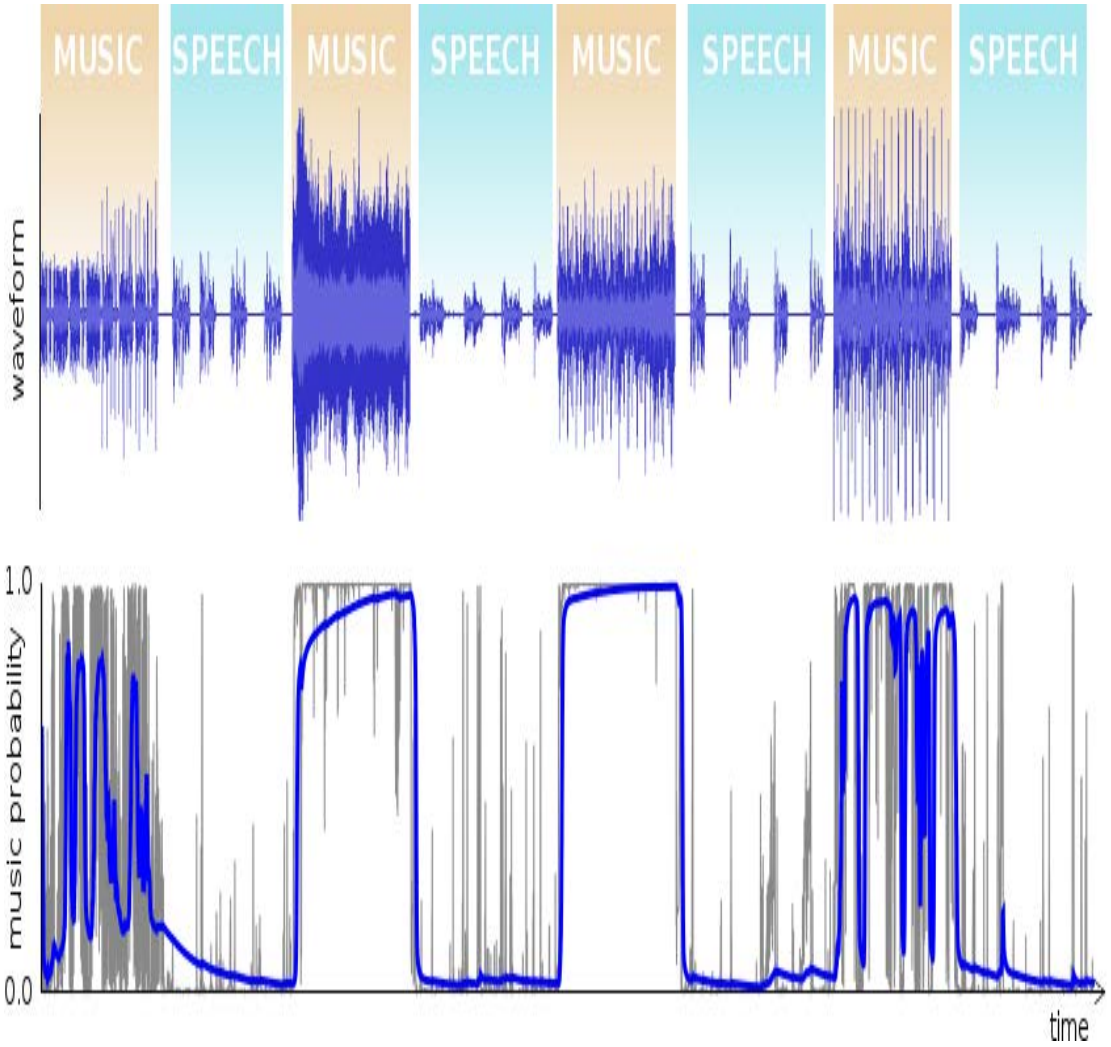
Ταξινόμηση Ήχου: Ομιλία / Μουσική



Παναγιώτης Νικολαΐδης pnikolaid@auth.gr ΑΕΜ: 8206

Χαράλαμπος Παπαδιάκος charaldp@auth.gr ΑΕΜ: 8302

Περιεχόμενα

ΜΕΡΟΣ Α’

1 Εισαγωγή

Αντικείμενο της παρούσας εργασίας αποτελεί η ταξινόμηση δεδομένων ήχου ως ομιλία ή ως μουσική.

Η εργασία αυτή χωρίζεται σε δύο μέρη. Στο πρώτο μέρος (ενότητα 1 έως 3), θα διατυπωθεί αναλυτικότερα το πρόβλημα και οι ζητούμενοι στόχοι του ενώ θα γίνει μία σύντομη ανασκόπηση βασικών στοιχείων των Συστημάτων Αναγνώρισης Προτύπων. Στη συνέχεια, θα ακολουθήσει μία συνοπτική παρουσίαση των προγραμμάτων και εργαλείων που μπορούν να χρησιμοποιηθούν για αυτό το σκοπό. Με βάση τις δυνατότητες και τα πλεονεκτήματα που προσφέρει το κάθε πρόγραμμα, θα αποφασιστεί ποια θα χρησιμοποιηθούν.

Στο δεύτερο μέρος (4-6), παρουσιάζονται τα συστήματα που επιλέχθηκαν να υλοποιηθούν. Για κάθε ένα από αυτά τα συστήματα, παρουσιάζονται λεπτομερώς τα δομικά τους στοιχεία και στάδια καθώς και τα εργαλεία που χρησιμοποιήθηκαν για να υλοποιηθούν. Για τα συστήματα αυτά, εξετάζεται η επίδραση διαφόρων παραμέτρων στην συνολική τους επίδοση και με βάση αυτή την ανάλυση επιλέγονται οι καλύτερες τιμές των παραμέτρων του κάθε συστήματος. Τέλος, για τα συστήματα που προέκυψαν, γίνονται συγκρίσεις των επιδόσεων τους βάσει των οποίων επιλέγεται το καλύτερο σύστημα.

2 Διατύπωση Προβλήματος

Το πρόβλημα το οποίο ζητείται να επιλυθεί ορίζεται από τον διαγωνισμό “MIREX 2018” [1]. Ειδικότερα, το πρόβλημα ορίζεται ως “Music and Speech Detection” [2].

2.1 Περιγραφή

Η ανάγκη για τον εντοπισμό μουσικής ή / και ομιλίας είναι εμφανής σε πολλές διεργασίες ήχου που σχετίζονται με πραγματικά δεδομένα όπως αρχεία ηχογραφήσεων, εκπομπές και σε οποιοδήποτε υλικό το οποίο είναι πιθανό να εμπεριέχει ομιλία και μουσική. Ο διαχωρισμός του σήματος σε τμήματα ομιλίας και μουσική είναι ένα προφανές πρώτο βήμα προτού εφαρμοστούν αλγόριθμοι ομιλίας ή μουσικής. Τελευταία, η βιομηχανία που σχετίζεται με την εύρεση και τον εντοπισμό κλεμμένου υλικού, ολοένα και περισσότερα ενδιαφέρεται όχι μόνο για τον εντοπισμό μουσικής αλλά και για την εξακρίβωση για τον αν αυτή εμφανίζεται στο προσκήνιο (foreground) ή στο παρασκήνιο (background) [2]. Όπως γίνεται αντιληπτό, ο εντοπισμός τμημάτων μουσικής και ομιλίας σε ένα αρχείο ήχου είναι πολλές φορές αναγκαίος για την κατάλληλη χρήση αλγορίθμων. Αποτελεί ουσιαστικά, το πρώτο απαραίτητο βήμα για την επεξεργασία ενός αρχείου ήχου.

2.2 Music and Speech Detection Task

Όπως προαναφέρθηκε, στην παρούσα εργασία ζητείται να εντοπιστούν και να κατηγοριοποιηθούν τα τμήματα μουσικής και ομιλίας ενός αρχείου ήχου. Αυτό υποδεικνύει ότι σε ένα δοσμένο αρχείο ήχου, θα βρίσκονται ανάμεικτα και τμήματα ήχου και τμήματα μουσικής, τα οποία μάλιστα δύναται να αλληλεπικαλύπτονται. Ζητείται επομένως, να αναπτυχθεί αλγόριθμος ο οποίος θα βρίσκει τα τμήματα αυτά, το ξεκίνημα και το πέρας των οποίων δεν είναι εκ των προτέρων γνωστά και να τα ταξινομήσει στις παρακάτω δύο πιθανές κατηγόριες: μουσική ή ομιλία. Εφόσον αναπτυχθεί ο αλγόριθμος αυτός, η αποτελεσματικότητα του θα αξιολογηθεί με βάση τον εντοπισμό και την ταξινόμηση διαφόρων τμημάτων ήχου από διάφορα νέα αρχεία ήχου. Να σημειωθεί ότι ο διαγωνισμός παρέχει πληθώρα αρχείων ήχου τόσο για την διαδικασία εκπαίδευσης όσο και για την διαδικασία αξιολόγησης του αλγορίθμου μηχανικής μάθησης [2].

