειας είναι συγκεντρωμένο κάτω από την συχνότητα των 5640.53 Hz

• **Spectral Centroid**

Ο φασματικός μέσος όρος είναι το κέντρο «βαρύτητας» του φάσματος ενός σήματος. Η ακολουθία του φασματικού κεντροειδούς ποικίλλει για τμήματα ομιλίας, ενώ για ήχους ανθρώπινων κραυγών η απόκλιση του φασματικού κέντρου είναι σημαντικά χαμηλή. Το φασματικό κέντρο είναι ένας καλός προγνωστικός δείκτης της "φωτεινότητας" ενός ήχου [23] χρησιμοποιείται ευρέως στην ψηφιακή επεξεργασία ήχου και μουσικής ως αυτόματη μέτρηση της μουσικής χροιάς.

$$\mu_1 = \int x f(x) dx$$

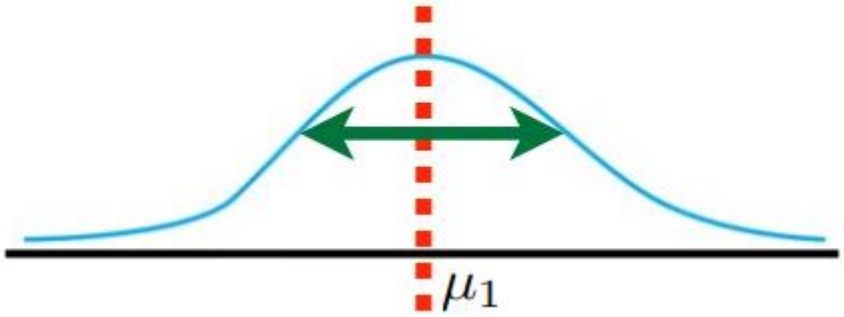


Εικόνα 6. 3 Μέση τιμή του φάσματος

• **Spread**

Τυπική Απόκλιση ή φασματική εξάπλωση περιγράφει τη μέση απόκλιση του χάρτη ρυθμού γύρω από το κέντρο του, που συνήθως σχετίζεται με το εύρος ζώνης του σήματος. Τα σήματα που μοιάζουν με θόρυβο έχουν συνήθως μεγάλη φασματική εξάπλωση, ενώ μεμονωμένοι τονικοί ήχοι με μεμονωμένες κορυφές θα έχουν ως αποτέλεσμα χαμηλή φασματική εξάπλωση.

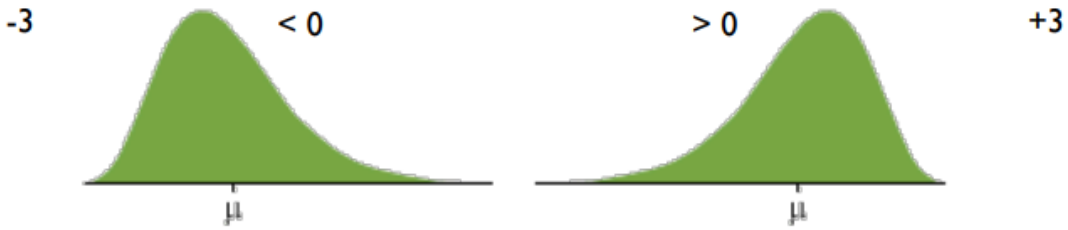
$$\sigma^2 = \mu_2 = \int (x - \mu_1)^2 f(x) dx$$



Εικόνα 6. 4 Τυπική απόκλιση σήματος

• **Skewness**

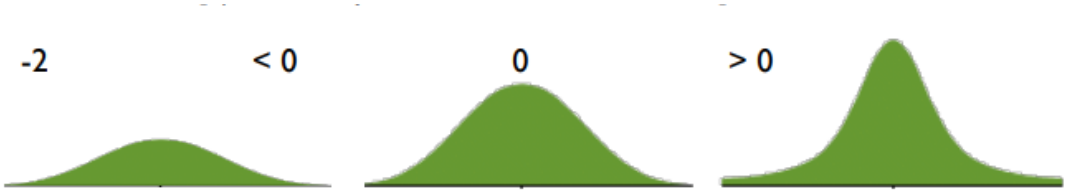
Μετράει την συμμετρία γύρω από το κεντροειδές. Η φασματική κλίση χρησιμοποιείται με άλλες φασματικές ροπές για να διακρίνει τον τόπο άρθρωσης. Για αρμονικά σήματα, υποδεικνύει τη σχετική ισχύ υψηλότερων και κατώτερων αρμονικών. Για παράδειγμα, στο σήμα τεσσάρων τόνων, υπάρχει μια θετική λοξή όταν κυριαρχεί ο κάτω τόνος και μια αρνητική λοξή όταν κυριαρχεί ο επάνω τόνος [29].



Εικόνα 6. 5 Στο πρώτο σχήμα οι τιμές είναι μαζεμένες στα αριστερά του μέσου όρου στο δεύτερο το αντίθετο

• **Kurtosis**

Η φασματική κύρτωση μετρά την επιπεδότητα του φάσματος σε σχέση με τη καμπύλη Gauss, του φάσματος γύρω από το κέντρο του. Αντίθετα, χρησιμοποιείται για να δείξει την αιχμή ενός φάσματος. Για παράδειγμα, καθώς ο λευκός θόρυβος αυξάνεται στο σήμα ομιλίας, η κύρτωση μειώνεται, υποδεικνύοντας ένα φάσμα λιγότερης αιχμής.



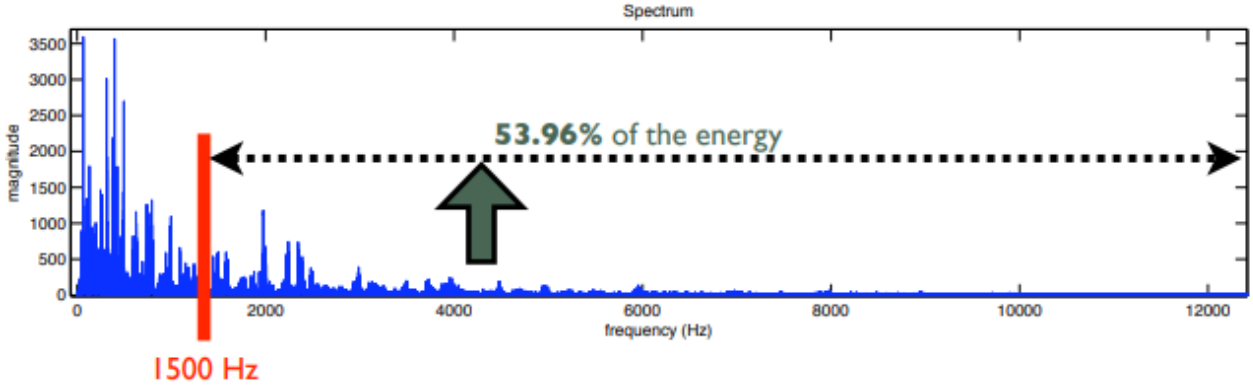
Εικόνα 6. 6 Χαμηλή κύρτωση, κανονική κύρτωση, υψηλή κύρτωση

• **Mel-Frequency Cepstral Coefficients**

Οι Συντελεστές Κλίμακας Mel χρησιμοποιούνται ευρέως στην αναγνώριση ομιλίας, αντιπροσωπεύουν τα φασματικά χαρακτηριστικά που βασίζονται στην κλίμακα συχνότητας Mel, είναι μια κλίμακα που δείχνει το πόσο ανθρώπινη είναι μια μουσική. Ο αριθμός-παράμετρος όπου παρατηρείται στον κώδικα είναι ο αριθμός των συντελεστών που χρησιμοποιήθηκαν για την εξαγωγή χαρακτηριστικών.

• **Brightness**

Με το χαρακτηριστικό γίνεται έλεγχος της φωτεινότητας του ήχου, δηλαδή τον τόνο. Αυτό μετράται βάσει το ποσοστό ενέργειας πάνω από μια επιλεγμένη συχνότητα. Η προεπιλεγμένη τιμή στο MIRToolbox είναι 1500Hz. Η τιμή 1000 Hz έχει προταθεί στο (Laukka, Juslin and Bresin, 2005) και η τιμή 3000 Hz έχει προταθεί στο (Juslin, 2000) οπότε δοκιμάστηκαν και αυτές οι τιμές ως παράμετροι.



Εικόνα 6. 7 Το 53.96% της ενέργειας είναι συγκεντρωμένο στις συχνότητες πάνω από 1500Hz

6.3 Κατάτμηση και εξαγωγή χαρακτηριστικών

Παρακάτω φαίνεται η εικόνα (εικόνα 6.8 για παράθυρα 0.5 δευτερολέπτων) της αλγοριθμικής λογικής που χρησιμοποιήθηκε. Όπως είναι ευδιάκριτο κάθε ηχητικό χαρακτηριστικό για κάθε τραγούδι εξήχθη στο εκάστοτε παράθυρο, οι τελικές υπολογισμένες τιμές των χαρακτηριστικών αποθηκεύονται σε ένα αρχείο xls.

```
%the current working folder is set as the path.
dirName=pwd;

%all wav files are inserted into a variable "files".
files = dir( fullfile(dirName,'*.wav') );

%the window is set for 1/2 seconds. The next two times we run the
%code for each genre we change it to 1 and 2 respectively.
win=0.5;

%counter used in the forloop.
k=1;

%one song at a time
for f=1:length(files)

    %creates the fullpath using the folder and the song name
    fullFileName = fullfile(dirName, files(f).name)

    %mirlength computes the duration of the file loaded through miraudio which
    %loades data from audio files
    duration=mirgetdata(mirlength(miraudio(fullFileName)));

    %loop for every window of the song
    for i=0:ceil((duration/win)-1)

        %the variable "audio" holds each window per loop
        audio=miraudio(fullFileName,'Extract',i*win,(i+1)*win);

        %a two dimentional array AF is created where every line holds the values
        %of the audio features seperated by each column as numbered below

        AF(k,1)=mirgetdata(mirlowenergy(audio));
        AF(k,2)=mirgetdata(mirzerocross(audio));
        AF(k,3)=mirgetdata(mirrolloff(audio,'Threshold',0.3));
        AF(k,4)=mirgetdata(mirrolloff(audio,'Threshold',0.5));
        AF(k,5)=mirgetdata(mirrolloff(audio,'Threshold',0.7));
        AF(k,6)=mirgetdata(mirrolloff(audio,'Threshold',0.9));
        AF(k,7)=mirgetdata(mircentroid(audio));
        AF(k,8)=mirgetdata(mirspread(audio));
        AF(k,9)=mirgetdata(mirskewness(audio));
        AF(k,10)=mirgetdata(mirkurtosis(audio));


        mfccs=mirgetdata(mirmfcc(audio));
        AF(k,11)= mfccs(1);
        AF(k,12)= mfccs(2);
        AF(k,13)= mfccs(3);
        AF(k,14)= mfccs(4);
        AF(k,15)= mfccs(5);
        AF(k,16)=mfccs(6);
        AF(k,17)=mfccs(7);
        AF(k,18)=mfccs(8);
        AF(k,19)=mfccs(9);
        AF(k,20)=mfccs(10);
        AF(k,21)=mfccs(11);
        AF(k,22)=mfccs(12);
        AF(k,23)=mfccs(13);
        AF(k,24)=mirgetdata(mirbrightness(audio,'CutOff',1000));
        AF(k,25)=mirgetdata(mirbrightness(audio,'CutOff',3000));

        %counter raised by one so we can get to the next time window
        k=k+1;
    end
end

%export of AF array into an xls file titled "Features"
xlswrite('Features.xls', AF);
```

Εικόνα 6. 8 Κώδικας όπου εισήχθη στο Matlab

Αφού εξήχθησαν τα χαρακτηριστικά για κάθε είδος μουσικής σε αρχείο xls συγχωνεύτηκαν σε ένα ενιαίο αρχείο κατάλληλο για επεξεργασία-ταξινόμηση στο Matlab, όπου σαν στήλες είχε τα ονόματα των χαρακτηριστικών και στο τέλος κάθε γραμμής είχε την επισημείωση(label) που καθορίζει το είδος (εικόνα 6.9). Επαναλήφθηκε η διαδικασία τρεις φορές για κάθε είδος κατάτμησης (1/2 δευτερόλεπτου, 1 δευτερολέπτου και 2 δευτερολέπτων).

S	T	U	V	W	X	Y	Z
mfccs9	mfccs10	mfccs11	mfccs12	mfccs13	brightness(1000)	brightness(3000)	label
0,377253	0,091687	-0,04974	0,02226	-0,11874	0,914918549	0,569453532	rap
0,093422	0,028054	-0,03155	-0,01421	0,207526	0,552960059	0,4167436	rap
0,210161	0,06524	-0,05407	0,210668	0,002236	0,510528526	0,438613336	rap
0,116683	-0,31083	0,018568	0,402683	0,132322	0,458265565	0,350129716	rap
-0,02641	-0,1596	0,165342	0,390497	0,067133	0,634235915	0,476547446	rap
0,111017	0,053474	0,116643	0,083827	0,194244	0,72928941	0,509645553	rap
0,006767	-0,2133	0,059712	0,415986	0,162393	0,45035898	0,33711889	rap
0,250434	-0,0735	-0,06842	0,205492	0,003803	0,489879206	0,416542877	rap
0,263573	-0,18734	-0,19721	0,164878	0,222492	0,45158313	0,347053813	rap
-0,00058	-0,18741	-0,0368	0,308032	0,302279	0,60385616	0,443051294	rap
0,135648	0,084425	-0,02383	-0,05279	0,289959	0,728720365	0,499067107	rap
0,295773	-0,07548	-0,19079	0,085365	0,107021	0,543153173	0,395733222	rap
0,225179	-0,04637	-0,03973	0,242458	0,097782	0,48944807	0,413698149	rap
0,125916	-0,33813	-0,04441	0,370574	0,132184	0,44449952	0,342953218	rap
0,075726	-0,12194	0,095829	0,420689	0,311125	0,444417815	0,33126376	rap
0,102644	0,061429	0,049057	0,129009	0,203669	0,729927564	0,500341388	rap
-0,02218	-0,2816	0,072704	0,417489	0,119341	0,466255205	0,346982984	rap
0,17603	-0,05775	0,017765	0,188085	-0,03532	0,510577742	0,434686689	rap
0,235377	-0,23011	-0,23935	0,112508	0,266404	0,458496576	0,354941203	rap
0,052581	-0,31208	-0,08595	0,330955	0,12575	0,58928054	0,447964623	rap

Εικόνα 6. 9 Στιγμιότυπο αρχείου xls όπου αποτελεί την βάση αληθείας

6.4 Επίλογος

Στο κεφάλαιο αυτό γίνεται η περιγραφή της διαδικασίας της κατάτμησης. Πρώτα ο λόγος για τον οποίο επιλέχθηκαν κάποια συγκεκριμένα χαρακτηριστικά και στη συνέχεια περιγράφονται τα χαρακτηριστικά αυτά. Για κάθε ένα από αυτά τα χαρακτηριστικά δίνεται μια εικόνα ώστε να γίνει πιο κατανοητή η λειτουργία του. Δίνεται επίσης ο κώδικας που εισήχθη στο Matlab και το Excel αρχείο στο οποίο εξήχθησαν τα χαρακτηριστικά.

Κεφάλαιο 7: Πειράματα μηχανικής μάθησης που εκπονήθηκαν

7.1 Εισαγωγή

Σε αυτό το κεφάλαιο μιλάμε για την διαδικασία της ταξινόμησης: τους τρόπους επικύρωσης καθώς και τους αλγόριθμους που χρησιμοποιήθηκαν

Τώρα που έχει ολοκληρωθεί η διαδικασία της εξαγωγής ηχητικών παραμέτρων καθώς και επισημείωσης τους στα διαφορά παράθυρα μπορούμε να προχωρήσουμε σε πειράματα μηχανικής μάθησης. Τα πειράματα πραγματοποιήθηκαν στο MATLAB R2021a με την χρήση του εξειδικευμένου εργαλείου Classification Learner όπου ουσιαστικά είναι εφαρμογές εκμάθησης μοντέλων Εποπτευόμενης Μηχανικής Μάθησης (Supervised Machine Learning). Στα πειράματα αυτά χρησιμοποιήθηκε η εκάστοτε βάση αληθείας (οι τιμές των εξαγόμενων χαρακτηριστικών μαζί με τις επισημειώσεις).

7.2 Τρόποι επικύρωσης (validation)

Ένα κρίσιμο ερώτημα που έχρηζε απάντηση είναι πώς θα επικυρωθούν τα μοντέλα αυτά. Στην εκπαίδευση μοντέλων μέσω εποπτευόμενης Μηχανικής Μάθησης για να δημιουργηθούν τα μοντέλα ταξινόμησης, πρέπει να διαχωριστούν τα δεδομένα εισόδου σε δεδομένα εκπαίδευσης (training set) έτσι ώστε να εκπαιδευτεί ο αλγόριθμος κατηγοριοποίησης και σε δεδομένα ελέγχου (testing set) έτσι ώστε να προσδιοριστεί το ποσοστό επιτυχίας του αλγορίθμου αυτού [30].

Κατά κύριο λόγο στη διαδικασία εποπτευόμενης μηχανικής μάθησης χρησιμοποιούνται δύο τρόποι επικύρωσης. Ο ένας είναι ο k-fold validation ο οποίος χωρίζει τα δεδομένα σε k υποσύνολα εκ των οποίων τα k-1 χρησιμοποιούνται για εκπαίδευση και το ένα χρησιμοποιείται για έλεγχο ενώ η συνολική διαδικασία πραγματοποιείται k φορές. Στο τέλος κάθε block έχει χρησιμοποιηθεί για testing και είναι γνωστό ποια μέθοδος πήγε καλύτερα στην ταξινόμηση. Στα εν λόγω πειράματα χρησιμοποιήθηκε k=5 και k=10. Το k=10 χρησιμοποιήθηκε γιατί είναι μια γενικευμένη τεχνική [31] και το k=5 για ενδεχόμενο γενίκευσης της εκπαίδευσης. Όσο πιο λίγο γίνεται η εκπαίδευση αποφεύγεται ο κίνδυνος της υπέρ-εκπαίδευσης. Όσο πιο πολύ διαιρούνται τα δεδομένα σε folds παραμαθαίνει το μοντέλο τα συγκεκριμένα δεδομένα και όταν έρθει ένα άγνωστο δεδομένο δεν μπορεί να το κατηγοριοποιήσει.