3 Προσχέδιο Συστήματος

3.1 Θεωρητικό Υπόβαθρο

Το απαιτούμενο θεωρητικό υπόβαθρο για την διεκπεραίωση αυτής της εργασίας ανήκει στο πεδίο της Αναγνώρισης Προτύπων. Ο συγγραφέας του άρθρου [3] παρουσιάζει πολλά εισαγωγικά στοιχεία σχετικά με την Αναγνώριση Προτύπων, τα επιμέρους στάδιά της και τον σχεδιασμό ενός συστήματος αναγνώρισης προτύπων.

Παρακάτω παρατίθεται ένα γενικό σχεδιάγραμμα των δομικών στοιχείων ενός συστήματος αναγνώρισης προτύπων, όπως αυτό παρουσιάζεται στο άρθρο [3]:

Εικόνα 1: Σύστημα αναγνώρισης προτύπων

Όπως φαίνεται παραπάνω, ένα τέτοιο σύστημα αποτελείται από 6 στοιχεία-στάδια:

1. Απόκτηση Αρχικών Δεδομένων
2. Προεπεξεργασία
3. Εξαγωγή Χαρακτηριστικών
4. Επιλογή Χαρακτηριστικών
5. Επιλογή Μοντέλου Ταξινόμησης και Εκπαίδευση
6. Αξιολόγηση

Συνεπώς, για την επιτυχή ολοκλήρωση της εργασίας πρέπει να υλοποιηθούν όλα τα παραπάνω. Στις επόμενες ενότητες αυτού του κεφαλαίου θα προσδιοριστεί ο τρόπος με τον οποίο θα υλοποιηθεί κάθε ένα από αυτά τα στάδια. Αφότου παρουσιαστούν τα διαθέσιμα εργαλεία που μπορούν να χρησιμοποιηθούν για το εκάστοτε στάδιο, τελικά θα επιλέγεται το καταλληλότερο βάσει ορισμένων πλεονεκτημάτων που παρουσιάζει έναντι των υπολοίπων.

3.2 Απόκτηση Αρχικών Δεδομένων & Προεπεξεργασία

Ένα από τα σημαντικότερα στάδια της Αναγνώρισης Προτύπων είναι η απόκτηση αρχικών δεδομένων για την εκπαίδευση του αλγορίθμου. Τα δεδομένα αυτά θα πρέπει οπωσδήποτε να είναι επαρκή σε πλήθος ώστε να μπορεί να εξαχθούν αρχικά όρια εκμάθησης. Το ακριβές μέγεθος δεδομένων που απαιτούνται δεν είναι δυνατό να βρεθεί μέσω κλειστών τύπων. Ωστόσο, ένας ευρέως γνωστός εμπειρικός κανόνας ορίζει ότι ο αριθμός των δεδομένων που θα χρησιμοποιηθούν για την μηχανική μάθηση θα πρέπει να είναι τουλάχιστον 10 φόρες μεγαλύτερος από των αριθμό των χαρακτηριστικών της κάθε κλάσης. Ένα δεύτερο γνώρισμα που πρέπει να διαθέτουν τα δεδομένα είναι να αποτελούν αντιπροσωπευτικά του γενικού συνόλου. Αυτό το γνώρισμα είναι μερικές φορές ιδιαίτερα δύσκολο να ικανοποιηθεί και εναπόκειται στην κρίση του σχεδιαστή να επιλέξει με κατάλληλο τρόπο τα δεδομένα αυτά [3].

Η προεπεξεργασία αφορά κυρίως την αφαίρεση θορύβου από τα αρχικά δεδομένα, την κανονικοποίηση των χαρακτηριστικών τους ώστε να αυτά κυμαίνονται σε μικρό εύρος και τέλος την αφαίρεση ακραίων περιπτώσεων στα αρχικά δεδομένα [3].

Σε αυτή την εργασία, τα δύο αυτά βήματα έχουν ήδη ολοκληρωθεί από τον διαγωνισμό “MIREX 2018”. Συγκεκριμένα, παρέχονται έτοιμα δεδομένα ήχου, σε μορφή αρχείων που περιέχουν ανά τμήματα μουσική ή ομιλία [2]. Τα αρχεία αυτά είναι αρκετά μεγάλα σε μέγεθος με αποτέλεσμα να ικανοποιούν τον εμπειρικό κανόνα επάρκειας που αναφέρθηκε παραπάνω. Για να εξακριβωθεί αν αποτελούν αντιπροσωπευτικά, θα γίνει αναπαραγωγή αυτών των αρχείων ώστε να γίνει αντιληπτό αν εμπεριέχουν ομιλία και μουσική που πράγματι θυμίζει ομιλία ή μουσική, ελέγχεται δηλαδή με αυτό τον τρόπο αν υπάρχουν ακραίες ηχογραφήσεις σε αυτά τα αρχεία.

Σε ό,τι αφορά την προεπεξεργασία για την αντιμετώπιση του θορύβου, θεωρείται ότι ο θόρυβος έχει αποσβεστεί. Επιπλέον, ακόμα και στην περίπτωση που κάτι τέτοιο δεν ισχύει, δεν δύναται να εντοπιστεί ο θόρυβος διότι δεν είναι γνωστά τα χαρακτηριστικά των διατάξεων που μετέτρεψαν τον ήχο σε ηλεκτρονικά αρχεία και τυχόν δυσλειτουργίες τους που να έχουν ως αποτέλεσμα την προσθήκη θορύβου συγκεκριμένου συχνοτικού περιεχομένου.

Συμπερασματικά, τα δεδομένα στην ιστοσελίδα [2], κρίνονται κατάλληλα για χρήση στην εκπαίδευση του αλγορίθμου μηχανικής μάθησης.

3.3 Εξαγωγή Χαρακτηριστικών

Η εξαγωγή των χαρακτηριστικών ενός ηχητικού σήματος αποτελεί ένα από τα σημαντικότερα αρχικά βήματα για την αποτελεσματική μετέπειτα ταξινόμηση του περιεχομένου του από ένα μοντέλο. Η χρησιμότητα της εξαγωγής χαρακτηριστικών έγκειται, στην μείωση του όγκου δεδομένων τα οποία θα επεξεργαστεί το μοντέλο πρόβλεψης άρα στην μείωση της πολυπλοκότητας. Ο τρόπος με τον οποίο μπορεί να επιτευχθεί το παραπάνω είναι μέσω του διαχωρισμού του σήματος σε

παράθυρα και την εξαγωγή χαρακτηριστικών σε περιοχές δειγμάτων της τάξης 512 - 4096. Το μέγεθος των παραθύρων αλλά και η επικάλυψη μεταξύ τους μπορεί να μεταβάλλει την ευαισθησία στις μεταβολές των χαρακτηριστικών αλλά και την ακρίβεια της χρονικής οριοθέτησης κάθε αποσπάσματος μουσικής ή ομιλίας.

Από όλες τις διαθέσιμες επιλογές προγραμμάτων οι οποίες μπορούν να επιλύσουν το πρόβλημα της εξαγωγής χαρακτηριστικών, θα χρησιμοποιηθούν οι εξής δύο: MIRtoolbox του MATLAB και Sonic Visualizer [4], [5]. Οι δυνατότητες του Sonic Visualiser είναι περιορισμένες και η εξαγωγή χαρακτηριστικών σχετικά αργή συγκριτικά με άλλα προγράμματα, ωστόσο το γραφικό περιβάλλον και η οπτικοποίηση των χαρακτηριστικών που δύναται να παρέχει είναι χρήσιμη για μια αρχική αξιολόγηση της χρησιμότητας τους. Συνεπώς, το πρόγραμμα Sonic Visualizer θα χρησιμοποιηθεί μόνο για μία πρώτη εκτίμηση και εικόνα των χαρακτηριστικών. Το βασικό πρόγραμμα εξαγωγής χαρακτηριστικών θα είναι το πακέτο εργαλείων MIRtoolbox του προγράμματος MATLAB. Η ευρύτητα των δυνατοτήτων του προγραμματιστικού περιβάλλοντος MATLAB, σε συνδυασμό με τις επιλογές που παρέχονται από το πακέτο εργαλείων MIRtoolbox, μπορούν να πετύχουν μια αυτοματοποιημένη διαδικασία επεξεργασίας πολλαπλών αρχείων ήχου, κάτι απαιτείται λόγω της μορφής και του πλήθους των δεδομένων που δίνονται.

3.4 Επιλογή Χαρακτηριστικών

Η επιλογή των χαρακτηριστικών τα οποία θα χρησιμοποιηθούν για την εκπαίδευση του μοντέλου και μετέπειτα για την ταξινόμηση των δεδομένων, επηρεάζει άμεσα την επίδοση του συστήματος αναγνώρισης προτύπων. Τα χαρακτηριστικά τα οποία μπορούν εξαχθούν με την χρήση του MIRtoolbox, είναι τα εξής [4]:

- RMS Energy
- Spectral Brightness
- Spectral Roll Off
- Zero Crossing Rate
- Spectral Spread
- Spectral Skewness
- Spectral Kurtosis
- Spectral Flatness
- MFCCs
- Spectral Irregularity

Για την επιλογή των χαρακτηριστικών θα πραγματοποιηθεί μία αξιολόγηση τους και θα υπολογιστεί η συσχέτιση τους με την επιθυμητή έξοδο του ταξινομητή. Αυτό μπορεί να επιτευχθεί μέσω της επιλογής Select Attributes του προγράμματος Weka [6], όπως παρουσιάστηκε και στο μάθημα. Επίσης θα χρησιμοποιηθεί και ο αλγόριθμος αξιολόγησης χαρακτηριστικών $Relieff$, που υπάρχει και στο MATLAB. Ο αλγόριθμος $Relieff$ αποδίδει βάρη σε όλα τα χαρακτηριστικά μέσω επαναληπτικής μεθόδου, χρησιμοποιώντας κάποιο κατώφλι σχετικότητας ενός χαρακτηριστικού με την κλάση της εξόδου, έτσι μπορεί να επιδείξει τα

χαρακτηριστικά τα οποία έχουν την μεγαλύτερη συσχέτιση με την έξοδο του ταξινομητή [7].

Ο λόγος που πραγματοποιείται η διαδικασία επιλογής χαρακτηριστικών είναι η μείωση των πολλαπλών διαστάσεων του προβλήματος (dimensionality reduction). Είναι ευρέως γνωστό, πολλαπλά χαρακτηριστικά που οδηγούν σε πολλαπλές διαστάσεις, αυξάνουν υπέρμετρα την πολυπλοκότητα του αλγόριθμου εκμάθησης και δυσχεραίνουν δραστικά την ολική επίδοση του συστήματος αναγνώρισης προτύπων. Επιπλέον, υπάρχει ο κίνδυνος το σύστημα να μάθει και να προσαρμοστεί πλήρως και στον θόρυβο που υπάρχουν στα δεδομένα (overfitting). Για όλα τα παραπάνω, κρίνεται αναγκαία επιλογή ορισμένων χαρακτηριστικών, που διακρίνουν καλύτερα τα δεδομένα και όχι η χρήση όλων των διαθέσιμων από τα διάφορα προγράμματα [3].

3.5 Επιλογή Μοντέλου Ταξινόμησης και Εκπαίδευση

Η φύση του μοντέλου το οποίο θα χρησιμοποιηθεί για την εκπαίδευση και ταξινόμηση των δεδομένων του συγκεκριμένου προβλήματος επηρεάζει την ποιότητα πρόβλεψης του. Η γνώση του κατάλληλου είδους μοντέλου για την αποτελεσματική επίλυση του προβλήματος εκ των προτέρων είναι δύσκολη, για τον λόγο αυτό είναι πιθανό να εξεταστούν παραπάνω από μια επιλογές προγραμμάτων για την εκπαίδευση μοντέλων ταξινόμησης. Το βασικό πρόγραμμα που θα χρησιμοποιηθεί για τον παραπάνω σκοπό, είναι το Weka. Με την χρήση της δυνατότητας Classify μπορούν να χρησιμοποιηθούν διάφοροι τύποι ταξινομητών, όπως οι Bayesian, οι Rule Based ή τα δέντρα απόφασης [6]. Επιπλέον θα εξεταστεί η απόδοση των ασαφών μοντέλων με την χρήση του πακέτου εργαλείων Fuzzy Logic Toolbox του MATLAB [8].

3.6 Αξιολόγηση

Για την αξιολόγηση της ικανότητας πρόβλεψης του εκπαιδευμένου μοντέλου θα πρέπει να χρησιμοποιηθούν δείκτες οι οποίοι θα αντιπροσωπεύουν την ικανότητα του να λύσει το συγκεκριμένο πρόβλημα ταξινόμησης. Κάθε δείκτης προσδίδει μία διαφορετική οπτική της απόδοσης του μοντέλου. Ενδεικτικά, κάποιοι από αυτούς που μπορούν να εξαχθούν από το Weka είναι τα παρακάτω: tp_rate, precision, recall και f-measure [6]. Η εξαγωγή των δεικτών απόδοσης δίνουν την δυνατότητα σύγκρισης μεταξύ διάφορων ειδών μοντέλων και ταξινομητών.

ΜΕΡΟΣ Β’

4 Υλοποιημένα Συστήματα

4.1 Δεδομένα Εκπαίδευσης / Αξιολόγησης

Απαραίτητο στοιχείο οποιουδήποτε συστήματος αναγνώρισης πρότυπων είναι τα δεδομένα εκπαίδευσης και αξιολόγησης.
Τα δεδομένα που χρησιμοποιήθηκαν είναι τα παρακάτω:

- Δεδομένα εκπαίδευσης:
  - GTZAN: 64 λεπτά, 9.383 εγγραφές
  - Musan Corpus: 3 ώρες, 26.788 εγγραφές

- Δεδομένα αξιολόγησης:
  - GTZAN: 64 λεπτά, 9.383 εγγραφές
  - Musan Corpus: 3 ώρες, 26.788 εγγραφές
  - MIREX Examples 2015: 5 ώρες, 44.366 εγγραφές

Τα δεδομένα εκπαίδευσης είναι αναγκαίο να αποτελούνται από αντιπροσωπευτικά δείγματα μουσικής και ομιλίας. Επιπλέον, τα δείγματα κάθε κλάσης πρέπει να είναι ισάριθμα σε πλήθος. Με βάση τα παραπάνω, αξιοποιήθηκαν για την εκπαίδευση τα δεδομένα GTZAN και Musan Corpus. Αντιθέτως, η χρήση των δεδομένων MIREX Examples 2015 περιορίστηκε μόνο στην αξιολόγηση, λόγω του χαμηλού ποσοστού δειγμάτων ομιλίας (25%) και του υψηλού θορύβου που παρατηρήθηκε.

Για τα δεδομένα Musan Corpus, λαμβάνεται ένα μικρό μέρος τους επιλέγοντας τυχαία ορισμένα αρχεία μουσικής και ομιλίας από τους διάφορους υπό-φακέλους που τα απαρτίζουν για την αποφυγή μεγάλης υπολογιστικής πολυπλοκότητας. Ο συνολικός όγκος των δεδομένων που μπορούν να εξαχθούν από το συγκεκριμένο σετ δεδομένων απαιτεί υπερβολικά μεγάλη επεξεργαστική ισχύ και ο χρόνος επεξεργασίας τους θα ήταν απαγορευτικός. Επίσης κατά την επιλογή των αρχείων για το συγκεκριμένο σετ δεδομένων ελέγχεται πως η συνολική διάρκεια των αρχείων μουσικής και ομιλίας είναι ίση χρονικά, κάτι που στη συνέχεια αναμένεται να οδηγήσει σε παραπλήσια πλήθη εγγραφών σε κάθε κλάση.

4.2 Εξαγωγή Χαρακτηριστικών

Η εξαγωγή των χαρακτηριστικών γίνεται αυτοματοποιημένα μέσω της εργαλειοθήκης MIRtoolbox του MATLAB σε ξεχωριστά .csv αρχεία ώστε να είναι εφικτή η χρήση τους από διάφορες πλατφόρμες (MATLAB, WEKA κ.α.) .

Τα χαρακτηριστικά [9] τα οποία εξάγονται, παρόμοια με αυτά που βρίσκονται συχνά στην βιβλιογραφία είναι τα εξής:

1	RMS Energy
2	Roll Off 85
3	Roll Off 90
4-16	MFCCs (13)
17	Zero Crossing Rate
18	Spectral Flatness
19	Spectral Kurtosis
20	Spectral Brightness
21	Spectral Irregularity
22	Spectral Centroid

Σε αυτό το στάδιο, ως είδος παράθυρου ολίσθησης επιλέχτηκε το παράθυρο Hamming λόγω των ομαλών μεταβάσεων στα άκρα του, αποτρέποντας έτσι την εμφάνιση υψισύχνων τόνων που αλλοιώνουν το αρχικό σήμα. Εκτελέστηκαν πολλές εξαγωγές χαρακτηριστικών για διάφορες τιμές του μήκους παραθύρου ολίσθησης. Να σημειωθεί ότι είναι απαραίτητο το μήκος αυτό να είναι αρκετά μικρό ώστε να μπορεί να καταγραφούν γρήγορες εναλλαγές μεταξύ ομιλίας- μουσικής.

Για την εξαγωγή των χαρακτηριστικών υλοποιήθηκαν τα παρακάτω scripts:

1. extractFeaturesCsvMultiple<GTZAN/Musan>.*m*: Εξάγει χαρακτηριστικά από πολλαπλά αρχεία ομιλίας ή μουσικής επιλέγοντας το path στο οποίο αυτά περιέχονται και ενοποιεί όλες τις εγγραφές σε ένα αρχείο .csv. Οι ετικέτες των εγγραφών ως μουσική ή ομιλία τοποθετούνται με βάση το path από το οποίο διαβάζονται τα αρχεία στην συγκεκριμένη επανάληψη, δηλαδή την τιμή του index (i). Οι εγγραφές που περιέχουν NaN τιμές για οποιοδήποτε χαρακτηριστικό διαγράφονται.