Ο άλλος συνήθης τρόπος είναι το split ή αλλιώς Holdout Validation: ο διαχωρισμός σε δεδομένα έλεγχου και δεδομένα εκπαίδευσης, όπου η εκπαίδευση γίνεται μόνο μια φορά. Είναι μια χρήσιμη μέθοδος όταν έχουμε μεγάλο όγκο βάσης αληθείας(data-set) ή όταν θέλουμε να χτίσουμε ένα αρχικό μοντέλο [32]. Πραγματοποιήθηκαν μετρήσεις με ποσοστό δεδομένων ελέγχου (training set) 30% και 40%.

Σύμφωνα με όλα τα παραπάνω χρησιμοποιήσαμε συνολικά 4 μεθόδους επικύρωσης.
5-fold validation, 10-fold validation, 30% Holdout και 40% Holdout.

7.3 Αλγόριθμοι εκμάθησης μοντέλου

Στα προβλήματα ταξινόμησης κατά κύριο λόγο όπως έγινε σε [25],[23],[27],[30],[33] χρησιμοποιούνται σε συχνότητα ο k-Nearest Neighbor (kNN), τα Support Vector Machines(SVM), Δέντρα Αποφάσεων (Decision Trees), Νευρωνικά Δίκτυα(Neural Networks). Δεδομένου δε ότι εκτελείται μια διερεύνηση ταξινόμησης ηχητικού περιεχομένου στην ελληνική μουσική χρησιμοποιήθηκαν ποικίλοι αλγόριθμοι οι οποίοι είναι οι βασικοί αλγόριθμοι μηχανικής μάθησης που προαναφέρθηκαν με κάποιες παραλλαγές τους, διαθέσιμοι από την εφαρμογή Classification Learner του Matlab.

Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines - SVM)

•**Cubic SVM**

Ένα SVM που χρησιμοποίει τετραγωνικό πυρήνα

•**Fine Gaussian SVM**

Κάνει λεπτομερείς διακρίσεις μεταξύ των κλάσεων χρησιμοποιώντας Γκαουσιανό πυρήνα με την κλίμακα πυρήνα ρυθμισμένη στην τετραγωνική ρίζα των μεταβλητών εισόδου προς 4.

•**Medium Gaussian SVM**

Μεσαίες διακρίσεις, με κλίμακα πυρήνα ρυθμισμένη στην τετραγωνική ρίζα των μεταβλητών εισόδου.

•**Quadratic SVM**

Ένας γρήγορος και εύκολος στην ερμηνεία διακριτικός ταξινομητής που τοποθετεί ελλειπτικά παραβολικά ή υπερβολικά όρια μεταξύ των κλάσεων

Αλγόριθμοι κατηγοριοποίησης πλησιέστερου γείτονα(k-NN)

•**Fine KNN**

Ταξινομητής πλησιέστερου γείτονα ο οποίος πραγματοποιεί λεπτομερείς διακρίσεις μεταξύ των κλάσεων. Ο αριθμός των γειτόνων
έχει οριστεί σε 1.

•**Medium KNN**

Ταξινομητής πλησιέστερου γείτονα ο οποίος πραγματοποιεί λιγότερες διακρίσεις από των Fine KNN καθώς ο αριθμός των γειτόνων έχει οριστεί σε 10.

•**Weighted KNN**

Ταξινομητής πλησιέστερου γείτονα μέτριας λεπτομέρειας διακρίσεως μεταξύ των κλάσεων, χρησιμοποιώντας ένα βάρος απόστασης με αριθμός των γειτόνων να έχει οριστεί σε 10.

•**Cosine KNN**

Μέσες διακρίσεις μεταξύ κλάσεων, χρησιμοποιώντας μια μετρική απόστασης συνημιτόνου. Ο αριθμός των γειτόνων έχει οριστεί σε 10.

•**Cubic KNN**

Μέσες διακρίσεις μεταξύ των τάξεων, με χρήση μετρικής κυβικής απόστασης. Ο αριθμός των γειτόνων έχει οριστεί σε 10.

Δέντρα Αποφάσεων (Decision Trees)

•**Bagged Trees**

Τα «Συσκευασμένα Δέντρα» είναι Δέντρα Αποφάσεως τα οποία χρησιμοποιούν τον αλγόριθμο «τυχαίο δάσος» του Breiman.

Νευρωνικά Δίκτυα (Neural Networks)

•**Narrow Neural Network**

Ένα νευρωνικό δίκτυο με μια πλήρως συνδεδεμένη στρώση μεγέθους 10.

•**Medium Neural Network**

Ένα νευρωνικό δίκτυο με μια πλήρως συνδεδεμένη στρώση μεγέθους 25.

•**Wide Neural Network**

Ένα νευρωνικό δίκτυο με μια πλήρως συνδεδεμένη στρώση μεγέθους 100.

•**Bilayered Neural Network**

Ένα νευρωνικό δίκτυο με δυο πλήρως συνδεδεμένες στρώσεις μεγέθους 10 και 10. Τα δεδομένα προκύπτουν από τη τελευταία πλήρως συνδεδεμένη στρώση.

•**Trilayered Neural Network**

Ένα νευρωνικό δίκτυο με τρεις πλήρως συνδεδεμένες στρώσεις μεγέθους 10,10 και 10. Τα δεδομένα προκύπτουν από τη τελευταία πλήρως συνδεδεμένη στρώση.

7.4 Επίλογος

Στο έβδομο κεφάλαιο φαίνεται πως μετά την εξαγωγή των ηχητικών παραμέτρων το επόμενο βήμα είναι τα πειράματα μηχανικής μάθησης. Για τα πειράματα αυτά χρησιμοποιήθηκε το εργαλείο Classification Learner. Τα δεδομένα εισόδου έπρεπε να χωριστούν σε δεδομένα εκπαίδευσης και δεδομένα ελέγχου. Μέσα από την χρήση διάφορων τρόπων επικύρωσης παρατηρήθηκε ότι η εκπαίδευση δεν πρέπει να ξεπερνάει ένα συγκεκριμένο όριο. Τέλος, απαριθμούνται οι μηχανές διανυσμάτων υποστήριξης, οι αλγόριθμοι κατηγοριοποίησης πλησιέστερου γείτονα και τα νευρωνικά δίκτυα που χρησιμοποιήθηκαν.

Κεφάλαιο 8: Αποτελέσματα πειραμάτων

8.1 Εισαγωγή

Σε αυτό το κεφάλαιο βλέπουμε τα αποτελέσματα των αλγορίθμων καθώς και τι παρατηρούμε από τα δεδομένα που προκύπτουν

Ένα πολύ σύνηθες κριτήριο που χρησιμοποιείται για να εξεταστεί η αποτελεσματικότητα του μοντέλου σε πειράματα αναγνώρισης, είναι η απόδοση κατηγοριοποίησης (Performance Rate). Η απόδοση αναγνώρισης ορίζεται ως το κλάσμα με αριθμητή τον αριθμό τον δειγμάτων που ταξινομήθηκαν ορθώς και παρανομαστή το σύνολο των δεδομένων εισόδου στο μοντέλο [30] .Στην σχέση 8.1 βλέπουμε τον μαθηματικό τύπο όπου P(%) είναι η απόδοση, N_{CCS} ο αριθμός δειγμάτων οπού ταξινομήθηκαν σωστά(Correctly classified samples) και N το σύνολο των δειγμάτων εισόδου.

$$\boldsymbol{P}(\%) = \frac{\boldsymbol{N_{CCS}}}{\boldsymbol{N}} \times \mathbf{100} \qquad (8.1)$$

Ένα άλλο κριτήριο που είναι εξίσου σημαντικό και άξιο ανάλυσης είναι το ποσοστό εσφαλμένης ταξινόμησης (False Rate) το οποίο είναι το ποσοστό του αριθμού των δειγμάτων που κατηγοριοποιήθηκαν εσφαλμένα ως προς το συνολικό αριθμό εισόδων στο μοντέλο. Συγκεκριμένα μας ενδιαφέρει το μερικό ποσοστό σφάλματος το οποίο βλέπουμε πως ορίζεται στην (8.2)

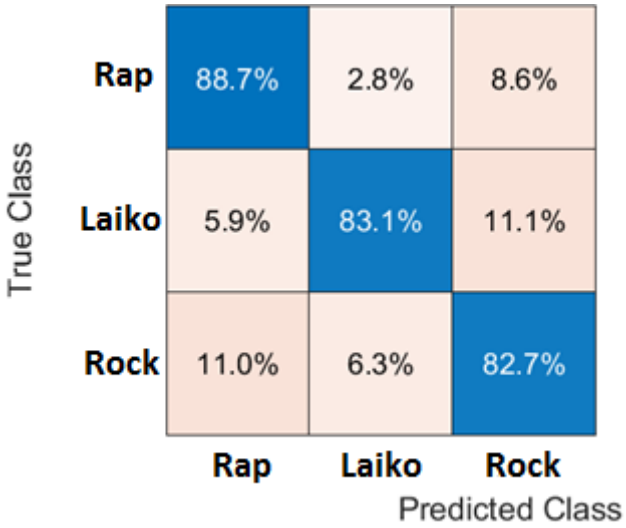
$$\boldsymbol{FR}_{\frac{Ci}{Cj}}(\%) = \boldsymbol{FR}\left(\frac{\boldsymbol{Ci}}{\boldsymbol{Cj}}\right) = \frac{\boldsymbol{N_{FN}}(\frac{\boldsymbol{Ci}}{\boldsymbol{Cj}})}{\boldsymbol{N_{Ci}}} * \mathbf{100} \qquad (8.2)$$

Το μερικό ποσοστό σφάλματος FR για την κλάση Ci ως προς την κλάση Cj ισούται με τον λόγο του αριθμού των δειγμάτων Nfn (Ci/Cj) της κλάσης Ci που λανθασμένα κατηγοριοποιήθηκαν στη κλάση Cj προς τον συνολικό αριθμό δειγμάτων Nci όπου περιλαμβάνονται στη κλάση Ci.

Τα σφάλματα αποτελούν εξίσου σημαντική πληροφορία διότι μπορούμε να ανιχνεύσουμε την λανθασμένη κατηγοριοποίηση μεταξύ δυο κλάσεων.

8.2 Μήτρα σύγχυσης (confusion matrix)

Τα ποσοστά της απόδοσης καθώς και τον σφαλμάτων εξάγονται μέσω της μήτρας σύγχυσης (confusion matrix) όπου προκύπτει μετά από όλη την διαδικασία εκπαίδευσης.



Εικόνα 8. 1 Παράδειγμα Μήτρας Σύγχυσης

Στην εικόνα 3.1 φαίνεται την μήτρα σύγχυσης όπου παράχθηκε από εκπαίδευση μοντέλου με τον αλγόριθμο Cubic SVM σε παράθυρο ανάλυσης 1 δευτερολέπτου με μέθοδο επικύρωσης 5 Cross Validation. Κάθετα είναι οι πραγματικές κλάσεις και οριζόντια είναι οι κλάσεις που προβλέφθηκαν. Στην μπλε διαγώνιο παρατηρείται το ποσοστό των δειγμάτων που ταξινομήθηκαν ορθώς. Στα υπόλοιπα πεδία υπάρχουν τα μερικά ποσοστά σφάλματος, παραδείγματος χάρη 11% των δειγμάτων ταξινομήθηκαν εσφαλμένα ως Rap ενώ η πραγματική τους κλάση ήταν Rock.

8.3 Αποτελέσματα Αλγορίθμων

Ακολουθούν όλα τα αποτελέσματα αλγορίθμων σε πίνακες, οι αποδόσεις καθώς και τα σφάλματα για όλους τους αλγόριθμους σε όλα τα παράθυρα ανάλυσης με όλους τους τρόπους επικύρωσης.

	P	P(rap)	P(laiko)	P(rock)	E(l>rock)	E(l>rap)	E(rock>rap)	E(rock>l)	E(rap>rock)	E(rap>l)
Fine KNN	87,00%	91,00%	85,20%	84,80%	8,90%	5,90%	10,70%	4,50%	6,80%	2,20%
Cubic SVM	86,80%	91,00%	84,50%	84,90%	10,20%	5,30%	9,60%	5,60%	6,70%	2,30%
Weighted KNN	85,87%	92,60%	80,80%	84,20%	11,60%	7,60%	12,70%	3,10%	5,90%	1,50%
Medium Gaussian SVM	85,53%	90,80%	82,20%	83,60%	12,80%	5,00%	10,50%	6,00%	7,30%	1,80%
Cosine KNN	84,50%	91,20%	84,90%	77,40%	8,10%	7,00%	14,80%	7,80%	5,60%	3,20%
Medium Neural Network	84,37%	88,80%	82,10%	82,20%	13,20%	4,70%	9,00%	8,80%	8,00%	3,20%
Wide Neural Network	83,90%	89,10%	81,70%	80,90%	12,50%	5,70%	10,40%	8,70%	7,90%	3,00%
Medium KNN	83,50%	92,40%	77,90%	80,20%	12,80%	9,30%	15,70%	4,10%	6,00%	1,60%
Bagged Trees	83,27%	90,20%	81,30%	78,30%	11,80%	6,90%	13,80%	7,90%	7,00%	2,80%
Quadratic SVM	82,97%	89,10%	78,90%	80,90%	15,40%	5,70%	11,70%	7,40%	8,40%	2,50%
Fine Gaussian SVM	82,40%	82,20%	94,00%	71,00%	2,90%	3,10%	6,90%	22,10%	3,20%	14,70%
Cubic KNN	82,30%	92,00%	76,00%	78,90%	13,80%	10,20%	16,80%	4,30%	6,20%	1,80%
Bilayered Neural Network	81,87%	87,80%	79,80%	78,00%	14,90%	5,30%	12,00%	10,00%	8,90%	3,30%
Trilayered Neural Network	81,73%	87,00%	79,40%	78,80%	15,50%	5,10%	10,90%	10,30%	9,80%	3,30%
Narrow Neural Network	81,13%	86,90%	78,30%	78,20%	16,70%	5,00%	12,30%	9,60%	10,40%	2,70%

Εικόνα 8. 2 Παράθυρα 2 δευτερολέπτων, 5 Cross Fold Validation

	P	P(rap)	P(laiko)	P(rock)	E(l>rock)	E(l>rap)	E(rock>rap)	E(rock>l)	E(rap>rock)	E(rap>l)
Fine KNN	87,90%	91,40%	86,50%	85,80%	8,10%	5,50%	10,30%	3,90%	6,50%	2,10%
Cubic SVM	87,30%	91,70%	84,90%	85,30%	9,90%	5,10%	9,10%	5,60%	6,30%	2,10%
Weighted KNN	86,50%	93,00%	81,70%	84,80%	11,50%	6,80%	12,00%	3,20%	5,70%	1,30%
Medium Gaussian SVM	86,10%	91,60%	82,70%	84,00%	12,20%	5,10%	10,30%	5,70%	6,70%	1,60%
Cosine KNN	84,97%	91,70%	86,30%	76,90%	7,20%	6,50%	14,60%	8,50%	5,20%	3,00%
Wide Neural Network	84,80%	89,10%	83,70%	81,60%	11,00%	5,30%	9,30%	9,10%	7,50%	3,40%
Medium KNN	84,50%	92,50%	80,30%	80,70%	11,30%	8,40%	15,10%	4,30%	5,90%	1,70%
Medium Neural Network	84,30%	90,10%	81,80%	81,00%	12,80%	5,40%	10,00%	9,00%	7,20%	2,70%
Bagged Trees	83,50%	90,40%	81,40%	78,70%	11,90%	6,70%	13,30%	8,00%	7,00%	2,60%
Cubic KNN	83,37%	92,30%	78,80%	79,00%	11,80%	9,40%	16,50%	4,50%	6,10%	1,60%
Quadratic SVM	83,20%	89,20%	79,40%	81,00%	15,20%	5,40%	11,50%	7,60%	8,30%	2,50%
Fine Gaussian SVM	82,87%	82,40%	94,20%	72,00%	2,80%	3,00%	6,50%	21,50%	3,30%	14,40%
Trilayered Neural Network	82,13%	87,50%	80,00%	78,90%	14,60%	5,40%	10,60%	10,50%	9,20%	3,30%
Bilayered Neural Network	81,97%	87,70%	79,10%	79,10%	15,30%	5,60%	11,30%	9,50%	9,30%	3,00%
Narrow Neural Network	81,07%	86,60%	77,50%	79,10%	17,10%	5,40%	11,70%	9,30%	10,60%	2,80%