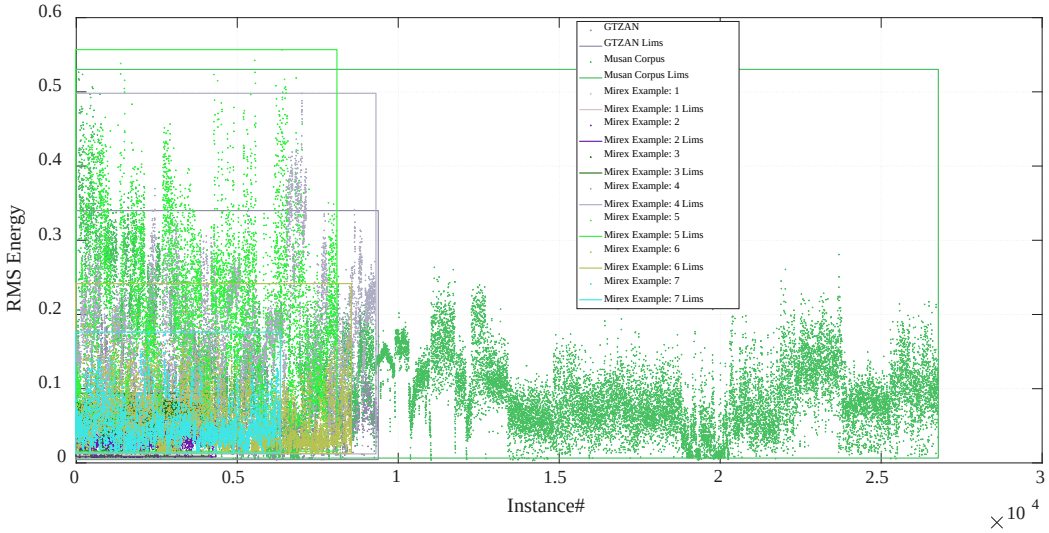
2. extractFeaturesCsvMixedMirexExamples.m: Παρόμοια λειτουργία με το παραπάνω με τη διαφορά ότι εξάγει χαρακτηριστικά από πολλά αρχεία ήχου με ανάμεικτα κομμάτια ομιλίας και μουσικής σε ξεχωριστά αρχεία .csv το καθένα. Στα ανάμεικτα αρχεία ήχου η διαδικασία τοποθέτησης ετικέτας σε κάθε εγγραφή είναι ελαφρώς πολυπλοκότερη. Με τη χρήση των αρχείων .csv που περιέχουν πληροφορίες σχετικά με τα χρονικά όρια των τμημάτων μουσικής και ομιλίας εξάγεται ο πίνακας (a) ο οποίος περιλαμβάνει το τμήμα μουσικής/ομιλίας του .csv στο οποίο ανήκει η κάθε εγγραφή με μια τιμή 1 ή 0. Αθροίζοντας κατά στήλες το πίνακα αυτό, είναι εφικτό να εξεταστεί εάν μια συγκεκριμένη εγγραφή ανήκει σε ένα ή παραπάνω τμήματα μουσικής η ομιλίας ή εάν δεν ανήκει κάπου. Για απλούστευση της διαδικασίας διατηρούνται οι εγγραφές που ανήκουν μόνο σε ένα τμήμα μουσικής ή ομιλίας και η ετικέτες τους αποδίδονται με βάση την $3^{η}$ στήλη του αρχείου .csv που περιλαμβάνει τα δεδομένα timestampData.

4.3 Κανονικοποίηση Χαρακτηριστικών

Πριν από την κανονικοποίηση των δεδομένων, να σημειωθεί ότι ελέγχεται αν τα δεδομένα περιέχουν εξωκείμενες τιμές (outliers) βάσει των τιμών των εξαγόμενων χαρακτηριστικών. Συγκεκριμένα υπολογίζεται η μέση τιμή $mean_{val}$ και η διακύμανση std_{dev} κάθε χαρακτηριστικού σε κάθε σετ δεδομένων και διαγράφονται οι εγγραφές οι οποίες βρίσκονται εκτός του διαστήματος $mean_{val} \pm 5 * std_{dev}$. Η τιμή $outlierMultiplier = 5$ επιλέγεται καθώς έτσι γίνεται ένα ικανοποιητικό φιλτράρισμα χωρίς να διαγράφεται μεγάλο μέρος των δεδομένων (περίπου 0.5-1.5%). Με τα παραπάνω επιτυγχάνεται αποτελεσματικότερη εκπαίδευση του μοντέλου ταξινόμησης καθώς τα δεδομένα εκπαίδευσης είναι πλέον πιο αντιπροσωπευτικά.

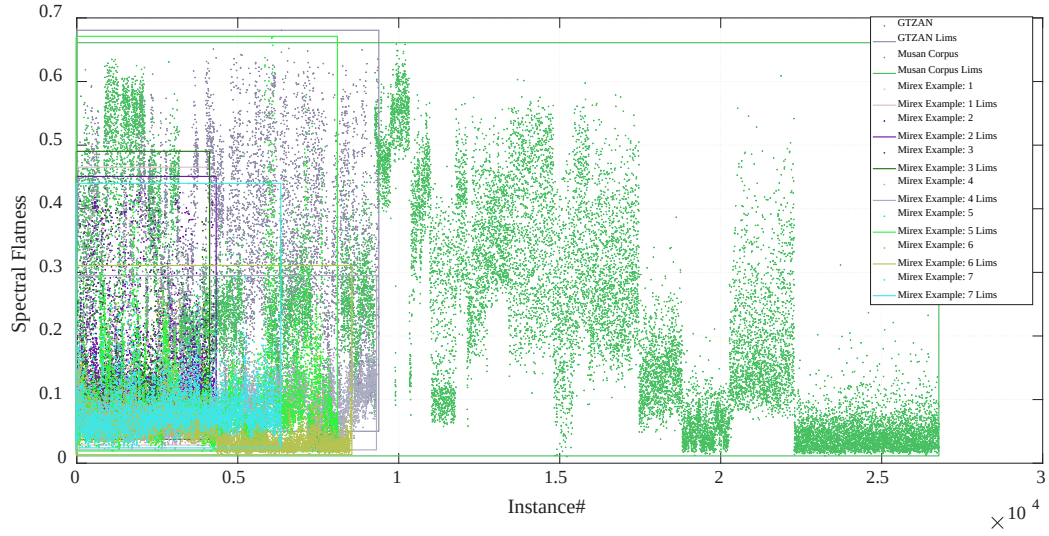
Επόμενο βήμα αποτελεί η κανονικοποίηση των τιμών των χαρακτηριστικών. Είναι προτιμητέο οι τιμές των εξαγόμενων χαρακτηριστικών εκπαίδευσής να βρίσκονται σε κοινό εύρος (π.χ. 0 έως 1) ώστε να εκπαιδευτεί σωστά το μοντέλο ταξινόμησης. Είναι επίσης φανερό ότι σε περίπτωση που χρησιμοποιηθεί διαφορετικό σετ δεδομένων για αξιολόγηση, τα εξαγόμενα χαρακτηριστικά του, θα πρέπει και αυτά να κανονικοποιηθούν με τον ίδιο τρόπο που έγινε η κανονικοποίηση των δεδομένων εκπαίδευσης. Αυτό είναι απαραίτητο διότι χρησιμοποιώντας κανονικοιποιημένα δεδομένα εκπαίδευσης, το μοντέλο ταξινόμησης εκπαιδεύεται ως προς μία συγκεκριμένη αναφορά και επομένως, η αναφορά αυτή πρέπει να διατηρηθεί ίδια και στην κανονικοποίηση των δεδομένων αξιολόγησης.