Εικόνα 8. 3 Παράθυρο 2 δευτερολέπτων, 10 Cross Validation

	P	P(rap)	P(laiko)	P(rock)	E(l>rock)	E(l>rap)	E(rock>rap)	E(rock>l)	E(rap>rock)	E(rap>l)
Fine KNN	87,37%	91,30%	85,50%	85,30%	8,80%	5,70%	10,00%	4,60%	7,00%	1,70%
Cubic SVM	86,50%	90,20%	84,00%	85,30%	10,60%	5,30%	8,30%	6,30%	7,30%	2,50%
Weighted KNN	85,73%	92,30%	80,80%	84,10%	12,10%	7,20%	12,20%	3,60%	6,30%	1,30%
Medium Gaussian SVM	85,60%	90,70%	82,70%	83,40%	12,30%	5,00%	10,40%	6,20%	7,40%	1,90%
Medium Neural Network	84,30%	88,30%	82,70%	81,90%	12,80%	4,50%	8,80%	9,30%	7,70%	4,00%
Cosine KNN	84,27%	90,20%	82,90%	79,70%	9,70%	7,40%	13,20%	7,10%	6,90%	2,90%
Wide Neural Network	84,13%	88,70%	82,30%	81,40%	13,80%	3,80%	8,90%	9,80%	6,20%	5,10%
Quadratic SVM	83,53%	89,40%	80,00%	81,20%	14,80%	5,20%	11,40%	7,40%	8,30%	2,20%
Bilayered Neural Network	83,13%	88,10%	81,60%	79,70%	12,90%	5,50%	12,10%	8,20%	8,90%	2,90%
Medium KNN	82,93%	90,10%	76,10%	82,60%	14,70%	9,20%	12,90%	4,50%	8,20%	1,70%
Fine Gaussian SVM	82,77%	81,80%	93,80%	72,70%	3,30%	2,90%	6,90%	20,40%	4,00%	14,30%
Bagged Trees	82,40%	89,70%	80,10%	77,40%	12,10%	7,90%	14,70%	8,00%	7,00%	3,30%
Cubic KNN	82,10%	89,90%	74,80%	81,60%	15,40%	9,80%	13,80%	4,60%	8,20%	1,90%
Trilayered Neural Network	82,03%	87,20%	77,70%	81,20%	16,40%	6,00%	10,00%	8,80%	10,50%	2,30%
Narrow Neural Network	81,30%	85,50%	80,90%	77,50%	14,70%	4,50%	11,60%	11,00%	10,50%	4,00%

Εικόνα 8. 4 Παράθυρο 2 δευτερολέπτων, 30 Holdout Validation

	P	P(rap)	P(laiko)	P(rock)	E(l>rock)	E(l>rap)	E(rock>rap)	E(rock>l)	E(rap>rock)	E(rap>l)
Cubic SVM	86,17%	90,90%	83,40%	84,20%	10,30%	6,30%	9,50%	6,40%	7,00%	2,10%
Fine KNN	85,53%	90,70%	83,30%	82,60%	10,50%	6,20%	12,60%	4,80%	7,00%	2,30%
Medium Gaussian SVM	85,20%	90,70%	81,60%	83,30%	12,70%	5,70%	10,00%	6,70%	7,70%	1,60%
Weighted KNN	84,70%	92,00%	79,50%	82,60%	12,20%	8,30%	13,80%	3,60%	6,20%	1,70%
Medium Neural Network	83,37%	87,30%	80,40%	82,40%	14,00%	5,60%	8,00%	9,60%	9,20%	3,50%
Quadratic SVM	83,37%	89,40%	78,80%	81,90%	14,70%	6,50%	10,50%	7,60%	7,90%	2,70%
Cosine KNN	83,13%	89,30%	82,50%	77,60%	9,40%	8,10%	14,00%	8,30%	6,80%	3,90%
Wide Neural Network	83,10%	87,10%	82,00%	80,20%	12,70%	5,30%	9,40%	10,30%	8,00%	4,90%
Bagged Trees	82,00%	88,90%	80,00%	77,10%	12,10%	7,90%	14,00%	8,90%	7,90%	3,30%
Medium KNN	81,97%	89,40%	74,40%	82,10%	14,80%	10,80%	13,80%	4,10%	8,60%	2,00%
Fine Gaussian SVM	81,87%	81,70%	93,10%	70,80%	3,00%	3,90%	7,90%	21,30%	3,30%	15,00%
Bilayered Neural Network	81,67%	86,50%	78,10%	80,40%	15,80%	6,00%	9,10%	10,50%	10,60%	2,90%
Trilayered Neural Network	81,23%	85,80%	81,90%	76,00%	12,30%	5,80%	10,60%	13,40%	8,80%	5,40%
Cubic KNN	81,17%	88,70%	73,80%	81,00%	14,50%	11,70%	15,20%	3,80%	9,00%	2,20%
Narrow Neural Network	80,77%	88,00%	76,40%	77,90%	17,20%	6,40%	12,30%	9,80%	8,70%	3,20%

Εικόνα 8. 5 Παράθυρο 2 δευτερολέπτων, 40 Holdout Validation

	P	P(rap)	P(laiko)	P(rock)	E(l>rock)	E(l>rap)	E(rock>rap)	E(rock>l)	E(rap>rock)	E(rap>l)
Cubic SVM	84,83%	88,70%	83,10%	82,70%	11,10%	5,90%	11,00%	6,30%	8,60%	2,80%
Weighted KNN	84,00%	88,50%	80,20%	83,30%	12,20%	7,50%	13,10%	3,60%	9,50%	2,00%
Fine KNN	83,90%	85,80%	83,30%	82,60%	9,80%	6,90%	12,20%	5,10%	10,80%	3,40%
Wide Neural Network	83,37%	86,90%	83,00%	80,20%	11,30%	5,70%	10,70%	9,10%	8,70%	4,40%
Medium Gaussian SVM	82,67%	87,40%	80,30%	80,30%	13,10%	6,60%	13,50%	6,10%	9,80%	2,80%
Cosine KNN	82,43%	88,70%	84,10%	74,50%	7,60%	8,30%	17,20%	8,30%	7,50%	3,90%
Medium KNN	82,10%	88,10%	78,50%	79,70%	11,90%	9,60%	15,80%	4,50%	9,50%	2,40%
Fine Gaussian SVM	81,60%	85,10%	91,80%	67,90%	3,60%	4,70%	12,70%	19,40%	3,60%	11,40%
Bagged Trees	81,13%	87,70%	79,90%	75,80%	12,20%	7,80%	15,90%	8,30%	8,90%	3,40%
Medium Neural Network	80,83%	86,10%	79,10%	77,30%	14,00%	6,90%	13,40%	9,30%	9,80%	4,10%
Cubic KNN	80,63%	87,30%	76,10%	78,50%	13,30%	10,60%	17,00%	4,50%	10,00%	2,70%
Quadratic SVM	79,77%	85,90%	76,20%	77,20%	16,50%	7,40%	15,00%	7,80%	11,00%	3,10%
Trilayered Neural Network	78,53%	83,90%	76,90%	74,80%	16,50%	6,60%	14,60%	10,70%	12,30%	3,70%
Bilayered Neural Network	77,73%	82,80%	77,20%	73,20%	16,20%	6,70%	14,70%	12,20%	12,40%	4,70%
Narrow Neural Network	77,07%	83,00%	75,50%	72,70%	16,80%	7,70%	16,50%	10,80%	12,30%	4,70%

Εικόνα 8. 6 Παράθυρο 1 δευτερολέπτου, 5 Cross Fold Validation

	P	P(rap)	P(laiko)	P(rock)	E(l>rock)	E(l>rap)	E(rock>rap)	E(rock>l)	E(rap>rock)	E(rap>l)
Cubic SVM	85,00%	88,90%	83,20%	82,90%	10,90%	5,90%	11,00%	6,00%	8,60%	2,50%
Weighted KNN	84,33%	88,70%	80,70%	83,60%	12,00%	7,30%	12,90%	3,50%	9,30%	2,00%
Fine KNN	84,30%	86,40%	83,50%	83,00%	9,90%	6,60%	11,80%	5,20%	10,20%	3,30%
Wide Neural Network	83,93%	87,50%	83,80%	80,50%	11,00%	5,30%	10,60%	8,90%	8,50%	4,10%
Medium Gaussian SVM	83,17%	87,80%	80,80%	80,90%	12,60%	6,50%	13,20%	6,00%	9,60%	2,60%
Cosine KNN	82,87%	88,80%	84,60%	75,20%	7,50%	7,90%	16,80%	8,00%	7,20%	4,00%
Medium KNN	82,60%	88,20%	79,50%	80,10%	11,60%	8,90%	15,50%	4,40%	9,30%	2,60%
Fine Gaussian SVM	81,97%	84,90%	91,80%	69,20%	3,80%	4,50%	11,60%	19,10%	3,80%	11,30%
Bagged Trees	81,57%	87,70%	80,00%	77,00%	12,10%	7,90%	15,00%	8,00%	8,80%	3,40%
Cubic KNN	81,30%	87,60%	77,50%	78,80%	12,40%	10,00%	16,50%	4,70%	9,70%	2,70%
Medium Neural Network	80,73%	85,90%	79,30%	77,00%	14,10%	6,60%	12,90%	10,10%	10,10%	4,00%
Quadratic SVM	79,93%	86,00%	76,30%	77,50%	16,40%	7,30%	14,90%	7,60%	11,00%	3,00%
Trilayered Neural Network	78,30%	83,40%	76,90%	74,60%	16,40%	6,70%	14,70%	10,70%	12,50%	4,10%
Bilayered Neural Network	78,27%	83,80%	76,70%	74,30%	16,80%	6,50%	15,00%	10,70%	12,10%	4,10%
Narrow Neural Network	77,07%	83,10%	74,30%	73,80%	18,10%	7,60%	16,10%	10,10%	13,00%	3,90%

Εικόνα 8. 7 Παράθυρο 1 δευτερολέπτου, 10 Cross Fold Validation

	P	P(rap)	P(laiko)	P(rock)	E(l>rock)	E(l>rap)	E(rock>rap)	E(rock>l)	E(rap>rock)	E(rap>l)
Cubic SVM	84,07%	87,40%	81,80%	83,00%	12,30%	5,80%	10,70%	6,30%	9,40%	3,20%
Weighted KNN	82,53%	86,40%	78,40%	82,80%	13,50%	8,10%	13,00%	4,20%	11,20%	2,40%
Wide Neural Network	82,40%	87,00%	80,50%	79,70%	13,30%	6,20%	11,90%	8,30%	9,60%	3,40%
Fine KNN	82,37%	83,30%	82,00%	81,80%	11,10%	6,90%	12,30%	5,80%	12,70%	3,90%
Medium Gaussian SVM	81,67%	85,80%	78,90%	80,30%	14,90%	6,30%	13,20%	6,40%	11,20%	3,00%
Cosine KNN	81,57%	87,30%	83,10%	74,30%	8,70%	8,20%	17,10%	8,60%	8,50%	4,20%
Medium Neural Network	80,90%	85,40%	78,90%	78,40%	14,50%	6,50%	12,90%	8,70%	10,90%	3,60%
Medium KNN	80,83%	86,40%	77,00%	79,10%	12,70%	10,20%	15,90%	5,00%	11,00%	2,60%
Fine Gaussian SVM	80,43%	83,80%	90,90%	66,60%	4,00%	5,10%	12,90%	20,50%	4,70%	11,40%
Bagged Trees	80,23%	87,10%	79,00%	74,60%	12,50%	8,50%	16,60%	8,80%	9,20%	3,70%
Cubic KNN	79,63%	85,60%	75,40%	77,90%	13,50%	11,10%	16,80%	5,30%	11,40%	3,00%
Quadratic SVM	79,47%	85,30%	75,80%	77,30%	16,60%	7,60%	14,70%	8,00%	11,80%	2,90%
Trilayered Neural Network	78,23%	81,00%	76,80%	76,90%	17,10%	6,10%	13,00%	10,10%	14,70%	4,30%
Bilayered Neural Network	77,47%	81,70%	74,10%	76,60%	19,50%	6,40%	13,20%	10,20%	14,30%	4,00%
Narrow Neural Network	76,90%	83,30%	73,70%	73,70%	16,90%	9,30%	17,00%	9,40%	13,00%	3,70%

Εικόνα 8. 8 Παράθυρο 1 δευτερολέπτου, 30 Holdout Validation

	P	P(rap)	P(laiko)	P(rock)	E(l>rock)	E(l>rap)	E(rock>rap)	E(rock>l)	E(rap>rock)	E(rap>l)
Cubic SVM	83,60%	87,50%	81,70%	81,60%	12,10%	6,20%	11,20%	7,20%	9,60%	3,00%
Weighted KNN	82,70%	86,80%	78,40%	82,90%	13,70%	8,00%	13,10%	4,00%	10,80%	2,40%
Fine KNN	82,57%	84,50%	81,10%	82,10%	11,50%	7,40%	12,30%	5,60%	11,90%	3,60%
Wide Neural Network	82,23%	84,50%	82,20%	80,00%	12,30%	5,50%	10,10%	9,90%	9,60%	5,80%
Medium Gaussian SVM	81,37%	85,70%	79,10%	79,30%	14,10%	6,80%	14,00%	6,70%	11,30%	3,00%
Cosine KNN	81,23%	87,00%	83,50%	73,20%	8,60%	7,90%	17,50%	9,40%	8,20%	4,70%
Medium KNN	80,90%	86,40%	77,10%	79,20%	13,60%	9,40%	15,90%	4,90%	10,60%	3,00%
Fine Gaussian SVM	79,97%	85,70%	90,50%	63,70%	3,60%	5,90%	16,20%	20,10%	3,70%	10,70%
Bagged Trees	79,90%	87,10%	78,10%	74,50%	13,30%	8,60%	16,50%	9,00%	9,40%	3,50%
Cubic KNN	79,67%	85,90%	75,50%	77,60%	14,00%	10,50%	16,90%	5,40%	11,00%	3,20%
Medium Neural Network	79,63%	86,00%	77,10%	75,80%	15,50%	7,40%	14,40%	9,90%	10,70%	3,30%
Quadratic SVM	79,53%	84,80%	76,50%	77,30%	16,10%	7,40%	14,70%	8,00%	11,70%	3,50%
Bilayered Neural Network	77,90%	81,90%	76,90%	74,90%	16,40%	6,80%	14,20%	11,00%	13,60%	4,50%
Trilayered Neural Network	77,80%	83,10%	77,20%	73,10%	15,10%	7,70%	15,40%	11,50%	12,30%	4,60%
Narrow Neural Network	76,30%	80,20%	77,00%	71,70%	16,30%	6,70%	16,30%	12,00%	13,60%	6,30%

Εικόνα 8. 9 Παράθυρο 1 δευτερολέπτου, 40 Holdout Validation

	P	P(rap)	P(laiko)	P(rock)	E(l>rock)	E(l>rap)	E(rock>rap)	E(rock>l)	E(rap>rock)	E(rap>l)
Cubic SVM	83,07%	86,80%	82,50%	79,90%	10,20%	7,30%	12,80%	7,30%	9,50%	3,70%
Weighted KNN	82,83%	87,40%	79,00%	82,10%	11,80%	9,20%	13,30%	4,60%	9,60%	3,00%
Wide Neural Network	82,47%	86,00%	83,70%	77,70%	9,60%	6,80%	12,70%	9,60%	8,90%	5,10%
Fine KNN	81,67%	84,00%	80,90%	80,10%	10,60%	8,50%	12,90%	7,00%	11,00%	5,10%
Medium Gaussian SVM	81,30%	86,70%	81,00%	76,20%	10,30%	8,70%	16,30%	7,50%	9,50%	3,80%
Medium KNN	81,30%	86,90%	78,50%	78,50%	11,00%	10,50%	16,10%	5,40%	9,60%	3,60%
Cosine KNN	80,77%	88,40%	82,50%	71,40%	7,00%	10,50%	19,40%	9,20%	6,70%	5,00%
Cubic KNN	80,10%	86,10%	76,80%	77,40%	11,70%	11,40%	16,90%	5,80%	10,00%	3,90%
Bagged Trees	79,13%	87,70%	79,60%	70,10%	9,80%	10,60%	19,60%	10,30%	7,70%	4,60%
Medium Neural Network	78,27%	83,10%	79,20%	72,50%	12,90%	7,90%	16,20%	11,30%	11,50%	5,40%
Quadratic SVM	77,53%	83,90%	76,50%	72,20%	13,70%	9,80%	18,30%	9,50%	11,50%	4,50%
Fine Gaussian SVM	77,33%	95,00%	80,30%	56,70%	2,60%	17,10%	34,70%	8,60%	2,70%	2,30%
Trilayered Neural Network	75,60%	81,00%	77,40%	68,40%	13,40%	9,20%	19,00%	12,60%	12,60%	6,30%
Bilayered Neural Network	75,23%	81,40%	75,60%	68,70%	14,90%	9,50%	18,70%	12,60%	12,70%	5,90%
Narrow Neural Network	74,20%	80,90%	74,90%	66,80%	15,00%	10,10%	20,80%	12,50%	13,10%	6,00%

Εικόνα 8. 10 Παράθυρο 1/2 δευτερολέπτου, 5 Cross Fold Validation

	P	P(rap)	P(laiko)	P(rock)	E(l>rock)	E(l>rap)	E(rock>rap)	E(rock>l)	E(rap>rock)	E(rap>l)
Cubic SVM	83,33%	87,10%	82,80%	80,10%	9,90%	7,30%	12,70%	7,30%	9,40%	3,50%
Weighted KNN	83,43%	87,80%	80,00%	82,50%	11,20%	8,80%	12,90%	4,70%	9,30%	2,90%
Wide Neural Network	82,53%	86,30%	83,20%	78,10%	9,80%	7,00%	12,20%	9,70%	8,50%	5,20%
Fine KNN	82,10%	84,30%	81,60%	80,40%	10,10%	8,20%	12,70%	6,90%	10,90%	4,80%
Medium Gaussian SVM	81,57%	86,90%	81,20%	76,60%	10,10%	8,60%	16,10%	7,40%	9,40%	3,70%
Medium KNN	81,80%	87,50%	78,90%	79,00%	10,50%	10,60%	15,50%	5,50%	9,30%	3,20%
Cosine KNN	81,20%	89,30%	82,30%	72,00%	6,90%	10,80%	19,20%	8,80%	6,40%	4,30%
Cubic KNN	80,60%	86,90%	77,10%	77,80%	11,30%	11,60%	16,50%	5,70%	9,70%	3,40%
Bagged Trees	79,90%	88,10%	80,50%	71,10%	9,50%	10,00%	19,00%	9,90%	7,50%	4,40%
Medium Neural Network	78,27%	83,70%	79,30%	71,80%	12,00%	8,70%	16,50%	11,70%	10,30%	6,00%
Quadratic SVM	77,70%	83,90%	76,80%	72,40%	13,60%	9,60%	18,20%	9,40%	11,60%	4,50%
Fine Gaussian SVM	79,03%	94,90%	83,80%	58,40%	2,50%	13,70%	32,40%	9,20%	2,60%	2,50%
Trilayered Neural Network	75,60%	80,80%	76,30%	69,70%	14,60%	9,10%	18,20%	12,10%	13,30%	5,80%
Bilayered Neural Network	75,03%	81,50%	76,10%	67,50%	13,70%	10,20%	20,00%	12,50%	12,30%	6,20%
Coarse KNN	74,20%	80,80%	65,30%	76,50%	20,90%	13,80%	19,10%	4,40%	15,80%	3,40%

Εικόνα 8. 11 Παράθυρο 1/2 δευτερολέπτου, 10 Cross Fold Validation

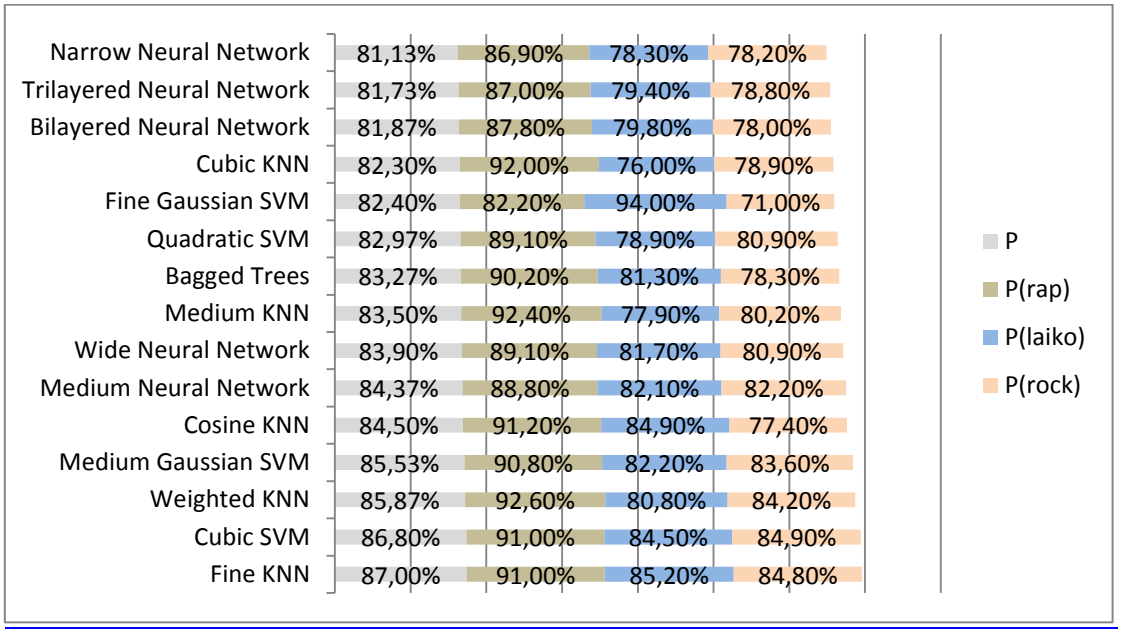
	P	P(rap)	P(laiko)	P(rock)	E(l>rock)	E(l>rap)	E(rock>rap)	E(rock>l)	E(rap>rock)	E(rap>l)
Cubic SVM	82,60%	86,30%	82,80%	78,70%	9,60%	7,60%	13,70%	7,60%	9,90%	3,90%
Weighted KNN	82,20%	86,60%	78,70%	81,30%	11,90%	9,40%	13,70%	5,00%	10,20%	3,10%
Wide Neural Network	81,90%	85,20%	83,70%	76,80%	9,70%	6,50%	12,20%	10,90%	9,10%	5,80%
Fine KNN	81,03%	83,30%	80,80%	79,00%	10,20%	9,10%	14,00%	7,00%	11,20%	5,60%
Medium Gaussian SVM	81,00%	85,90%	81,10%	76,00%	9,90%	9,00%	16,60%	7,40%	10,40%	3,80%
Medium KNN	80,87%	86,50%	77,90%	78,20%	10,90%	11,20%	16,20%	5,60%	10,10%	3,40%
Cosine KNN	80,23%	88,60%	81,60%	70,50%	6,80%	11,60%	20,70%	8,80%	6,70%	4,70%
Cubic KNN	79,53%	85,60%	76,20%	76,80%	11,40%	12,30%	17,30%	5,90%	10,70%	3,70%
Bagged Trees	78,73%	87,30%	79,20%	69,70%	9,80%	11,00%	19,50%	10,80%	8,20%	4,50%
Medium Neural Network	78,57%	82,70%	79,30%	73,70%	13,20%	7,50%	15,60%	10,70%	12,30%	4,90%
Quadratic SVM	77,33%	83,50%	76,50%	72,00%	13,30%	10,10%	18,60%	9,40%	11,90%	4,60%
Fine Gaussian SVM	76,03%	95,20%	78,10%	54,80%	2,00%	19,90%	37,20%	8,10%	2,70%	2,20%
Bilayered Neural Network	75,17%	80,90%	75,90%	68,70%	13,30%	10,80%	19,70%	11,60%	13,10%	6,00%
Trilayered Neural Network	74,60%	78,80%	76,40%	68,60%	14,50%	9,10%	20,10%	11,30%	14,60%	6,60%
Narrow Neural Network	74,20%	80,10%	74,50%	68,00%	15,80%	9,70%	19,80%	12,10%	14,00%	5,90%

Εικόνα 8. 12 Παράθυρο 1/2 δευτερολέπτου, 30 Holdout Validation

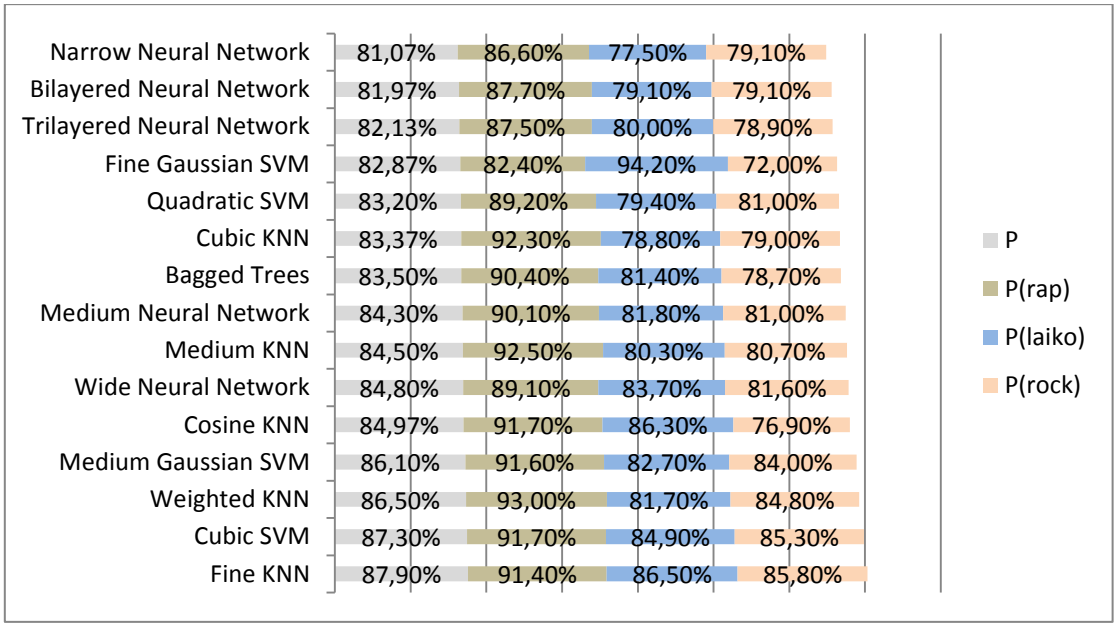
	P	P(rap)	P(laiko)	P(rock)	E(l>rock)	E(l>rap)	E(rock>rap)	E(rock>l)	E(rap>rock)	E(rap>l)
Cubic SVM	82,37%	86,80%	81,80%	78,50%	10,80%	7,40%	13,60%	7,80%	9,20%	4,00%
Weighted KNN	81,23%	86,20%	77,70%	79,80%	13,20%	9,20%	15,00%	5,10%	10,60%	3,10%
Wide Neural Network	80,80%	84,70%	82,30%	75,40%	10,30%	7,40%	14,30%	10,40%	9,30%	6,00%
Medium Gaussian SVM	80,13%	86,80%	79,70%	73,90%	11,40%	8,90%	18,40%	7,70%	9,20%	3,90%
Medium KNN	79,83%	84,60%	77,40%	77,50%	13,20%	9,40%	16,20%	6,30%	11,50%	3,90%
Fine KNN	79,77%	81,70%	79,90%	77,70%	11,10%	9,00%	14,70%	7,60%	12,10%	6,20%
Cosine KNN	79,13%	86,80%	81,60%	69,00%	8,50%	10,00%	20,60%	10,40%	7,70%	5,50%
Cubic KNN	78,73%	83,60%	76,00%	76,60%	13,70%	10,40%	16,60%	6,80%	11,90%	4,40%
Bagged Trees	77,97%	87,50%	78,90%	67,50%	10,40%	10,80%	22,00%	10,40%	7,80%	4,70%
Medium Neural Network	77,30%	83,50%	78,60%	69,80%	13,10%	8,30%	18,00%	12,20%	9,90%	6,50%
Quadratic SVM	76,70%	84,00%	75,20%	70,90%	14,70%	10,10%	19,90%	9,20%	11,30%	4,70%
Bilayered Neural Network	75,30%	81,30%	75,50%	69,10%	15,40%	9,10%	18,70%	12,10%	12,80%	5,90%
Trilayered Neural Network	75,00%	82,40%	76,70%	65,90%	13,50%	9,90%	20,90%	13,30%	11,50%	6,00%
Narrow Neural Network	73,70%	81,30%	75,20%	64,60%	14,90%	9,90%	22,10%	13,30%	11,40%	7,30%
Fine Gaussian SVM	73,53%	95,60%	73,70%	51,30%	2,70%	23,60%	41,20%	7,40%	2,30%	2,00%

Εικόνα 8. 13 Παράθυρο 1/2 δευτερολέπτου, 40 Holdout Validation

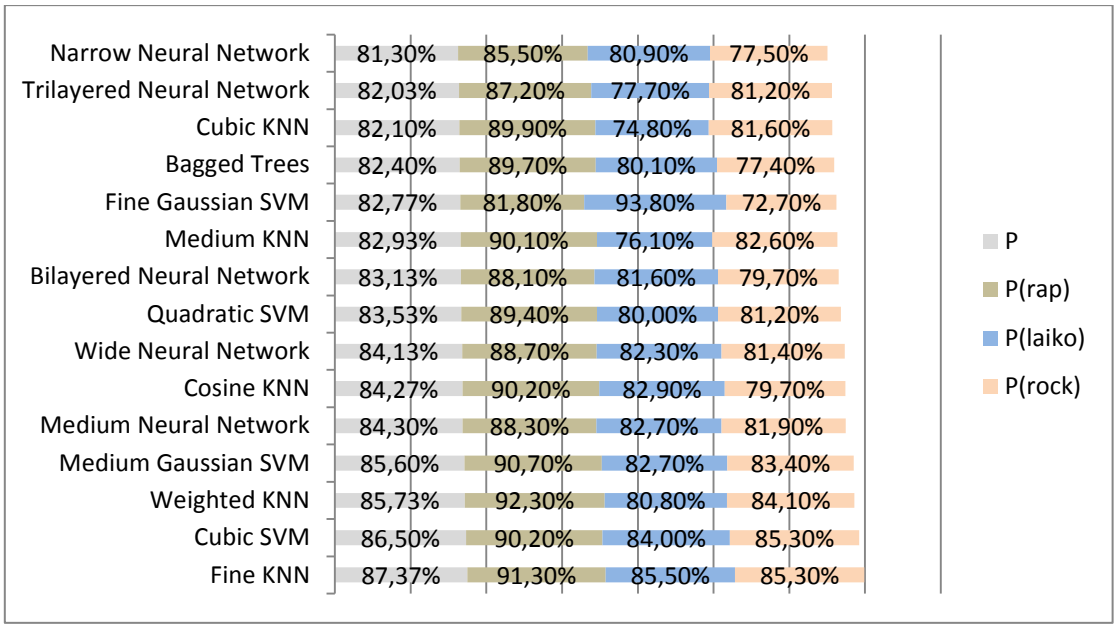
Ακολουθούν τα ποσοστά αποδόσεων σε μορφή ραβδογραμμάτων.



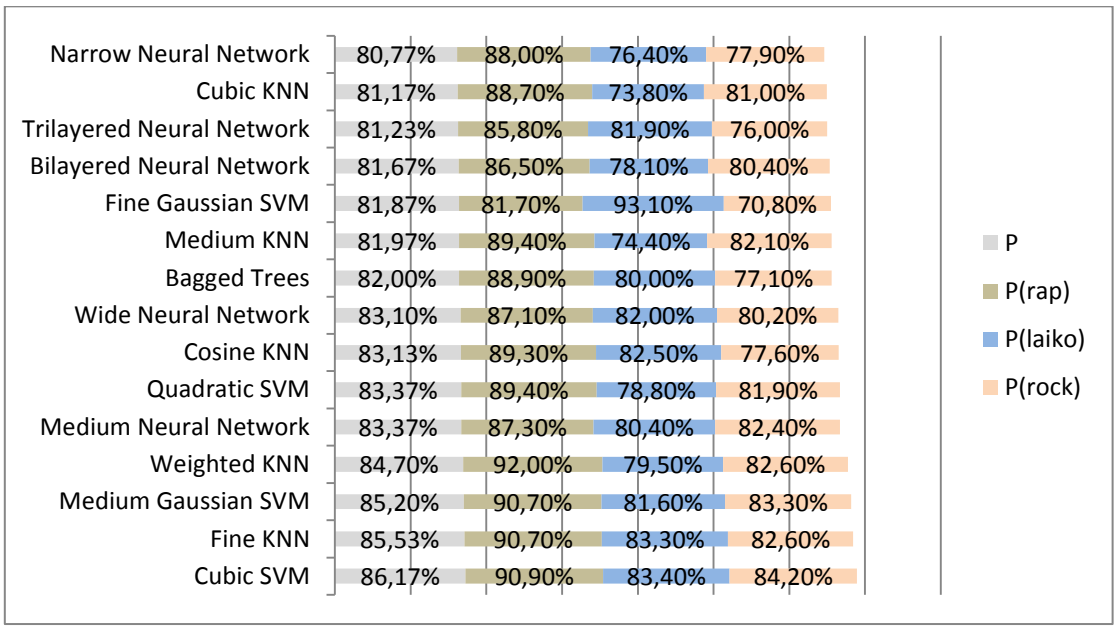
Εικόνα 8. 14 Παράθυρο 2 δευτερολέπτων, 5 Cross Fold Validation



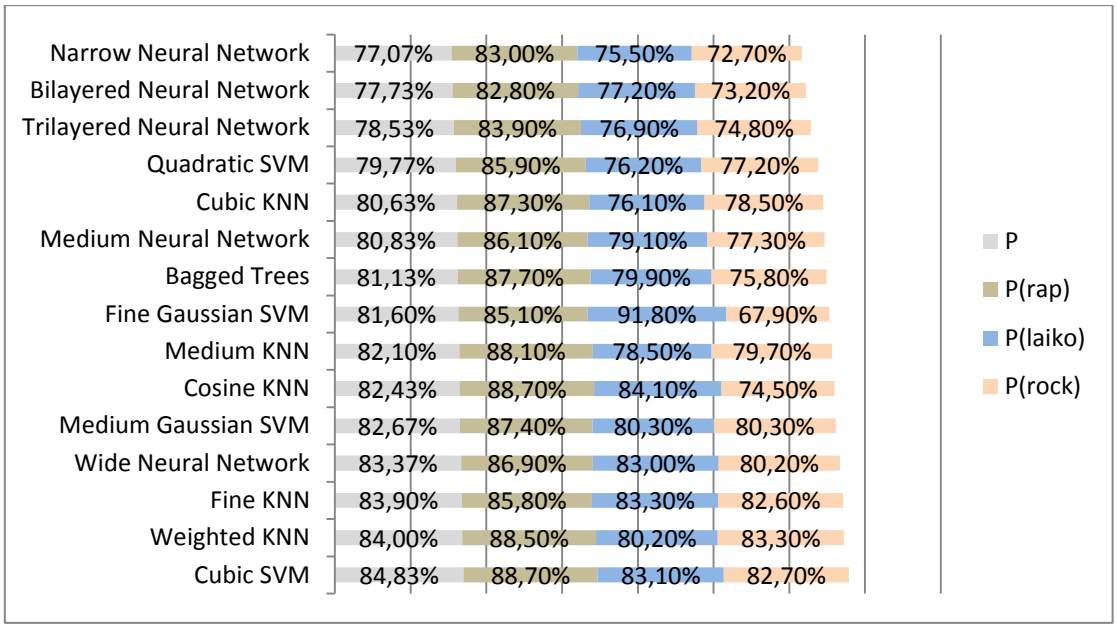
Εικόνα 8. 15 Παράθυρο 2 δευτερολέπτων, 10 Cross Fold Validation