Ωστόσο, για μερικά χαρακτηριστικά, ακολουθείται μία λίγο διαφορετική διαδικασία κανονικοποίησης. Για παράδειγμα, το χαρακτηριστικό RMS Energy επηρεάζεται όχι μόνο από το είδος του ήχου (ομιλία ή μουσική), αλλά και από την ένταση ηχογράφησης. Επομένως, διαφορετικές τιμές έντασης ηχογράφησης των σετ δεδομένων, μπορούν να αλλοιώσουν τις τιμές του χαρακτηριστικού και άρα πρέπει να γίνει ξεχωριστά κανονικοποίηση σε κάθε σετ δεδομένων για το συγκεκριμένο χαρακτηριστικό. Η επίδραση των διαφορετικών τιμών έντασης ηχογράφησης γίνεται αντιληπτή στο παρακάτω διάγραμμα, όπου φαίνονται ξεκάθαρα οι έντονες διαφορές στο εύρος τιμών του χαρακτηριστικού RMS Energy ανάμεσα στα σετ δεδομένων:

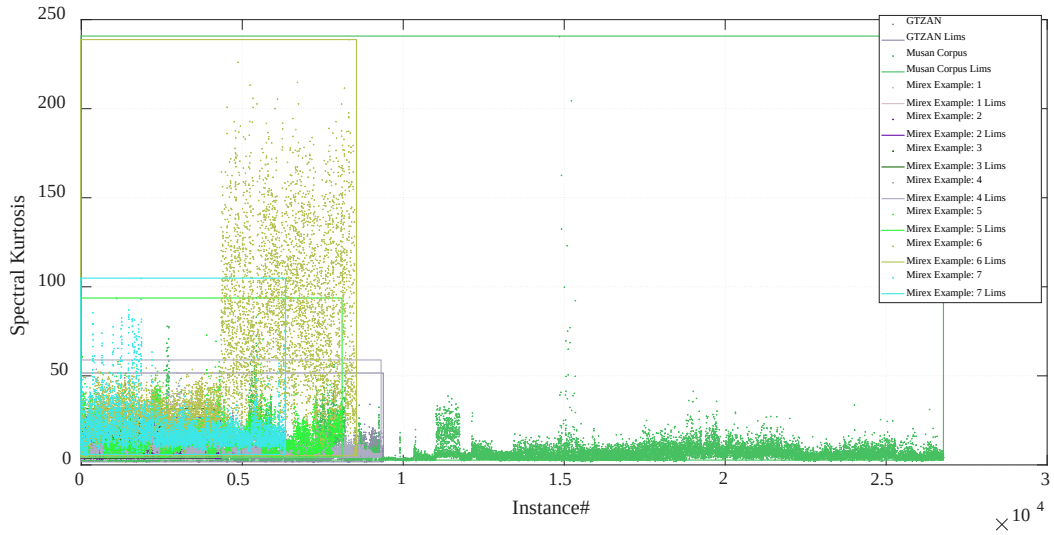


Εικόνα 2: Εύρος τιμών του χαρακτηριστικού RMS Energy

Παρόμοια μεγάλες διαφορές στα εύρη τιμών ενός χαρακτηριστικού ανάμεσα σε διάφορα σετ δεδομένων παρατηρήθηκαν για τα χαρακτηριστικά Spectral Flatness και Spectral Kurtosis:

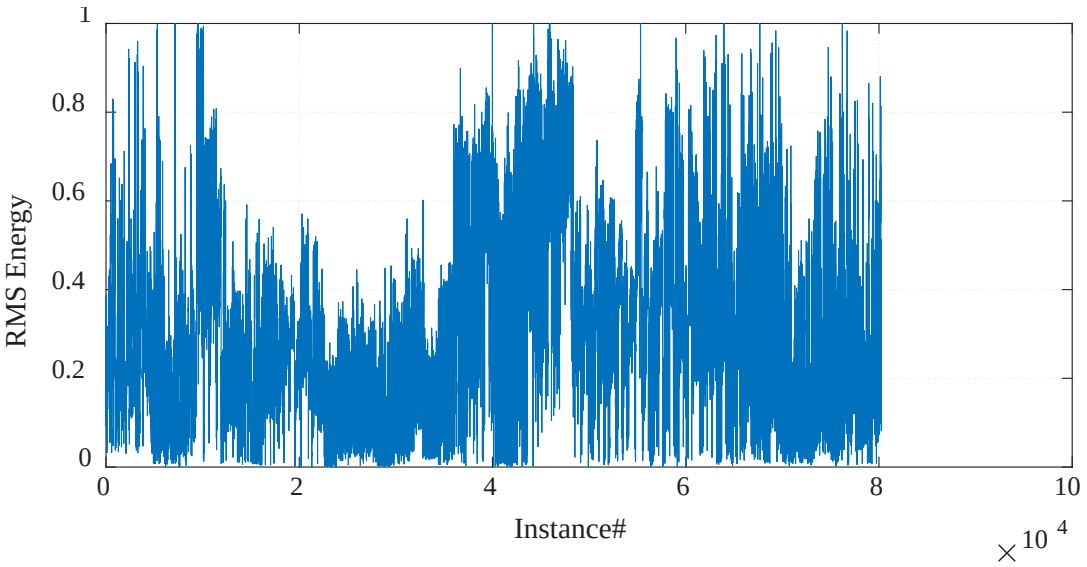


Εικόνα 3: Εύρος τιμών του χαρακτηριστικού Spectral Flatness



Εικόνα 4: Εύρος τιμών του χαρακτηριστικού Spectral Kurtosis

Λαμβάνοντας υπόψη την προηγούμενη συζήτηση για μεγάλες μεταβολές στα εύρη τιμών ανά σετ δεδομένων, μπορεί να υποπτευθεί κανείς ότι τα μεταβολές αυτές για τα χαρακτηριστικά Spectral Flatness και Spectral Kurtosis, ίσως οφείλονται στον θόρυβο. Συνεπώς, και για τα χαρακτηριστικά Spectral Flatness και Spectral Kurtosis θα ακολουθηθεί η ίδια διαδικασία κανονικοποίησης με αυτή του χαρακτηριστικού RMS Energy. Σε αυτό το σημείο να τονιστεί ότι, ακολουθώντας αυτή τη μέθοδο, παρατηρήθηκε βελτίωση στην επίδοση διαφόρων μοντέλων ταξινόμησης ως προς accuracy. Η εξομάλυνση των μεταβολών στα εύρη τιμών, με αυτή την διαδικασία κανονικοποίησης για τα προαναφερθέντα χαρακτηριστικά, απεικονίζεται παρακάτω:



Εικόνα 5: Νέο εύρος τιμών του χαρακτηριστικού RMS Energy

4.4 Ταξινόμηση με Ασαφές Νευρωνικό Δίκτυο

Το πρώτο μοντέλο ταξινόμησης που εξετάζεται είναι το Ασαφές Νευρωνικό Δίκτυο. Ωστόσο προτού χρησιμοποιηθεί το μοντέλο ταξινόμησης, γίνεται επιλογή των πιο περιγραφικών χαρακτηριστικών από αυτά που αναφέρθηκαν στην ενότητα 4.2 μέσω του αλγορίθμου Relieff [10]. Η χρήση πολλών χαρακτηριστικών έχει ως συνέπεια την δραματική αύξηση της πολυπλοκότητας στην εκπαίδευση. Η συνέπεια αυτή είναι γνωστή και ως κατάρα της διαστασιμότητας [11].

Η υλοποίηση του μοντέλου ταξινόμησης γίνεται με τη χρήση του Fuzzy Logic Toolbox του MATLAB. Η επιλογή των χαρακτηριστικών μέσω του αλγορίθμου Relieff, η εκπαίδευση και αξιολόγηση αυτού του μοντέλου υλοποιούνται με τα παρακάτω scripts:

1. *classificationFuzzy.m*: Συνάρτηση που δέχεται ως ορίσματα τα εξαγόμενα χαρακτηριστικά (data), τον αριθμό των καλύτερων features που επιλέγονται (NF), τον αριθμό των κανόνων του Ασαφούς Νευρωνικού Δικτύου (NR), το πλήθος των πτυχών (K_folds) της διασταυρωμένης επικύρωσης και την μέθοδο ομαδοποίησης (clustering) των κανόνων με την οποία αρχικοποιείται το μοντέλο Substractive Clustering (SC) ή Fuzzy C-Means (FCM). Αρχικά εκτελεί τον αλγόριθμο Relieff για την αξιολόγηση όλων των χαρακτηριστικών και επιλέγει τα (NF) καλύτερα από αυτά. Στη συνέχεια αρχικοποιεί ένα Ασαφές Νευρωνικό Δίκτυο για κάθε πτυχή (fold), με την χρήση των συναρτήσεων genfis2/ genfis3 και το εκπαιδεύει με τη χρήση της συνάρτησης anfis. Από τις παραμέτρους οι οποίες επιστρέφονται από την συνάρτηση anfis επιλέγεται να χρησιμοποιηθεί αυτή που περιέχει στιγμιότυπο (TSK_Model_HD) του εκπαιδευμένου μοντέλου (δηλαδή από μια συγκεκριμένη εποχή) το οποίο παρουσιάζει το ελάχιστο τετραγωνικό σφάλμα στα δεδομένα ελέγχου, έτσι ελέγχεται πως το εκπαιδευμένο μοντέλο δεν έχει υποστεί υπερεκπαίδευση. Τέλος, υπολογίζει την ακρίβεια (accuracy) των μοντέλων μέσω διασταυρωμένης επικύρωσης. Η συνάρτηση πέρα από την συνολική επίδοση του μοντέλου (accuracy) επιστρέφει την επίδοση κάθε επιμέρους πτυχής (fold) της διασταυρωμένης επικύρωσης του εκπαιδευμένου μοντέλου, ως προς την ακρίβεια στο διάνυσμα (foldsAccuracy). Επίσης επιστρέφονται τα μοντέλα κάθε πτυχής της διασταυρωμένης επικύρωσης για την αξιολόγηση τους σε διαφορετικά δεδομένα στην δομή (TSK_Model_HD_Opt). Για την μετέπειτα χρήση των εκπαιδευμένων μοντέλων σε άλλα δεδομένα απαιτείται το διάνυσμα των καλύτερων χαρακτηριστικών που βρέθηκαν με τη χρήση του αλγορίθμου Relieff το οποίο και επιστρέφεται στο διάνυσμα (rank).