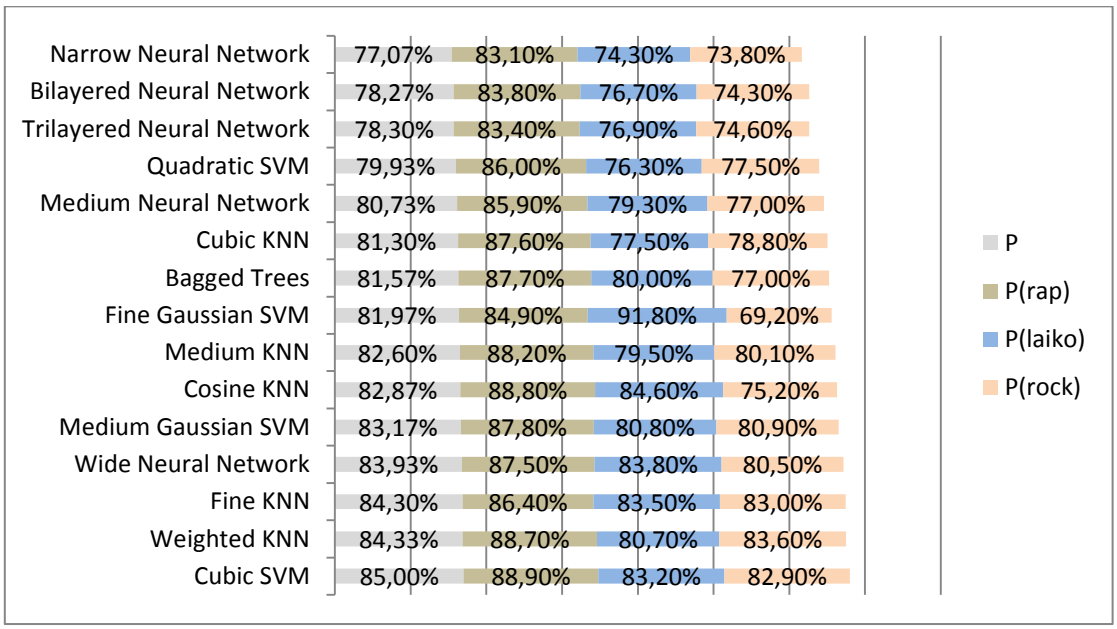
Εικόνα 8. 16 Παράθυρο 2 δευτερολέπτων, 30 Holdout Validation



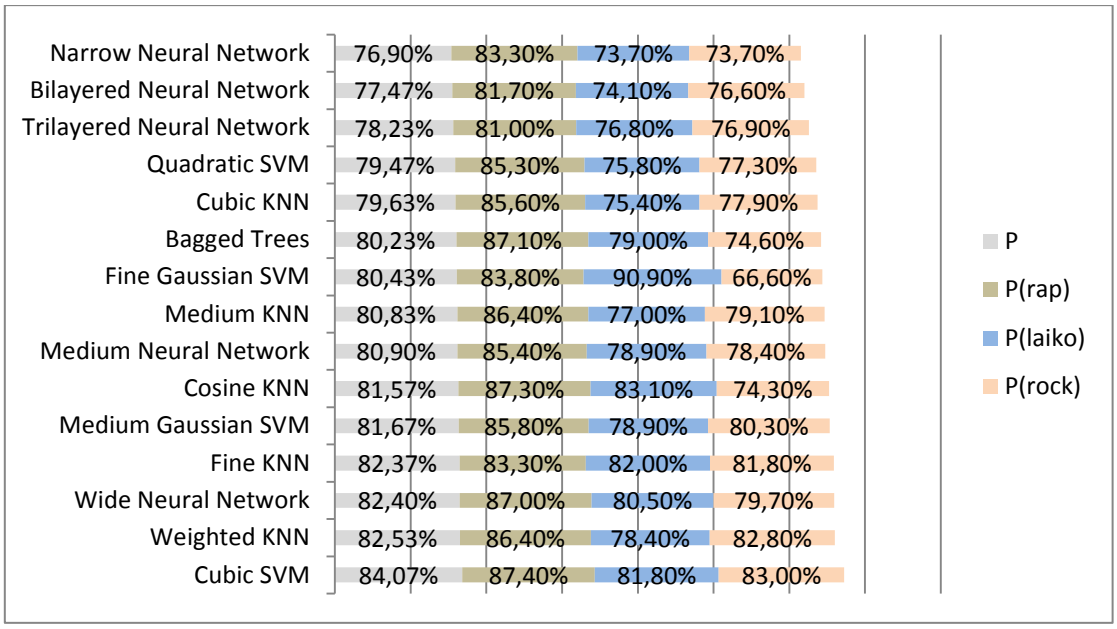
Εικόνα 8. 17 Παράθυρο 2 δευτερολέπτων, 40 Holdout Validation



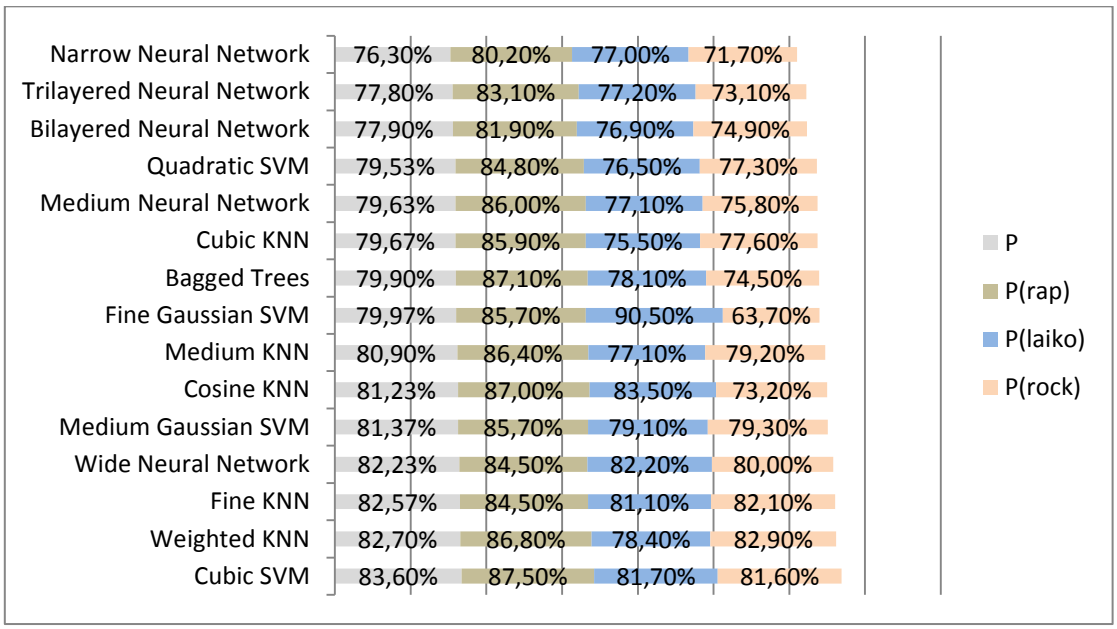
Εικόνα 8. 18 Παράθυρο 1 δευτερολέπτου, 5 Cross Fold Validation



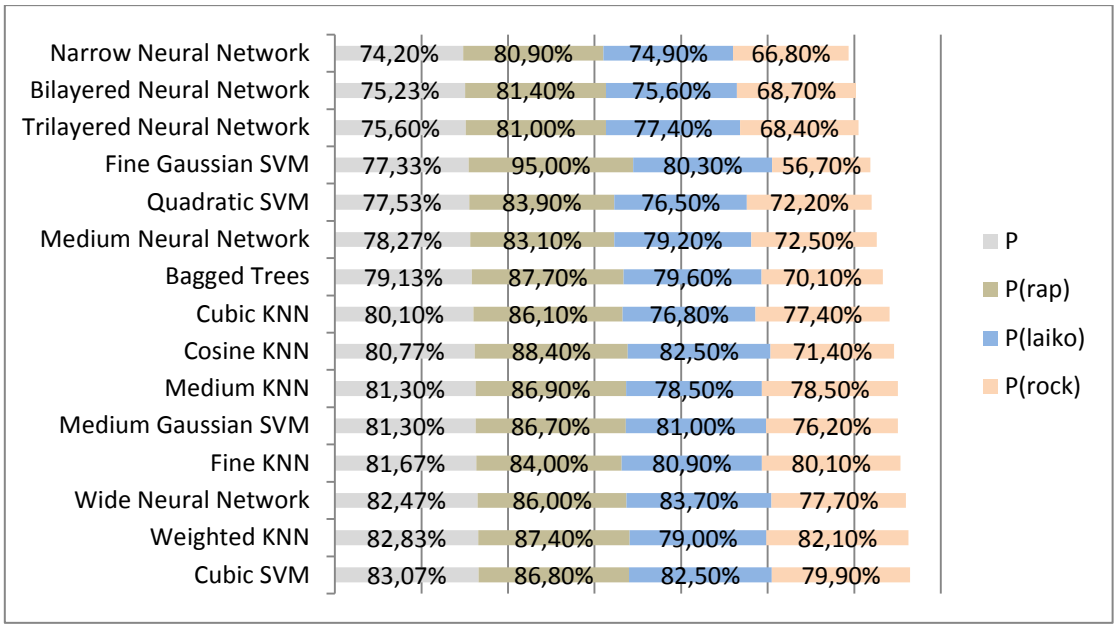
Εικόνα 8. 19 Παράθυρο 1 δευτερολέπτου, 10 Cross Fold Validation



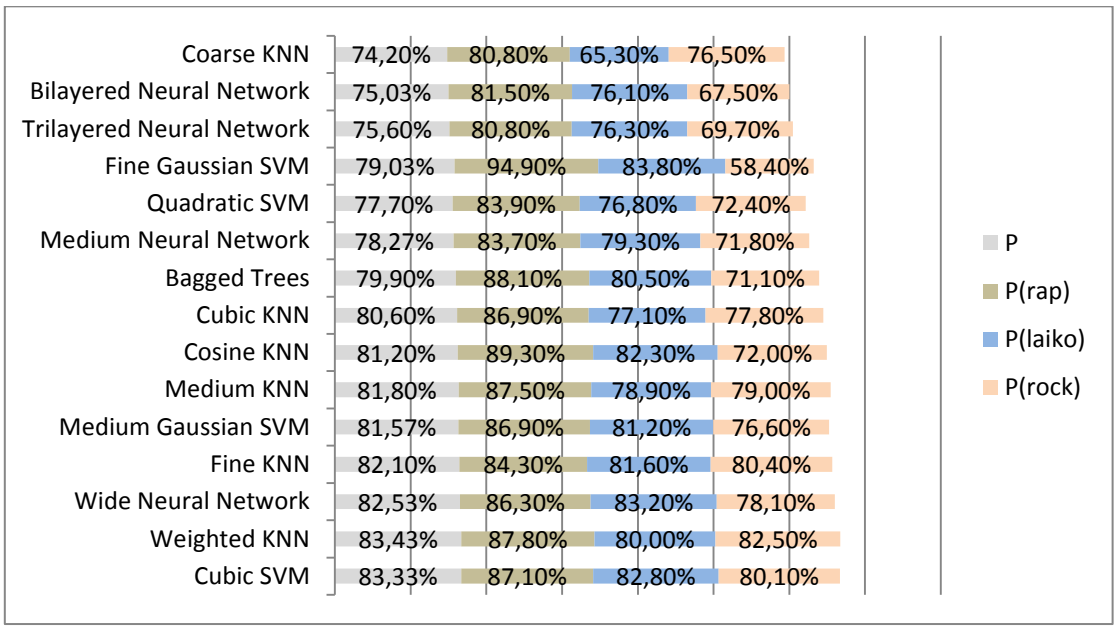
Εικόνα 8. 20 Παράθυρο 1 δευτερολέπτου, 30 Holdout Validation



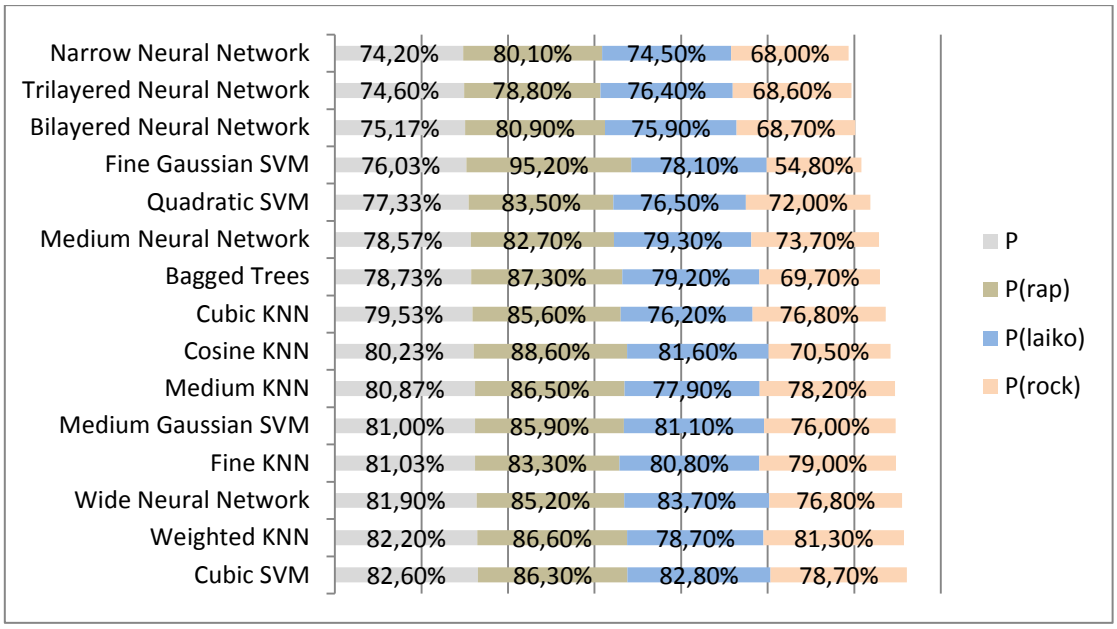
Εικόνα 8. 21 Παράθυρο 1 δευτερολέπτου, 40 Holdout Validation



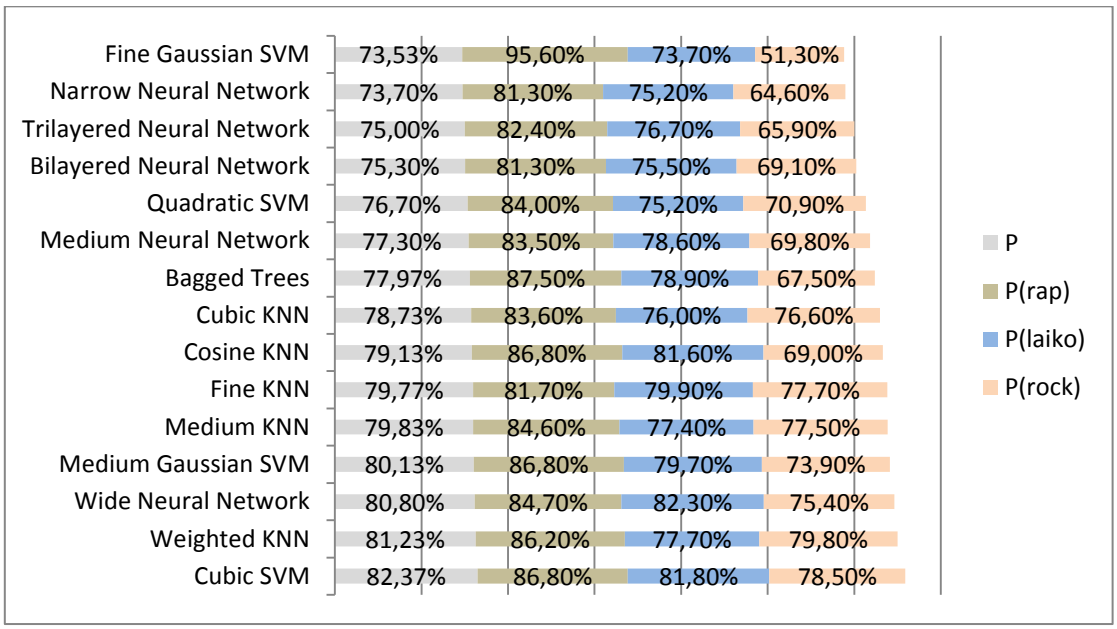
Εικόνα 8. 22 Παράθυρο 1/2 δευτερολέπτου, 5 Cross Fold Validation



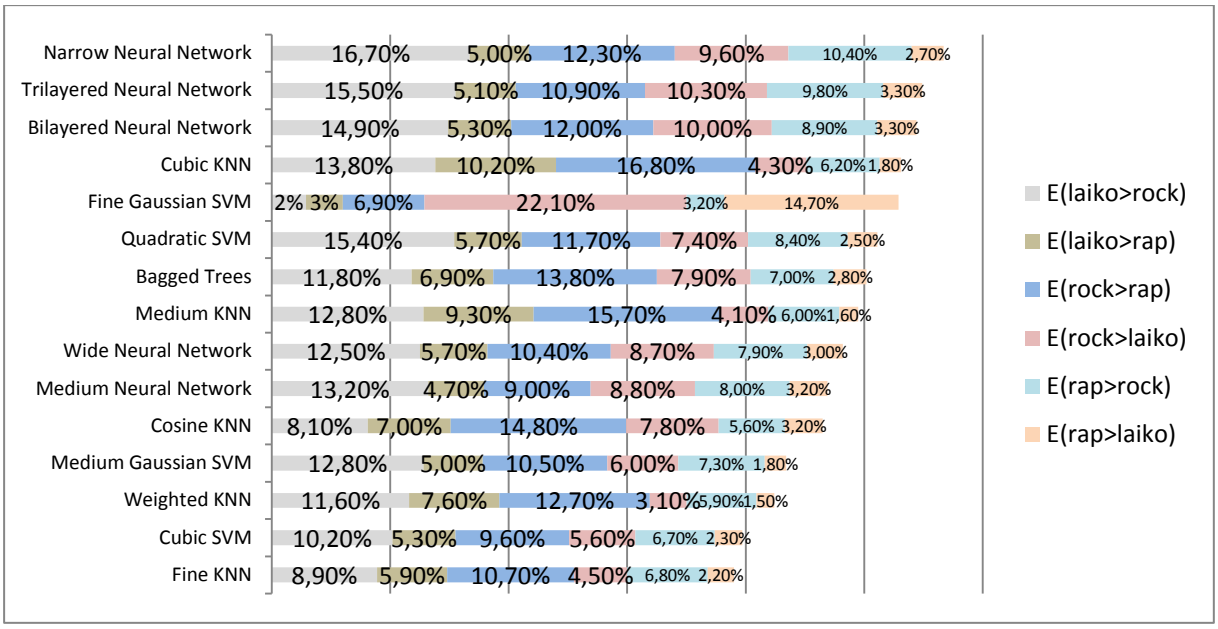
Εικόνα 8. 23 Παράθυρο 1/2 δευτερολέπτου, 10 Cross Fold Validation