2. *trainEvalMultiDataset.m:* Σε αυτό το script, εκτιμάται η απόδοση του συστήματος σε συνθήκες παρόμοιες με αυτές του διαγωνισμού. Γίνεται δηλαδή, η εκπαίδευση μοντέλων με κάποιο συγκεκριμένο σετ δεδομένων και έπειτα εξάγονται οι δείκτες αξιολόγησης των μοντέλων για τις προβλέψεις τους σε διαφορετικά σετ δεδομένων. Πιο συγκεκριμένα καλείται

δύο φορές η συνάρτηση *classificationFuzzy* με ορίσματα για τα δεδομένα εκπαίδευσης μοντέλων με τη χρήση δεδομένων GTZAN και Musan Corpus. Στη συνέχεια εξάγεται η πρόβλεψη (pred###) κάθε πτυχής των 2 εκπαιδευμένων μοντέλων χρησιμοποιώντας ως δεδομένα αξιολόγησης τα 2 εναπομένοντα τα οποία είναι άγνωστα προς το εκάστοτε μοντέλο. Η πρόβλεψη (pred###) εξάγεται με τη συνάρτηση evalfis και την διακριτοποίηση της. Με τη χρήση των προβλέψεων κάθε πτυχής των μοντέλων και τις πραγματικές κλάσεις (actual###), εξάγεται ο πίνακας προβλέψεων τους (ConfusionMatrix###) και πλέον με τη χρήση αυτών μπορεί να υπολογιστεί η συνολική επίδοση της διασταυρωμένης επικύρωσης ως (accuracy###).

4.5 Ταξινόμηση με διαθέσιμα μοντέλα του Weka

Χρησιμοποιείται επίσης το πρόγραμμα Weka όπου υπάρχει η δυνατότητα εξέτασης πολλαπλών μοντέλων ταξινόμησης. Στο πρόγραμμα Weka δίνονται ως οι είσοδοι τα αρχεία .csv που εξήχθησαν μέσω των scripts *extractFeaturesCsvMultipleFiles.m* και *extractFeaturesCsvMixedFiles*. Στη συνέχεια, χρησιμοποιείται ο μετασχηματισμός PCA για την μείωση της διαστασιμότητας των χαρακτηριστικών. Συγκεκριμένα, επιλέγονται τα μετασχηματισμένα χαρακτηριστικά τα οποία περιλαμβάνουν τουλάχιστον το 95% της συνολικής διακύμανσης. Ο αριθμός των μετασχηματισμένων χαρακτηριστικών που απομένουν είναι 15 και αναμένεται πως περιέχουν το μεγαλύτερο ποσοστό της χρήσιμης πληροφορίας. Τέλος, η αξιολόγηση γίνεται με 5-πτυχη διασταυρωμένη επικύρωση.

Τα μοντέλα τα οποία εμφάνισαν αξιοσημείωτη επίδοση στα δεδομένα εκπαίδευσης που χρησιμοποιήσαμε είναι τα εξής:

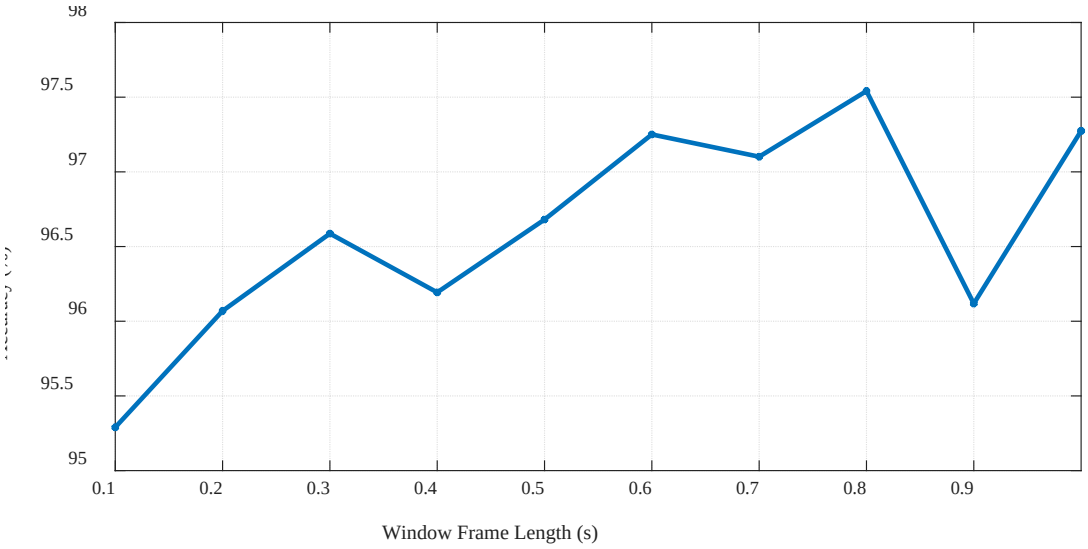
- Rep Tree: Καμία επιπλέον παραμετροποίηση
- IBk (kNN): Η παράμετρος k των κοντινότερων γειτόνων επιλέχθηκε στους 4 γείτονες.
- Multi-Layer Perceptron: Καμία επιπλέον παραμετροποίηση
- Random Committee: Ένας meta-classifier ο οποίος εφαρμόζει την τεχνική σύνθεσης μοντέλων (Ensenbles). Συγκεκριμένα, επιλέχθηκε το μοντέλο Random Tree χωρίς κάποια επιπλέον παραμετροποίηση.
- Random Forest: Καμία επιπλέον παραμετροποίηση

5 Αποτελέσματα

5.1 Ασαφές Νευρωνικό Δίκτυο

Τα scripts σε MATLAB (βλ. ενότητα 4.2-4.4) που χρησιμοποιούνται για την υλοποίηση του Συστήματος Αναγνώρισης Προτύπων με Ασαφές Νευρωνικό Δίκτυο, δίνουν την δυνατότητα αυτοματοποιημένων διαδικασιών. Επομένως, είναι εύκολη η εξέταση του συστήματος μέσω επαναλήψεων FOR για διάφορες παραμέτρους και η εξαγωγή των διαγραμμάτων που προκύπτουν. Οι επαναλήψεις αυτές για διαφορετικές τιμές μίας παραμέτρου γίνονται χρησιμοποιώντας ένα μικρό μέρος του σετ δεδομένων GTZAN, τόσο για εκπαίδευση όσο και για αξιολόγηση. Ο λόγος που επιλέγεται ένα μέρος του σετ δεδομένων είναι η μείωση της υπολογιστικής πολυπλοκότητας. Στα διαγράμματα που θα ακολουθήσουν, γίνεται προσπάθεια εύρεσης μίας υπό-βέλτιστης τιμής μίας παραμέτρου. Για τον σκοπό αυτό, συγκρίνουμε την επίδοση του συστήματος ως προς την ακρίβεια αξιολόγησης (accuracy), για διάφορες τιμές της εκάστοτε παραμέτρου, διατηρώντας όλες τις υπόλοιπες σταθερές.

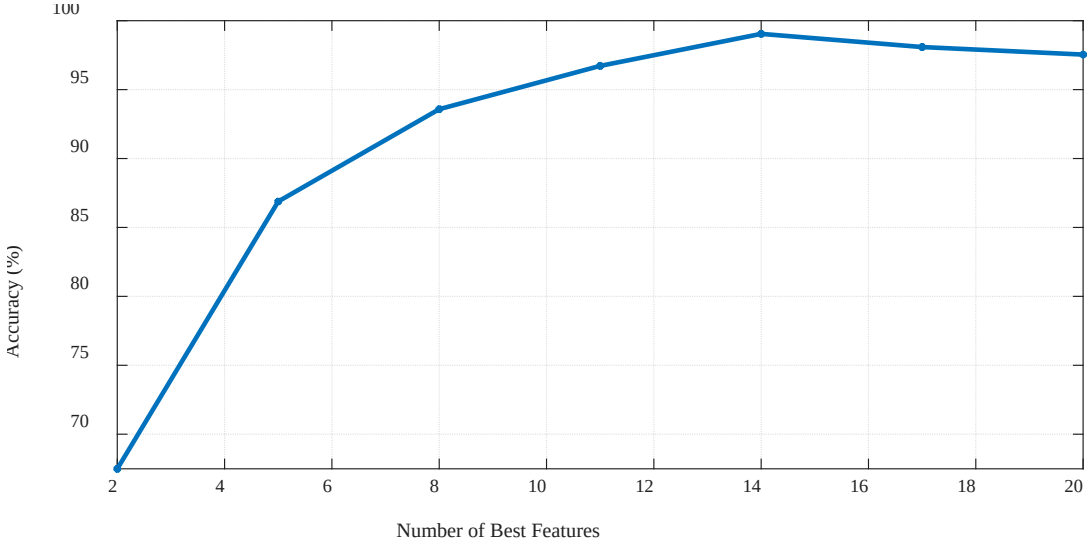
Στο παρακάτω διάγραμμα φαίνεται η επίδοση του μοντέλου για διάφορες τιμές της διάρκειας παραθύρου ολίσθησης σε δευτερόλεπτα:



Εικόνα 5: Επίδοση σαφούς μοντέλου συναρτήσει του μήκους παραθύρου

Παρατηρείται βέλτιστη τιμή μήκους παραθύρου $w = 0{,}8\ sec$ επιτυγχάνοντας 97.5% ακρίβεια αξιολόγησης. Επιπλέον, παρατηρείται μία αυξητική τάση της ακρίβειας αξιολόγησης όσο το παράθυρο μεγαλώνει. Αυτό είναι αναμενόμενο διότι μεγαλύτερα παράθυρα επιτυγχάνουν μεγαλύτερη φασματική αναλυτικότητα. Ωστόσο, η χρονική αναλυτικότητα μειώνεται όσο αυτά μεγαλώνουν. Παρόλα αυτά, θεωρούμε τα 0.8s ως μία αποδεκτή τιμή χρονικής αναλυτικότητας, καθώς είναι αρκετά σπάνια η περίπτωση όπου ένα μουσικό κομμάτι έχει διάρκεια μικρότερη

από 0.8s ή ένας άνθρωπος μιλάει για λιγότερο από 0.8s. Επομένως, στα επόμενα διαγράμματα επιλέγεται ως μήκος παραθύρου ολίσθησης τα 0.8s.
Στο επόμενο διάγραμμα φαίνεται η ακρίβεια που επιτυγχάνεται στο ασαφές μοντέλο για διάφορες τιμές του αριθμού των καλύτερων χαρακτηριστικών σύμφωνα με την κατάταξη του αλγορίθμου αξιολόγησης χαρακτηριστικών Relieff:



Εικόνα 6: Επίδοση ασαφούς μοντέλου για διάφορες τιμές του αριθμού χαρακτηριστικών

Από το παραπάνω διάγραμμα προκύπτει ότι μέγιστη ακρίβεια επιτυγχάνεται όταν ο αριθμός των καλυτέρων χαρακτηριστικών είναι $NF = 14$. Επιπλέον, εφόσον η ακρίβεια σταθεροποιείται για τιμές μεγαλύτερες της $NF = 14$, είναι φανερό ότι η επιλογή αυτών των τιμών δεν έχει κανένα νόημα καθώς όχι μόνο δεν βελτιώνεται η επίδοση του συστήματος αλλά και η πολυπλοκότητα αυξάνεται. Επομένως, για την εξαγωγή του επόμενου διαγράμματος θα επιλεχθούν $w = 0{,}8\ sec$ και $NF = 14$.

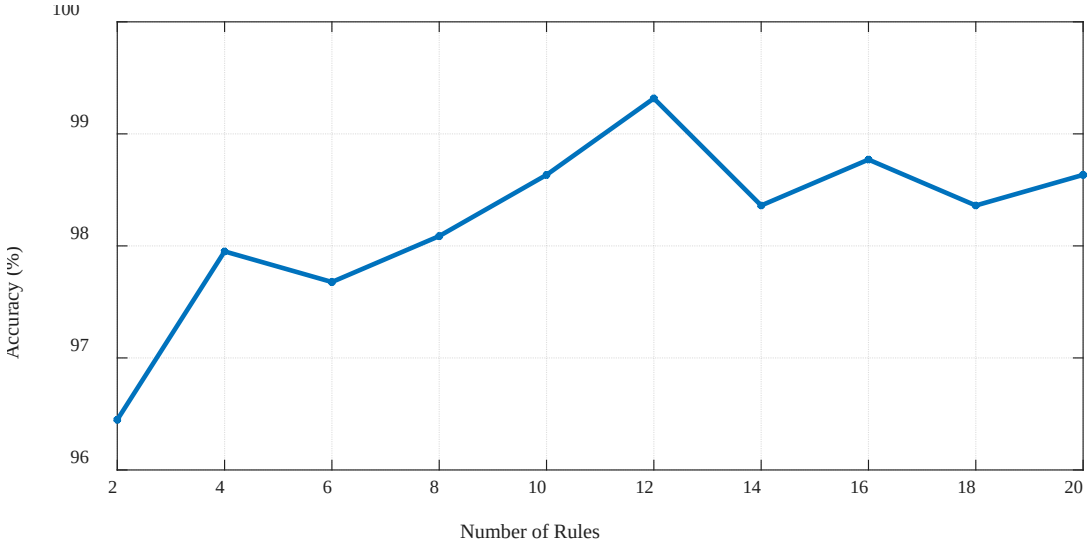
Αυτά τα δεκατέσσερα καλύτερα χαρακτηριστικά, όπως προέκυψαν από την κατάταξη του αλγόριθμου Relieff είναι τα παρακάτω:

Κατάταξη	Χαρακτηριστικά
1	RMS Energy
2	MFCC7
3	MFCC3
4	MFCC4
5	MFCC5
6	ZCR
7	MFCC6
8	MFCC9

9	MFCC8
10	MFCC2
11	MFCC10
12	MFCC11
13	MFCC12
14	MFCC1

Σε αυτό το σημείο να αναφερθεί ότι η κατάταξη του χαρακτηριστικού RMS Energy δεν αποτελεί λόγος προβληματισμού λόγω της διαδικασίας κανονικοποίησής του όπως αυτή περιεγράφηκε στην ενότητα 4.3.

Στη συνέχεια, υπολογίζεται η ακρίβεια του ασαφούς μοντέλου για διάφορες τιμές του αριθμού κανόνων με βάση τους οποίους εξάγεται η εκτίμηση του μοντέλου:



Εικόνα 7: Επίδοση ασαφούς μοντέλου για διάφορες τιμές του αριθμού κανόνων

Με βάση το παραπάνω διάγραμμα και για παρόμοιους λόγους σχετικά με επίδοση και πολυπλοκότητα συστήματος, επιλέγεται ως βέλτιστη τιμή του αριθμού κανόνων η τιμή $NR = 12$. Επομένως οι υπό-βέλτιστες τιμές των παραμέτρων του ασαφούς μοντέλου έχουν βρεθεί ως $w = 0{,}8\ sec,\ NF = 14$ και $NR = 12$. Σε αυτό το σημείο η παραμετροποίηση του ασαφούς μοντέλου έχει ολοκληρωθεί.

Με τις παραπάνω τιμές, εξήχθησαν οι παρακάτω επιδόσεις του ασαφούς μοντέλου ως προς δύο διαφορετικές μεθόδους Clustering αρχικοποίησης του μοντέλου:

Μέθοδος Αρχικοποίησης	**Ακρίβεια**
FCM	84.33%
SC	84.29%

Πίνακας 1: Αποτελέσματα ασαφούς μοντέλου με χρήση του σετ δεδομένων GTZAN

Μέθοδος Αρχικοποίησης	Ακρίβεια
FCM	97.69%
SC	95.85%

Πίνακας 2: Αποτελέσματα ασαφούς μοντέλου με χρήση του σετ δεδομένων Musan Corpus

Από τους παραπάνω πίνακες φαίνεται πως δεν μπορούμε να αποφανθούμε για το ποια μέθοδος αρχικοποίησης είναι η πιο αποτελεσματική και επομένως στη συνέχεια θα αξιολογηθούν και οι δύο.

Τέλος, η τελική αξιολόγηση του ασαφούς νευρικού δικτύου θα γίνει συγκρίνοντας την επίδοση του, για τις δύο μεθόδους αρχικοποίησης, χρησιμοποιώντας διαφορετικά σετ δεδομένων για εκπαίδευση και αξιολόγηση. Στο ασαφές μοντέλο εφαρμόζεται διασταυρωμένη επικύρωση στα δεδομένα εκπαίδευσης ώστε να αποφευχθεί το φαινόμενο της υπερεκπαίδευσης. Στους παρακάτω πίνακες φαίνονται οι επιδόσεις για χρήση ως δεδομένων εκπαίδευσης είτε το σετ δεδομένων GTZAN είτε το σετ δεδομένων Musan Corpus, ενώ για αξιολόγηση χρησιμοποιείται οποιοδήποτε από τα σετ δεδομένων GTZAN ή Musan Corpus ή Mirex Examples:

Ταξινομητής:	**Δεδομένα Εκπαίδευσης:**	**Δεδομένα Αξιολόγησης:**	**Ακρίβεια Εκπαίδευσης:**	**Ακρίβεια Αξιολόγησης:**
Fuzzy FCM (MATLAB)	GTZAN	Musan Corpus	84.83 %	79.93%
		Mirex Examples		67.18%
	Musan Corpus	GTZAN	97.69%	74.32%
		Mirex Examples		74.25%
Fuzzy SC (MATLAB)	GTZAN	Musan Corpus	84.29%	87.82%
		Mirex Examples		67.67%
	Musan Corpus	GTZAN	95.85%	73.79 %
		Mirex Examples		69.83%

Πίνακας 3: Επίδοση ασαφούς μοντέλου για διαφορετικά σετ δεδομένων

Αρχικά, από τους πίνακες παρατηρείται ότι η ακρίβεια μειώνεται δραματικά όταν χρησιμοποιείται διαφορετικό σετ αξιολόγησης. Επομένως προκύπτει η αναγκαιότητα δοκιμής του συστήματος σε διαφορετικό σετ αξιολόγησης ώστε τα η υπολογισμένη ακρίβεια αξιολόγησης να εκτιμάται πλησιέστερα σε αυτήν που θα έχει στην πράξη το σύστημα όταν θα χρησιμοποιηθεί σε άγνωστα, προφανώς, αρχεία ήχου.