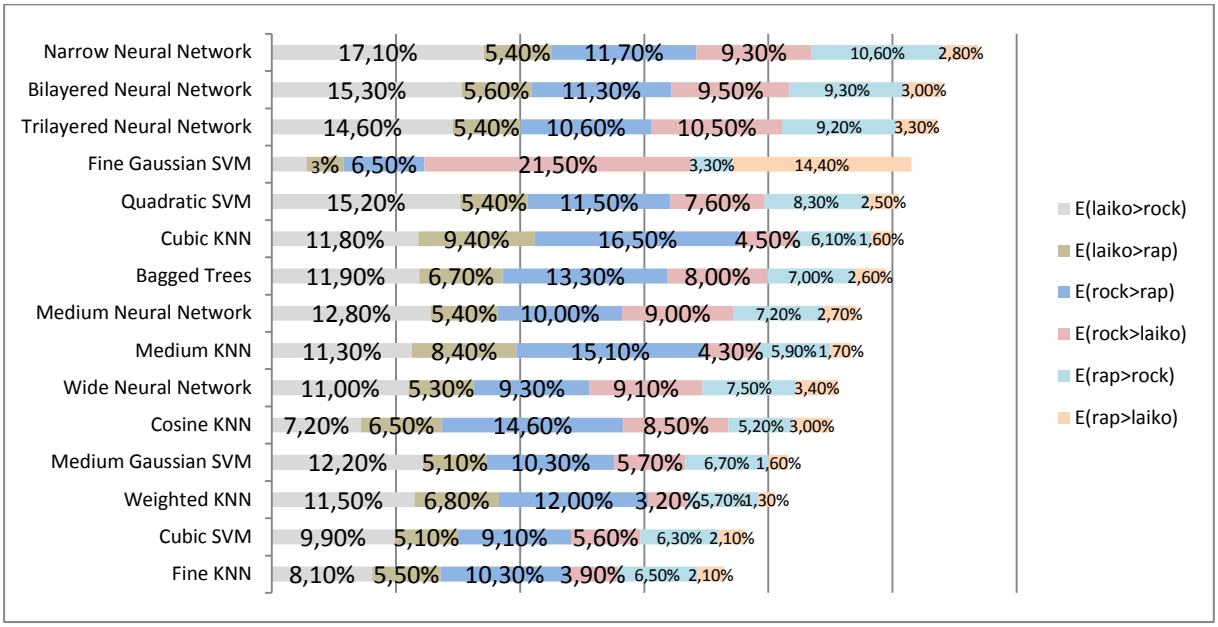
Εικόνα 8. 24 Παράθυρο 1/2 δευτερολέπτου, 30 Holdout Validation



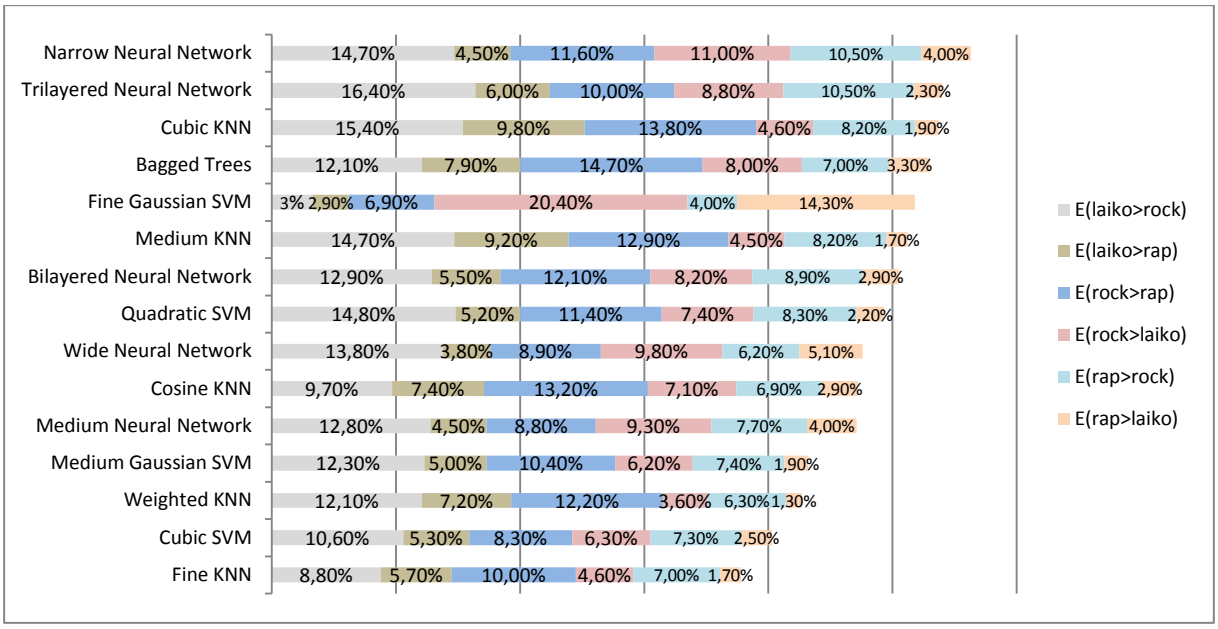
Εικόνα 8. 25 Παράθυρο 1/2 δευτερολέπτου, 40 Holdout Validation



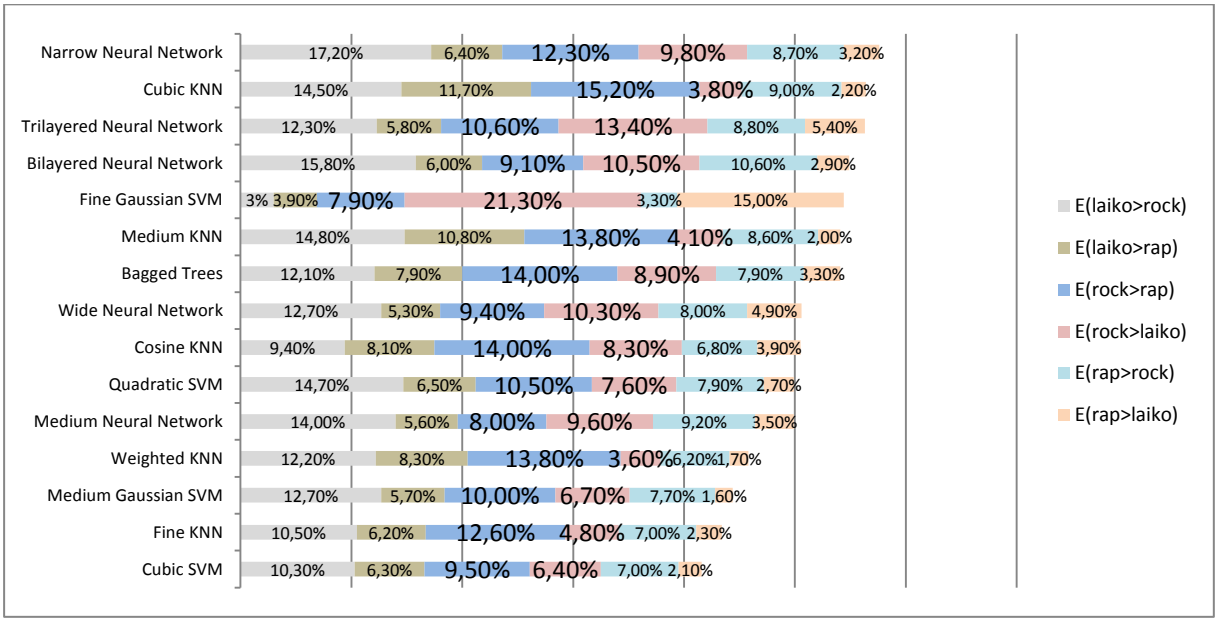
Εικόνα 8. 26 Ποσοστό σφαλμάτων σε παράθυρο 2 δευτερολέπτων, 5 Cross Fold Validation



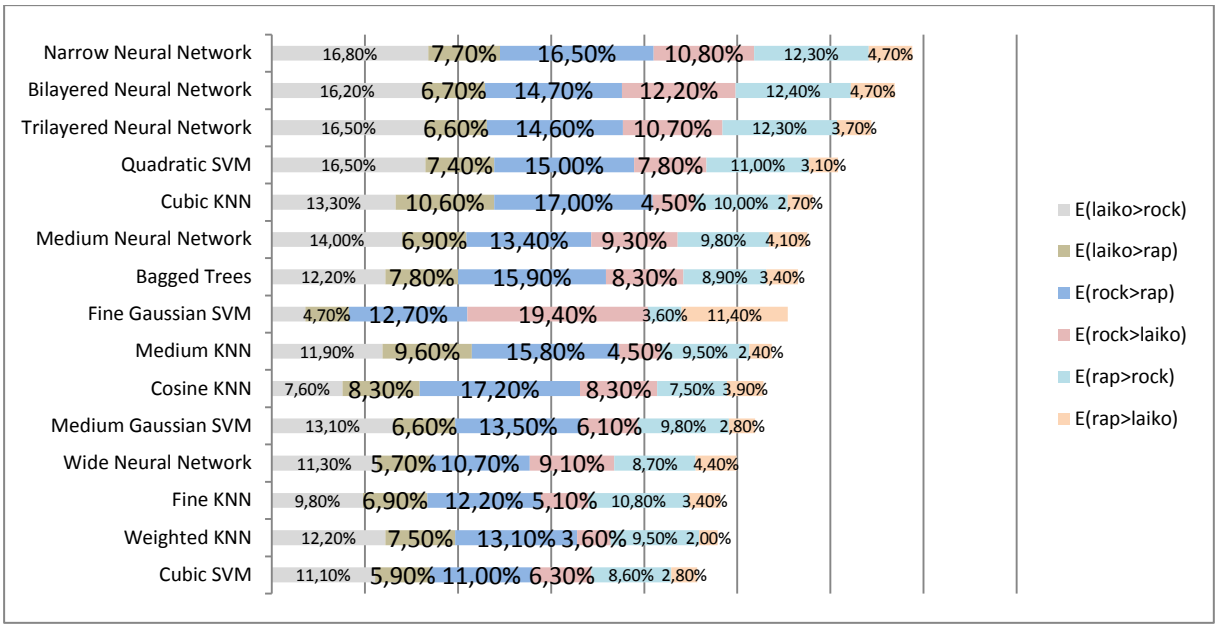
Εικόνα 8. 27 Ποσοστό σφαλμάτων σε παράθυρο 2 δευτερολέπτων, 10 Cross Fold Validation



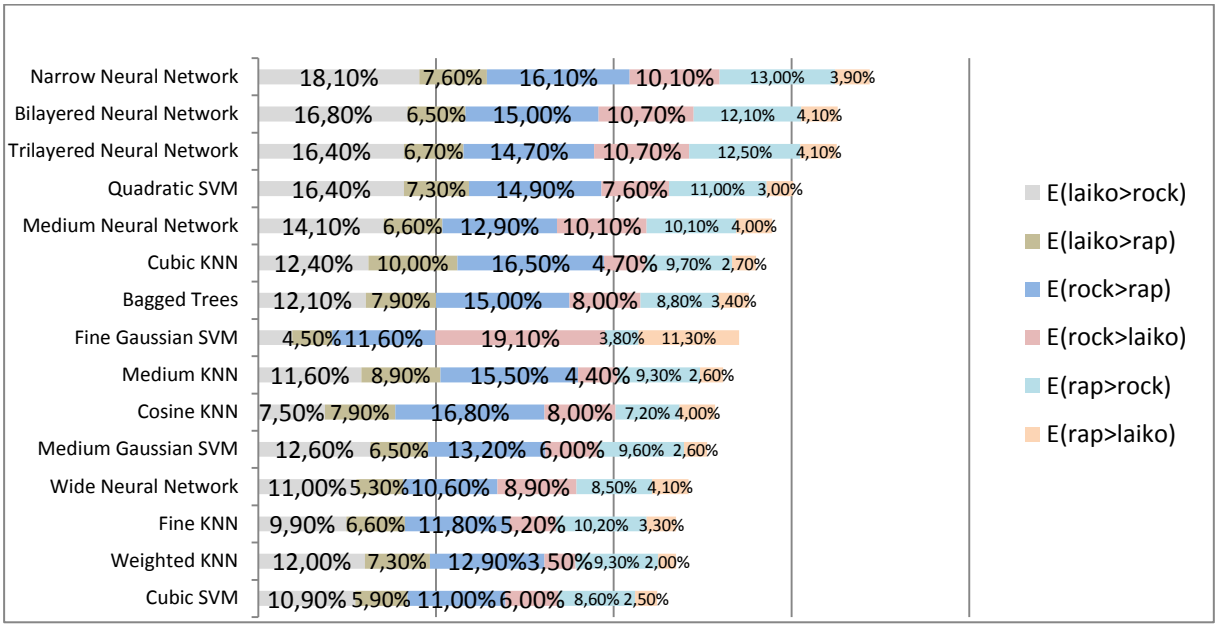
Εικόνα 8. 28 Ποσοστό σφαλμάτων σε παράθυρο 2 δευτερολέπτων, 30 Holdout Validation



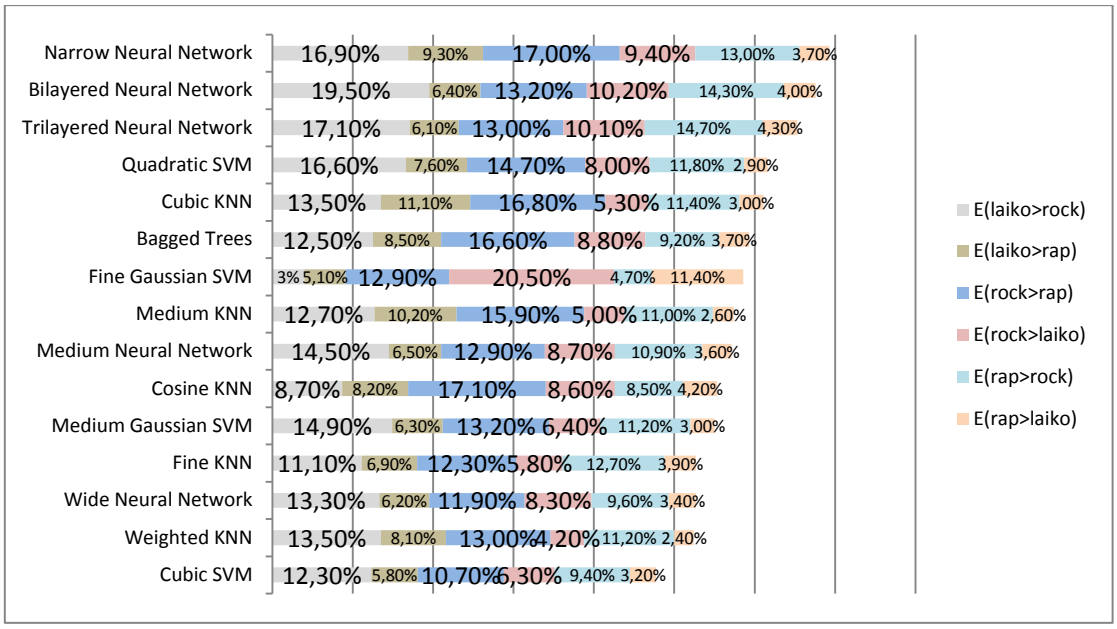
Εικόνα 8. 29 Ποσοστό σφαλμάτων σε παράθυρο 2 δευτερολέπτων, 40 Holdout Validation



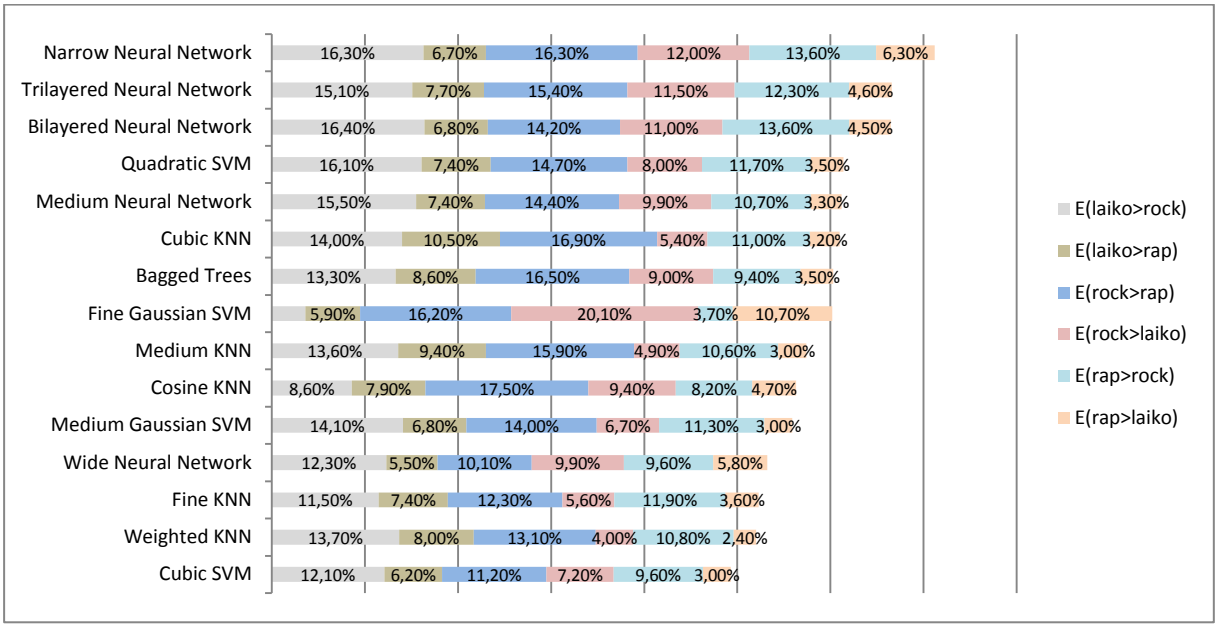
Εικόνα 8. 30 Ποσοστό σφαλμάτων σε παράθυρο 1 δευτερολέπτου, 5 Cross Fold Validation



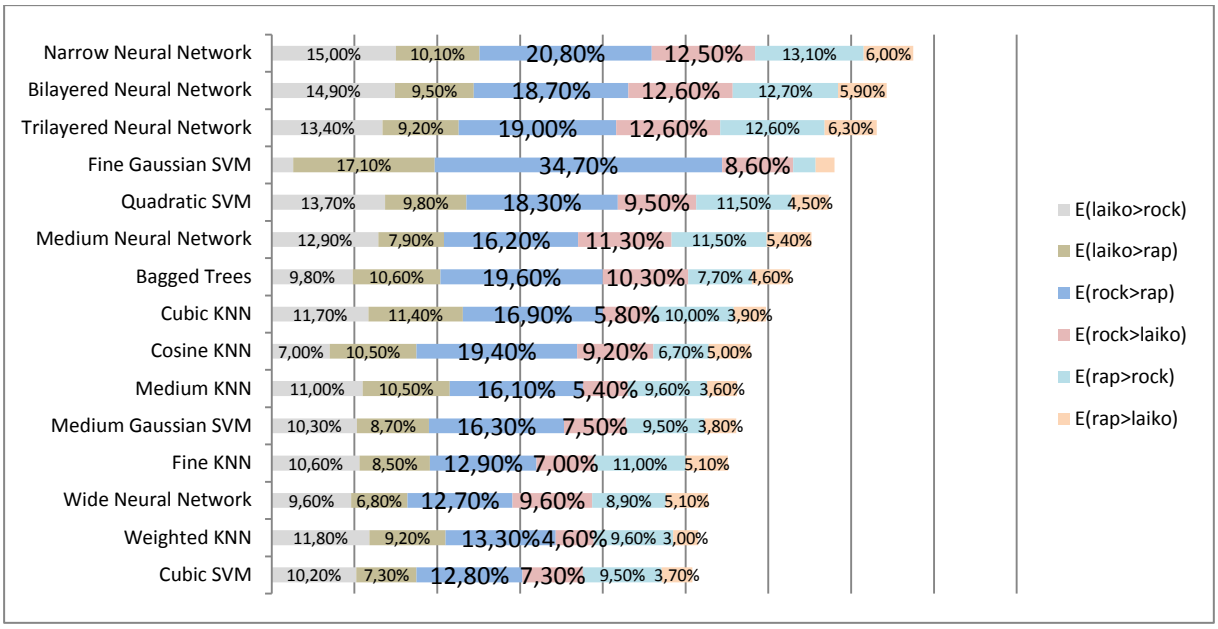
Εικόνα 8. 31 Ποσοστό σφαλμάτων σε παράθυρο 1 δευτερολέπτου, 10 Cross Fold Validation.



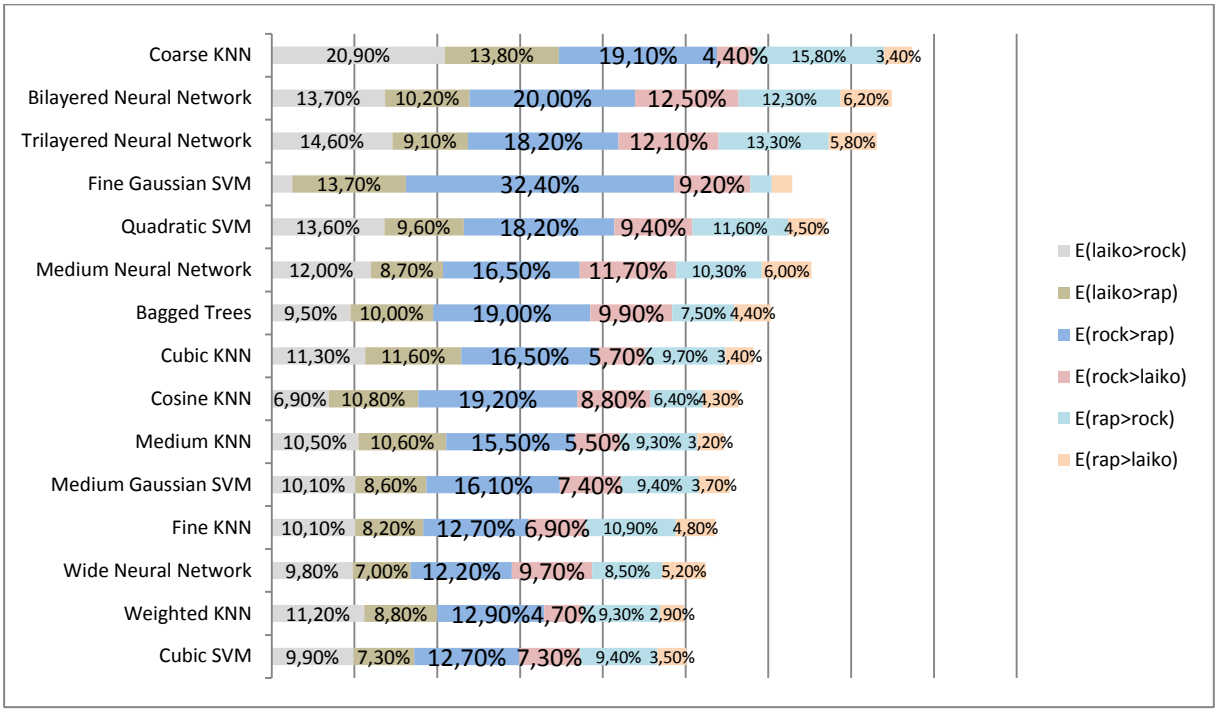
Εικόνα 8. 32 Ποσοστό σφαλμάτων σε παράθυρο 1 δευτερολέπτου, 30 Holdout Validation.



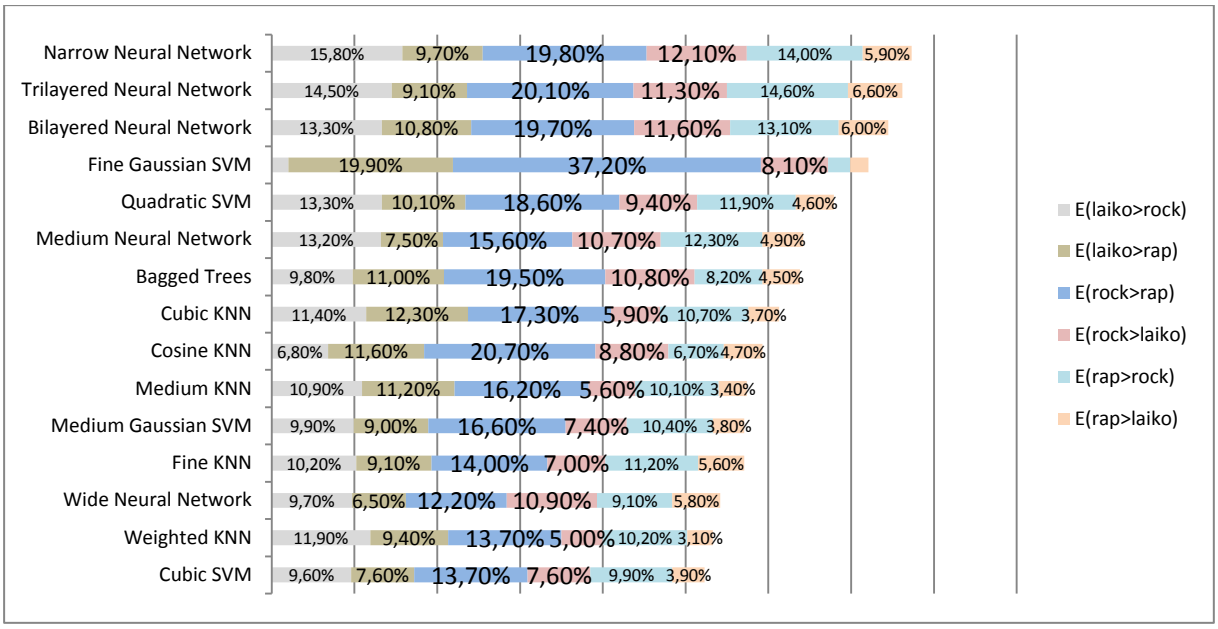
Εικόνα 8. 33 Ποσοστό σφαλμάτων σε παράθυρο 1 δευτερολέπτου, 40 Holdout Validation



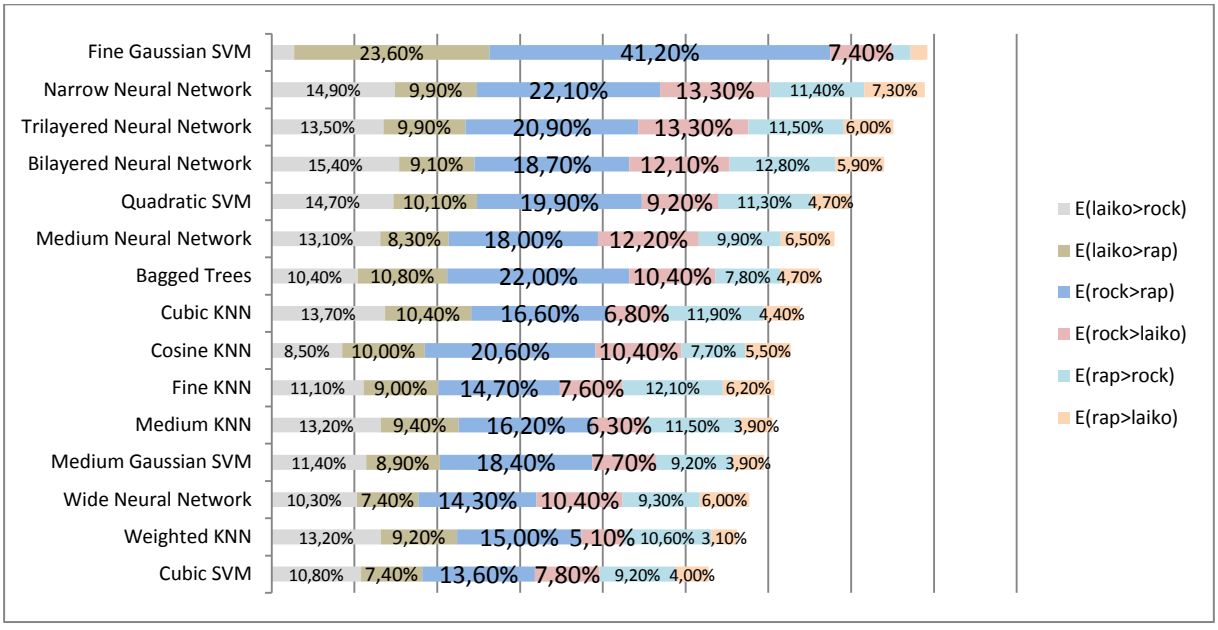
Εικόνα 8. 34 Ποσοστό σφαλμάτων σε παράθυρο 1/2 δευτερολέπτου, 5 Cross Fold Validation



Εικόνα 8. 35 Ποσοστό σφαλμάτων σε παράθυρο 1/2 δευτερολέπτου, 10 Cross Fold Validation



Εικόνα 8. 36 Ποσοστό σφαλμάτων σε παράθυρο 1/2 δευτερολέπτου, 30 Holdout Validation



Εικόνα 8. 37 Ποσοστό σφαλμάτων σε παράθυρο 1/2 δευτερολέπτου, 40 Holdout Validatio

8.4 Σχολιασμός αποδόσεων αλγορίθμων και σφαλμάτων τους

Σε γενικές γραμμές παρατηρείται ότι οι αποδόσεις των αλγορίθμων σε όλα τα παράθυρα και σε όλους τους τρόπους επικύρωσης κυμαίνονται στο 70%-90%, ποσοστό πολύ ικανοποιητικό και ρεαλιστικό όσον αφορά την ακρίβεια σε πειράματα μηχανικής μάθησης [34].

Η μεγαλύτερη συνολική απόδοση επετεύχθη με τον αλγόριθμο Fine KNN σε παράθυρο 2 δευτερολέπτων και μέθοδο επικύρωσης 10 Cross Fold Validation με ποσοστό απόδοσης 87,90%.

Όσο αφορά τις μεγαλύτερες μερικές αποδόσεις, δηλαδή την καλύτερη απόδοση για κάθε είδος ελληνικής μουσικής, βρέθηκε για το ελληνικό ραπ 95.60% με τον Fine Gaussian SVM σε παράθυρο ½ δευτερολέπτου με μέθοδο επικύρωσης 40% Holdout Validation. Για την ελληνική λαϊκή μουσική 94,20% με τον Fine Gaussian SVM σε παράθυρο 2 δευτερολέπτων και μέθοδο επικύρωσης 10 Cross Fold Validation. Τέλος για το ελληνικό Ροκ 85,80% με τον Fine KNN σε παράθυρο 2 δευτερολέπτων με μέθοδο επικύρωσης 10 Cross Fold Validation.

Σχετικά με την γενική σύγκριση των παραθύρων ανάλυσης παρατηρούνται οι καλύτερες συνολικές αποδόσεις να υπάρχουν στο παράθυρο των 2 δευτερολέπτων μετά του 1 δευτερολέπτου και τέλους του ½ δευτερολέπτου ανεξαρτήτως των μεθόδων επικύρωσης. Οι μέσοι όροι των αποδόσεων του 1 και ½ δευτερολέπτου είναι πιο κοντά σχετικά ενώ στο παράθυρο των 2 δευτερολέπτων παρατηρήθηκε αύξηση κατά 3 μονάδες.

Για τις διαφορετικούς μεθόδους επικύρωσης σε συγκεκριμένο παράθυρο γενικά δεν παρατηρήθηκαν μεγάλες αποκλίσεις από την μια μέθοδο επικύρωσης στην άλλη. Ο 10 Cross Fold Validation είχε περίπου 1% καλύτερη απόδοση από τον 5 Cross Fold Validation. Παρόμοια απόκλιση στην απόδοση είχαν τα αποτελέσματα στους τρόπους επικύρωσης Holdout ή Split με ελάχιστα καλύτερη απόδοση να έχει ο 30% Holdout Validation από τον 40% Holdout Validation. Τέλος, όσον αφορά την σύγκριση μεταξύ των δυο μεθόδων επικύρωσης στο κάθε παράθυρο, οι Cross Fold Validation είχε ελάχιστα καλύτερα αποτελέσματα (τάξεως 1%) από τους Holdout Validation.

Γενικότερα, σαν καλύτερες αποδόσεις σε όλους τους αλγόριθμους και τα παράθυρα παρατηρήθηκε να έχει το ραπ, μετά το λαϊκό και τέλος το ροκ. Ο πιθανότερος λόγος που παρουσιάστηκαν αυτές οι αποδόσεις είναι πως με τον όρο ραπ περιγράφουμε μια πολύ μικρότερη γκάμα τραγουδιών σε σχέση με το ροκ ή το λαϊκό. Το dataset που επιλέχθηκε στην ελληνική ραπ ήταν μεταξύ της περιόδου 1995-2005 όπου τα τραγούδια είχαν πολλά περισσότερα κοινά χαρακτηριστικά όπως ο ρυθμός ή η φωτεινότητα του ήχου πράγμα που καθιστά πολύ πιο πετυχημένη την εκπαίδευση και έλεγχο ενός μοντέλου και την ελαχιστοποίηση σφαλμάτων. Σε αντίθεση η ταμπέλα ροκ και λαϊκό έχει πολύ μεγαλύτερο εύρος διαφορετικότητας ηχητικών χαρακτηριστικών. Το rock επιλέχθηκε από τη δεκαετία του 80 έως την δεκαετία του 10, περίοδος στην οποία ένα είδος μουσικής ενώ, κρατάει το ίδιο όνομα, τα τραγούδια από την αρχή της περιόδου μέχρι το τέλος της απέχουν παρασάγγας. Επίσης, θα θεωρηθεί ροκ και μια μπαλάντα με αργό ρυθμό και πλήκτρα αλλά και ένα πιο γρήγορα ρυθμικό κομμάτι με ηλεκτρικές κιθάρες και δυνατά κρουστά. Όσο αφορά το ελληνικό λαϊκό τραγούδι που ήταν από την περίοδο 1940 με 1990 περίπου υπάρχουν σαφείς διαφορές στην ποιότητα των κομματιών καθώς σε όλη αυτήν την περίοδο οι μέθοδοι και οι τεχνολογίες ηχογράφησης γνώρισαν μεγάλη πρόοδο και αλλαγές.

Όσο αφορά τα σφάλματα κατηγοριοποίησης τα αποτελέσματα είναι αρκετά ικανοποιητικά με μέσο όρο γύρω στο 15%, πλην κάποιον υπερβολικών αποκλίσεων.

Τα μεγαλύτερα ποσοστά σφαλμάτων σε όλα τα παράθυρα και τρόπους επικύρωσης παρουσιαστήκαν στα δείγματα που κατηγοριοποιήθηκαν σαν ελληνικό ραπ, ενώ η πραγματική κλάση ήταν ελληνικό ροκ. Τα μικρότερα ποσοστά σφαλμάτων παρουσιάστηκαν στα δείγματα που λανθασμένα κατηγοριοποιήθηκαν σαν ελληνικό λαϊκό ενώ ήταν ραπ.

Ο αλγόριθμος με τα πιο ασταθή ποσοστά σφαλμάτων στα περισσότερα πειράματα ήταν ο Fine Gaussian SVM με 41,2% σφάλμα κατηγοριοποίησης σε ραπ ενώ η αληθινή κλάση ήταν ροκ και σφάλμα 2% κατηγοριοποίησης σε λαϊκό ενώ η αληθινή κλάση ήταν το ραπ. Άξιο παρατήρησης είναι ότι, ενώ αυτός ο αλγόριθμος έτυχε να εμφανίσει κάποια από τα μεγαλύτερα ποσοστά σφαλμάτων, ο ίδιος εμφάνισε τα μεγαλύτερα ποσοστά επιτυχίας στην κατηγοριοποίηση του ραπ και του λαϊκού σε συγκεκριμένα παράθυρα και τρόπους επικύρωσης.