Επιπλέον, δεδομένου ότι το σετ δεδομένων Mirex Examples πρότεινεται από το διαγωνισμό ως σετ δεδομένων αξιολόγησης, το καλύτερο μοντέλο ταξινόμησης προκύπτει αυτό με μέθοδο αρχικοποίησης FCM και δεδομένα εκπαίδευσης Musan Corpus. Τέλος, παρατηρούμε ότι η εκπαίδευση με Musan Corpus οδηγεί πάντα σε καλύτερη ταξινόμηση των δεδομένων Mirex Examples. Επομένως προκύπτει ότι τα δεδομένα Musan Corpus είναι πιο αντιπροσωπευτικά των δεδομένων Mirex Examples [2], γεγονός αναμενόμενο καθώς το Musan Corpus είναι αρκετά μεγαλύτερο από το GTZAN και η ποιότητα του είναι πλησιέστερη με αυτή του Mirex Examples (θόρυβος κτλ.).

5.2 Μοντέλα Ταξινόμησης Weka

Παρομοίως με πριν, αξιολογούνται διάφορα μοντέλα ταξινόμησης διαθέσιμα στο πρόγραμμα Weka:

Μοντέλο Ταξινόμησης	Ακρίβεια
Rep Tree	81.61%
IBk (kNN)	92.79%
Multi-Layer Perceptron	85.14%
Random Committee	88.88%
Random Forest	93.38%

Πίνακας 4: Αποτελέσματα Weka για δεδομένα εκπαίδευσης GTZAN

Μοντέλο Ταξινόμησης	Ακρίβεια
Rep Tree	93.66%
IBk (kNN)	98.22%
Multi-Layer Perceptron	96.80%
Random Committee	97.37%
Random Forest	97.76%

Πίνακας 5: Αποτελέσματα Weka για δεδομένα εκπαίδευσης Musan Corpus

Από τους παραπάνω πίνακες, προκύπτει πως τα δύο αποτελεσματικότερα μοντέλα ταξινόμησης είναι τα IBk και Random Forest. Για αυτά τα δύο επικρατέστερα μοντέλα ταξινόμησης, όπως και προηγουμένως, παρουσιάζεται ο πίνακας αξιολόγησής τους για διαφορετικά σετ δεδομένων και εκπαίδευσης:

Ταξινομητής:	**Δεδομένα Εκπαίδευσης:**	**Δεδομένα Αξιολόγησης:**	**Ακρίβεια Αξιολόγησης:**
Random Forest (Weka)	GTZAN	Musan Corpus	89.28%
		Mirex Examples	65.07%
	Musan Corpus	GTZAN	73.90%
		Mirex Examples	68.82%
IBk (Weka)	GTZAN	Musan Corpus	88.01%
		Mirex Examples	65.56%
	Musan Corpus	GTZAN	70.23%
		Mirex Examples	72.61%

Πίνακας 6: Επίδοση μοντέλων του Weka για διαφορετικά σετ δεδομένων

Ακολουθώντας το ίδιο σκεπτικό με πριν, ως καλύτερος ταξινομητής προκύπτει το μοντέλο ταξινόμησης IBk για δεδομένα εκπαίδευσης Musan Corpus.

5.3 Επιλογή Καλύτερου Ταξινομητή

Από όλη την προηγούμενη ανάλυση που προηγήθηκε στις ενότητες 5.1 και 5.2 καταλήγουμε στον παρακάτω πίνακα, που συγκρίνει το καλύτερο μοντέλο ταξινόμησης του ασαφούς συστήματος με το καλύτερο διαθέσιμο μοντέλο του προγράμματος Weka:

Ταξινομητής:	**Δεδομένα Εκπαίδευσης:**	**Δεδομένα Αξιολόγησης:**	**Ακρίβεια Αξιολόγησης:**
Fuzzy FCM (MATLAB)	Musan Corpus	Mirex Examples	74.25%
IBk (Weka)	Musan Corpus	Mirex Examples	72.61%

Πίνακας 7: Τελική σύγκριση των δύο καλύτερων ταξινομητών

Καταληκτικά, ως καλύτερο μοντέλο ταξινόμησης και αυτό που επιλέγεται για τα δύο ζητούμενα tasks του διαγωνισμού MIREX 2018 είναι το Ασαφές Νευρωνικό Δίκτυο με μέθοδο αρχικοποίησης FCM.

6 Σύνοψη

Σε αυτή την εργασία, στο πρώτο μέρος της, διατυπώθηκε αρχικά το ζητούμενο πρόβλημα που ζητείται όπως αυτό προκύπτει από τις οδηγίες του διαγωνισμού MIREX 2018: Music and/or Speech Detection. Για το δοθέν πρόβλημα προέκυψε ότι η αντιμετώπιση του καλεί την σχεδίαση ενός Συστήματος Αναγνώρισης Προτύπων. Με βάση αυτή την παρατήρηση παρουσιάστηκαν μερικές βασικές αρχές της θεωρίας Αναγνώρισης Προτύπων καθώς και βασικά στοιχεία των Συστημάτων Αναγνώρισης Προτύπων. Στη συνέχεια, παρουσιάστηκαν διάφορα εργαλεία που μπορούν να υλοποιήσουν το ζητούμενο σύστημα και εκτιμήθηκε ένα προσχέδιο αυτού.

Στο δεύτερο μέρος, παρουσιάστηκαν λεπτομερώς τα συστήματα που υλοποιήθηκαν. Ακολούθησε ανάλυση των παραπάνω συστημάτων που αφορούσε τις επιδόσεις τους ως προς διάφορες παραμέτρους τους και με βάση αυτήν, τα συστήματα παραμετροποιήθηκαν κατάλληλα. Έπειτα, έγινε σύγκριση όλων των συστημάτων και επιλέχθηκε το καλύτερο.

Ως καλύτερο σύστημα αναγνώρισης προτύπων, προέκυψε το Ασαφές Νευρωνικό Δίκτυο με μέθοδο αρχικοποίησης FCM με τιμές παραμέτρων του συστήματος $w = 0{,}8\ sec$, $NF = 14$ και $NR = 12$ με χρήση δεδομένων εκπαίδευσης το σετ δεδομένων Musan Corpus. Το σύστημα πέτυχε 74.25% ακρίβεια αξιολόγησης στην ταξινόμηση του σετ δεδομένων Mirex Examples. Τέλος, να σημειωθεί ότι, εφόσον ταξινόμηση έγινε σε επίπεδο παραθύρου, ολοκληρώθηκαν επιτυχώς και τα δύο ζητούμενα tasks.

Ο κώδικας των scripts, τα .csv αρχεία που περιέχουν τις τιμές των χαρακτηριστικών που εξήχθησαν για διάφορα δεδομένα, καθώς και άλλα χρήσιμα αρχεία βρίσκονται στο repository: https://github.com/charaldp/SoundAndImageTech/.

ΑΝΑΦΟΡΕΣ

[1] «MIREX 2018: Main Page,» [Ηλεκτρονικό].

[2] «MIREX 2018: Music and/or Speech Detection,» [Ηλεκτρονικό].

[3] P. Robi, «Pattern Recognition,» *Wiley Encyclopedia of Biomedical Engineering,* 2006.

[4] «MIRtoolbox - User's Manual,» [Ηλεκτρονικό].

[5] «Documentation-Sonic Visualiser,» [Ηλεκτρονικό].

[6] «Weka,» [Ηλεκτρονικό].

[7] «Relieff,» [Ηλεκτρονικό].

[8] «Fuzzy Logic Toolbox,» [Ηλεκτρονικό].

[9] L. Vrysis , N. Tsipas, C. Dimoulas και G. Papanikolaou, «Extending Temporal Feature Integration for Semantic Audio Analysis.,» *Proceedings in Audio Engineering Society Convention,* (2017)..

[10] I. Kononenko, E. Simec και M. Robnik-Sikonja, «Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF,» *Applied Intelligence,* τόμ. 7, αρ. 1, pp. 39--55, 1997.

[11] L. Vrysis, N. Tsipas, C. Dimoulas και G. Papanikolaou, «Crowdsourcing Audio Semantics by Means of Hybrid Bimodal Segmentation with Hierarchical Classification,» *Journal of the Audio Engineering Society,* τόμ. 64, pp. 1042-1054, 12 Δεκέμβριος 2016.