Όσο αφορά τα παράθυρα ανάλυσης και τους τρόπους επικύρωσης, αναλογικά με τα ποσοστά επιτυχίας, καλύτερο παράθυρο ήταν αυτό των 2 δευτερολέπτων μετά του ενός δευτερολέπτου και τέλος τα περισσότερα σφάλματα είχε το παράθυρο ½ δευτερολέπτου. Αντίστοιχα ο καλύτερος τρόπος επικύρωσης ήταν ο 10 Cross Fold Validation σε σχέση με τους υπόλοιπους.

Κεφάλαιο 9: Αξιολόγηση χαρακτηριστικών

9.1 Εισαγωγή

Σε αυτό το κεφάλαιο γίνεται η αξιολόγηση των ηχητικών χαρακτηριστικών που χρησιμοποιήθηκαν για την ταξινόμηση των δεδομένων. Με απλά λόγια, αξιολόγηθηκε ποια ήταν τα πιο σημαντικά χαρακτηριστικά που συνέβαλλαν στον σωστό διαχωρισμό των μουσικών ειδών.

9.2 Αξιολόγηση χαρακτηριστικών μέσω του WEKA

Παρόλο που ολοκληρώθηκαν τα πειράματα μηχανικής μάθησης κρίθηκε αναγκαίο να εξεταστεί η αποδοτικότητα των χαρακτηριστικών. Όπως είναι προφανές, δεν αποδίδουν όλα τα χαρακτηριστικά την ίδια συνεισφορά στη ταξινόμηση κατηγοριών ελληνικής μουσικής. Για να γίνει αξιολόγηση χαρακτηριστικών όσο αφορά την αποδοτικότητα τους χρησιμοποιήθηκε το WEKA που είναι ένα λογισμικό μηχανικής μάθησης. Για την διαδικασία κατάταξης χρησιμοποιήθηκαν δύο αλγόριθμοι,ο InfoGain, ο οποίος μετράει πως κάθε χαρακτηριστικό συμβάλλει στην μείωση της εντροπίας και έτσι ταξινομεί με φθίνων τρόπο τα χαρακτηριστικά όπου έχουν το μεγαλύτερο κέρδος πληροφορίας και ο OneR ο οποίος δομεί το μονοεπίπεδο δέντρο αποφάσεως για κάθε χαρακτηριστικό εισόδου και ταξινομεί πάλι με φθίνουσα σειρά τα χαρακτηριστικα τα οποία έχουν τα μεγαλύτερα ποσοστά σωστής ταξινόμησης.

9.3 Αποτελέσματα αξιολόγησης χαρακτηριστικών

Ακολουθούν όλα αποτελέσματα αξιολόγησης χαρακτηριστικών για όλα τα παράθυρα ανάλυσης.

	Παράθυρο 2 δευτερολέπτων		**Παράθυρο 1 δευτερόλεπτου**		**Παράθυρο 1/2 δευτερόλεπτου**	
#	**InfoGain**	**OneR**	**InfoGain**	**OneR**	**InfoGain**	**OneR**
1	Rolloff (0.3)	Spread	Spread	Spread	Rolloff (0.3)	Spread
2	Spread	Rolloff (0.3)	Rolloff (0.3)	Rolloff (0.5)	Spread	Rolloff (0.3)
3	Rolloff (0.5)	Rolloff (0.5)	Mfccs 3	Rolloff (0.3)	Mfccs 3	Rolloff (0.5)
4	Rolloff (0.9)	Mfccs 3	Rolloff (0.9)	Mfccs 3	Rolloff (0.9)	Rolloff (0.9)
5	Mfccs 3	Skewness	Rolloff (0.5)	Zerocross	Rolloff (0.5)	Mfccs 3
6	Kurtosis	Brightness (3000)	Mfccs13	Rolloff (0.9)	Mfccs13	Zerocross
7	Skewness	Zerocross	Kurtosis	Mfccs 1	Zerocross	Mfccs 13
8	Brightness (3000)	Brightness (1000)	Zerocross	Skewness	Mfccs1	Lowenergy
9	Mfccs 1	Rolloff (0.9)	Mfccs 1	Kurtosis	Kurtosis	Mfccs 1
10	Brightness (1000)	Mfccs 1	Skewness	Lowenergy	Skewness	Kurtosis
11	zerocross	kurtosis	Brightness (3000)	mfccs13	Rolloff (0.7)	Rolloff (0.7)
12	Mfccs 13	Lowenergy	Mfccs 2	Brightness (1000)	Brightness (3000)	Mfccs 4
13	Lowenergy	Mfccs 13	Brightness (1000)	Brightness (3000)	Mfccs 4	Brightness (1000)
14	Mfccs 2	Centroid	Lowenergy	Rolloff (0.7)	Brightness (1000)	Skewness
15	Rolloff (0.7)	Rolloff (0.7)	Mfccs 4	Mfccs 4	Mfccs 12	mfccs12
16	Mfccs 4	Mfccs 4	Rolloff (0.7)	Mfccs 12	Mfccs 2	Mfccs 6
17	Centroid	Mfccs 6	Centroid	Mfccs 6	Centroid	Mfccs 10
18	Mfccs 6	Mfccs 2	Mfccs 12	Centroid	Mfccs 6	Brightness (3000)
19	Mfccs 12	Fccs 10	Mfccs 6	Mfccs 2	Mfccs 10	Mfccs 2
20	Mfccs 10	Mfccs 12	Mfccs 10	Mfccs 10	Mfccs 11	Mfccs 11
21	Mfccs 11	Mfccs 9	Mfccs 11	Mfccs 9	Lowenergy	Centroid
22	Mfccs 8	Mfccs 11	Mfccs 8	Mfccs 11	Mfccs 8	Mfccs 9
23	Mfccs 9	Mfccs 8	Mfccs 9	Mfccs 7	Mfccs 9	Mfccs 8
24	Mfccs 7	Mfccs 7	Mfccs 7	Mfccs 8	Mfccs 7	Mfccs 5
25	Mfccs 5	Mfccs 5	Mfccs 5	Mfccs 5	Mfccs 5	Mfccs 7

Εικόνα 9. 1 Κατάταξη αξιολόγησης χαρακτηριστικών σε όλα τα παράθυρα

9.4 Συμπεράσματα

Παρατηρείται σε όλα τα παράθυρα ανάλυσης και με χρήση όλων των αλγορίθμων επικύρωσης να έχουν περισσότερη αξία πληροφορίας κυρίως τα χαρακτηριστικά της ενεργειακής συχνότητας αποκοπής (Spectral roll-off) για 30% της ενέργειας και η τυπική απόκλιση (Spread). Ακολουθούν τα Roll-off για 30% 50% και 90% και το Mfccs για 3 συντελεστές. Περίπου στην μέση της κατάταξης είναι τα υπόλοιπα φασματικά χαρακτηριστικά και στα χαμηλότερα σημεία mfcss με διάφορους συντελεστές.

Κεφάλαιο 10: Συμπεράσματα και μελλοντικές κατευθύνσεις

10.1 Συμπεράσματα

Αφού συλλέχθηκαν με αυστηρές παραμέτρους 50 τραγούδια για κάθε είδος ελληνικής μουσικής για καθένα από τα τρία είδη και βρέθηκαν μέσα από εκτενή μελέτη ποια χαρακτηριστικά πρέπει να εξαχθούν, εκπαιδεύτηκαν διάφορα μοντέλα Εποπτευόμενης Μηχανικής Μάθησης στα παράθυρα 2 , 1 και ½ δευτερολέπτων με τους τρόπους επικύρωσης 5, 10 Cross Validation και 30%, 40% Holdout Validation. Τα αποτελέσματα ήταν αρκετά ικανοποιητικά καθώς ήταν παρόμοιας τάξης με όμοια πειράματα ταξινόμησης ξένης μουσικής. Καλύτερα ποσοστά κατηγοριοποίησης τάξης 80%-95% είχε το ελληνικό ραπ και το ελληνικό λαϊκό τραγούδι και η ελληνική ροκ είχαν 75%-90%. Τα σφάλματα, επίσης, πλην ελαχίστων περιπτώσεων, ήταν σε ικανοποιητικά ποσοστά τάξης 10%-15%. Οι αλγόριθμοι SVM και KNN φάνηκαν να έχουν τις καλύτερες αποδόσεις, ενώ ακολούθησαν τα δέντρα και τα νευρωνικά δίκτυα. Επίσης, ο καλύτερος συνδυασμός παραθύρου και τρόπου επικύρωσης φάνηκε να είναι το παράθυρο 2 δευτερολέπτων με μέθοδο επικύρωσης 10 Cross Fold Validation όσο αφορά την γενική κατηγοριοποίηση των τριών ειδών μουσικής. Εάν, παρόλα αυτά, εξεταστεί εάν ένα ηχητικό απόσπασμα ανήκει σε συγκεκριμένο είδος, θα γινόταν χρήση για το ελληνικό ραπ του Fine Gaussian SVM σε παράθυρο ½ δευτερολέπτου με μέθοδο επικύρωσης 40% Holdout Validation. Για την ελληνική λαϊκή μουσική, του Fine Gaussian SVM σε παράθυρο 2 δευτερολέπτων και μέθοδο επικύρωσης 10 Cross Fold Validation και για το ελληνικό Ροκ του Fine KNN σε παράθυρο 2 δευτερολέπτων με μέθοδο επικύρωσης 10 Cross Fold Validation.

10.2 Μελλοντικές κατευθύνσεις

Ο τομέας της μηχανικής μάθησης και συγκεκριμένα των διαφόρων χρήσεων στον ήχο, όπως ηχητική αναγνώριση ή διαφοροποίηση ειδών μουσικής, είναι ένα πεδίο ανεξερεύνητο και θεωρούμε ότι είμαστε ακόμα στα αρχικά στάδια ανάπτυξης σαν επιστήμη.

Εμπνευσμένοι από την εκπόνηση αυτής της πτυχιακής εργασίας σκοπός είναι στο μέλλον να την εξελίξουμε περαιτέρω. Θα ήταν ενδιαφέρον να επαναληφθεί το ίδιο πείραμα με περισσότερα είδη ελληνικής μουσικής όπως ποπ ή τραπ ή λιγότερα γνωστά είδη όπως μινιμαλιστική ηλεκτρονική μουσική ή ελληνικό heavy metal.

Θα ήταν αρκετά κερδοφόρο σε γνώση να γίνουν επίσης πειράματα με βάση αληθείας ήχους από πραγματικές συνθήκες όπως ζωντανές εκτελέσεις τραγουδιών χωρίς να υπάρχει αποθορυβοποίηση και αυστηρές συνθήκες ηχογράφησης.

Μια καινοτόμα ιδέα που θα μπορούσε να υλοποιηθεί μετά το πέρας των σπουδών μας είναι να χρησιμοποιηθεί η γνώση από αυτήν την εργασία, ώστε να δημιουργηθεί μία εφαρμογή κινητού η οποία σε σχεδόν πραγματικό χρόνο θα μπορεί να κατηγοριοποιεί είδη ελληνικής μουσικής.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] K. Doshi, «Audio Deep Learning Made Simple (Part 1): State-of-the-Art Techniques,» Towards Data Science, 11 Φεβρουάριου 2021.

[2] Γ. Τ. Κώστας Παναγιωτάκης, «Τμηματοποίηση ήχου και κατηγοροιοποίηση σε μουσική και ομιλία,» 24 Νοεμβρίου 2000.

[3] D. Gerhard, «Audio Signal Classification: History and Current Techniques,» 2003.

[4] D. Gerhard, «Audio Signal Classification: An Overview,» School of Computing Science, Simon Fraser University.

[5] S. G. M. T. O. Lei Chen, «Mixed Type Audio Classification with Support Vector Machine,» IEEE International Conference on Multimedia and Expo, 2006.

[6] F. Y. W. W. Syed Zubair, «Dictionary learning based sparse coefficients for audio classification with max and average pooling,» Digital Signal Processing, Μαΐου 2013.

[7] Loris Nanni, Yandre M. G. Costa, Rafael L. Aguiar, Rafael B. Mangolin, Sheryl Brahnam & Carlos N. Silla Jr., «Ensemble of convolutional neural networks to improve animal audio classification,» EURASIP Journal on Audio, Speech, and Music Processing, 26 Μαΐου 2020.

[8] Harb, H., Chen, L., «A general audio classifier based on human perception motivated model,» Multimed Tools Appl, 2007.

[9] H. Harb, L. Chen and J. . -Y. Auloge, «Mixture of experts for audio classification: an application to male female classification and musical genre recognition,» IEEE International Conference on Multimedia and Expo, 2004.

[10]Saadia Zahid, Fawad Hussain, Muhammad Rashid, Muhammad Haroon Yousaf, and Hafiz Adnan Habib, «Optimized Audio Classification and Segmentation Algorithm by Using Ensemble Methods,» 2015.

[11] «Μουσική,» [Ηλεκτρονικό]. Available: https://el.wikipedia.org/wiki/%CE%9C%CE%BF%CF%85%CF%83%CE%B9%CE%BA%CE%AE?fbclid=IwAR2QaMd7rojPVqjWIsSg-rtgapPHP0RcJrs0DNPsVowmmADl7djHxV2LEGo.

[12] «Η ΙΣΤΟΡΙΑ ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΜΟΥΣΙΚΗΣ,» [Ηλεκτρονικό]. Available: https://slideplayer.gr/slide/3648708/?fbclid=IwAR2u9TBZgGLVCaqbtRadWBZJX2PUdzUbqtYCV5JANQNCLa6IYiftgjXglCc.

[13] «Το αστικό λαϊκό τραγούδι,» [Ηλεκτρονικό]. Available: https://www.musicheaven.gr/html/modules.php?name=News&file=article&id=3424&fbclid=IwAR2Fpdu0dmOmaHqRTraiyEr-eL89XoNXloOlsT09Jmxr3aJv6mwJCSnJlmQ.

[14] «Ελληνικό Hip – Hop,» [Ηλεκτρονικό]. Available: https://www.musicportal.gr/greek_hiphop_music/?lang=el&fbclid=IwAR3NIHr2wyj9S6PBQtCezWPLT59p7uX5L8vvruTTggOs_ZEaU0DQJmdNNYc.

[15] Ashutosh Kulkarni, Deepak Iyer, Srinivasa Rangan Sridharan , «Audio Segmentation,» Stanford University, Stanford.

[16] Pablo Gimeno, Ignacio Viñals, Alfonso Ortega, Antonio Miguel & Eduardo Lleida , «Multiclass audio segmentation based on recurrent neural networks for broadcast domain data,» EURASIP Journal on Audio, Speech, and Music Processing, 2020.

[17] Deepanway Ghosal, Maheshkumar H. Kolekar, «Music Genre Recognition using Deep Neural Networks and Transfer Learning,» Indian Institute of Technology Patna, India, 2018.

[18] Mrs. Teena Varma, Glenn Mendonza, Kevin Paulson, Jovin Maliyakal, Gururaj Parulekar, «Music Genre Recognition and Classification,» Journal of Engineering Sciences, India, 2020.

[19] Benedikt S. Vogler, Amir Othman, «Music Genre Recognition,» 2016.

[20] «ΠΑΡΑΔΟΣΙΑΚΑ ΛΑΙΚΑ ΜΟΥΣΙΚΑ ΟΡΓΑΝΑ,» [Ηλεκτρονικό]. Available: http://paizomemousiki.blogspot.com/2017/04/blog-post.html.

[21] «Η ROCK ΜΟΥΣΙΚΗ,» [Ηλεκτρονικό]. Available: http://lyk-klas-arsak.att.sch.gr/wp-content/uploads/2012/03/9H-Istoria-ths-Rock-mousikis.pdf.

[22] «Η Rap μουσική,» [Ηλεκτρονικό]. Available: https://www.eimaifoititis.gr/arthra/texnes-politismos/h-rap-mousikh.

[23] G. Tzanetakis and P. Cook, «Musical genre classification of audio signals,» IEEE Transactions on Speech and Audio Processing, Ιουλίου 2002.

[24] N. Scaringella, G. Zoia, Member, «Automatic genre classification of music,» IEEE.

[25] Dalwon Jang, Minho Jin and C. D. Yoo, «Music genre classification using novel features and a weighted voting method,» IEEE International Conference on Multimedia and Expo, 2008.

[26] N. Scaringella, G. Zoia, «Automatic genre classification of music,» IEEE.

[27] Lee, Chang-Hsing & Shih, Jau-Ling & Yu, Kun-Ming & Lin, Hwai-San, «Automatic Music Genre Classification Based on Modulation Spectral Analysis of Spectral and Cepstral Features,» 2009.

[28] Annesi, Paolo & Basili, Roberto & Gitto, Raffaele & Moschitti, Alessandro & Petitti, Riccardo, «Audio Feature Engineering for Automatic Music Genre Classification,» 2007.

[29] «Spectral Descriptors,» [Ηλεκτρονικό]. Available: https://www.mathworks.com/help/audio/ug/spectral-descriptors.html#SpectralDescriptorsExample-3.

[30] Κ. Ρ. Γ., «Εφαρμογή αλγορίθμων μηχανικής εκμάθησης για εξόρυξη και κατηγοριοποίηση πληροφοριών περιεχομένου στα οπτικοακουστικά μέσα,» Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Θεσσαλονίκη, 2015.

[31] L. R. Olsen, «Multiple-k: Picking the number of folds for cross-validation,» 14 Νοεμβρίου 2021. [Ηλεκτρονικό]. Available: https://cran.r-project.org/web/packages/cvms/vignettes/picking_the_number_of_folds_for_cross-validation.html.

[32] E. Allibhai, «Hold-out vs. Cross-validation in Machine Learning,» 3 Οκτωβρίου 2018. [Ηλεκτρονικό]. Available: https://medium.com/@eijaz/holdout-vs-cross-validation-in-machine-learning-7637112d3f8f.

[33] Panagakis, Yannis & Benetos, Emmanouil & Kotropoulos, C, «Music Genre Classification: A Multilinear Approach,» 2008.

[34] « How to know if your machine learning model has good performance» [Ηλεκτρονικό]. Available: https://www.obviously.ai/post/machine-learning-model-performance?fbclid=IwAR0J5h90NCfcE3EV0i7jGcBklzr9zAd27sH3LICx-THLOhnqT65RA58PAUU